ABSTRACT

VERIFICATION OF EMOTION RECOGNITION FROM FACIAL EXPRESSION

by
Yanjia Sun

Analysis of facial expressions is an active topic of research with many potential applications, since the human face plays a significant role in conveying a person's mental state. Due to the practical values it brings, scientists and researchers from different fields such as psychology, finance, marketing, and engineering have developed significant interest in this area. Hence, there are more of a need than ever for the intelligent tool to be employed in the emotional Human-Computer Interface (HCI) by analyzing facial expressions as a better alternative to the traditional devices such as the keyboard and mouse.

The face is a window of human mind. The examination of mental states explores the human's internal cognitive states. A facial emotion recognition system has a potential to read people's minds and interpret the emotional thoughts to the world. High rates of recognition accuracy of facial emotions by intelligent machines have been achieved in existing efforts based on the benchmarked databases containing posed facial emotions. However, they are not qualified to interpret the human's true feelings even if they are recognized. The difference between posed facial emotions and spontaneous ones has been identified and studied in the literature. One of the most interesting challenges in the field of HCI is to make computers more human-like for more intelligent user interfaces.

In this dissertation, a Regional Hidden Markov Model (RHMM) based facial emotion recognition system is proposed. In this system, the facial features are extracted from three face regions: the eyebrows, eyes and mouth. These regions convey relevant

information regarding facial emotions. As a marked departure from prior work, RHMMs for the states of these three distinct face regions instead of the entire face for each facial emotion type are trained. In the recognition step, regional features are extracted from test video sequences. These features are processed according to the corresponding RHMMs to learn the probabilities for the states of the three face regions. The combination of states is utilized to identify the estimated emotion type of a given frame in a video sequence. An experimental framework is established to validate the results of such a system. RHMM as a new classifier emphasizes the states of three facial regions, rather than the entire face. The dissertation proposes the method of forming observation sequences that represent the changes of states of facial regions for training RHMMs and recognition. The proposed method is applicable to the various forms of video clips, including real-time videos. The proposed system shows the human-like capability to infer people's mental states from moderate level of facial spontaneous emotions conveyed in the daily life in contrast to posed facial emotions. Moreover, the extended research work associated with the proposed facial emotion recognition system is forwarded into the domain of finance and biomedical engineering, respectively. CEO's fear facial emotion has been found as the strong and positive predictor to forecast the firm stock price in the market. In addition, the experiment results also have demonstrated the similarity of the spontaneous facial reactions to stimuli and inner affective states translated by brain activity. The results revealed the effectiveness of facial features combined with the features extracted from the signals of brain activity for multiple signals correlation analysis and affective state classification.

VERIFICATION OF EMOTION RECOGNITION FROM FACIAL EXPRESSION

by
Yanjia Sun

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Electrical Engineering

Helen and John C. Hartmann Department of
Electrical and Computer Engineering

January 2016

.

APPROVAL PAGE

VERIFICATION OF EMOTION RECOGNITION FROM FACIAL EXPRESSION

Yanjia Sun

Dr. Ali N. Akansu, Dissertation Advisor — Date
Professor of Electrical and Computer Engineering, NJIT

Dr. John D. Carpinelli, Committee Member — Date
Professor of Electrical and Computer Engineering, NJIT

Dr. James E. Cicon, Committee Member — Date
Assistant Professor of Finance, University of Central Missouri

Dr. Yun Q. Shi, Committee Member — Date
Professor of Electrical and Computer Engineering, NJIT

Dr. Richard A. Haddad, Committee Member — Date
Professor Emeritus of Electrical and Computer Engineering, NJIT

BIOGRAPHICAL SKETCH

Author: Yanjia Sun

Degree: Doctor of Philosophy

Date: January 2016

Undergraduate and Graduate Education:

- Doctor of Philosophy in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2016

- Master of Arts in Industrial Management and Technology,
  Central Michigan University, Mount Pleasant, MI, 2009

- Bachelor of Science in Electrical Engineering and Automation,
  Shanghai Institute of Technology, Shanghai, P. R. China, 2007

Major: Electrical Engineering

Presentations and Publications:

Y. Sun and A. N. Akansu, "Facial expression recognition with regional hidden Markov model," *Electronic Letters*, vol. 50, no. 9, pp. 671-673, Apr. 2014.

Y. Sun and A. N. Akansu, "Automatic inference of mental states from spontaneous facial expressions," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014.

Y. Sun, J. E. Cicon, and A. N. Akansu, "The power of fear: facial emotion analysis of CEOs to forecast firm performance," *IEEE International Conference on Information Reuse and Integration (IRI),* San Francesco, CA, Aug. 2014.

Y. Sun, H. Ayaz, and A. N. Akansu, "Neural correlates of affective context in facial expression analysis: a simultaneous EEG-fNIRS study," *IEEE Global Conference on Signal and Information Processing (GlobalSIP),* Orlando, FL, Dec. 2015.

谨以此论文献给我的父母，孙宝生先生和严玉兰女士，挚爱的妻子, 王欢和可爱的儿子, 孙泽天。

To my dear father

Baosheng Sun,

To my dear mother

Yulan Yan,

To my beloved wife

Huan Wang,

And

To my lovely son

William Z. Sun

ACKNOWLEDGMENT

First of all, I would like to express my deepest gratitude to Prof. Ali. N. Akansu who not only served as my dissertation advisor, providing an inspiring research environment but also gave me all his support, encouragement and guidance throughout my doctoral studies. He taught me how to figure out the problem in the more efficient way and how to professionally communicate with others in my life.

Next, I wish to express my sincere appreciation to Prof. John D. Carpinelli, who injected the significant elements into my life: confidence, courage, and patience. I learned from him how to enjoy life and keep peaceful attitude to face dilemma.

I must extend my gratitude to Prof. James E. Cicon, although he moved to University of Central Missouri. He spent a lot of time to discuss the ideas and the details in term of my research work. My special thanks are given to Prof. Yun Q. Shi and Prof. Richard A. Haddad actively participating in my committee.

I am very grateful to Prof. Leonid Tsybeskov, Prof. Durgamadhab Misra and the staff of the ECE department, especially Jacinta Y. Williams, Tanita S. Turner, and Barbara Feltz for their helps in solving many administrative issues I met during my years at NJIT.

The last but the most, I would like to thank my parents, Baosheng Sun and Yulan Yan, my wife, Huan Wang, my son, William Z. Sun, and parents-in-law, Yuejian Wang and Dongxiao Yang, for their endless love, support and patience during my research. Without their psychological support, this dissertation would not be completed.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF TABLES
(Continued)

Table **Page**

LIST OF FIGURES

LIST OF FIGURES
(Continued)

CHAPTER 1

INTRODUCTION

Emotion as an important element in people's daily communication shapes the life contexts. Although speech itself is often sufficient for communication, non-verbal communication can be useful in synchronizing the dialog and to make the dialog smoother. Facial expression has been regarded as one of the most efficient and instinctive ways to reflect human feelings and intentions in the form of nonverbal communication. Due to the practical values, research on emotion has increased significantly over the past two decades with contribution in many fields including psychology, neuroscience, medicine, history, sociology, and computer science. The scientific study of the facial expression of emotion began with Charles Darwin's The Expression of Emotions in Man and Animals [1]. Facial expression conveys how the individual spontaneously interprets and reacts to any stimuli (e.g. video, image, question, etc.) in the daily life. It also reveals the responses the individual evokes in others. Although often momently appearing on the face and dynamically changed, facial expressions indeed are the mirror reflecting the human inner mind in the aspect of psychology.

Analysis of facial expressions is always an attractive topic since the human face plays a significant role in conveying the mental states [2]. According to the recent studies [3] in psychology field, they empirically revealed that people reliably infer others preferences from spontaneous facial expressions. Meanwhile, automatic facial emotion recognition for human-computer interaction (HCI) designs is also a most important research field, instead of the traditional interface devices such as the keyboard and mouse concentrating on the transmission of explicit messages whereas overlooking implicit emotions conveyed by the users. Importantly, sometimes response to user's emotions by the computer, comparing with the human's reactions, is perceived as more natural, persuasive, and trusting. One of the interesting challenges in the community of HCI is how to make computers be more human-like for intelligent user interfaces. To make the information of facial expressions available for application on HCI, several efforts have been made on automatic analysis of spontaneous facial expressions in [4, 5, 6]. Psychologists Ekman and Friesen made seminal contributions to the field of emotion recognition in [7] in order to make identification standard. They defined the universal recognition of six emotions. Namely, they are Anger, Disgust, Fear, Happiness, Sadness and Surprise. The high recognition accuracy of facial emotions has been achieved in many existing researches based on the benchmarked databases containing posed facial emotions. Their purposes focus on the optimal performance of their methods rather than their plausibility in the psychological aspect. That is, the emotions shown in still images or image sequences are elicited by asking subjects to perform a series of exaggerate emotional expressions. However, they are not qualified to interpret the human's true feelings even if they are recognized. As a survey of affect recognition methods indicated in [8], the difference between posed facial emotions and spontaneous ones has been found in [9, 10]. In addition, real life emotions are often complex and involve several simultaneous emotions. They may occur either as the quick succession of different emotions, superposed emotions, or the overacting of an emotion. These phenomena are referred as blended emotions that produce several simultaneous facial expressions. The resulting facial expressions are not identical according to the type of blending.

The other main challenge trigging the researchers' interest is whether the information expressed on the face can be accurately read as it is implicitly driven by human inner affective states. Lucey et al. [11] show the automatic pain detection from the captured expressions on the face of the participant who were suffering from shoulder pain. The relationship between the facial responses and ad effectiveness studied in [12] demonstrates that ad liking can be predicted accurately from webcam facial responses. The part of human brain has been demonstrated to involve in the expression of emotions in the field of neuroscience [13]. It is worth to note the importance of signals captured from the brain in Central Nervous System (CNS) that have been proved to provide informative characteristics in responses to the emotional states. The development of brain-computer interface (BCI) opens a new way to diversify the communication between computer and human. Functional near-infrared spectroscopy (fNIRS) [14] and electroencephalography (EEG) [15] show the strong reliability of detecting the activity in human brain although they measure different physiological responses to the affective state changes. Instead, the multimodal methodology has been proposed and achieved satisfactory performance.

The main focus of dissertation is on further delving into the theory of facial emotion recognition as well as empirical conclusion in psychology so as to make the performance of recognition more accurately. The attention is also put on the implementation of facial emotions recognition system so as to apply the findings into the challenges in the various fields such as psychology, finance and biomedical engineering. Specifically, contributions of the dissertation include the following.

First, a region-based model is proposed to analyze the facial expression associated with the empirical conclusion drawn in the psychology flied. The performance of recognizing human facial emotions is more accurate comparing to that of the state-of-the-art mechanism. In addition, the further study demonstrates the ability of the automatic facial expression system using the proposed model to imitate the human to examine people's internal feeling through actual interaction.

Second, CEO's facial expressions during interviews are found to provide the significant information for predicting the firm performance. Fear emotion is the only one type which shows positive and significant explanatory power with respect to the firm's announcement returns.

Third, people's spontaneous facial expressions are demonstrated to be able to convey the inner affective states translated by the brain activity. The proposed method for the combination of fNIRS signals and EEG signals outperforms using either fNIRS or EEG when the stimuli are videos or images. In addition, the finding shows that the recognition performance of these three modal methods with respect to video-content stimuli is better than that of image-content stimuli.

Further details in terms of above three points are provided as below. Outline of the dissertation is discussed at the end of this chapter.

1.1 Automatic Facial Emotion Recognition System

Faces are important signals for humans because they inform about fundamental aspects of social communication such as gender, race, social status, and emotional states. Surveys related on facial emotion recognition can be found in [16, 17, 18]. In general, the procedure can be distinguished by three phases. Initially, the system must detect the human face in the scene. In the second phase, facial expression information is extracted from the facial image or image sequences. The final phase determines the categories of extracted features for recognition. Most existing approaches of facial features extraction are either appearance-based or geometric-based. The appearance-based techniques use texture of facial skins including wrinkles, bulges and furrows as information for facial features extraction. The typical methods include Discrete Cosine Transform (DCT) coefficients used in [19], Haar-like features proposed in [20] and Gabor filters utilized in [21]. Facial Action Coding System (FACS) proposed in [22] involves identifying the facial muscle movements which are individual or in groups. The variations in the face cause by the muscle movements are named as Action Units, which form the different facial expressions as proposed in [23]. Alternative to appearance features, the geometric-based techniques extract facial feature points or the shapes of the significant face regions like eyebrows, eyes, and mouth to form a feature vector as proposed in [24, 25].

Classification following features extraction provides classified features for recognition. It is the final stage of a facial emotion recognition system. Methods of classification can be generally divided into static ones and dynamic ones. Static-based classification, based on the information of the input image, takes advantage of Support Vector Machine [26], Neural Network [27], Bayesian Network [28], and Linear Discriminant Analysis [29] are used in many literatures for facial expression recognition. Dynamic classifiers like Hidden Markov Model (HMM) [30] utilize temporal information to analyze facial expressions. Existing researches employ the emotion-specific HMMs to train and test the facial features extracted from the whole face or several specific face regions in the image such as [31, 32, 33]. However, configured cues based on parts of the face play a more important role than holistic features for recognition of facial emotion. The view is widely supported by many investigators in [34, 35, 36]. Practically, this implies that facial emotions are represented for recognition as descriptions of part and part relations. In other word, an emphasis on classifying local features of a face might effectively improve the recognition performance.

For analysis of spontaneous facial expressions, several databases have been proposed. However, as the survey in [8] indicated, existing databases more or less expose some flaws. Firstly, video collection needs to be constrained so as to minimize the bias. That is, the subjects cannot know that they are doing experiments for any purpose. In [37], subjects' spontaneous expressions were gathered through letting them describe the computer games after playing, while the psychologist was asked to stay with subjects in the same room in the experiments. Any of psychologist's behaviors may bias the subjects' facial expressions. In [38], the subject's face and hands are attached markers for collecting experimental data during recording videos. Although some databases tried to conceal the experimental purpose by hiding the camera such as [39], the subjects still could be aware that they were doing experiments since they were asked to watch emotional videos in a locked room. Subjects in [5] tried to convince the interviewers they were telling the truth while they probably lied to the interviewers. Subjects in such paradigm were most likely to subjectively express exaggerate or opposite emotions in order to persuade the interviewers. In addition, another critical factor is to objectively label the videos with emotion types, whereas the video clips are usually labeled by the researchers who are the designers of the database such as proposed in [40].

In this dissertation, a Regional Hidden Markov Model (RHMM) based facial emotion recognition system is proposed. Facial features are extracted from three face regions as eyebrows, eyes and mouth. These regions convey relevant information regarding facial emotions. As a marked departure from prior work, RHMMs are trained for the states of these three distinct face regions rather than being modeled the entire face for each facial emotion type. In the recognition step, facial features in these regions are extracted from test video sequences. These facial features are processed by the corresponding RHMMs in order to measure probabilities for face regions. The highest combination of probabilities for three face regions is utilized to estimate the emotion type of a given frame in a video sequence. The proposed system is explained in more detail in Chapter 3.

The contributions of the system are summarized as follows:

First, a new classifier emphasizing states of three face regions rather than modeling emotion types based on the entire face is used in the proposed system.

Second, the proposed way of forming observation sequences fed to RHMMs is robust to the various forms of video clips for training and recognition, especially in the real-time scene, other than the fixed form of videos for designated HMMs starting from neutral emotion to the peak and fading to neutral one as proposed in [41, 42, 43].

Third, the proposed system shows its human-like capability of inferring other people's mental states from moderate facial spontaneous expressions, instead of merely considering its optimal performance in some specific experimental environment.

1.2 The Impact of CEO Facial Expression on Firm Performance Forecasting

The CEO of the firm disseminates information about the firm through many different avenues. These include quarterly reports, annual shareholder letters, periodic press releases, telephone conference calls, and multimedia interviews [44, 45]. There is a growing finance literature which analyzes how the market reacts to the textual information in reports, letters, press releases and other financial documents [46]. However, there is limited research on how markets react to non-textual information such as that provided by a subject's voice and/or appearance [47, 48]. Recently interest in non-verbal information has begun to emerge in the finance literature. For instance, the stress levels in the voices of managers during telephone conversations are analyzed in [49]. The emotional expressiveness in traders' faces is analyzed in [50]. The "attractiveness" or "beauty" in pictures taken of CEOs is analyzed in [51].

Managerial emotional states are little studied in the finance literature. However, it is an important issue because only a small fraction of the information which managers deliver is contained in their verbal message [52]. The authors in [49] observe that significant additional information is conveyed by managerial facial expressions, postures, and gestures. They add to the literature by showing that a manager, who knows that their firm is going to underperform, betrays this information with the stress in their voice. Other studies explore yet other sources of information such as the facial emotional state of analysts [50] or the effect of the physical attractiveness of the CEO on firm stock price performance [51]. Our study adds to this important and growing literature stream by examining both the determinants of the CEO's facial emotions and how the market reacts to these emotions.

This study is the first to provide evidence of the role that CEO facial emotions play as a source of information about the firm's financial performance. Although an individual can display a range of emotions, fear is widely recognized as the greatest motivator. Thus it is not surprising to find that when a CEO appears to be fearful under interrogation, that the market perceives this fear as a signal that the CEO will work harder to increase firm value. Indeed, the finding shows that the performance is significantly higher for those firms in which the CEO exhibited fear during his interview.

1.3 Analysis of Human Emotion Associated with Brain Activity

The neural mechanisms of emotion processing have been a key research area in cognitive neuroscience and psychiatry specifically due to clinical applications such as mood disorders [53, 54]. Prefrontal cortex (PFC) has been implicated with the emotion regulation and functional neuroimaging has been used to investigate neural correlates of emotion processing within this cortical region [55, 56, 57, 58]. Neurophysiological changes are found to be induced by non-consciously perceived emotional stimuli [59]. Noninvasive and portable neuroimaging techniques, Functional near-infrared spectroscopy (fNIRS) [60] and electroencephalography (EEG) [61] both have been utilized for emotion recognition although they measure different physiological responses to the mental state changes. For details of both methods of monitoring brain activity, please refer to Chapter 6.

Neurophysiological mechanism is demonstrated to infer human emotions by using non-invasive sensor modalities. fNIRS and EEG are vigorously developed as non-invasive and mobile systems [62, 63]. Therefore, they have been the leading technologies for affective state recognition. The recent study reported in [64] combines fNIRS and EEG with autonomic nervous system (ANS) including skin conductance responses (SCR) and heart rate (HR) for emotion analysis. Although ANS objectively measures the biological dynamics to reflect the emotional changes, its lack of mobility prevents it to be practical for the recognition of affective states.

These studies based on various modalities offer methods to measure inner affective reaction to stimuli. However, they cannot assist people to make rapid and effective social judgment on the affective expressions in real life scenarios. The inner affective states translated by human brain activity and consistent with those expressed on the face are worth to be explored. It may be of interest of the researchers and scientists for the various applications, such as the mind-controlled auto car, the treatment of autism, etc. The empirical findings reported in [3] show that people's spontaneous facial information implies their preferences per relevant stimuli.

This dissertation provides new insights for the exploration and analysis of spontaneous facial affective expression associated with the simultaneous brain activity in the form of hemodynamic and electrophysiological changes. To the best of our knowledge, this is the first attempt to detect the affective states by jointly using fNIRS, EEG and video capture of facial expressions. The main contribution demonstrates that human spontaneous facial expressions convey the affective states translated by the brain activity. The experimental results are found on the premise that the perceiver has no knowledge of stimuli prior to the experiment. The spontaneous facial expressions of the unaware perceiver can be triggered by the emotional stimulus [65]. Moreover, the neural activity changes are found due to non-conscious perception of emotional stimuli. To assess the plausibility of our findings, the perceiver's expressiveness is examined. The results indicate that there was no perceiver with exaggerative expression over all the trials performed. In addition, the findings reveal that the video-content stimuli more readily induce the perceiver's affect states than the image-content stimuli. This can be explained as dynamic (video) stimulus provides more contextual information than the static (image) one.

Compared to the static (image) stimuli, dynamic ones indicate enhanced emotion in the specific brain activation patterns as also shown in [66].

1.4 Organization of This Dissertation

The dissertation is organized as follows. In Chapter 2 an effective method to improve the performance of an automatic facial emotion recognition system is given. Firstly, the related work is reviewed and then the drawback of the existing algorithms is stated in Section 2.1. With the support of empirical findings in psychology field, the proposed method to bring the high performance of system is discussed in Section 2.2. An explicit statistical model for finding the hidden states introduced by Rabiner [67] is revisited in Section 2.3.

In Chapter 3, the system based on the proposed method in Chapter 2 for facial expression recognition that utilizes two different types of features extracted from face is provided. The process of feature extraction, the training and testing of the proposed model, and the performance evaluation are discussed in detail in sub-sections of Section 3.1. A facial expression database and the procedure to form observation sequences for training and testing the proposed model are provided in the same section. To improve the performance of facial expression recognition in person-independent case, the method of extracting geometric facial features is given in Section 3.2. Analogously, the structure of Section 3.2 also includes the training and testing of the proposed model and the performance evaluation. In addition, the comparison of the proposed method in this section, the one proposed in Section 3.1, and the state-of-the-art methods is reported in the same section.

In Chapter 4, the capability of the geometric-based system proposed in Chapter 3.2

is explored to infer people's mental states despite derived from spontaneous facial expressions. Furthermore, the system is compared with the human to see whether it is able to interact naturally with the user, similar to the way human-human interaction takes place. Initially, the means of referring to the emotional reaction to stimuli and objectively reflecting a person's mental state is provided in Section 4.1. Then, the experimental protocol of assessing the system is designed in Section 4.2. Two phases separately for the providers who express spontaneous facial expressions and for perceivers who evaluate the affective states are discussed in the same section. A face dataset with labeled affective states for the system evaluation is provided in Section 4.3. In addition, the labels of video clips in the dataset are assessed in order to ensure that perceivers remain consistent in terms of the emotion types labeled in the video clips. In Section 4.4, the comparison of performance of inferring human mental states by perceivers and the proposed system is reported in detail. The method of measuring affective stats by the system and the method of examining the reliability of experimental results are discussed in the same section.

In Chapter 5, one hypothesis that CEO facial expressions have effect on firm's market value is first proposed. The related work on the effect of facial emotions of traders and employees on the firm performance is stated in Section 5.1. The collected dataset of CEO interview videos for the study are provided in Section 5.2. The following section describes the merit of the proposed automatic facial expression recognition system that benefits to analyze CEO facial reaction to the various questions raised by the interviewer during the interview. The proposed method for modeling the relationship of CEO facial expressions and the firm performance in the market after CEO interview is presented in Section 5.3. The pre-processing procedure for baseline adjustment in terms of the detected

CEO’s facial expressions, the preliminary analysis of correlation of CEO facial expressions and the firm prices, the regression model to explore the effect CEO facial expression on the firm performance, and the assessment on the reliability of the experimental results are discussed in the same section.

In Chapter 6, the study explores the verification of the human facial expressions registered as video signal by simultaneous analysis of affective states through multimodal brain activity measures. The related work of existing methods of investigating the brain signals implicated with emotion processing is reviewed in Section 6.1. In Section 6.2, the protocol of experiment is described in detail. The experimental apparatus of measuring brain activity and facial emotion is brought in Section 6.3. Functional Near Infrared Spectroscopy (fNIRS) and Electroencephalography (EEG), two non-intrusive and safe brain monitoring methods, are introduced and compared in terms of their different functionalities. The methods of pre-processing captured signals, selecting feature, and evaluating human inner spontaneous affective states through brain activity are proposed in Section 6.4. The experimental performance using signals of three modalities of fNIRS, EEG, and face emotion recognition is analyzed. The results that explain the findings are further evaluated in the same section.

The dissertation is concluded in Chapter 7. The summary of the contributions of the dissertation and future work are discussed in the same chapter.

CHAPTER 2

RIGIONAL HIDDEN MARKOV MODEL

Apart from their usefulness in allowing us to identify who he or she is, faces inform us about people's emotions and intentions. A facial expression is one or more motions of the muscles beneath the skin of the face. These movements convey the emotional state of an individual to observers. For face recognition, the current consensus is that the holistic features are particularly important. Different from face identification, face expression is represented and recognized in terms of their individual facial regions and features. Action units (AUs) represent the muscular activity that produces facial appearance changes. Human emotion consists of thousands of expressions, which differ in subtle changes due to a few facial features or blending of several emotions. The Facial Action Coding System (FACS) [22] decomposes the facial behavior into 46 action units (AUs), each of which is anatomically related to the contraction of a specific set of facial muscles movement. The study provides substantial support for the idea that face recognition and facial emotion recognition are dissociable cognitive functions, and this explains why these two facial features are suggested to be separately employed for specific purpose.

In this chapter, the rationale of facial emotion recognition, associated with existing work in the psychology field, is discussed. A Markov chain based face model with hidden states is proposed that follows the human-like behavior to recognize the facial emotions. A classic method for classification [67] is revisited and an efficient derivative method to model face for emotion recognition is proposed. The mathematical steps required for the derivation of Hidden Markov Model (HMM) and its application is also presented in the chapter.

2.1 Problem Statement

Classification following features extraction provides classified features for recognition. It is the final stage of a facial emotion recognition system. Methods of classification can be generally divided into static ones and dynamic ones. Static-based classification, based on the information of the input image, takes advantage of Support Vector Machine [26], Neural Network [27], Bayesian Network [28], and Linear Discriminant Analysis [29] are used in many literatures for facial expression recognition. Dynamic classifiers like Hidden Markov Model (HMM) [30] utilize temporal information to analyze facial expressions. Existing researches employ the emotion-specific HMMs to train and test the facial features extracted from the whole face or several specific face regions in the image such as [31]. However, configured cues based on parts of the face play a more important role than holistic features for recognition of facial emotion. The view is widely supported by many investigators in [34]. Practically, this implies that facial emotions are represented for recognition as descriptions of part and part relations. In other word, an emphasis on classifying local features of a face might effectively improve the recognition performance.

Table 2.1 States of Three Face Regions

<table>
<tr><th colspan="2">Face Regions</th><th>Observable States</th></tr>
<tr><td colspan="2">Eyebrows</td><td>raise, fall, neutral</td></tr>
<tr><td colspan="2">Eyes</td><td>open, close, neutral</td></tr>
<tr><td rowspan="2">mouth</td><td>Mouth</td><td>open, close, neutral</td></tr>
<tr><td>lips corners</td><td>up, down, pull, pucker, neutral</td></tr>
</table>

Table 2.2 Library of Emotion Types and States of Face Regions

Emotion Type	**States of Face Regions**
Anger	Eyebrows fall, eyes close, mouth close, and lips corners pucker
Disgust	Eyebrows fall, eyes close, mouth close and lips corners pull
Sadness	Eyebrows raise, eyes close, mouth close and lips corners down
Happiness	Eyebrows neutral, eyes neutral, mouth open, and lips corners up
Surprise	Eyebrows raise, eyes open, mouth open, and lips corners neutral
Fear	Eyebrows raise, eyes open, mouth open, and lips corners pull

2.2 Sates of Face Regions

The information of facial emotion is mainly represented in the motions and deformations of eyebrows, eyes, and mouth regions of a human face. Therefore, the combination of motions of these three regions can be used to estimate a facial emotion. The approximate regions are estimated for eyebrows, eyes and mouth based on the size of the frame and the prior knowledge of the visual system so that these regions can be automatically extracted from the frames of a face video. All possible states of each region are shown in Table 2.1. It is found that the mouth region is slightly different from the eyebrows and eyes regions.

Besides the mouth itself, the corners of the lips are also an important feature used in facial emotion recognition. In the proposed method to train HMM, there are 14 HMMs to be created where one corresponds to each state of the facial regions. In Table 2.2, the relationship of emotions and states of regions is tabulated to identify the states associated with each emotion.

2.3 Hidden Markov Model

Hidden Markov Model (HMM) is a statistical Markov model with observable states and hidden states [67]. It assumes the system is modeled as Markov chain based process. The hidden states are not observed but they can be estimated by the given observable state through modeling the system. For the traditional Markov model, the assumption is the observable output in any given state is deterministic. However, it is limited to model the system with random character in the nature. That is, the output lies in the latent states while it is related with the observable states. In that way, it is necessary that the observable event is a probabilistic function of the state. HMM is derived from the Markov model and beyond it with an underlying stochastic process that cannot be directly observed, but can only be estimated through another set of stochastic processes that is represented by the observable sequence with given symbols.

Before the description of HMMs, the elements of an HMM by specifying the following parameters are defined:

N: the number of states in the model. The state of the model at time *t* is given by q_t,

$$S = \{S_i\}, \quad 1 \le i \le N \tag{2.1}$$

and

$$q_t \in S, \quad 1 \le t \le T \tag{2.2}$$

where T is the length (number of frames) of the output observable symbol sequence.

M: the number of distinct observable symbols per state. o_k is defined as one of all possible observable symbols for each state O_i at time t, then

$$O_i = \{o_k\}, \quad 1 \le i \le N, 1 \le k \le M \tag{2.3}$$

π : An N-element vector indicates the initial state probability.

$$\begin{aligned} &\pi = \{\pi_i\}, \quad 1 \le i \le N \\ &where\ \pi_i = P(q_1 = S_i) \end{aligned} \tag{2.4}$$

π_i indicates the initial state probability of each state.

A: An N x N matrix specifies the state-transition probability that the state will transit from state S_i to state S_j.

$$\begin{aligned} &A = \{a_{ij}\}, \\ &where\ a_{ij} = P\left[q_{t+1} = S_j | q_t = S_i\right], \quad 1 \le i \le N, 1 \le j \le N \end{aligned} \tag{2.5}$$

and

$$0 \le a_{ij} \le 1, \quad s.t. \sum_{j=1}^{N} a_{ij} = 1, \quad 1 \le i \le N \tag{2.6}$$

B: An *M x N* matrix represents the probability that the system will generate the observable symbol o_k at state S_i and at time *t*.

$$\begin{aligned} &B = \{b_i(k)\}, \triangleright \\ &where\ b_i(k) = P[o_k \text{ at } t \mid q_t = S_i], \quad 1 \le i \le N, \quad 1 \le k \le M \end{aligned} \tag{2.7}$$

and

$$0 \le b_i(k) \le 1, \quad s.t. \sum_{k=1}^{M} b_i(k) = 1, \quad 1 \le i \le N \tag{2.8}$$

The complete parameter set of the discrete HMM is represented as $\lambda = (\pi, A, B)$.

Three basic problems in HMM design will be discussed that are helpful to understand and apply HMM for specific training and recognition processes.

1. **Probability Evaluation** $P(O \mid \lambda)$ **:** find the probability of an observable sequence $O = \{o_1, ..., o_T\}$ given an HMM parameter set $\lambda = (\pi, A, B)$. In order to take use of an HMM for facial expression recognition, it is necessary to calculate the probability of observable sequence $P(O \mid \lambda)$ given the parameter of specific emotion model $\lambda = (\pi, A, B)$ in which the highest probability of observable face feature sequence $O = \{o_1, ..., o_T\}$ determine the facial emotion.

The most straightforward way to calculate the observable probability $P(O|\lambda)$ is by enumerating every possible latent sequence of length *T*, so there are N^T possible combinations of sequence where *N* is the total number of states. Therefore, the probability of the observable sequence $O = \{o_1, ..., o_T\}$ is to find each possible sequence of the hidden states, and sum these probabilities. The probability of $P(O|\lambda)$ is the summation of this joint probability over all N^T possible state sequences.

$$P(O|\lambda) = \sum_{i=1}^{N} P(o_1, ..., o_T, q_t = S_i | \lambda) \tag{2.9}$$

Calculating the probability in this manner is computationally expensive, particularly with large models or long sequences, and one might find that it is able to use the time invariance of the probabilities to reduce the complexity of the problem. Fortunately, it is able to use a more efficient procedure called the *Forward-Backward algorithm* [67] to overcome this limitation. First, a partial probability $\alpha_t(i)$ is defined, which is the probability of forwardly reaching an intermediate state in the entire sequence. One might calculate the probability of reaching an intermediate state as the sum of all possible paths to that state.

$$\alpha_t(i) = P(O,\ q_t = S_i | \lambda) \tag{2.10}$$

The probability of observing the sequence given the HMM is the sum of these final partial probabilities that is the sum of all possible paths.

$$\begin{cases} \alpha_1(i) = \pi_i b_i(1), \quad 1 \le i \le N \quad when\ t = 1 \\ \alpha_{t+1}(j) = b_j(k) \sum_{i=1}^{N} \alpha_t(i) a_{ij}, \quad 1 \le t \le T, 1 \le j \le N, 1 \le k \le M \end{cases} \tag{2.11}$$

Note that computation complexity in this way is much less expensive than summing the probabilities for all sequences. The step of backward procedure is similar with the forward one, except that it starts from the last state given the initial estimation. Hence, I do not get the procedure into detail.

2. **Decoding State Sequence**: find the sequence of hidden states that most probably generate an observable sequence given a HMM.

Similarly, there is the straightforward way to find the most probable sequence of hidden states. It is to list all possible sequences of hidden states and finding the probability of the observed sequence for each of the combinations. The most probable sequence of hidden states is the combination that maximizes $P(Q | O, \lambda)$. However, high computational complexity drives us to find a more efficient way. *Viterbi algorithm* [67] is used to find the single best state sequence $Q = \{q_1, ..., q_T\}$ given the observable symbol sequence and the HMM parameters.

One direct way might be considered is recursively to find the most probable sequence of hidden states given an observation sequence and a HMM. First, the partial probability $\delta_t(i)$ is defined, which is the probability of reaching an intermediate state in the entire sequence.

$$\delta_t(i) = \max_{q_1, \dots q_{t-1}} P(Q = S_i, O | \lambda) \tag{2.12}$$

Then, the probability of reaching an intermediate state is calculated followed by the sum of all possible paths to that state.

$$\begin{cases} \delta_1(i) = max[\pi_i b_i(1)], \quad 1 \le i \le N \quad when\ t = 1 \\ \delta_j(t+1) = max[b_j(k)\delta_j(t)a_{ij}], \quad 1 \le t \le T, 1 \le j \le N, 1 \le k \le M \end{cases} \tag{2.13}$$

3. **Parameter Estimation**: given a sequence of observations, generate a HMM in order to maximize the output probability of observable symbol sequence $P(O \mid \lambda)$.

The problem can be thought as system modeling. The intrinsic character of system is adjusted by training a set of observable sequences. After that, the modeled system can be used to recognize other sequences of observable sequence for classification. There is, however, no efficient way to optimize the model parameter set that globally maximizes the probability of the symbol sequence. Instead, the *Baum-Welch method* [67] is used for finding the maximum likelihood of model parameter set so that its likelihood function is locally maximized through an iterative procedure.

Assume $\xi_t(i, j)$ is the probability of being in state S_i at time *t*, and state S_j at time *t+1*, given the model and the observation sequence.

$$\xi_t(i,j) = P(q_t = S_i,\ q_{t+1} = S_j \mid O, \lambda), \quad 1 \le i \le N, \ \ 1 \le j \le N \tag{2.14}$$
$$= \frac{P(q_t = S_i,\ q_{t+1} = S_j, O \mid \lambda)}{P(O \mid \lambda)}$$
$$= \frac{\alpha_t(i) a_{ij} b_j(k)}{\sum_{i=1}^{N} \alpha_t(i) a_{ij} b_j(k)}$$

$\gamma_t(i)$ is notated as the probability of being in state S_i at time *t*, given the observation sequence *O*, and the model λ:

$$\gamma_t(i) = P(q_t = S_i \mid O, \lambda) = \sum_{j=1}^{N} \xi_t(i,j) \tag{2.15}$$

New HMM is $\hat{\lambda} = (\hat{\pi}, \hat{A}, \hat{B})$, and old HMM is $\lambda = (\pi, A, B)$.

$$\bar{\pi} = \text{expected frequency (number of times) in state } \mathrm{S}_i \text{ at time (t=1)} = \gamma_1(i) \tag{2.16}$$
$$\bar{a}_{ij} = \frac{\text{expected number of transitions from state } \mathrm{S}_i \text{ to state } \mathrm{S}_j}{\text{expected number of transitions from state } \mathrm{S}_i} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$
$$\bar{b}_j(k) = \frac{\text{expected number of times in state } \mathrm{S}_j \text{ and observing sysmbol } \mathrm{v}_k}{\text{expected number of times in state } \mathrm{S}_j} = \frac{\sum_{\substack{t=1 \\ \text{s.t. } O_t = v_k}}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$

The new model $\hat{\lambda}$ is compared with the old one λ by computing the distance score that reflects the statistical similarity of the HMM. If the distance score is higher than a predefined threshold ε , the entire training loop is repeated using the new model parameters by replace λ with $\hat{\lambda}$ recursively until

$$\left|\log[P(O\,|\,\lambda)] - \log[P(O\,|\,\bar{\lambda}]\right| < \varepsilon, \quad \varepsilon \text{ is the threshold} \tag{2.17}$$

Otherwise, the process of training the HMM is stopped and the new model parameters are saved.

2.4 Chapter Summary

In this chapter, the rationale of recognizing facial emotion is discussed. A Regional Hidden Markov Model (RHMM), associated with the psychology knowledge, for improving facial emotion recognition performance is proposed. To do that, facial features are extracted from three face regions as eyebrows, eyes and mouth. These regions convey relevant information regarding facial emotions. As a marked departure from prior work, RHMMs are trained for the states of these three distinct face regions rather than modeling the entire face for each facial emotion type. In the recognition step, facial features in these regions are extracted from test video sequences. These facial features are processed by the corresponding RHMMs in order to measure probabilities for face regions. The highest combination of probabilities for three face regions is utilized to estimate the emotion type of a given frame in a video sequence. A RHMM based facial emotion recognition system is proposed in the following chapter.

CHAPTER 3

FACIAL EMOTION RECOGNITION SYSTEM

Research on facial emotion recognition has been active due to its potential applications in various fields including education, military, business, and medical treatment. Six basic expressions such as Anger, Disgust, Sadness, Happiness, Surprise, and Fear have been widely utilized for facial emotion recognition [7]. Early versions of automated facial expression recognition systems for both still images and videos have become available in the market. The recognition methods for still images are relatively simple so that its recognition rate is negatively impacted. In video sequences, although the processing cost is significantly higher, the correlation between adjacent image frames provide more information for facial emotion recognition.

Classification is the crucial stage in a facial emotion recognition system. Methods of classification can be generally divided into static and dynamic ones. Static-based classification method takes use of the information of the input image while dynamic classifiers like Hidden Markov Model (HMM) [30] utilize temporal information to analyze facial emotions. Existing researches employ the emotion-specific HMMs to train and test the facial features extracted from the whole face or several specific face regions in the image. However, Emotions are comprised of its mapping on various regions of a face. Configured cues based on parts of the face plays a more important role than holistic features for recognition of facial emotion. In the other words, an emphasis on classifying local features of a face might effectively improve the recognition performance.

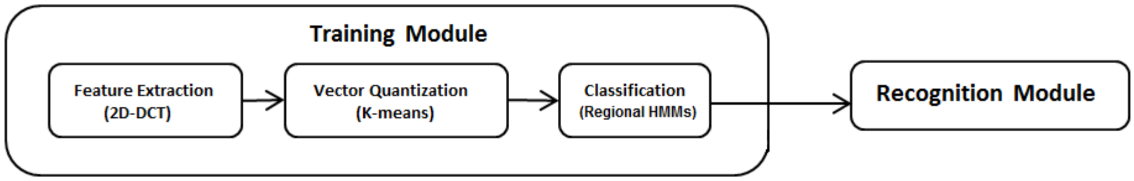

Figure 3.1 Block diagram of a facial emotion recognition system.

In this chapter, a Regional Hidden Markov Model (RHMM) based facial emotion recognition system is proposed. The proposed method employs three facial regions and extracts features independently for each region. It then classifies regional features by states of each region, because facial emotions express through the combination of local changes on face images. It classifies features more finely than using emotion type specific HMM.

3.1 System Based on Appearance Technique Using RHMM

The proposed system consists of a training module and a recognition module, as displayed in Figure 3.1. In the training module, there are three sections. They are feature extraction, vector quantization (VQ) and classification sections. 2D-DCT coefficients are used as features and they are mapped as the observation sequences after VQ [68]. As a marked departure from prior work, RHMMs are trained for the states of three distinct facial regions. They are eyebrows, eyes and mouth regions. Most significant information relevant to facial expressions appears in these three regions of a human face. In the recognition module, these regions are extracted from test sequences. They are processed by the corresponding trained RHMMs to calculate probabilities of observable sequences. The combination of the calculated highest probabilities of sates of the three distinct facial regions is utilized to recognize a facial emotion types.

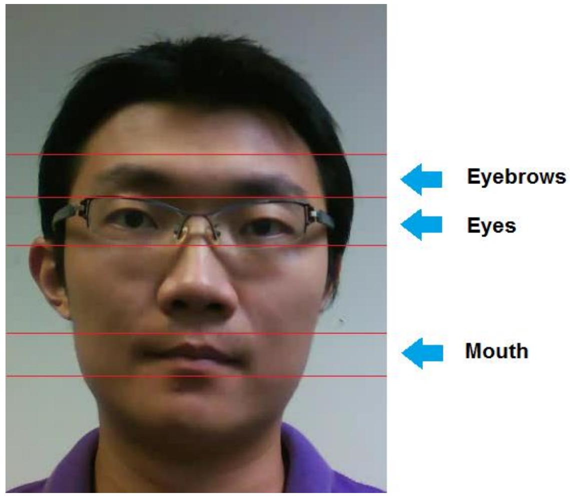

Figure 3.2 Feature extraction from specific face region.

3.1.1 Appearance Feature Extraction

The feature extraction is an important step for training HMMs. They are obtained from 2D-DCT coefficient vectors of facial regions for each frame after vector quantization.

Given a face image frame, the whole regions of eyebrows, eyes, and mouth in rectangles are extracted (see Figure 3.2) and used to form a group of pixel values. Instead of using pixels of each region [69], it is better to represent its features by two-dimension Discrete Cosine Transform (2D-DCT) coefficients in order to exploit spatial correlation.

Employing 2D-DCT coefficients is preferred over using pixel values for various reasons. First, many 2D-DCT coefficients tend to be small and even close to zero. The larger DCT coefficients are generally clustered in the low frequency range representing the most significant features of a face image. Therefore, keeping only large coefficients reduces the dimension of the observation sequences, hence, the implementation cost. In addition, DCT coefficients are less susceptible to image noise than pixels. Its advantage can be intuitively observed in Figure 3.3.

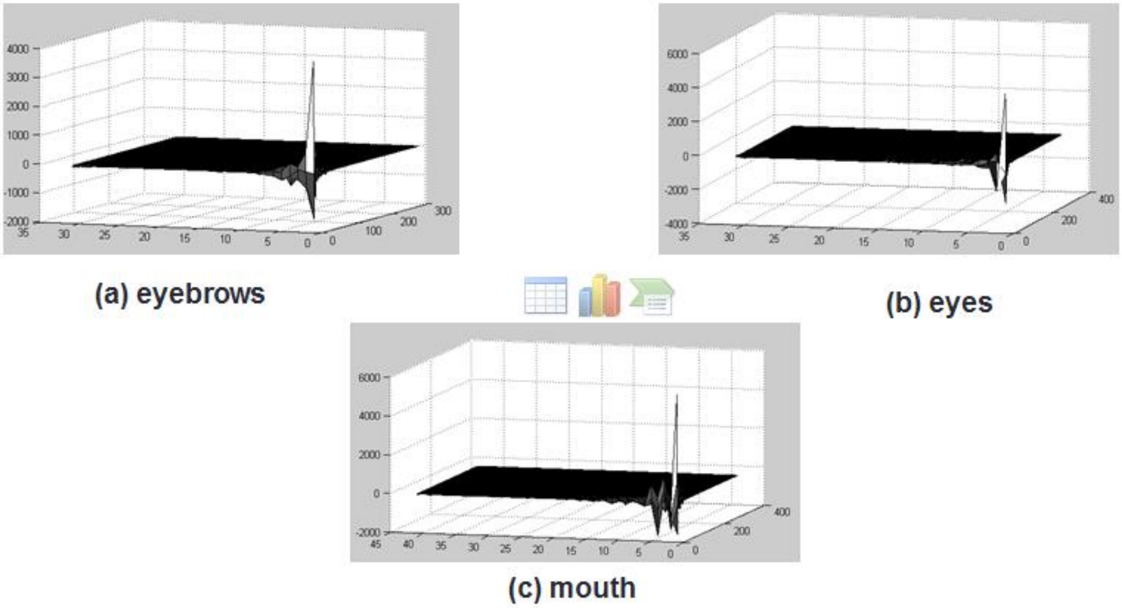

Figure 3.3 8 by 8 2D-DCT coefficients extracted from three face regions. (a) eyebrows, (b) eyes, (c)mouth. X axis represents the width size of the region and Y axis represents the height size of the region.

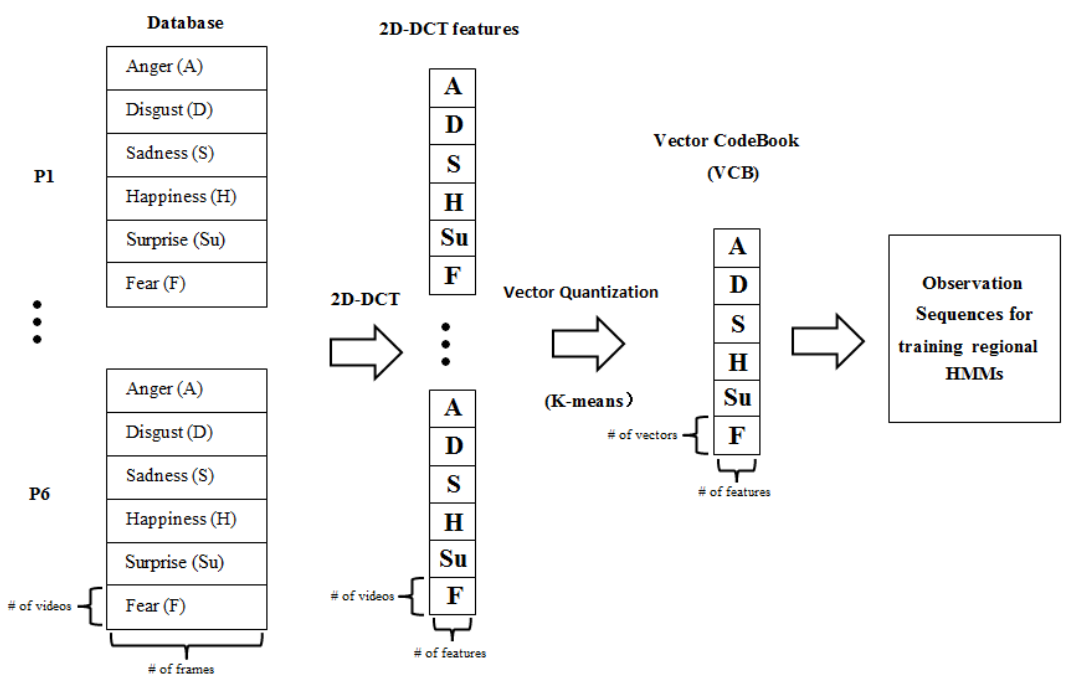

Figure 3.4 Generation of observation sequences.

For the experiments, 2D-DCT coefficient vectors of size 8 by 8, 32 by 32 and 64 by 64 in each region are respectively used as facial features. They are classified by facial regions and then vector quantized using Vector Cookbook (VCB) to form observation sequences as highlighted in Figure 3.4. Observation sequences of each region are classified for the defined states of facial regions in order to train RHMM proposed in Chapter 2**.** The procedure to form observation sequences is the same for each state of all regions. To illustrate this function, the fall state of the eyebrows region is used in this example.

1. Each video is labeled with one type emotion per person consists of N^{v}_{frame} image frames.
2. *n* by *n* 2D-DCT coefficients in the low frequency (the top-left corner) are extracted from the eyebrow region in i^{th} frame of the j^{th} video in the database and then are reshaped as a feature vector:

$$v^{eyebrows}_{i,j} = \left[1,...,n^2\right]^T, \quad i = 1,...,N^{v}_{frame}, \; j = 1,...,N_{video} \tag{3.1}$$

3. All feature vectors are quantized through k-means algorithm [70] to form a *M*-size Vector Codebook $VCB_{eyebrows}$. The size of each vector in the Vector Codebook is 1 by n^2

$$GVB_{eyebrows} = [v_1{}',...,v_M{}']^T \tag{3.2}$$

4. To form the l^th observation sequence for training RHMM of the fall state of

eyebrows, project each feature vector (frame) of a video showing eyebrows fall into $VCB_{eyebrows}$ to generate an observation symbol o_i, $i=1,...,N_{frame}^{v}$ by selecting the closest cluster.

$$O_l = [o_1,...,o_{N_{frame}^{v}}], \quad l=1,...,N_{video}^{fall} \tag{3.3}$$

where N_{frame}^{v} is the number of videos showing eyebrows fall. It contains Anger and Disgust emotion types as defined.

3.1.2 Training RHMM and Recognition

Each state of each face region is represented by a distinct RHMM. A set of observation sequences based on a state of a face region is used to train an *N*-state RHMM as summarized below.

First, initialize a RHMM as $\lambda=(\pi,A,B)$. $\pi=\{\pi_i\}$ is the initial state distribution that contains the probability of the model being in each state at time *t*=1. *where* $\pi_i = P(q_1 = S_i)$. $A=\{a_{ij}\}$ is the state transition probability distribution. $B=\{b_i(k)\}$ is the observation probability distribution represented by the probability density function (pdf) at time *t* given the state *i* of the model. One might choose Gaussian distribution to represent the pdf type.

$$B = b_j(k) \sim N(\mu_j, \Sigma_j), \quad j=1,...,N \tag{3.4}$$

where μ_j is the mean, Σ_j is the covariance. As described in Section II-B, each observation sequence is divided into several parts according to the number of states of a RHMM. For example, there are three states (fall, raise, and neutral) in the eyebrows regions. The first state of the RHMM is related to the first third of an observation sequence, the second state to the second third, and last state is related to the last third of this observation sequence. The observation parts related to one state are used to estimate the parameters of the observation probability distribution.

Second, after setting initial model parameters, the RHMM is trained using the *Baum-Welch algorithm* [67] to re-estimate initial parameters. This step estimates the parameters that maximize the probabilities of observation sequences for the given RHMM. The new model is compared with the old model by computing the distance score. If the distance score is higher than a predefined threshold, the entire training loop is repeated using the new model parameters. Otherwise, the process of training the HMM is stopped and the new model parameters are saved.

In the recognition step, each frame of the test videos goes through the same process in order to form the observation sequences O_i corresponding to different facial regions. *i* is one of three face regions (the mouth region containing mouth and the corners of the lips). The probability of the observation sequence for the related RHMM $P(O_i \mid \lambda_j)$ is calculated using the *forward-backward algorithm* [67], where *j* is the state of a face region. A prototype emotion is expressed by combining states of the all facial regions. The probability of an emotion type P_k is calculated by summing the probabilities of the states of different facial regions according to Table 2.2.

$$P_k = \sum P(O_i \mid \lambda_j) \tag{3.5}$$

where k is one of the six basic emotion types. The face frame is recognized as the facial emotion that yields the highest measured probability, $\max(P_k)$.

3.1.3 System Performance Evaluation

For the performance tests of the proposed facial expression recognition system, a video database of six people with six basic emotions is created (see sample frames in Figure 3.5). The database contains 360 video sequences recorded using a SONY DSC-W530 digital camera. Each one has 100 frames with 640 × 480 pixels resolution and 30 frames per second. For each emotion type, 10 video sequences per person are recorded. Videos are recorded with sufficient lighting in a lab environment. The person is asked to create specified facial expressions for recording. Any video sequence shows only the frontal face of a person. Each sequence is required to begin from the Neutral state to some emotion and finally come back to the Neutral state. It benefits to record emotions and emphasis the motion of facial regions.

Three sets of experiments are conducted. First two experiments are in person dependent case. For each emotion types, five sequences per person are used as the training set, and the rest are utilized for performance testing purposes. First one is on the frontal faces using 2D-DCT and RHMMs. Second one on the face involving in-plane rotation, which is analyzed as follows. In addition, the third experiments for person independent case are also conducted for performance comparisons.

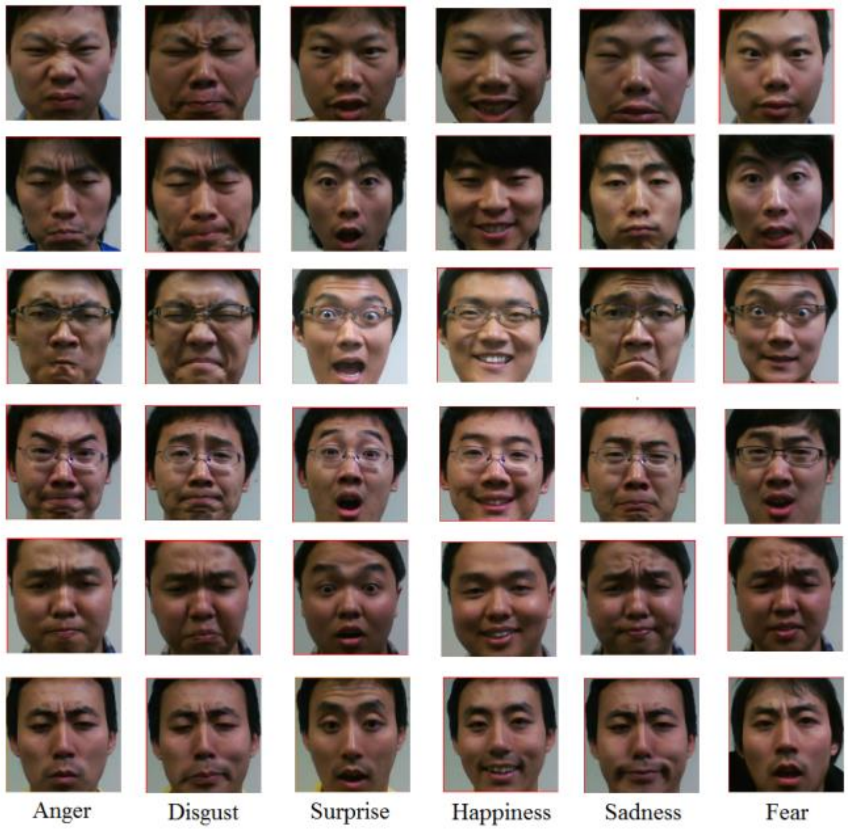

Figure 3.5 Sample images of generated video database.

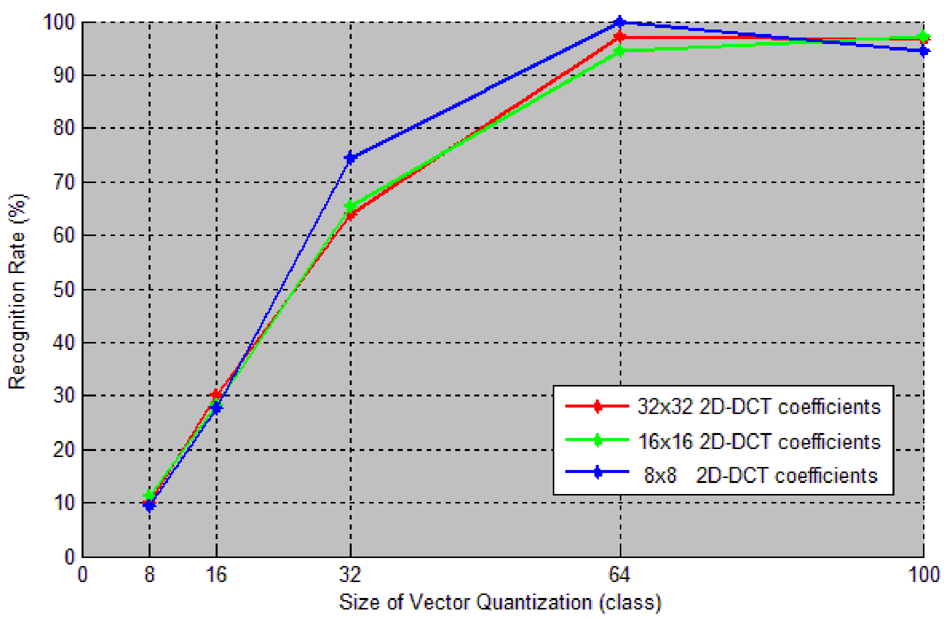

Figure 3.6 Facial emotion recognition performance of the proposed method for person dependent case with various sizes of vector codebook and 2D-DCT.

Table 3.1 Missed Recognition Rates of Proposed Method with Various Emotion Types (Person Dependent Case)

2D-DCT	VCB Size	Anger	Disgust	Surprise	Fear	Happiness	Sadness	Total	Recognition Rate
8x8	8	18	25	30	30	30	30	163	9.44%
	16	5	28	30	29	13	25	130	27.78%
	32	0	18	10	0	5	13	46	74.44%
	64	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**100%**
	100	0	5	0	0	0	5	10	94.44%
16x16	8	15	25	30	30	30	30	160	11.11%
	16	5	30	30	29	15	20	129	28.33%
	32	0	20	15	5	5	17	62	65.56%
	64	0	5	0	0	0	5	10	94.44%
	100	0	0	0	0	0	5	5	97.22%
32x32	8	15	27	30	30	30	30	162	10%
	16	5	26	30	30	10	25	126	30%
	32	0	20	15	10	5	15	65	63.89%
	64	0	0	0	0	0	5	5	97.22%
	100	0	0	0	0	0	6	6	96.67%
Total		**63**	**229**	**220**	**193**	**143**	**231**	**1079**	

Table 3.2 Performance of State of the Art

Type of Classifier	The State of The Art	Performance
Emotion-specific HMMs	S. Das and Koichi [71]	93%
	A. Sánchez et al.[72]	92.42%
	J. Buenaposada et al. [73]	89%
Emotion-specific Support Vector Machine (SVM)	I. Kotsia and I. Pitas [74]	99.7%
	T. Wu et al. [75]	97.81%

A. Performance based on 2D-DCT and RHMM

Frames of videos are extracted to form observation sequences through the calculation of 2D-DCT coefficient vectors after VQ operation. Considering the background in each vide frame, the face is extracted to an array of size 400 × 300 pixels. 8 × 8 DCT coefficients in the low frequency range are used. All DCT vectors are VQ quantized with VCB size of 64.

Performance of up to a 100% recognition rate for person dependent cases is achieved. The test results also show that the system performance is sensitive to the size of VCB as displayed in Figure 3.6. The system performance sharply decreases when VCB size is reduced. The recognition rate is over 94.44% when VCB size is 64 or larger. A similar performance trend is observed for various DCT sizes.

The recognition performance of the proposed system is tabulated in Table 3.1. The performance of the proposed methodology in person-dependent case is quite satisfactory when compared with the performance of the state-of-the-art in the literature shown in Table 3.2.

Note that these results are for person dependent case where test sets share same people with training sets but not videos. They indicate the relationship between recognition rate and states of facial regions. In the last row, missed recognition rates for 6 emotion types are listed. The number of misses in the Anger state is far less than the others. The Sadness state has the most misses although it is close to the Disgust, Surprise and Fear states. The trend of recognition misses may be explained with the help of Table 2.2 where emotions are defined using states of facial regions. Since the pull motion of lip corners is similar to that of the down lip corners, the Sadness, the Disgust and the Fear state are more difficult to classify. The Fear sate is defined almost the same as the Surprise state except for the motion of lip corners. Hence, they both have high recognition misses. Due to the definition of the Happiness state being close to that of the Neutral state, and the entire sequence is required to begin from the Neutral state to some emotion and finally come back to the Neutral state. Its recognition miss is not low. The motion of the lip corners pucker in the Anger state is apparently different from that of lips neutral, pull and down. It can be simply classified even its definition is close to the Disgust state.

B. Comparison of 2D-DCT and KLT

The experiment is conducted based on the eigenanalysis using Karhunen-Loeve Transform (KLT) [68]. The performance of recognition up to 97.78% is achieved when 50 KLT coefficients and 64-size Global Vector Book are used. There is no significant gain in the recognition rate when the size of KLT coefficients is larger than 50. The performance trend is similar to that based on 2D-DCT transform with various sizes of Global Vector Book (see Figure 3.7).

C. Performance Impact of Rotation

One might compensate for face images rotated by in-plane rotation to test performance using the proposed method. Image frames in a facial video sequence are randomly rotated (except the first frame) in 2D plane. Three facial tracking points are marked by hand in the first frame (two on the outer eye corners and one on the nose as the anchor) and tracked in the following frames. Several sample frames of such sequence are displayed in Figure 3.8. The rotation angle of the frame is estimated by calculating the slope of the line passing through two outer eye corners. Then, the corresponding frontal face is formed going through the expression recognition described in section 6.1 using VCB size of 64. Figure 3.9 displays performance comparison for rotated faces of ±90, ±60, ±30, ±1 and 0 degrees with and without rotation estimator prior to recognition. This observation highlights the good performance of the proposed method after rotation compensation.

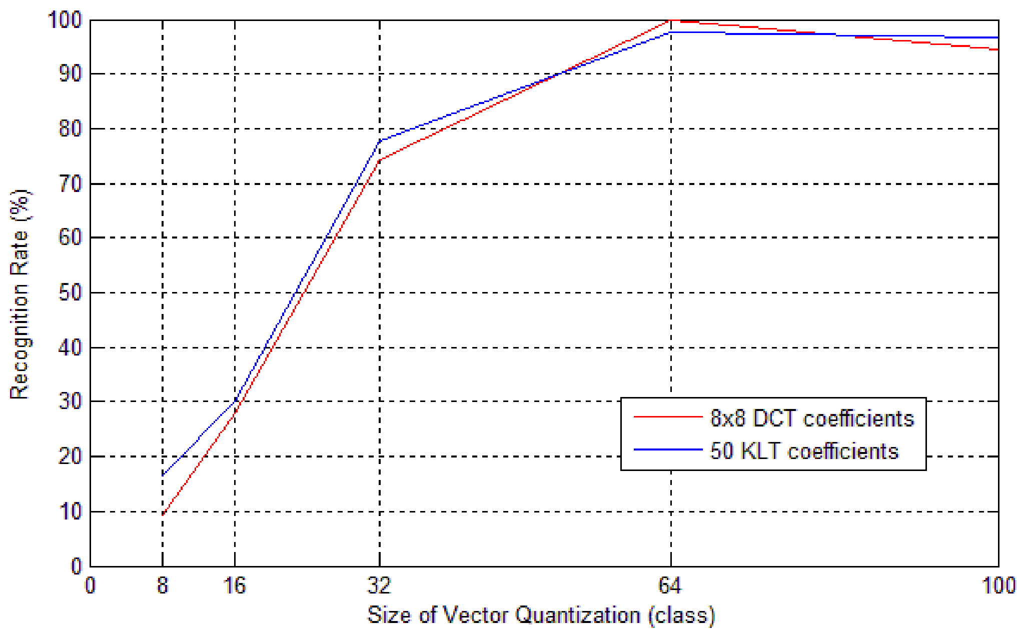

Figure 3.7 Performance comparison using 2D-DCT features and KLT ones.

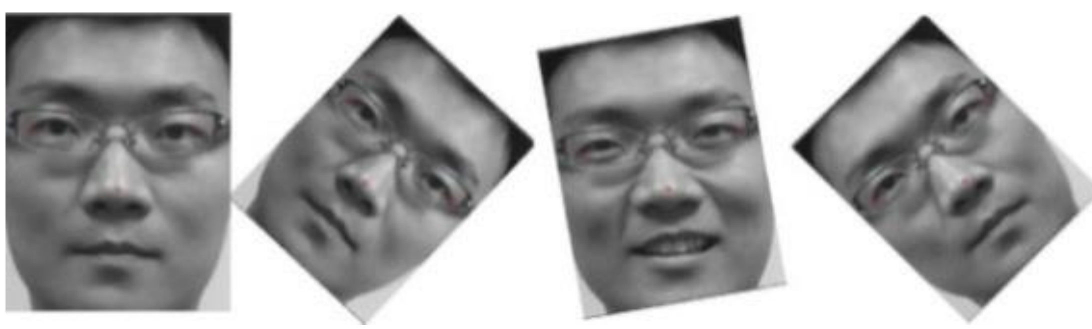

Figure 3.8 Sample image frames of in-plane rotated faces along with three tracking points.

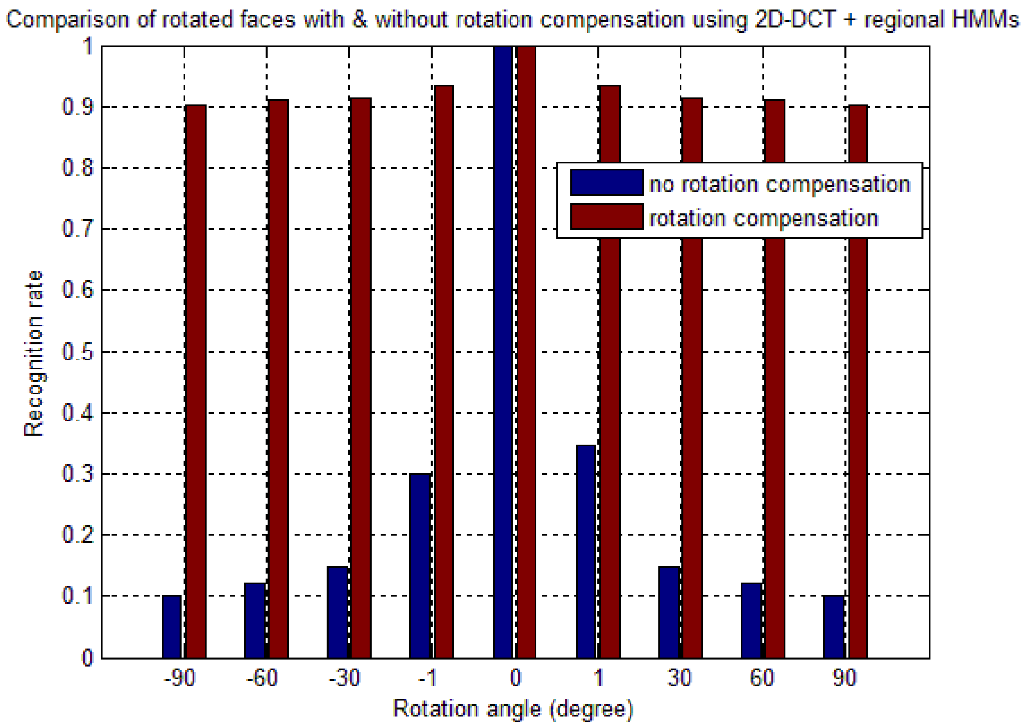

Figure 3.9 Performance comparison for in-plane rotated faces of ±90, ±60, ±30, ±1 and 0 degrees with and without rotation estimator prior to recognition.

Table 3.3 Missed Recognition Rates of Facial Emotion Types for Person Independent Case

2D-DCT	**VCB size**	**A**	**D**	**SU**	**F**	**H**	**SA**	**Recognition Rate**
8x8	64	18	30	12	30	13	30	26.11%

D. Person Independent Case

For the person independent case, all facial emotion videos of three people to train the system are selected. Then, the performance of the system is tested for three people who were not involved in the training step. All the videos used in this experiment are the same as those used in prior experiments. The resulting performance calculations are tabulated in

Table 3.3. The abbreviations A, D, SU, F, H, and SA represent Anger, Disgust, Surprise, Fear, Happiness, and Sadness, respectively. The performance for the person independent case is not satisfactory at this point. The main reason for such drastic performance degradation is due to the fact that spectral characteristics are utilized like DCT coefficient vectors as observables. They depend on personal facial features where spatial information is lost. The spatial signatures utilizing for the person independent facial emotion recognition system will be discussed in the next section.

3.2 System Based on Geometric Technique Using RHMM

3.2.1 Geometric Features Extraction

To describe the states of face regions, 41 facial feature points are identified on each frame of video, as shown in Figure 3.10. They are comprised of 10 points on the eyebrows region, 12 points on the eyelids, 8 points on the mouth, 10 points on the corners of the lips, and one anchor feature point on the nose. The 2D coordinates of facial feature points in various face regions are extracted to form corresponding observation sequences for classification. The facial feature points are tracked using a constrained local model [76]. It utilizes a non-parametric method to represent the distribution of candidate locations by using an optimization method called constrained mean-shifts. It outperforms the other methods for deformable model fitting [77, 78, 79].

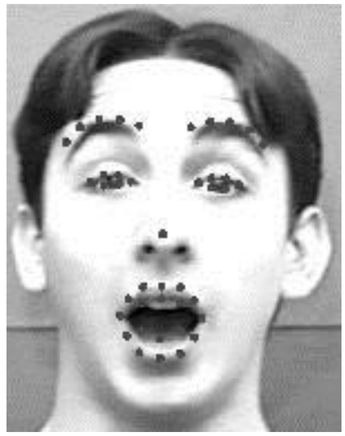

Figure 3.10 The facial feature points on a face image.

Source: T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," IEEE International Conference on Automatic Face and Gesture Recognition, pp. 46-53, Mar. 2000.

3.2.2 Observation Sequence Generation

During the training phase, 14 different RHMMs, labeled λ_1 to λ_{14} are generated, each representing one of the 14 observable states of the three facial regions. The formation of the observation sequences is the first step in the process of training the RHMMs. The accuracy and validity of the observation sequences determine the resulting classification and indirectly dictate the recognition performance. In the proposed method, the observation sequences are formed using the coordinates of the facial feature points in a 2D representation.

Observation sequences of each region are classified for the defined states in order to train RHMMs as proposed in Chapter 2. The procedure to form observation sequences is the same for each state of all regions. To illustrate this function, the fall state of the eyebrows region is used in this example.

First, calculate the relative distance $d_i, \quad i = 1, ..., N_{eyebrows}$ between each feature point in the eyebrows region and the anchor point on the nose for each frame of a video in the database. $N_{eyebrows}$ the number of facial feature points in the eyebrows region.

Second, then, form

$$\begin{aligned} S_{eyebrows,n} &= \{d_1, d_i, ..., d_{N_{eyebrows}}\} \\ S^{fall}_{eyebrows,m} &= \{d_1, d_i, ..., d_{N_{eyebrows}}\} \end{aligned}, \quad n = 1, ..., N_{frame}, m = 1, ..., N^{fall}_{frame} \tag{3.6}$$

where $S_{eyebrows,n}$ represents a set of relative distances in the eyebrows region of the n^{th} frame in the database N_{frame}. $S^{fall}_{eyebrows,m}$ represents another set of relative distances in the fall state of eyebrows region for the frame of N^{fall}_{frame}. N^{fall}_{frame} is the number of frames showing the falling eyebrows. Both the Anger and Disgust emotion types can be indicated by falling eyebrows.

Third, project $S^{fall}_{eyebrows,m}$ onto $S_{eyebrows,n}$ to measure the distance ε_j, $j = 1, ..., N_{frame}$. In some cases, ε_j may be small, even zero if $S^{fall}_{eyebrows,m}$ is projected onto the same state of the eyebrows region.

Fourth, the feature vector for the m^{th} frame of N^{fall}_{frame} is formed as

$$\begin{aligned} S_{eyebrows,n} &= \{d_1, d_i, ..., d_{N_{eyebrows}}\} \\ S^{fall}_{eyebrows,m} &= \{d_1, d_i, ..., d_{N_{eyebrows}}\} \end{aligned}, \quad n = 1, ..., N_{frame}, m = 1, ..., N^{fall}_{frame} \tag{3.7}$$

Fifth, group the fall feature vectors to obtain the training matrix for eyebrows fall as

$$Tr^{fall}_{eyebrows} = \begin{bmatrix} V^{fall}_{eyebrows,1} \\ \vdots \\ V^{fall}_{eyebrows,m} \\ \vdots \\ V^{fall}_{eyebrows,N^{fall}_{frame}} \end{bmatrix} \tag{3.8}$$

It should be noted that the images in the database are classified in a unique order to train RHMMs for different facial regions. To train RHMMs for the eyebrows region, images are classified following its states as fall, raise, and neutral. In the same way, the eyes, moth, and lip corners regions are classified based on the states shown in Table 2.2. The advantage of this manipulation is to always form the observation sequences as unique ordered sequences, no matter what emotion type is shown in the frame of a video. Thus, the proposed system adapts to process various form of videos and avoids defining specific forms of videos, such as starting from the neutral emotion to the peak and then fading back to the neutral emotion. In addition, it brings benefits to the model selections. Selecting the right model for the observation sequences is crucial for successful classification.

One type of model, called the ergodic model, allows any of its states to reach all other states. Therefore, this type of model has more degrees of freedom but requires more parameters. The other model type is the left-to-right model. It either stays in its current state or goes directly to the next state, and does not rely on the previous states. It not only uses few parameters for training but also models a natural sequential event always starting and ending in two fixed states. Given the specific form of observation sequences in the framework, the left-to-right model is regarded as the optimal model for training and recognition RHMM. Although Cohen et al. [30] indicated that the ergodic model would reduce to the left-to-right model with infinite training data, it is very difficult to obtain infinite images for training in practice. Furthermore, several efforts [41, 42] have successfully applied the left-to-right model for facial emotion recognition. The structure of the RHMMs is shown in Figure 3.11, where $S_i, \quad i = 1, \ldots, N$ represents the state of a RHMM and N is the number of states. Due to the observation sequences formed in the unique order for each state of all face regions, the observation probability distribution of each RHMM is unique.

3.2.3 System Performance Evaluation

A. Setup

To test the performance of the proposed facial emotion recognition system, the widely referenced Cohn-Kanade database [80] is used to recognize the six universal facial emotions listed in Table II. Each sequence begins from a neutral face and then reaches the peak of an emotion type. An emotion is labeled on a video clip based on the last frame.

Experimental results only provide the average recognition rate of all facial emotions, and do not reflect the result of each individual emotion. To handle this problem, confusion matrices [81] are employed in the experiments. In the confusion matrix, the emotion types detected by the system are shown in the columns and the actual emotion types are listed in the rows. A, D, F, H, SA, and SU represent Anger, Disgust, Fear, Happiness, Sadness, and Surprise, respectively. The accuracy of each emotion type is shown in the confusion matrix entries where the detected and actual emotion types are the same.

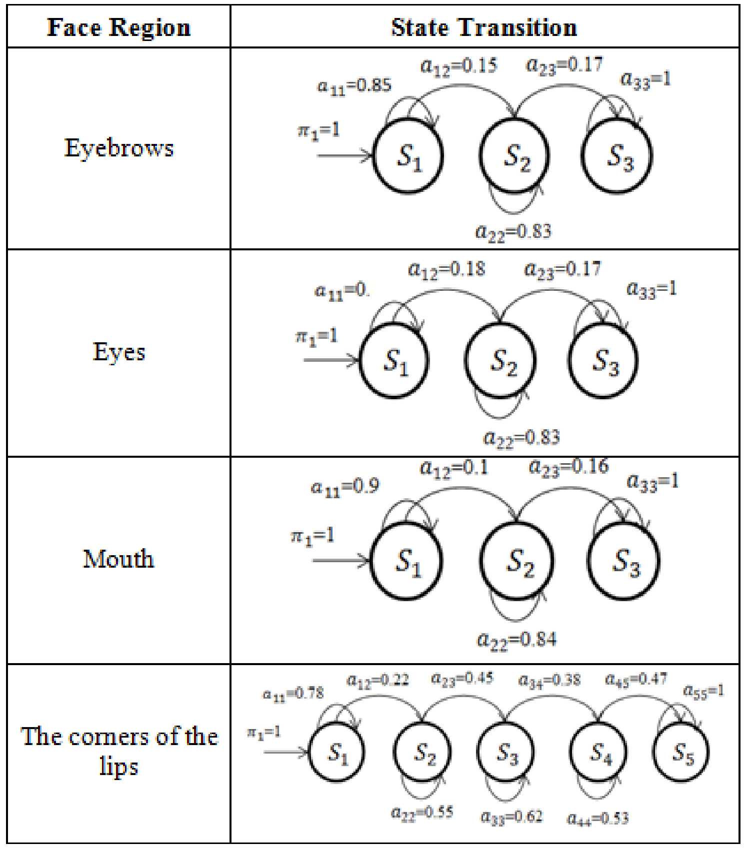

Figure 3.11 The structure of RHMMs for different face regions.

B. Comparison with the Approach using Emotion-specific HMMs in the Person-dependent Case

Two sets of experiments are performed for the person-dependent case. In the first set, the proposed RHMMs are trained for the states of the facial regions. 180 video sequences and 30 different subjects from the Cohn-Kanade database are used such that one sequence per subject per universal emotion type is included. Half the video sequences from each of 30 subjects are utilized for training, and the rest are used for testing. 5-fold cross-validation tests are run on the entire dataset. Next, another experiment is conducted for a single HMM representing an emotion type for an entire frame of an image sequence. Facial features in all regions are combined as observation sequences. Seven different HMMs are derived, each representing a specific emotion type, the six basic types and Neutral.

The recognition performances of person-dependent cases for the proposed RHMM and emotion-specific HMM systems are tabulated in Tables 3.4 and 3.5, respectively. The former performs 2.6% better than the latter tested on the same database. The recognition rates for each emotion type are shown in boldface. The other numbers in a row represent the probabilities of erroneous detection along with the corresponding incorrect emotion type. The recognition performance is based on the last frame of a sequence due to the nature of the database. As the tables show, the emotion types Surprise, Happiness and Neutral achieve the highest recognition rates for both systems. Sadness has the lowest recognition rate and is most often confused with Neutral, Anger and Disgust. The main reason for this observation is that the motion of the eyes and mouth regions in Sadness are close to those for Anger and Disgust. The motions of the corners of the lips among Neutral, Disgust and Sadness are also similar. In addition, the performance of the proposed system is also quite satisfactory when compared with that of the current state-of-the-art facial emotion recognition systems for Cohn-Kanade database. More specifically, in [71, 72, 73], the reported recognition rates are 93%, 92.42%, and 89%, respectively. Some existing algorithms performed well for the Cohn-Kanade database using Support Vector Machine (SVM) as a classifier, but the features are just extracted from the head and tail of an image sequence instead of the entire sequence [74, 75]. Thus, the results cannot reflect actual facial variations, since the facial expressions are dynamic and continuous and include the relative weak expressions in the transition. Experiments conducted in [75] demonstrate that the recognition rates sharply decrease when considering the early stage of emotional change with flat facial information using SVM.

Table 3.4 Recognition Performance (%) of the Proposed Method Using RHMMs in Person-dependent Case (Average Recognition Rate is 94.28%)

	A	**D**	**F**	**H**	**N**	**SA**	**SU**
A	**90**	6.67	0	0	0	6.67	0
D	3.33	**86.66**	0	0	0	10	0
F	0	0	**96.67**	3.33	0	0	0
H	0	0	0	**100**	0	0	0
N	0	0	0	0	**100**	0	0
SA	0	6.67	3.33	0	3.33	**86.66**	0
SU	0	0	0	0	0	0	**100**

Table 3.5 Recognition Performance (%) of Single HMM per Emotion Type in Person-dependent Case (Average Recognition Rate is 91.90%)

	A	**D**	**F**	**H**	**N**	**SA**	**SU**
A	**83.33**	6.67	3.33	0	0	6.67	0
D	3.33	**83.33**	0	3.33	0	10	0
F	0	0	**96.67**	3.33	0	0	0
H	0	0	0	**100**	0	0	0
N	0	0	0	0	**100**	0	0
SA	3.33	10	3.33	0	3.33	**80**	0
SU	0	0	0	0	0	0	**100**

Figure 3.12 displays the recognition rates of emotion types as a function of frames (time) in a video sequence. Seven facial emotions (including Neutral) in a frame are represented as probabilities, all of which are normalized to one. This representation offers to identify the dominant facial emotion in each frame of a video.

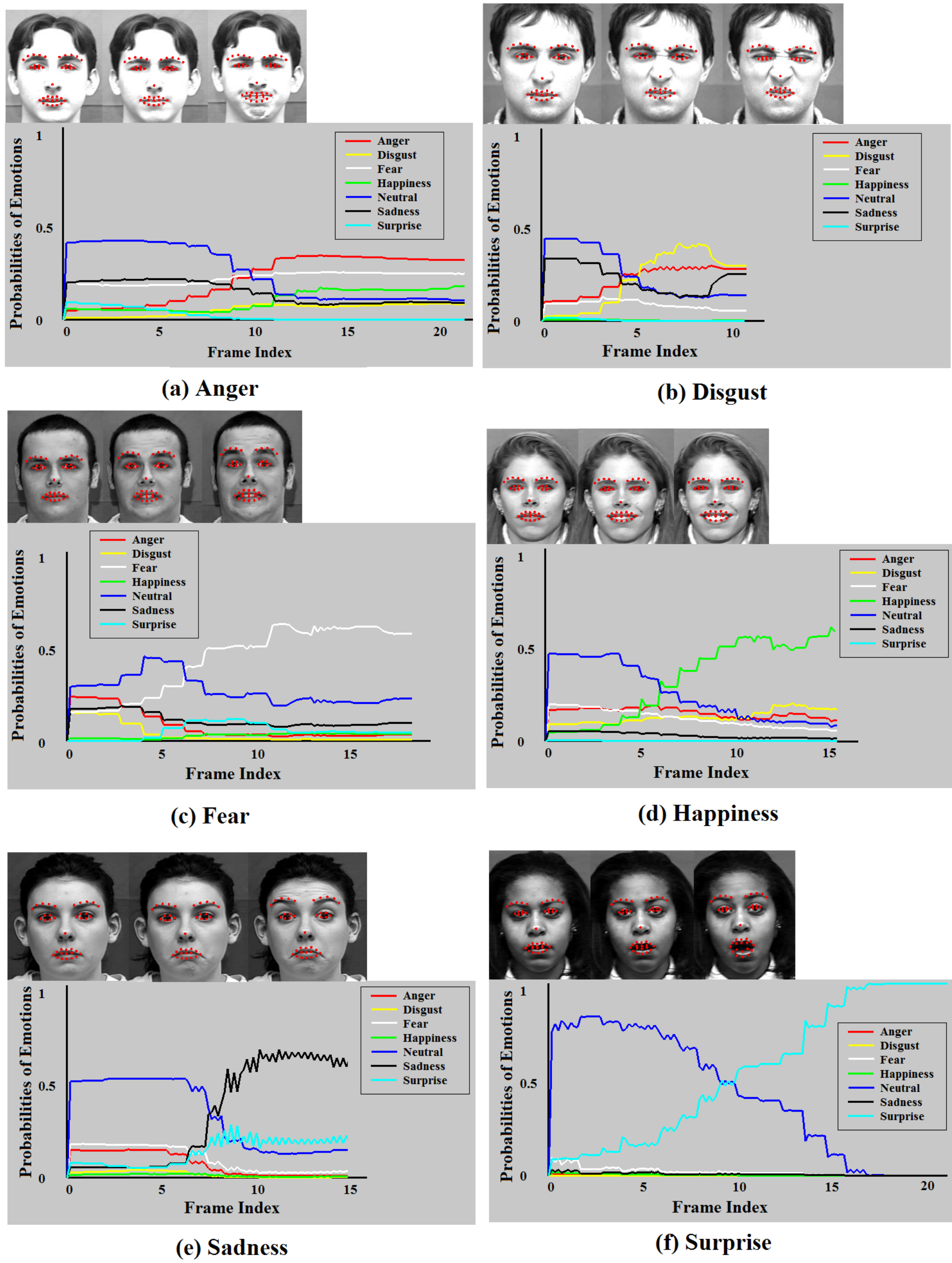

Figure 3.12 Recognition rates of emotion types as a function of frame index in a video sequence.

Table 3.6 Recognition Performance (%) of the Proposed Method Using RHMMs in Person-independent Case (Average Recognition Rate is 86.67%)

	A	**D**	**F**	**H**	**N**	**SA**	**SU**
A	**76.67**	10	3.33	0	0	10	0
D	3.33	**83.33**	0	3.33	0	10	0
F	0	0	**86.66**	3.33	0	3.33	6.67
H	0	0	3.33	**96.67**	0	0	0
N	0	0	0	0	**100**	0	0
SA	3.33	16.67	10	0	6.67	**63.33**	0
SU	0	0	0	0	0	0	**100**

Table 3.7 Recognition Performance (%) of Single HMM per Emotion Type in Person-independent Case (Average Recognition Rate is 84.28%)

	A	**D**	**F**	**H**	**N**	**SA**	**SU**
A	**73.33**	10	3.33	0	0	13.34	0
D	6.67	**80**	0	3.33	0	13.34	0
F	0	0	**83.33**	6.67	0	3.33	6.67
H	0	0	3.33	**93.33**	0	0	3.33
N	0	0	0	0	**100**	0	0
SA	3.33	23.33	6.67	0	6.67	**60**	0
SU	0	0	0	0	0	0	**100**

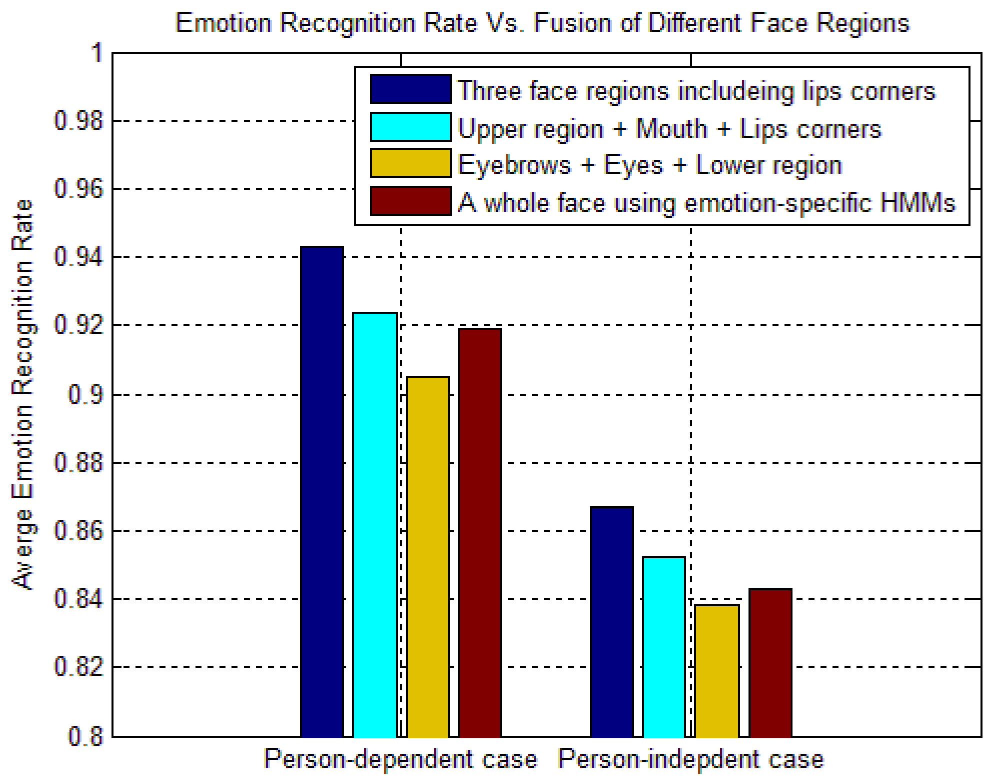

Figure 3.13 The performance of fusion of face regions.

C. Comparison with the Approach using Emotion-specific HMMs in Person-independent Case

To test the performance of the proposed method for new subjects, two sets of person-independent experiments are also conducted, in which subjects used in the training set do not appear in the testing set. 24 out of the 30 subjects used in the person-dependent case are utilized for training, with the remaining 6 subjects used for performance tests. Two sets of experiments are also run using 5-fold cross validation. One is to classify facial features using the proposed RHMM framework for the states of face regions. The other is to classify the combination of local features using the emotion-specific HMM for a single frame of a video sequence.

The recognition results of facial emotions for the person-independent case are listed in Tables 3.6 and Table 3.7. As would be expected, the performance in this case is lower than that of the person-dependent case. Nevertheless, it is still satisfactory when compared with the state-of-the-art performance in person-independent cases. In [79], [82], the recognition rates are 62.6% and 73.22%, respectively. The best reported recognition rate so far [83] is 88.9%.

To further compare RHMM and emotion-specific HMM, additional experiments are performed to see the importance of face regions to the facial emotion recognition. First, to merely highlight the mouth region, the states of the mouth and lips corners regions are modeled separately as RHMMs, but the combined states of the upper region containing eyebrows and eyes are trained. The states of the upper region are defined based on Table 2.2 as Eyebrows fall plus Eyes close, Eyebrows raise plus Eyes close, Eyebrows neutral plus Eyes neutral, and Eyebrows raise plus Eyes open. Next, the states of the eyebrows and eyes regions are merely emphasized by modeling the states of the overall lower region including the mouth and the corners of the lips. The corresponding states are mouth close plus lips corners pucker, mouth close plus lips corners pull, mouth close plus lips corners down, mouth open plus lips corner up, mouth open plus lips corners neutral, and mouth open plus lips corners pull. The experimental results in Figure 3.13 show that the performance of the proposed method is better than any method that combines facial regions. The recognition rate declines with the fusion of states of different face regions. The lower face region conveys more significant expressions than the upper region.

3.3 Chapter Summary

The appearance-based and the geometric based facial emotion recognition systems are proposed. Both show the promising performance in the person-dependent case comparing to the start of the art. The appearance-based features for the person independent case are not satisfactory. The main reason for such drastic performance degradation is due to the fact that spectral characteristics are utilized like DCT coefficient vectors as observables where spatial information is lost. The geometric-based features exhibit the strength at this point.

CHAPTER 4

AUTOMATIC INFERENCE OF MENTAL STATES FROM SPONTANEOUS FACIAL EXPRESSIONS

The high recognition accuracy of facial emotions has been achieved in many existing research, based on the benchmarked databases which contain posed facial emotions, as mentioned above. That is, the emotions shown in still images or image sequences are elicited for research purpose by asking subjects to perform a series of exaggerate emotional expressions. However, they are not qualified to interpret people's true feelings even if they are recognized. As a survey of affect recognition methods indicated in [8], the difference between posed facial emotions and spontaneous ones has been found in [9, 10]. The purpose of recognition of facial emotions not just limited to recognize the per-defined emotions. Instead, the application of technologies to the real life is what people are pursuing.

One of the interesting challenges in the community of human-computer interaction is how to make computers be more human-like for intelligent user interfaces. Human face plays a significant role in conveying mental states [2]. Casual expressions in the face at one moment due to some stimuli imply mental states even if people sometimes try to control their outer expressions. To make the information of facial expressions available for application on Human Computer Interface (HCI), several efforts have been made on automatic analysis of spontaneous facial expressions in [4, 5], not merely confined to pre-defined emotions. However, litter empirical evidence exists that the computer intelligently involves in the deep exploration of people's mental states behind the authentic facial emotions.

In this chapter, the capability of the geometric-based system proposed in Section 3.2 is explored to infer people's mental states despite derived from spontaneous facial expressions. Furthermore, the system is compared with the human to see whether it is able to interact naturally with the user, similar to the way human-human interaction takes place.

4.1 Human Mental State

A person's face is the window of the person's underlying mental state, that is, an internal cognitive states reacting to the external reality. The present experiments examine whether the proposed system has capability to automatically infer a person's mental state (e.g. positive affect and negative affect) based on spontaneous facial expressions. Affect [84] referred to the emotion elicited by stimuli and objectively reflects a person's mental state. Affect has been generally classified in the field of psychology using positive and negative dimensions [85]. Based on the basic facial emotion types, Happiness is coded as the positive affect and Anger, Disgust, Fear, and Sadness as the negative affect. In addition, recent studies in psychology field [3] empirically revealed that people reliably infer others' preferences from spontaneous facial expressions. To assess the plausibility of the proposed system, its performance results are compared with the human's inference of other people's mental states by analyzing the same datasets.

4.2 Experiment Protocol

The experiment consists of two phases. In the first phase, target video clips are collected which show participants' frontal-view facial emotions only. Target participants instinctively convey their facial information elicited by stimuli from real-life activities, such as conversations and events. In the subsequent phase, both the human perceivers and the proposed system separately judge target participants' affects by watching these video clips. Finally, the performance of the system is compared with the perceivers' data to determine the ability of the system to determine the affect being displayed.

4.2.1 First Phase for Targets

In this experiment, 24 professional actors (mean age=29.167, SD=9.907, 11 female) are selected as participants to express five prototype emotions, namely, Anger, Disgust, Fear, Happiness, and Sadness. Since Surprise can be revealed as a result of positive affect or negative affect, it is not used in this experiment. Their facial emotions are extracted from video clips of the TV series in which the actors perform. In this novel paradigm, subjects express their facial emotions triggered by real environment or events without bias. A total of 120 video clips with resolution 640x352 are extracted, each of which starts and ends with a single emotion per person. Each video has the duration of 1 to 3 seconds. In the video clips, the auditory information is filtered and the motions of the mouth due to speaking or chewing are isolated as well. Only motions of the mouth that express facial emotions are used. It thereby avoids unnecessarily disturbing the judgments of both the perceivers and the system.

4.2.2 Second Phase for Perceivers and System

15 graduate students (mean age=24.933, SD=5.483, 1 female) served as perceivers. Both the perceivers and the system are assigned to watch 120 videos clips. Since the lengths of

video clips are different, a video clip is replayed until a perceiver recognizes the type of affect. Once they have finished watching a video clip, the perceivers evaluate the type of affect (positive or negative) and the degree of the affect using a 21-point scale ranging from (-10=extremely negative, +10=extremely positive). Analogously, the system provides its affect rating in the range between -1 to +1. To calculate the similarity between ratings of the system and the perceivers, the perceivers' ratings are scaled between -1 to +1. Both the perceivers and the system do not know the emotion (affect) labeled on the video clips. The advantage of this paradigm is that they naturally give the final ratings in the absence of prior knowledge.

4.3 Evaluation of the Collected Video Clips

It cannot ensure that perceivers remain consistent in terms of the emotion types labeled in the video clips. To deal with this problem all perceivers are asked to independently evaluate all video clips after watching them. Each video is classified as an explicit video if the number of perceivers inferring the same facial affect as that labeled on the video clip is over 50% of all perceivers; otherwise, it is classified as an ambiguous video. A total of 13 of the 120 videos are classified as ambiguous videos. The remaining 107 explicit video clips are utilized for the experiments. Some samples are shown in Figure 4.1.

4.4 Human Mental Inference

4.4.1 Inference of System

Inference of targets' mental states by the system is based on facial emotion recognition. Since all collected video clips only show one type of emotion, emotions are recognized by

Figure 4.1 Samples of spontaneous facial expressions.
Source: http://youtube.com

the system by calculating the largest probability $P_v^{Emotion}$ of the emotion type among Anger, Disgust, Fear, Happiness, Neutral, Sadness, and Surprise as shown in Equation 4.1,

$$P_v^{Emotion} = \max(P_i), \quad v = 1,...,N_{video}, \; i = 1,..,7 \tag{4.1}$$

where N_{video} is the total number of videos. P_i is the probability of one emotion, which is obtained by summing the overall probabilities P_{ij} of this emotion in each frame of a video clip. P_{ij} is normalized to one. P_i is calculated as

$$P_i = \sum_{j=1}^{N_{frame}} p_{ij}, \;\; subject\ to \; \sum_{i=1}^{7} P_{ij} = 1 \tag{4.2}$$

where N_{frame} is the number of fames of a video.

The ratings of both the positive affect P_{pos} and the negative affect P_{neg} for a video are separately generated by the system by normalizing the probability of Happiness and the largest probability among Anger, Disgust, Fear and Sadness. The larger affect rating

P_v^{Affect} of the two affect values is selected as the affect recognized in the corresponding video clip.

$$P_v^{Affect} = \max(P_{pos},\ P_{neg}),\ v = 1,\ldots,N_{video}, \tag{4.3}$$

$$\text{subject to}\begin{cases} P_{pos} = \dfrac{P_{Happiness}}{P_{Happiness} + \max(P_{Anger}, P_{Disgust}, P_{Fear}, P_{Sadness})} \\ P_{neg} = \dfrac{\max(P_{Anger}, P_{Disgust}, P_{Fear}, P_{Sadness})}{P_{Happiness} + \max(P_{Anger}, P_{Disgust}, P_{Fear}, P_{Sadness})} \end{cases}.$$

If P_v^{Affect} is a positive affect, the rating is a positive value. Otherwise, the system generates a negative rating.

Accuracy is measured by the percentage of targets' emotions and affects judged by the system to be the same as that labeled on a video clip. The emotion recognition rate of the system is 76.64% and the affect recognition rate reaches a higher value of 85.98%. This is due to the fact that some emotion types are misrecognized as others but are nevertheless affiliated to the same affect type. For example, although Sadness is mistakenly recognized as Disgust, the affect is still classified as a negative one. The results indicate that the proposed system performs well to infer a person's mental state from their spontaneous facial expressions.

4.4.2 Comparison Between System and Human Perceiver

The work is determined if the proposed system shows human-like behaviors during interaction with the users. To do this, the affect recognition results of the automated system are compared with those of the human perceivers. First, the accuracy of the perceivers' inference is measured by the percentage of the targets' affect recognized by perceivers that match that labeled on a video clip. The results of each perceiver's inference are shown in Table 4.1. P_i, $i = 1,...,15$ represents the perceiver. Accuracy of all perceivers' affect judgments is 90.47%.

Next, the degree of similarity between ratings of the system and that of the perceivers is estimated. Similarity scores are calculated by the percentage of the affect recognitions of the system consistent with the perceivers' affect judgments within all videos. For this purpose, the one-sample t-test is used that assesses whether the sample mean is statistically different from the population mean. The similarity score for positive affect (mean=97.39%, SD=5.14, t(14)=35.68, p<0.001) is higher than that for negative affect (mean=76.90%, SD=6.84, t(14)=15.24, p<0.001), where is the probability of obtaining a test statistic close to the one that is observed, given that the null hypothesis is true. It is in line with the fact that the system is more sensitive to Happiness emotion. The similarity score is also calculated for overall affects (mean=81.31%, SD=5.22, t(14)=23.24, p<0.001). The results of one-sample t-test show that similarity scores of inferences between the system and perceivers are significantly above 50% for the two types of affects. That is, the proposed system is able to act as the human to infer people's mental states from their spontaneous facial expressions estimated from a video sequence. To examine the reliability of the results, it is examined whether some potential factors may result in the plausible performance in the following sections.

Table 4.1 Diffusion Coefficients and Molecular Diameters of Non-electrolytes Accuracy (%) of Human Perceivers' Inferences

P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**
86.92	96.26	96.26	87.85	92.52	94.39	81.31	82.24
P9	**P10**	**P11**	**P12**	**P13**	**P14**	**P15**	
92.52	92.52	91.59	79.44	94.39	94.39	94.39	

4.4.3 The Impact of Targets' Expressiveness and Perceivers' Judgments

First, one might speculate whether some targets that are particularly expressive influence the similarity scores is speculated. For example, the readability of some targets' facial expressions being stronger than that of others is more likely to cause perceivers and the system to easily judge their mental states. In that case, the similarity scores might be higher. To do this, 24 tests are performed, iteratively removing each target. The result shows that all similarity scores are significantly high (p<0.001).

Next, one might see if some perceivers' specific ratings for targets' facial expressions have an impact on the similarity scores. Likewise, 15 tests are performed iteratively removing each perceiver. The result shows that all similarity scores are significantly high (p<0.001).

In summary, neither the expressiveness of all targets or perceivers' judgments tends to impact the recognition rates of affects.

4.4.4 Influence of Targets' Spontaneous Facial Expressions

Apart from the personal factors, the degree of the targets' facial affects for different stimuli should also be considered. To test whether the degree of targets' facial expressions in the video clips might impact the similarity scores, the mean and standard deviation of the absolute values of all perceivers' ratings are calculated which are scaled as 0.46 out of 1 and 0.069, respectively. The mean and standard deviation of the absolute values of the system's ratings are 0.5027 out of 1 and 0.0612 respectively. Through further exploration, one might find that the expressions within five categories are fairly close: perceivers (M_{anger} =4.03, $M_{disgust}$ =4.64, M_{fear} =4.30, $M_{happiness}$ =5.8, $M_{sadness}$ =4.40) and the system (M_{anger} =0.40, $M_{disgust}$ =0.48, M_{fear} =0.52, $M_{happiness}$ =0.53, $M_{sadness}$ =0.55). The results indicate that the facial expressions in the video clips tend to be moderate instead of extreme.

In addition, one might also want to see whether some targets' spontaneous facial expressions predict similarity of inference between the system and perceivers. For each emotion type, the correlation between (a) recognition rates of the system toward each target and (b) similarity scores toward each target is calculated. Fisher's z transform is used to evaluate for sample correlations in the range of -1 and +1 to a near Gaussian population. Fisher's *z* transformation applied to the sample correlation coefficient is utilized to examine the value of the population correlation coefficients between two sets of variables. A positive correlation coefficient indicates that affect ratings provided by the system predict the similarity score, whereas a negative correlation coefficient indicates the opposite. A value of zero explains that affect ratings provided by the system do not in any way correlate to the similarity score. The correlations for each emotion type are close to zero (M_{anger} =-0.23, $M_{disgust}$ =-0.15, M_{fear} =0.02, $M_{happiness}$ =-0.18, $M_{sadness}$ =0.08, *Mean*=-0.10, *SD*=0.16, $t(4)$=1.38, p>0.1). The results explain that the correlations for the emotion types are not significantly higher or lower than zero. It explains that the human-like ability of the system to infer people's mental states is quite independent of the recognized spontaneous emotion types.

4.5 Chapter Summary

To explore the human’s mental states, the experiments are conducted on people’s spontaneous facial expressions using the proposed system. The results plausibly show that the system is able to infer people’s mental states despite derived from spontaneous, moderate-emotional facial information. In the further studies, the system with the human is compared to see whether the system is able to interact naturally with the user, similar to the way human-human interaction takes place. The results indicate that the system performs above chance to match the human’s inferences. Thereby, it demonstrates the ability of the system to imitate the human to examine people’s internal feeling through actual interaction.

CHAPTER 5

ANALYSIS OF CEO FACIAL EXPRESSION AND FIRM PERFORMANCE FORECASTING

Facial expressions have been regarded as one of the most efficient and instinctive way to telegraph feelings and intentions in the form of nonverbal communication [86]. Moreover, the psychology literature [3] empirically demonstrates that people can infer others' internal states through spontaneous facial expressions. The CEO plays a vital role in the firm and provides important facial information during the interview, which links to the firm's current market situation and future planning. This information permits shareholders to make better informed investment decisions.

In this chapter, the hypothesis is proposed that CEO facial expressions have effect on firm's market value. Information conveyed by the six universal facial emotions common to all human beings is brought attention. They are Anger, Disgust, Fear, Happiness, Sadness, and Surprise. Psychologist Ekman [87] provided evidence that these six facial emotions were universal to all cultures, population groups, age groups, and genders. This is an important finding because most modern companies are governed by a wide and diverse group of people. Thus, a system that is able to measure these emotions is of interest by the investors and provides information which does not need to be conditioned on culture, age, or gender.

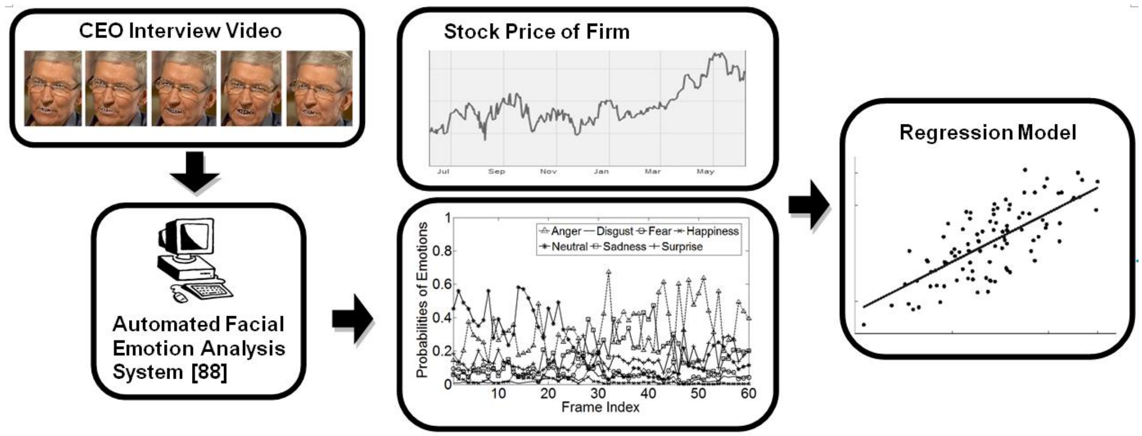

Figure 5.1 Framework to forecast the market value of firm based on CEO's measured facial emotions.

High rates of recognition accuracy of facial emotions by intelligent machines have been achieved in existing technological efforts based on the benchmarked databases containing posed facial emotions. However, they are not qualified to interpret the human's true feelings. A survey of affect recognition methods [8] indicated the difference between posed facial emotions and spontaneous ones**.** The automatic facial expression recognition system recently proposed in [88] demonstrated a capability to classify the spontaneous facial emotions. An important merit of the system, to infer people's mental states, was validated with human based evaluations. The system is used in this chapter. For more details, please refer to Section 3.2. The framework is shown in Figure 5.1.

Many psychologists viewed the emotion of fear as the greatest motivator [89]. Breaban and Noussair [50] showed that Fear was the most important motivating factor for stock traders. Less fearful CEOs often failed to reduce costs and had poor operating performance [90]. However, CEOs who were fearful often mask their fear from public observation making the measurement of fear difficult [91]. Thus, a revelation of Fear during a video interview is new information which the market utilizes to reevaluate the firm's stock price.

5.1 CEO Facial Expressions and Firm Performance

The CEO of the firm delivers significant information about the firm through various channels. There is much literature [52] analyzing how the market was related to the textual information in reports, letters and press releases. Speech is also a means that CEO provides the information, but it only contains part of the available information [92]. CEO’s emotional expressions provided incremental information which was priced by the market [49]. However, there is very limited research on how the markets react to non-verbal communication. Recently, interest in this non-verbal information has started to emerge. As [49] reported, emotion studies had analyzed the stress levels in the voices of managers during conference calls. Cook and Mobbs [51] showed that the attractiveness of the CEO as an important executive trait had an emotional effect on the firm announcement returns. Breaban and Noussair [50] discussed facial emotions of traders and its relationship to market behavior. Six basic emotions expressed on trader’s face are analyzed through the facial emotion recognition software. They found that a trader's loss-aversion is correlated with Fear emotion on his/her face, and that Fear emotion was the only explanatory variable. The other emotions were not priced by the market. They also found a negative correlation between Fear of traders (i.e., owners of the firm) and stock market returns. It is neutral that Fear on the face of an owner leads the owner to sell the firms stock which puts downward pressure on stock price. However, whether CEO delivers the same information to the market as the owner with Fear emotion is not explored. As Grove [89] and Bebchuk [90] showed, Fear on the face of an employee during the interview, undergoing an interrogation by the owner, motivated that employee to work harder. Therefore, A hypothesis is proposed that CEO’s Fear might result in upward pressure on the stock price.

5.2 Experiment Design

5.2.1 Data Collection

The interview videos of CEOs of Fortune 500 companies from 2006 through 2012 are obtained from YouTube and CNBC. They are utilized as samples for experimental analysis. Initially, 793 candidate videos are identified. These observations are manually matched to the Center for Research in Security Prices (CRSP) database and find available data for 720 videos. CRSP database consists of more than 20,000 stocks (inactive and active companies) from the NYSE, AMEX, and NASDAQ markets since 1926. It is found that 717 videos matching to Compustat database which is a database covering 99% of the world's total market capitalization with annual company data since 1952. In addition, the cumulative abnormal returns (CARs) of firms in 660 videos can be retrieved by running an EVENTUS event study on these observations. EVENTUS research software performs the state-of-the-art event study estimation and testing using the CRSP database or other stock return data. In this study, CEO’s facial expressions are solely focused on. Therefore videos containing other people, such as the interviewer or other interviewees are removed from the original. Due to the features of the system used for emotion recognition, the videos that cannot show CEO’s frontal face are also removed. In order to obtain the sufficient and effective CEO’s facial information, the videos less than 30 seconds are not considered in the experiment. Eventually, in total 531 videos are selected for the experiment.

5.2.2 Automatic Facial Expression Recognition System

The system utilizes Regional Hidden Markov Model (RHMM) as classifier to train the states of three face regions: the eyebrows, eyes and mouth, shown in Table 2.2, rather than modeling the entire face. Note from the table that the mouth region is slightly different from the eyebrows and eyes regions. In addition to the mouth itself, the lips corners are also important features. Considering a practical application, this system is able to classify frames as they come into analysis. As reported in [88] and can also be demoed in [93], the system not only outperformed the state-of-the-art [94] based on the posed expressions (Cohn-Kanade database [80]), but it also provided satisfactory performance on the spontaneous facial emotions. The capability of the system was validated to infer people's mental states based on spontaneous facial expressions reported similar with human-to-human interactions.

Considering the satisfied performance in the person-independent case, Cohn-Kanade database is used to train the system for CEO's facial expression analysis. In addition, the results of experiments verify that the system has the capability to infer mental states reported similar with human-to-human interactions from the spontaneous facial expressions. Please refer to [88] for more details.

This is the first study to automatically read CEO's face for predicting the firm's performance. The automatic facial emotion analysis system brings benefits to this study for three main reasons. First, the objective and measurable response to person's emotions by the system will be perceived as more natural, persuasive, and trusting. This allows us to plausibly analyze the measured data. Second, the system is able to recognize the six universal emotions. This means that the results of this chapter are well-suited to different population groups and cultures. The last but most important advantage is that the system automatically analyzes and measures the facial video data in the manner that is completely intuitive and prompt to the investors. It benefits them first to take actions on investment.

5.3 Model and Analysis

5.3.1 Data Pre-processing

Although the measured facial emotion types are common to different cultures and population groups [87], they can be influenced by various factors including race, time of day, and the general health. To control these factors, the measurements are calibrated over the first 5 seconds of the video interview. The video begins with the CEO in a quiescent state, during which the CEO is being introduced to the audience and mentally preparing him/herself to respond to the interviewer's questions. During this five-second calibration period, the median measured of all six types of emotion are computed. Then, these median measures are used to standardize the remaining frames of the video as follows:

$$EmotionType_{calibrated,i,k} = \frac{EmotionType_{raw,i,k} - EmotionType_{median,k}}{EmotionType_{median,k}}, \quad (5.1)$$
$$i = 1, \ldots, N_{frame}, \quad k = 1, \ldots, 6$$

where $EmotionType_{calibrated,i,k}$ is the k^{th} calibrated emotion type in the i^{th} frame of a video sequence. N_{frame} is the total number of frames in the video. $EmotionType_{calibrated,i,k}$ is used for the remainder of analysis. $EmotionType_{raw,i,k}$ is the k^{th} emotion type measured by the facial expression recognition system in the i^{th} frame of a video. $EmotionType_{median,k}$ is the calibration value for the k^{th} emotion type over the first 5 seconds of the video.

Finally, the average of the calibrated emotion type is utilized in a video as a value for each emotion value,

$$EmotionType_{j,k} = \sum_{i=1}^{N_{frame}} \frac{EmotionType_{calibreated,i,j,k}}{N_{frame}}, \quad j = 1,...N_{video} \tag{5.2}$$

where j is the j^{th} video, N_{video} is the number of videos.

5.3.2 Analysis of Correlations

In this section, the relationship between CEO's facial information and the firm performance is speculated. On one hand, CEOs interview videos are analyzed for the presence of facial features, specially, for the presence of six basic emotions. On the other hand, the firm performance is assessed using cumulative abnormal returns (CARs) of firms similar to that described in [52].

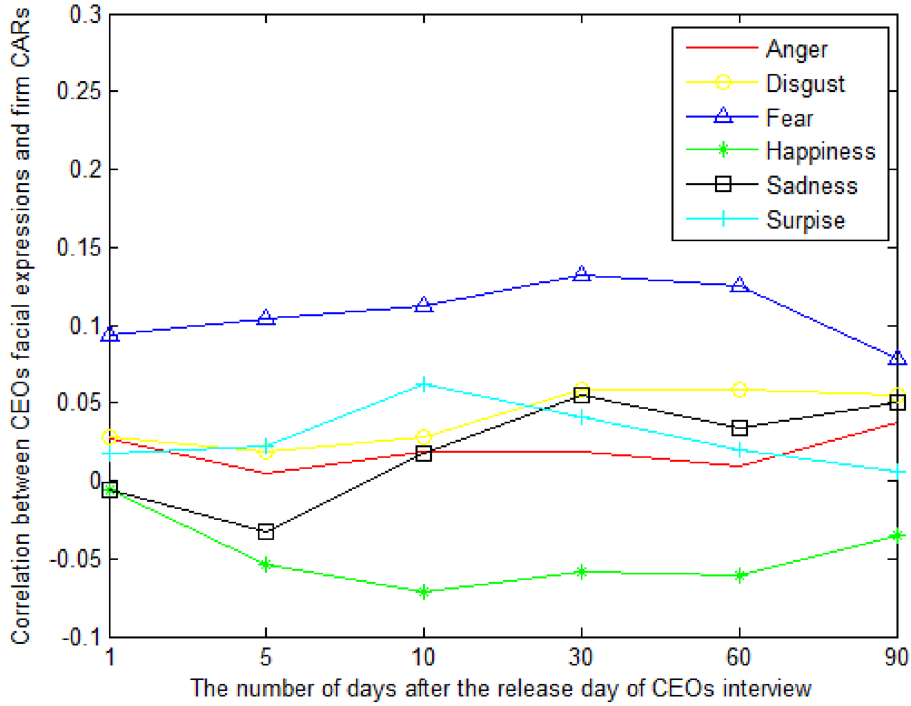

Figure 5.2 Correlation between dependent variable (CARs) and predictor variables (facial emotions) in the three months after the release day of CEO's interview.

Table 5.1 The Effect of CEO's Emotion Types on the Firm's CARs before Noise Suppression

	(-1,1)	**(-1,5)**	**(-1,10)**	**(-1,30)**	**(-1,60)**	**(-1,90)**
A	0.0043	0.0041	0.007	0.0104	0.0118	0.0232
D	0.0015	0.0049	0.0047	0.0153	0.0249	0.0214
F	0.01**	0.01*	0.01**	0.02**	0.03**	0.0256
H	-0.001	-0.006	-0.013	-0.020	-0.032	-0.037
SA	-0.001	-0.003	0.0020	0.0101	0.0099	0.0183
SU	0.0014	0.0020	0.0050	0.0031	0.0024	0.0009

Correlations between dependent variable (CARs) and predictor variables (facial emotions) are reported in Figure 5.2. The figure estimates the impact of various emotion types on the firm's market value. Fear emotion shows the highest and positive correlations with CARs whereas Happiness emotion shows negative correlation with CARs. The results respond to the proposed hypothesis that CEO's Fear emotion might put upward pressure on the firm performance. The happy CEO is likely not to be perceived as the right person to manage the firm and bring the expected performance [89, 90]. The correlation between CARs and CEO's facial emotions are not strong, since CEO dynamically conveys various facial expressions during the interviews. In this section, the relationship between the firm performance and CEO's various facial expressions during the interviews is preliminarily examined on. In order to support the hypothesis, the regression is modeled for further analysis in the following sections.

5.3.3 Model of Analysis

In this section, the regression is modeled as follows to explore whether one or more of six basic emotions have explanatory powers with respect to CARs.

$$\begin{aligned}\sum_{i=-1}^{n} ar_i &= \alpha + \beta_1 Anger + \beta_2 Disgust + \beta_3 Fear \\ &+ \beta_4 Happiness + \beta_5 Sadness + \beta_6 Surprise + \varepsilon\end{aligned} \tag{5.3}$$

where $\sum_{i=-1}^{n} ar_i$ is the CARs in n days, α is the intercept, $\beta_j, j = 1, \ldots, 6$ are the explanatory powers of six emotion types, and ε is residual. The results are reported in Table 5.1. The first row of the table shows six windows after the release day of CEOs' interviews, the first day (-1, 1), the first week (-1, 5), the first two weeks (-1, 10), the first month (-1, 30), the first two months (-1, 60), and the first three months (-1, 90). A, D, F, H, SA, and SU represent Anger, Disgust, Fear, Happiness, Sadness, and Surprise, respectively. The numerical values are the estimated coefficients for each type emotion over various windows. The statistical significance is denoted as, ** is 1%, * is 5%.

The findings in Table 5.1 show that estimated coefficients of Fear emotion, which is the only emotion, are positive and significant with respect to the firm's cumulative abnormal returns (coefficient = 0.01 in the first day, coefficient = 0.01 in the first week, coefficient = 0.01 in the first two weeks, coefficient = 0.02 in the first one month, and coefficient = 0.03 in the first two months). The result shows that the greater Fear on CEO before the market opens is, the more positively and significantly correlations with the firm's announcement returns exist. Although the significance of Fear's estimated coefficient in the first three months is larger than 5%, its estimated coefficient is still positively larger than that of other emotion types.

5.3.4 The Impact of Noise Expressed on CEO's Face

To ensure the reliability of the results obtained in previous section, it is examined whether the measured CEO's facial emotions contain measurement noise that probably interfere the experimental results. To do this, the noise of the measured facial emotion is filtered using eigenanalysis by using the Karhunen-Loève transform (KLT). Karhunen-Loève transform is a widely used technique of unsupervised learning to separate signal and noise [68].

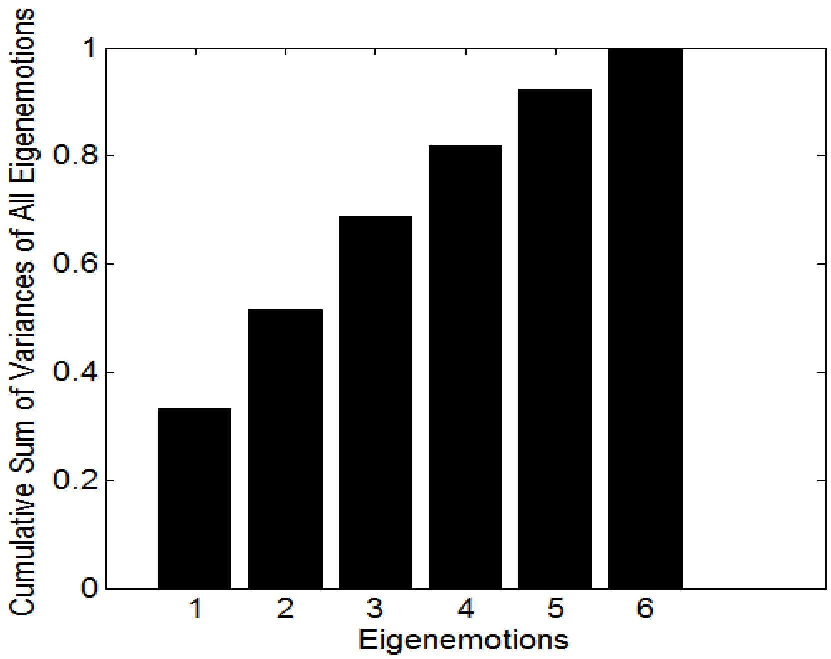

Figure 5.3 Cumulative sum of eigenvalues (variances).

Assuming that original facial emotions can be reconstructed by a linear combination of eigenemotions and the noise will lift them away from this linear relationship. Therefore, the least important eigenemotion vectors are dropped to eliminate the noise. Using KLT method, the covariance matrix of the measured CEO's facial emotions is decomposed into its eigenvectors coupled with the resulting eigenvalues, and it is the optimal orthonormal transform [68]. Most of the variances of the original facial emotions concentrate in the first few eigenemotions (principal components). Hence, the variance captured by the least significant eigenemotion vector is considered as noise which is rejected. Figure 5.3 displays the cumulative sum of eigenvalues (variances) of all eigenemotion vectors (eigenvectors). The cumulative sum of the first five eigenvalues is larger than 90% of total variances of original facial emotion.

Table 5.2 The Effect of CEO's Emotion Types on the Firm's CARs after Noise Suppression

	(-1,1)	**(-1,5)**	**(-1,10)**	**(-1,30)**	**(-1,60)**	**(-1,90)**
A	0.0069	0.0074	0.0165	0.0174	0.0173	0.0250
D	-0.001	0.0013	-0.005	0.0088	0.0199	0.0199
F	0.007*	0.009*	0.01**	0.020*	0.026*	0.0230
H	-0.002	-0.008	-0.019	-0.026	-0.037	-0.039
SA	-0.001	-0.002	0.0062	0.0133	0.0124	0.0191
SU	0	0	0	0	0	0

Table 5.3 R^2 after Noise Suppression over 6 Windows before and after Noise Suppression

	(-1,-1)	**(-1,5)**	**(-1,10)**	**(-1,30)**	**(-1,60)**	**(-1,90)**
R^2 before	0.059	0.044	0.065	0.057	0.072	0.055
R^2 after	0.058	0.044	0.0642	0.0554	0.0714	0.052

To filter out measurement noise, the eigenvector coupled with the smallest eigenvalue is set to zero. The regression is remodeled to explore the effect of CEO's facial emotions on the firm performance. The estimated coefficients are tabulated in Table 5.2. Comparing with Table 5.1, the estimated coefficients tabulated in Table 5.2 slightly increase except those of Surprise. However, the significance of all estimated coefficients does not show obvious changes. This observation implies the low-level measurement noise in data.

The goodness of fit of the regression model is important to evaluate a model. R^2 is calculated which is the measure of how close the predicted values are to the actual ones. The range of R^2 is between 0 and 1, and larger R^2 is the better goodness of fit of the model defined as

$$R^2 = 1 - \frac{\sum (Y - \hat{Y})^2}{\sum (Y - \bar{Y})^2} \tag{5.4}$$

where Y is the measured emotions, $\bar{Y}$ is the mean of Y, $\hat{Y}$ is the predicted emotions. Table 5.3 shows over 6 windows before and after noise suppression. The two-sample t-test is used that assesses whether the means of two samples are statistically different from each other. The result ($Mean_1$ =0.59, $Mean_2$ =0.58, $t(5)$=0.25, p>0.1) explains that has no significant variation before and after noise suppression over 6 windows. It indicates that the noise expressed on CEO's face does not need to be concerned.

5.3.5 Concern of Multicollinearity

Multicollinearity is a phenomenon in which two or more predictor variables (CEOs facial emotions) are highly correlated in the multiple regression model. Although it does not affect the overall fit of the model, it makes determining the effect of each predictor on dependent variable difficult. To examine multicollinearity, the variance inflation factor (VIF) is calculated for each predictor variable. VIF of 10 or above indicates multicollinearity issue exists.

Table 5.4 Variance Inflation Factor (VIF) of Each Emotion Type over 6 Windows

	A	**D**	**F**	**H**	**SA**	**SU**
VIF	1.32	1.29	1.19	1.30	1.19	1.23

$$VIF_k = \frac{1}{1-R_k^2}, \quad k = 1,\ldots,6 \tag{5.5}$$

where k is the k^{th} emotion type, R_k^2 is the measure of determination of a regression of the k^{th} predictor on all the other predictors. All VIFs shown in Table 5.4 are smaller than 10, suggesting that multicollinearity does not need to be concerned in the regression model.

5.3.6 Determinants of CEO's Fear Emotion

The determinants of CEO's Fear emotion is analyzed as

$$Fear_i = \alpha_i + \beta_1 Age + \beta_2 Tenure + \beta_3 Duality + \varepsilon_i \tag{5.6}$$

The results tabulated in Table 5.5 show that CEOs are more fearful when they have less control over the board, and that older CEOs convey more Fear on their face. For instance such as the estimated coefficient on CEO duality in the window (-1,1), it indicates that CEO who plays a dual role is 7% less fearful than other CEOs. The finding is in line with [95] that CEO duality has been blamed for poor performance and slow response to change in firms, and with the initial hypothesis that more CEO Fear positively affects the firm performance.

Table 5.5 Determinants of CEO's Fear Emotion over 6 Windows

	(-1,-1)	**(-1,5)**	**(-1,10)**	**(-1,30)**	**(-1,60)**	**(-1,90)**
Age	-0.0002	0.0005	0.001*	0.0009	0.0005	0.0008
Tenure	0.0001	-0.0002	0.0001	0.0006	0.0014	0.0004
Duality	-0.07**	-0.082*	-0.079*	-0.0845	-0.0855	-0.1325

5.4 Chapter Summary

This chapter demonstrates that, the CEO's emotion of fear, is the only emotion type which shows positive and significant explanatory power with respect to the firm's announcement returns. The market prices of the firms go up when CEOs convey Fear. The probability of multicollinearity is also examined. The result indicates that there is no strong correlation among the CEO's facial emotions measured from video. Several determinants provide evidence that Fear is less in younger CEOs, and is less when CEOs are also chairman of the board.

CHAPTER 6

ANALYSIS OF BRAIN ACTIVITY ASSOCIATED WITH FACIAL EXPRESSION

Expressions of human face are regarded as an important information source to show emotional changes, personal intents, and even states of health. Emotional states have been associated with information processing in the brain including the positive or negative affects induced by various stimuli. The goal of this study is to measure the correlation between spontaneous human facial affective expressions registered in video and the relevant brain activity. The study involves the simultaneous detections of various affective expressions by multi-modalities and the classification of spatio-temporal data with neural signature traits. The affective states of twelve male participants are investigated through three different sensors: functional near infrared spectroscopy (fNIRS) [14], electroencephalography (EEG) [61] signals, and video of facial expressions caused by images and videos used as stimuli for the experiments. These multi-modal measures are then related to the self-assessment. A method to jointly evaluate fNIRS and EEG signals for affective state detection is proposed. The experimental results reveal strong correlation between the spontaneous facial affective expressions in video and the affective states estimated by the relevant brain activity measured by fNIRS and EEG sensors. The proposed hybrid method improves performance over fNIRS or EEG only techniques. The findings also indicate that the video-content based stimuli trigger the perceiver's affect more effectively than image-content based stimuli. These findings suggest the joint utilization of video and brain signals, fNIRS and EEG for emotion analysis and affective brain-computer interface applications.

The major contributions of the chapter include:

First, this is the first attempt in exploring the relationship of human facial spontaneous affective states and relevant brain activity by simultaneous use of fNIRS, EEG and facial expressions registered in video.

Second, the spontaneous facial affective expressions recorded by a video camera are demonstrated to be in line with the affective states coded by brain activities. It is consistent with mobile brain/body imaging (MoBI) approaches [96].

Third, the experimental results show that the proposed technique outperforms methods using a subset of these signal types for the same task.

6.1 Overview of Related Work

Face is one of the information sources to infer people's mental states. Lucey et al. in [11] showed the automatic pain detection from the captured expressions on the face of the participants who were suffering from shoulder pain. The relationship between the facial responses and effectiveness of advertisement was studied in [12]. The authors demonstrated that likening of an ad can be predicted accurately from facial responses recorded by a webcam.

Affective computing has recently emerged as an interactive technology to quantify human affective states. There is much literature identifying facial expressions using tools such as the Facial Action Coding System [22]. Physiological reactions under various emotional states have been investigated. The electrical activity generated by motions of facial muscles due to the posed smile and frown is measured by electromyography (EMG) in [97]. It is noted that EMG signals are not strongly reflective of the true inner affective states.

Neurophysiological mechanism is demonstrated to infer human emotions by using non-invasive sensor modalities. fNIRS and EEG are vigorously developed as non-invasive and mobile systems [63]. Therefore, they have been the leading technologies for affective state recognition. The recent study reported in [64] combined fNIRS and EEG with autonomic nervous system (ANS) including skin conductance responses (SCR) and heart rate (HR) for emotion analysis. Although ANS objectively measures the biological dynamics to reflect the emotional changes its lack of mobility prevents it to be practical for the recognition of affective states.

These studies based on various modalities offer methods to measure inner affective reaction to stimuli. However, they cannot assist people to make rapid and effective social judgment on the affective expressions in real life scenarios. The inner affective states translated by human brain activity and consistent with those expressed on the face are worth to be explored. It may be of interest of the researchers and scientists for the various applications, such as the mind-controlled auto car, the treatment of autism, etc. The empirical findings reported in [3] showed that people's spontaneous facial information implies their preferences per relevant stimuli. Therefore, the hypothesis is proposed that the affective information translated by brain activity can be truly read through the human face. In this study, the spontaneous facial affective expressions are investigated by detecting the brain activity using both fNIRS and EEG. The block diagram of the system is displayed in Figure 6.1.

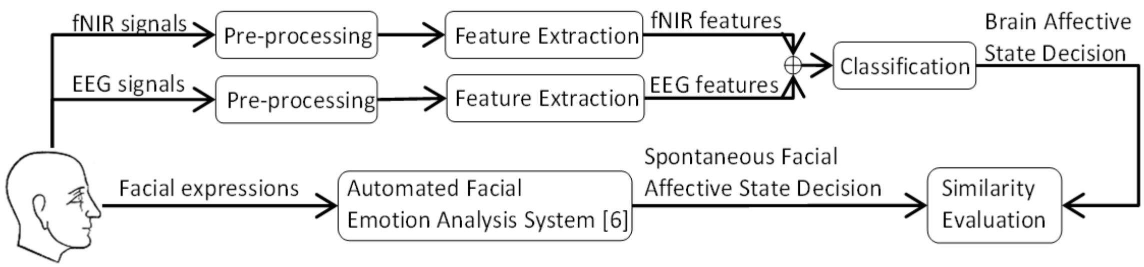

Figure 6.1 Block diagram of a system to evaluate spontaneous facial expression using the brain activity.

6.2 Design of Experiment

Twelve male participants (age: mean=27.58, SD=4.81) volunteered for the study. The perceiver (participant) was assigned to complete two tasks according to the experimental protocol shown in Figure 6.2. In the first task, each perceiver was asked to watch twenty videos with various emotional contents. Each video lasted 10-17 seconds such that that the perceiver can recognize the type of affect. After watching a video, the perceiver answered two simple questions (e.g. Were you able to watch the video carefully? What were you seeing?) in order to verify he understood the video content. In addition, perceiver was asked to evaluate the type of affect (positive or negative) and the degree of the affect using a ten-point Likert scale (ranging from 1 = extremely negative to 10 = extremely positive) in response to a set of affective states. It is worth to note that the perceiver did not know the video content in advance. The advantage of this paradigm is that perceiver naturally gives the final ratings in the absence of prior knowledge. In the second task, each perceiver was asked to observe twenty emotional images from Nencki Affective Picture System (NAPS) [98]. Each image was displayed for five seconds. Analogously, the perceiver answered two simple questions about the image content after observing the image.

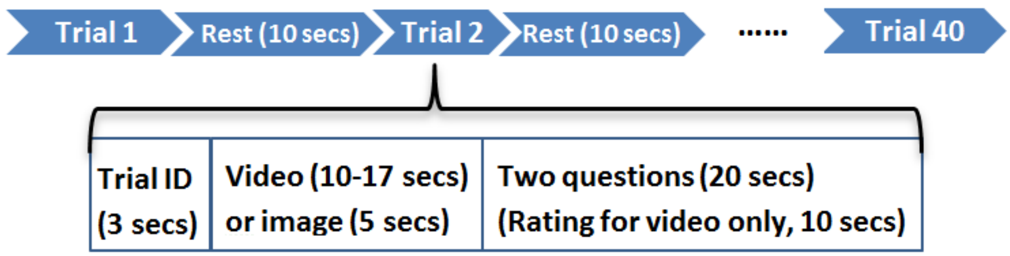

Fig. 6.2 Experimental protocol.

(a) (b)

Figure 6.3 (a) Experimental environment (b) brain measurement apparatus.

The perceiver was instructed the experimental procedure in detail before the experiment performed. Perceiver was seated on a comfortable chair facing a computer screen in a quiet room. Both fNIRS and EEG sensors were placed on the perceiver's forehead and scalp, respectively. During the experiment, perceiver's facial reaction to the stimulus was video recorded by a webcam. The experimental environment is shown in Figure 6.3 (a). The study has been approved by the Institutional Review Broad of New Jersey Institute of Technology. Before the experiment, each perceiver was asked to sign an agreement to participate in the study.

6.3 Experimental Apparatus

The neuroimaging systems used in this study consists of two commercial products: a wireless fNIR Model 1100W system and an Emotiv EPOC headset (see in Figure 6.3).

6.3.1 Functional Near Infrared Spectroscopy

Functional near infrared spectroscopy (fNIRS) is a non-invasive and safe neuroimaging method to quantify the changes of cerebral oxygenated and deoxygenated hemoglobin concentration in the cortex to assess the brain activity by using near infrared light attenuation. fNIRS measures cortical hemodynamic response similar to functional magnetic resonance imaging (fMRI) but without the many limitations and restrictions on the subject such as staying in supine position in a confined space or exposure to load noise [62]. fNIRS as a portable and cost effective functional neuroimaging modality is uniquely suitable to study cognition and emotion processing related brain activities due to relatively high spatial resolution and practical sensory setup [56].

A compact and battery-operated wireless fNIRS system, Model 1100W, developed by Drexel University (Philadelphia, PA) and manufactured by fNIRS Devices, LLC (www.fnirdevices.com) was used to monitor the prefrontal cortex of the perceiver as shown in Figure 6.4. The system measures cortical oxygenation changes during the task, and is composed of three modules: a sensor pad that holds light sources and detectors to enable a fast placement of four optodes, a control box for hardware, and a computer that runs COBI Studio software [99] and receives the data wirelessly. More information about the device was reported in [62].

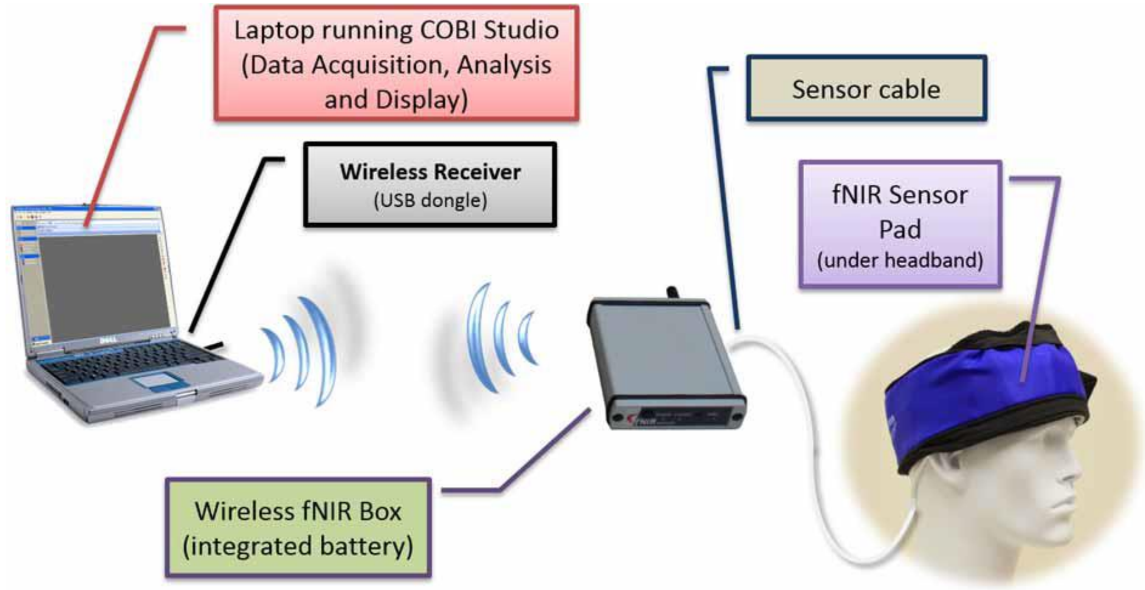

Figure 6.4 Components of fNIRS system: wireless transmitter, wireless box containing battery, and sensor pad [62].

6.3.2 Electroencephalography

Emotiv EPOC headset shown in Figure 6.5 acquires the 128 Hz Electroencephalography (EEG) signals by measuring electrical differences on the scalp and transmits the signals wirelessly to a Windows PC (www.emotiv.com). The system measures the scalp electrical potentials caused by the neurons firing. The cap has 14 electrodes located over 10-20 system positions AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 using 2 reference electrodes.

Figure 6.5 Emotiv EPOC headset: USB Dongle, EPOC headset, and electrodes.

Figure 6.6 Facial feature points on a face image.

6.3.3 Automatic Facial Emotion Recognition System

The automatic facial emotion recognition system proposed in [88] is used to identify the spontaneous facial affective states. As reported in [88], the system not only outperformed the state-of-the-art based on the posed expressions but also provided satisfactory performance on the spontaneous facial emotions. The system has been used to read CEO's facial expressions to forecast firm performance by only using recorded video signal [100].

It utilizes Regional Hidden Markov Model (RHMM) as its classifier to train the states of three face regions: the eyebrows, eyes and mouth. Since the biological information that describes a facial expression is mainly registered in the movement of these three regions as sensed and quantified in frames of a video sequence, it is natural to classify the states of each facial region rather than modeling the entire face. Considering a practical application, this system is able to classify frames as they come into analysis. To describe the states of face regions, 41 facial feature points are identified on each frame of video, as displayed in Figure 6.6. They are comprised of 10 feature points on the eyebrows region, 12 points on the eyelids, 8 points on the mouth, 10 points on the corners of the lips, and one anchor feature point on the nose. The 2D coordinates of facial feature points in various face regions are extracted to form corresponding observation sequences for classification. For more details, please refer to Section 3.2.

The system benefits this study for three main reasons. First, the objective and measurable response to a person's emotions by the system is perceived as more natural, persuasive, and trustworthy. This allows us to plausibly analyze the measured data. Second, the system is able to recognize the various affective states of interest. The last but the most important advantage is that the system automatically analyzes and measures the live facial video data in a manner that is intuitive and useful for different applications. It helps the user to take actions based on this analysis.

6.4 Model and Analysis

6.4.1 Stimuli Evaluation

Twenty emotional images used in the trials were obtained from the Nencki Affective

Picture System (NAPS) [98]. It is a novel, standardized, wide-range, high-quality, realistic picture database that is widely used for brain imaging studies. The positive-content images and negative-content ones according to the given valence values were selected. Each image is classified explicitly with the attached emotion type if over 50% of all perceivers express the same facial affect. Otherwise, it is classified as an ambiguous image that is discarded [88]. Eventually, all images are classified explicitly. Therefore, all of them are utilized for the experiments.

For the part of video-content experiment, twenty videos were selected from Youtube.com. They were categorized into the positive content (funny and happy) and the negative content classes (sad, angry, and disgust) through being independently watched and evaluated by all perceivers. To make sure the video contents consistent with perceivers' spontaneous affective states, the similar evaluation was done to classify all video clips as explicit video or ambiguous video. Eventually, all selected video clips are classified explicitly and used in the following experiments.

6.4.2 Data Pre-processing

Noise reduction in fNIRS signals is achieved by utilizing a low-pass filter with 0.1 Hz cutoff frequency [54]. Motion artifacts were eliminated prior to extracting the features from fNIRS signals by applying a fast Independent Component Analysis (ICA) [101]. The independent component was selected through modeling the hemodynamic response. The modeled hemodynamic response represents the expected hemodynamic response to the given stimulus calculated by convolving the stimulus function and a canonical hemodynamic response function (HRF). The HRF [102] consists of a linear combination of two Gamma functions as

$$h(t) = A\left(\frac{t^{\alpha_1 - 1}\beta_1^{\alpha_1} e^{-\beta_1 t}}{\Gamma(\alpha_1)} - c\frac{t^{\alpha_2 - 1}\beta_2^{\alpha_2} e^{-\beta_2 t}}{\Gamma(\alpha_2)}\right) \tag{6.1}$$

where A controls the amplitude, α and β control the shape and scale, respectively, and c determines the ratio of the response to undershoot. The t-test was used to select the independent component associated with the hemodynamic response. It is considered that the independent component with the smallest t-value is associated with the hemodynamic response to a given stimulus.

The EEG signals were passed through a low filter with 30 Hz cutoff frequency. The artifacts in the raw EEG signals were detected and removed using ICA in EEGLAB which is an open-source toolbox for analysis of single-trial EEG dynamics [103]. The average of the 5-second baseline brain signal before each trial was subtracted from the brain response data for the baseline adjustment. The perceiver was also instructed to minimize his head movements during the experiment in order to avoid the artifacts from the muscular activity at maximum level.

Inference of perceiver's facial affective expression by the automatic facial emotion recognition system is based on facial emotion recognition [88]. Happiness was coded as the positive affect and Anger, Disgust, Fear, and Sadness as the negative affect. Since Surprise could be revealed as a result of positive or negative affect, it was not used in this experiment. The ratings of both the positive affect and the negative affect for a video are separately generated by the system by normalizing the probability of Happiness and the largest probability among Anger, Disgust, Fear and Sadness. The larger affect rating of the two affect values is selected as the affect recognized in the corresponding video clip. Please refer to Section 4.4.1 for more details.

6.4.3 Analysis of Correlation

In a prior study, the relationship between the perceiver's spontaneous facial information and the concurrent inner affective states was highlighted [88]. On the one hand, the perceiver's recorded face videos were analyzed for the presence of facial features, especially for the presence of facial affective states. On the other hand, the mental states were evaluated using the perceiver's the given labels and self-assessment. The self-assessment has been widely used to measure the mental states in the literature [3, 88].

The correlation was calculated using the ground truth for all video and image trials (the given affective ratings for images and perceiver's self-assessment for videos as stimuli as stimuli) and the facial affective states measured by the automatic facial emotion recognition system. The result is summarized in Figure 6.7. The figure estimates the degree of the mental states delivered on the face over all perceivers. All perceivers' face affective states show the positive correlation with those reported by perceivers. The results comply with our hypothesis that the facial affective expression may reflect the affective state in the mind. However, it is likely that the self-assessment provided by the perceiver is derived from the perceiver's recall or from the second thoughts. In this section, the relationship between the facial affective expressions and mental states was examined. In order to support the hypothesis, the model for further analysis is built as detailed in the following sections.

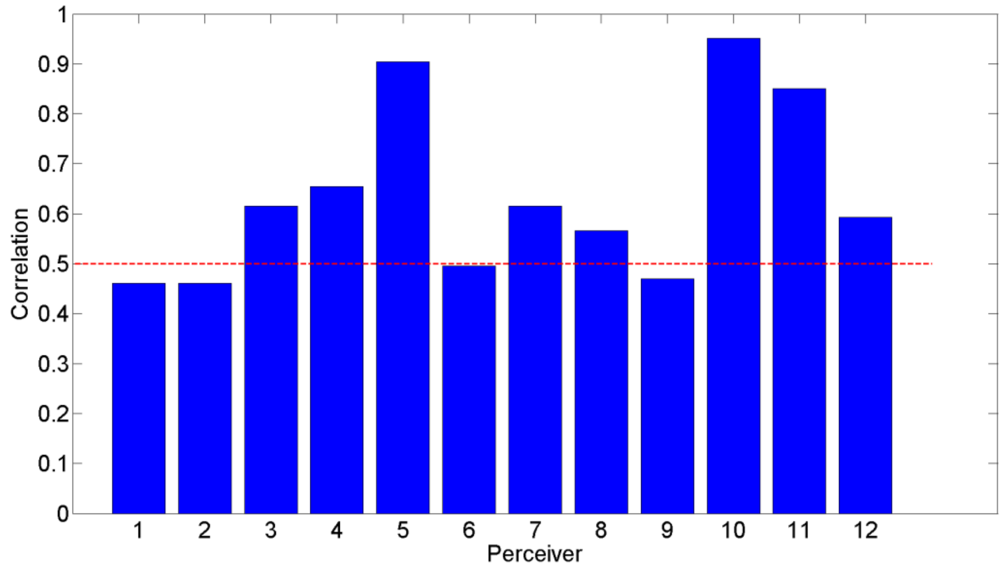

Figure 6.7 Correlation of the perceiver's facial affective states and the ground truth over all trials.

6.4.4 Feature Extraction

The recorded data of raw fNIRS light intensity at two wavelengths (730nm and 850nm) is converted to the relative changes in hemodynamic responses in terms of oxy-hemoglobin (Hbo) and deoxy-hemoglobin (Hbr) using the modified Beer-Lambert Law [99]. Total hemoglobin concentration changes (Hbt), the sum of Hbo and Hbr and is an estimate of total blood volume, and the difference of Hbo and Hbr, estimate of oxygenation change were also calculated for each optode. The mean, median, standard deviation, maximum, minimum, and the range of maximum and minimum of four hemodynamic response signals are calculated as features, giving 4×4×6=96 fNIRS features for each trial.

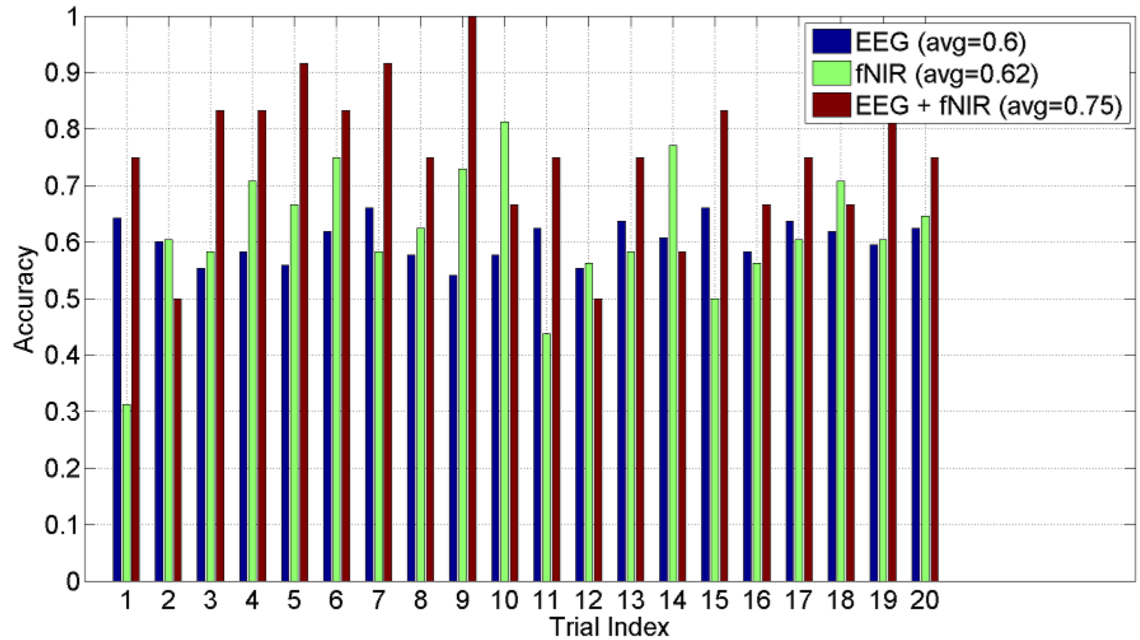

Figure 6.8 Performance comparison of the proposed method (EEG+fNIRS) and the ones where only EEG or fNIRS employed for all image-content trials.

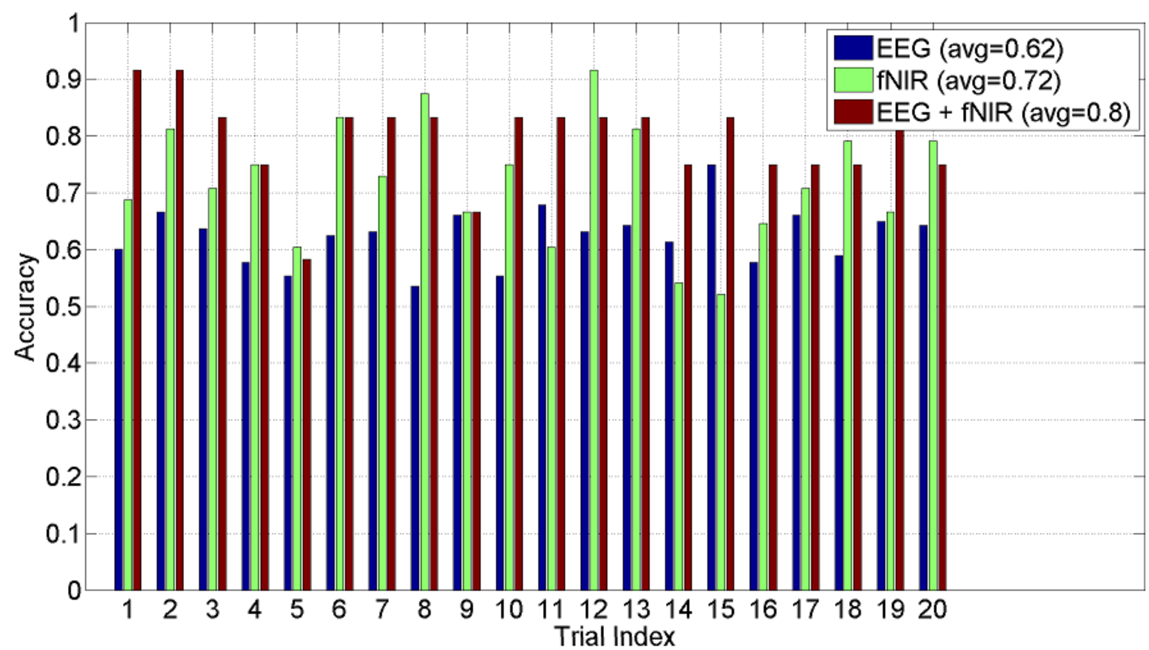

Figure 6.9 Performance comparison of the proposed method (EEG+fNIRS) and the ones where only EEG or fNIRS employed for all video-content trials.

The spectral power of EEG signals in different bands has been used for emotion analysis [104]. The logarithms of the power spectral density (PSD) for theta (4Hz $< f <$ 8Hz), slow alpha (8Hz $< f <$ 10Hz), alpha (8Hz $<f <$ 12Hz), and beta (12Hz $< f <$ 30Hz) bands are extracted from all 14 electrodes as features. In addition, the difference between the spectral power of all possible symmetrical pairs on the right and left hemisphere is extracted to measure the possible asymmetry in the brain activity due to the valance of emotional stimuli [105]. The asymmetry features is extracted from four symmetric pairs over four bands, AF3-AF4, F7-F8, F3-F4, and FC5-FC6. The total number of EEG features of a trial for 14 electrodes is $14\times4+4\times4=72$.

6.4.5 Affective State Detection

There were 40 trials (20 images and 20 videos) for each perceiver. The images were selected from the Nencki Affective Picture System (NAPS) [98] and the videos were collected from YouTube.com. Polynomial Support Vector Machine (SVM) was used for classification. fNIRS and EEG features were concatenated to form a larger features vector before feeding them into the model. For comparison, the univariate modality of either fNIRS or EEG features for recognition of the affective state is also applied. The leave-one-out approach for training and testing was applied. That is, the features extracted in nineteen trials over all perceivers were used for training and the remaining one is used for testing. The performance of the experiment was validated through twenty-time iterations.

The recognition performance for joint use of fNIRS and EEG along with the cases where only one is used is displayed in Figures 6.8 and 6.9 for the stimuli of image

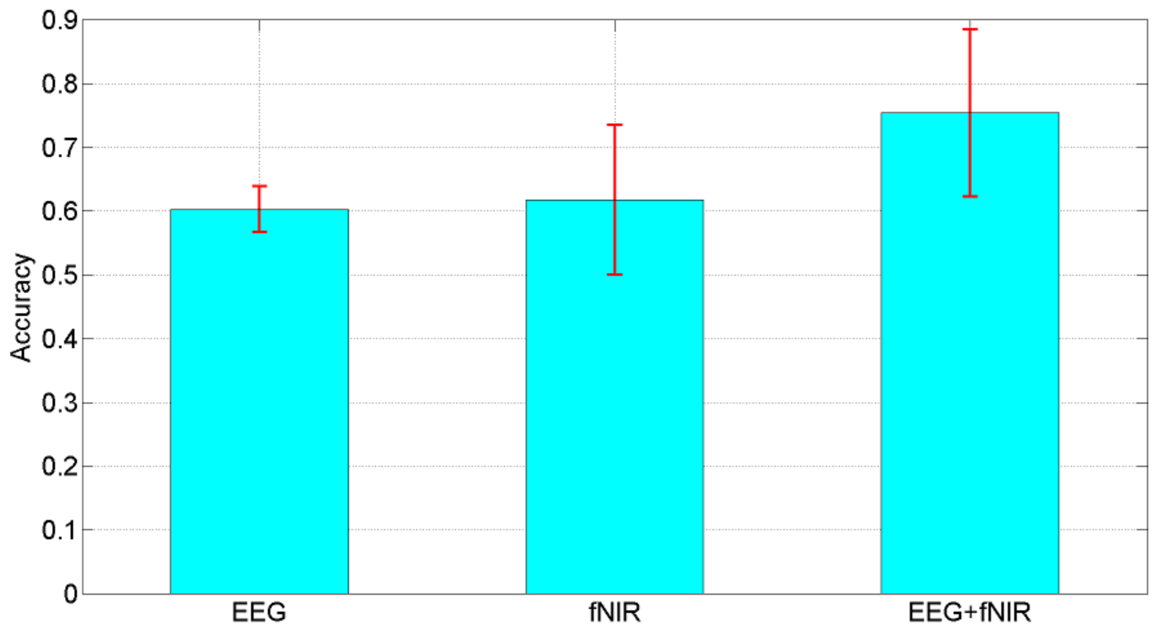

Figure 6.10 Average performances of the proposed fNIRS+EEG method along with EEG only and fNIRS only techniques for all image-content trials as displayed in error bars.

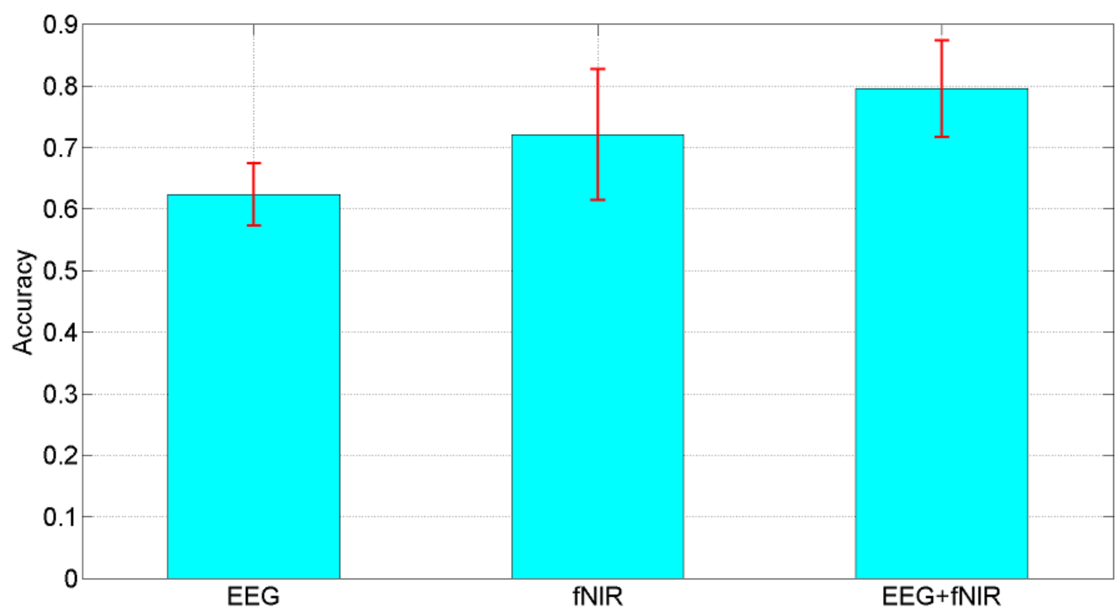

Figure 6.11 Average performances of the proposed fNIRS+EEG method along with EEG only and fNIRS only techniques for all video-content trials as displayed in error bars.

and video content types, respectively. The method jointly using fNIRS and EEG features outperforms the techniques where only one of them is utilized. This observation is consistent for almost all trials. In some trials, univariate modality of EEG or fNIRS shows higher performance. It was observed that some participants reflected strongly to the stimuli while some participants showed higher affective tolerance to the same stimuli. This finding is consistent with our pilot study. The average performances of these three methods are compared in Figures 6.10 and 6.11 for image-content and video-content stimuli, respectively. The proposed multi-modal method performs over 10% better than the single-modality methods for the same stimuli. It was also found that these three different methods recognize the affective response to video-content stimuli more accurately than those caused by image-content stimuli. It is likely that the dynamic (video) stimulus provides more contextual information than the static (image) one. The error bars show the variability of the performance for all trials.

6.4.6 Experiment Result Analysis

First, the reliability of the automatic facial emotion recognition system that is used to recognize the perceiver's facial reaction to the given stimulus is assessed. In order to achieve this, the affect recognition accuracy of the system is calculated by comparing the detected results with the perceiver's self-assessment and the actual labels. The average of all perceivers' facial inferences for each trial is tabulated in Tables 6.1 and 6.2 for image-content trials and video-content trials, respectively. $T_i,\ i = 1, ..., 20$ represents the trial. The affect recognition rate of the system reaches to 0.74 (SD=0.10) for image-content stimuli and 0.80 (SD=0.10) for video. The results are satisfactory and indicate that the system performs well to detect a person's spontaneous facial expressions.

Table 6.1 Average of Perceivers' Spontaneous Facial Affective Response to the Image-content Stimulus Each Trial

T1	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**
0.92	0.75	0.67	0.75	0.83	0.75	0.67	0.83	0.83	0.83
T11	**T12**	**T13**	**T14**	**T15**	**T16**	**T17**	**T18**	**T19**	**T20**
0.67	0.75	0.67	0.5	0.75	0.75	0.92	0.75	0.67	0.58

Table 6.2 Average of Perceivers' Spontaneous Facial Affective Response to the Video-content Stimulus Each Trial

T1	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**
0.83	0.67	0.83	0.59	0.92	0.83	0.67	0.83	0.92	1
T11	**T12**	**T13**	**T14**	**T15**	**T16**	**T17**	**T18**	**T19**	**T20**
0.75	0.83	0.83	0.83	0.83	0.83	0.67	0.92	0.67	0.83

Next, the degree of the similarity of spontaneous affective states expressed on the face and those coded by the brain activity was estimated. Similarity scores were calculated by the percentages of spontaneous facial affect recognized by the system same as the inner affective states translated by perceiver's brain signals. For this purpose, Wilcoxon Signed Rank Test was used where the affective state was classified as positive and negative. Wilcoxon Signed Rank Test to test the significance. The test results, (W=65, p=0.04) for image-content trials and (W=61, p=0.01) for video-content trials, indicates that the spontaneous facial affective states are consistent with those translated by human brain activity. That is, the spontaneous facial affective states can reflect the human true brain affective responses to the stimuli.

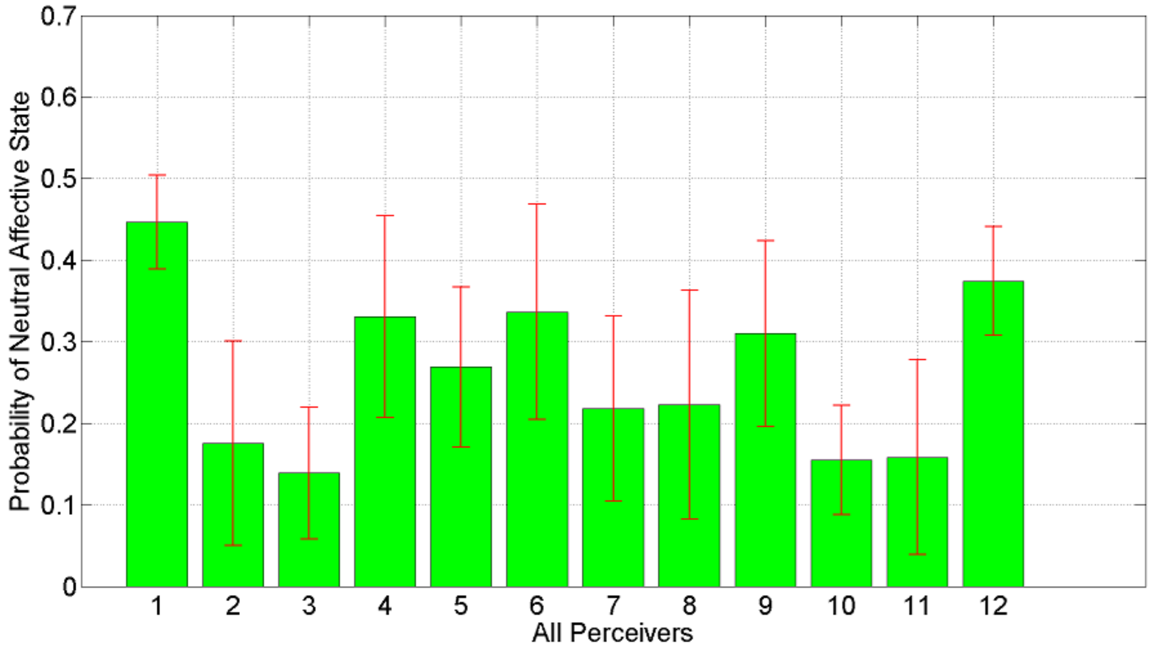

Figure 6.12 Average neutral affective state of each perceiver over all image-content trials displayed in the error bar.

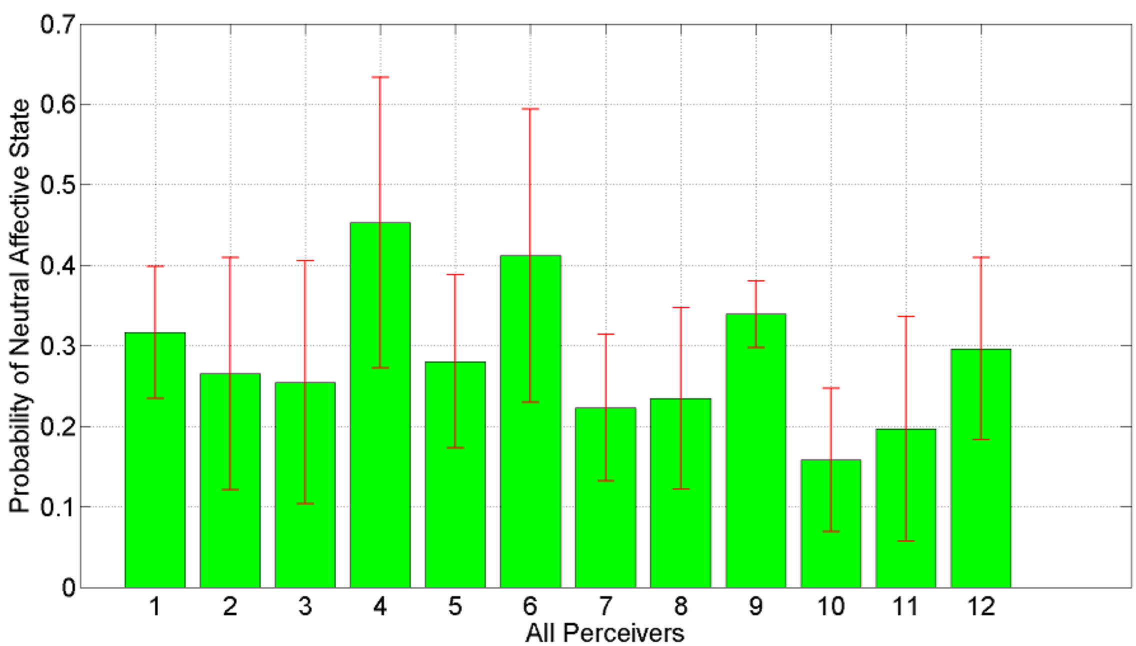

Figure 6.13 Average neutral affective state of each perceiver over all video-content trials displayed in the error bar.

To further check the reliability of our findings, it was examined whether some perceiver was expressive who might bias the average of all perceivers' facial expression for each trial. To do that, the perceiver's neutral affective state over trials was explored. During the experiment, the perceiver may express positive affect, negative affect, or neutral affect. The expressive perceiver is considered to be more prone to make the first two types of affective reactions to the stimuli rather than the neutral state. The average of a perceiver's neutral affective states over all trials for image-content stimuli and video-content stimuli is plotted as error bars in Figures 6.12 and 6.13, respectively. The one-sample t-test is used to assess whether the average of measured neutral states is close to the population mean ($\mu=0.3$) as there are three types of affective states. The results of one-sample t-test (Mean=0.26, SD=0.09, $t(11)=1.01$, $p>0.5$) for image-content trials and (Mean=0.29, SD=0.08, $t(11)=0.58$, $p>0.3$) for video-content trials explain that all perceivers are not expressive during the experiment.

6.5 Chapter Summary

This paper provides new insights for the exploration and analysis of spontaneous facial affective expression associated with the simultaneous brain activity in the form of hemodynamic and electrophysiological changes. Its main contribution is to demonstrate that human spontaneous facial expressions convey the affective states translated by the brain activity. The experimental results are found on the premise that the perceiver has no knowledge of stimuli prior to the experiment. The spontaneous facial expressions of the unaware perceiver can be triggered by the emotional stimulus [65]. Moreover, the neural activity changes are found due to non-conscious perception of emotional stimuli. To assess the plausibility of our findings, the perceiver's expressiveness was examined. The results indicate that there was no perceiver with exaggerative expression over all the trials performed. In addition, it was found that the video-content stimuli more readily induce the perceiver's affect states than the image-content stimuli. This can be explained as dynamic (video) stimulus provides more contextual information than the static (image) one. Compared to the static stimuli, dynamic ones indicate enhanced emotion in the specific brain activation patterns as also shown in [66].

In this study, the findings were derived from the combined analysis of hemodynamic and electrophysiological signals from the brain. The neural activities were measured by two non-invasive and complimentary neuroimaging techniques, fNIRS and EEG. The complementary nature of fNIRS and EEG has been reported in the literature with multimodality studies [63]. Particularly, both of them have received considerable attention on emotion inference and emotional mapping on the regional brain activities [55]. The proposed hybrid method for affective state detection jointly using fNIRS and EEG signals outperforms techniques that only employ fNIRS or EEG. It is due to the fact that they are known to contain complementary information [106] when the stimuli are videos or images.

On the other hand, the automated mechanism to measure facial reactions to different stimuli offers prompt, objective and accurate recognition performance in continuous time. The regional facial features are highlighted since they convey significant information relevant to expressions. It is natural to classify the states of each facial region rather than considering the holistic features of the entire face for recognition [34]. The experimental results support the hypothesis to show the high correlation between the recognized facial affective expressions and the ground truth for all trials (perceiver's self-assessment for videos as stimuli and the given affective ratings for images as stimuli).

To the best of the knowledge, this is the first attempt to detect the affective states by jointly using fNIRS, EEG and video capture of facial expressions. The experimental results offer strong evidence for the validation of the proposed method. In addition, the results highlight the reliability of spontaneous facial expression and its analysis as a promising methodology to serve for various applications in the future.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

7.1 Conclusion

Facial expression recognition by the system is regarded as an important way of building Human-computer Interface (HCI) as it can promote more efficient, accurate, and trusting performance than human's subjective impression. Furthermore, these merits allow people to plausibly analyze the measured data. The high recognition accuracy of facial emotions has been achieved in many existing researches based on the benchmarked databases containing posed facial emotions. However, their purposes focus on the optimal performance of their methods rather than their plausibility in the psychological aspect. In addition, subjects who participant is asked to perform a series of exaggerate emotional expressions during the experiment. These prototyped facial emotions cover a rather small part of our daily emotional displays. In this dissertation, according to the theory of facial expressions and the concerns of the pragmatic way of conveying emotions on the face, the contributions are summarized as bellow.

First, a region based model is proposed to analyze the facial expression associated with the empirical conclusion drawn by psychologists. Emotions are comprised of its mapping on various regions of a face. It is found that an emphasis on classifying local features of a face effectively improves the recognition performance. The performance of recognizing human facial emotions is more accurate than that of the state-of-the-art mechanism.

Second, a method of mapping prototyped emotions into affective states is applied to measure the various affective displays that people show in daily interpersonal interactions. The experimental findings show that the proposed facial expression system is able to infer people's mental states despite derived from spontaneous, moderate-emotional facial information. In the further studies, the system with the human is compared and demonstrates the system is able to interact naturally with the user, similar to the way human-human interaction takes place.

Third, the first attempt is done to read CEOs' facial expressions to infer their emotions during the interview for forecasting the firm performance. The findings confidently support that CEO's facial expressions during interviews provide the significant information for predicting the firm performance. Fear emotion is the only one type which shows positive and significant explanatory power with respect to the firm's announcement returns. The market prices of the firms go up when CEOs convey Fear. The probability of multicollinearity is examined. Several determinants provide evidence that Fear is less in younger CEOs, and is less when CEOs are also chairman of the board.

Fourth, people's spontaneous facial expressions are demonstrated to be able to convey the affective states translated by the brain activity. The proposed method for the combination of fNIR signals and EEG signals outperforms using either fNIR or EEG when the stimuli are videos or images. In addition, the finding shows that the recognition performance of these three modal methods with respect to video-content stimuli is better than that of image-content stimuli.

7.2 Future Work

Firstly, the system reliability is validated with human based experiments to infer people's mental states from facial expressions registered as video sequences in the dissertation. Real-time implementation of the proposed system may be of great interest for various applications spanning from health care to finance.

Secondly, an important contribution of the dissertation is that the proposed system is beneficial to analyze information that is non-quantitative and non-verbal. The social psychology studies the impact of managerial idiosyncratic behavior on the markets perception of firm performance. The future research can continue this line of research by applying facial emotion recognition to other market participants, such as the Federal Reserve Chairman, or the Chairman of the Bundesbank.

Finally, People's spontaneous facial expressions have been demonstrated to be consistent with those translated by the brain activity. The video sequences and images used in this study display short duration content although all perceivers stated that they were able to understand all stimuli. However, it is of interest to address how the perceiver reacts to the longer duration-content stimuli in future studies. The extension of this work might benefit to the specific applications that require the feedback of longer-duration content such as online education and entertainment.

REFERENCES

[1] C. Darwin, *The expression of the emotions in man and animals*, University of Chicago press, Chicago, IL, USA, 1965.

[2] R. El Kaliouby and P. Robinson, “Real-time inference of complex mental states from facial expression and head gestures,” *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 3, pp. 154, Jun. 2004.

[3] M. S. North, A. Todorov, and D. N. Osherson, “Inferring the preferences of others from spontaneous, low-emotional facial expressions,” *Journal of Experimental Social Psychology*, vol. 46, no. 6, pp. 1109-1113, Nov. 2010.

[4] G. C. Littlewort, M. S. Bartlett, and K. Lee, “Faces of pain: automated measurement of spontaneous facial expressions of genuine and posed pain,” *ACM Int’l Conf. Multimodal Interfaces*, pp. 15-21, Nov. 2007.

[5] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, “Recognizing facial expression: machine learning and application to spontaneous behavior,” *IEEE Int’l Conf. Computer Vision and Pattern Recognition*, pp. 568-573, Jun. 2005.

[6] M. F. Valstar and M. Pantic, “Fully automatic recognition of the temporal phases of facial actions,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 1, pp. 28-43, Jan. 2012.

[7] P. Ekman and W. V. Friesen, *Facial action coding system*, Consulting Psychology Press, Palo Alto, CA, USA, 1978.

[8] Z. Zeng , M. Pantic , G. I. Roisman, and T. S. Huang, “A survey of affect recognition methods: audio, visual, and spontaneous expressions,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39-58, Jan. 2009.

[9] M. Valstar, M. Pantic, Z. Ambadar, and J. F. Cohn, “Spontaneous versus posed facial behavior: automatic analysis of brow actions,” *ACM Int’l Conf. Multimodal Interfaces*, pp. 162-170, Nov. 2006.

[10] M. F. Valstar, H. Gunes, and M. Pantic, “How to distinguish posed from spontaneous smiles using geometric features,” *ACM Int’l Conf. Multimodal Interfaces*, pp. 38-45, Nov. 2007.

[11] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews, “Painful data: the UNBC-McMaster shoulder pain expression archive database,” *IEEE International Conference on Automatic Face and Gesture Recognition*, Santa Barbara, California, pp. 57-64, Mar. 2011

[12] D. McDuff, R. Kaliouby, J. Cohn, and R. Picard, "Predicting ad liking and purchase intent: large-scale analysis of facial responses to ads," *IEEE Trans. on Affective Computing*, vol. PP, no. 99, pp. 1-13, Dec. 2014.

[13] V. Zotev, R. Phillips, K. D. Young, W. C. Drevets, and J. Bodurka, "Prefrontal control of the amygdala during real-time fMRI neuro feedback training of emotion regulation," *PloS one*, vol. 8, no. 11, Nov. 2013.

[14] H. Ayaz, P. A. Shewokis, S. Bunce, K. Izzetoglu, B. Willems, and B. Onaral, "Optical brain monitoring for operator training and mental workload assessment," *NeuroImage*, vol. 59, no. 1, pp. 36-47, Jan. 2012.

[15] M. De Vos and S. Debener, "Mobile EEG: Towards brain activity monitoring during natural action and cognition," *International Journal of Psychophysiology*, vol. 91, no. 1, pp. 1-2, Jan. 2014.

[16] B. Fasel and J. Luttin, "Automatic facial expression analysis: survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259-275, Jan. 2003.

[17] M. Pantic and L. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424-1445, Dec. 2000.

[18] V. Bettadapura, "Face expression recognition and analysis: the state of the art," *Tech Report, arXiv:1203.6722*, Apr. 2012.

[19] L. Ma and K. Khorasani "Facial expression recognition using constructive feed forward neural networks," *IEEE Trans. on Systems, Man, and Cybernetics: Systems*, vol. 34, no. 3, pp. 1588-1595, Jun. 2004.

[20] P. Yang, Q. Liu and D. N. Metaxas, "Boosting encoded dynamic features for facial expression recognition," *Pattern Recognition Letter*, vol. 30, no. 2, pp. 132-139, Jan. 2009.

[21] M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," *IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 200-205, Apr. 1998.

[22] P. Ekman and W.V. Friesen, *Facial action coding system (FACS): Manual*, Consulting Psychologists Press, Palo Alto, CA, USA, 1978.

[23] Y. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97-115, Feb. 2001.

[24] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models - their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38-59, Jan. 1995.

[25] Y. Tong , Y. Wang , Z. Zhu and Q. Ji, "Robust facial feature tracking under varying face pose and facial expression," *Pattern Recognition*, vol. 40, no. 11, pp. 3195-3208, Nov. 2007.

[26] P. Michel and R. Kaliouby. "Real time facial expression recognition in video using support vector machines," *ACM International Conference on Multimodal Interaction*, pp. 258 -264, Nov. 2003.

[27] M. S. Bartlett, J. C. Hager, P. Ekman and T. J. Sejnowski, "Measuring facial expressions by computer image analysis," *Psychophysiology*, vol. 36, no. 2, pp. 253-264, Mar. 1999.

[28] I. Cohen, N. Sebe, F. Cozman, M. Cirelo, and T. Huang, "Learning Bayesian network classifiers for facial expression recognition both labeled and unlabeled data," *IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 595-601, Jun. 2003.

[29] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974-989, Oct. 2000.

[30] I. Cohen, N. Sebe, A. Garg, L. Chen, and T. Huang, "Facial expression recognition from video sequences: temporal and static modeling," *Computer Vision and Image Understanding*, vol. 91, no. 1, pp. 160-187, Jul. 2003.

[31] M. Yeasin, B. Bullot, and R. Sharma, "From facial expression to level of interests: a spatio-temporal approach," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 922-927, Jun. 2004.

[32] I. Cohen, A. Garg, and T. Huang, "Emotion recognition from facial expressions using multilevel HMM," *Conference on Neural Information Processing Systems*, Nov. 2000.

[33] T. Otsuka and J. Ohya, "Recognizing multiple persons' facial expressions using HMM based on automatic extraction of significant frames from image sequences," *IEEE Int'l Conf. Image Processing*, vol. 2, pp. 546-549, Oct. 1997.

[34] M. White, "Parts and wholes in expression recognition," *Cognition and emotion*, vol. 4, no. 1, pp. 39-60, 2000.

[35] C. Padgett and G. Cottrell, "Identifying emotion in static face images," *Joint Symposium on Neural Computation*, vol. 5, pp. 91-101, Nov. 1995.

[36] A. J. Calder, A. W. Young, J. Keane, and M. Dean. "Configural information in facial expression perception," *Journal of Experimental Psychology: Human perception and performance*, vol. 26, no. 2, pp. 527-551, Apr. 2000.

[37] R. Gajsek, V. Struc, F. Mihelic, A. Podlesek, L. Komidar, G. Socan, and B. Bajec "Multi-modal emotional database: AvID," *Informatica*, vol. 33, pp. 101-106, 2009.

[38] C. Busso, M. Bulut, C. C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "IEMOCAP: Interactive emotional dyadic motion capture database," *Journal of Language Resources and Evaluation*, vol. 42, no. 4, pp. 335-359, Dec. 2008.

[39] S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, and X. Wang, "A natural visible and infrared facial expression database for expression recognition and emotion inference," *IEEE Trans. on Multimedia*, vol. 12, no. 7, pp. 682-691, Nov. 2010.

[40] A. J. O'Toole, J. Harms, S. L. Snow, D. R. Hurst, M. R. Pappas, J. H. Ayyad, and H. Abdi, "A video database of moving faces and people," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 812-816, May 2005.

[41] Y. Zhu, L. C. De Silva, and C. C. Ko, "Using moment invariants and HMM in facial expression recognition," *Pattern Recognition Letter*, vol. 23, no. 1, pp.83-91, Jan. 2002.

[42] T. Otsuka and J. Ohya, "Recognition of facial expressions using HMM with continuous output probabilities," *IEEE Workshop on Robot and Human Communication*, pp.323-328, Nov. 1996.

[43] Y. Sun, M. Reale, and L. Yin, "Recognizing partial facial action units based on 3D dynamic range data for facial expression recognition," *IEEE Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 1-8, Sep. 2008.

[44] A. K. Davis, J. M. Piger, and L. M. Sedor, "Beyond the numbers: measuring the information content of earnings press release language," *Contemporary Accounting Research*, vol. 29, no. 3, pp. 845-868, Sep. 2012.

[45] P. C. Tetlock, M. Saar-Tsechansky, and S. Macskassy, "More than words: quantifying language to measure firms' fundamentals," *Journal of Finance*, vol. 63, pp. 1437-1467, May 2008.

[46] J. Brau, J. Cicon, and S. Ferris, "Soft strategic information and IPO underpricing," *working paper*, Brigham Young University, New Jersey Institute of Technology, University of Missouri, Oct. 2012.

[47] P. N. Juslin and K. R. Scherer, *Vocal expression of affect in the new handbook of methods in nonverbal behavior research*, Oxford press, Oxford, UK, 2005.

[48] M. Zuckerman, M. D. Amidon, S. E. Bishop, and S. D. Pomerantz, "Face and tone of voice in the communication of deception," *Journal of Personality and Social Psychology*, vol. 43, pp. 347-357, Aug. 1982.

[49] W. Mayew and M. Venkatachalam, “The power of voice: managerial affective states and future firm performance,” *The Journal of Finance*, vol. 67, no. 1, pp. 1-43, Jan. 2012.

[50] A. Breaban and C. Noussair, “Emotional states and market behavior,” *working paper*, Tilburg University, Jun. 2013.

[51] D. Cook and S. Moobs, “CEO selection and executive appearance,” *working paper*, University of Alabama, Jan. 2015.

[52] P. Brockman and J. Cicon, “The information content of management earnings forecasts: an analysis of hard versus soft information,” *Journal of Financial Research*, vol. 36, no. 2, pp. 147-173, Jun. 2013.

[53] A. R. Damasio, T. J. Grabowski, A. Bechara, H. Damasio, L. L. B. Ponto, J. Parvizi, and R. D. Hichwa, “Subcortical and cortical brain activity during the feeling of self-generated emotions,” *Nature Neuroscience*, vol. 3, no. 10, pp. 1049-1056, Oct. 2000.

[54] J. R. Gray, T. S. Braver, and M. E. Raichle, “Integration of emotion and cognition in the lateral prefrontal cortex,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 6, pp. 4115-4120, Mar. 2002.

[55] H. Doi, S. Nishitani, and K. Shinohara, “NIRS as a tool for assaying emotional function in the prefrontal cortex,” *Frontiers in human neuroscience*, vol. 7, no. 770, Nov. 2013.

[56] E. Glotzbach, A. Mühlberger, K. Gschwendtner, A. J. Fallgatter, P. Pauli, and M. J. Herrmann, “Prefrontal brain activation during emotional processing: a functional near infrared spectroscopy study (fNIRS),” *Open Neuroimaging Journal*, vol. 5, pp. 33-39, May 2011.

[57] K. Tai and T. Chau, “Single-trial classification of NIRS signals during emotional induction tasks: towards a corporeal machine interface,” *Journal of NeuroEngineering and Rehabilitation*, vol. 6, no. 39, Nov. 2009.

[58] F. Dolcos, K. S. LaBar, and R. Cabeza, “Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: an event-related fMRI study,” *Neuroimage*, vol. 23, no. 1, pp. 64-74, Sep. 2004.

[59] M. Tamietto and B. De Gelder, “Neural bases of the non-conscious perception of emotional signals,” *Nature Reviews Neuroscience*, vol. 11, no. 10, pp. 697-709, Oct. 2010.

[60] H. Ayaz, P.A. Shewokis, S. Bunce, K. Izzetoglu, B. Willems, and B. Onaral, “Optical brain monitoring for operator training and mental workload assessment,” *NeuroImage*, vol. 59, no. 1, pp. 36-47, Jan. 2012.

[61] M. De Vos and S. Debener, “Mobile EEG: towards brain activity monitoring during natural action and cognition,” *International Journal of Psychophysiology*, vol. 91, no. 1, pp. 1-2, Jan. 2014.

[62] H. Ayaz, B. Onaral, K. Izzetoglu, P. A. Shewokis, R. McKendrick, and R. Parasuraman, “Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development,” *Frontiers in Human Neuroscience*, no. 7, pp. 1-13, Dec. 2013.

[63] Y. Liu, H. Ayaz, A. Curtin, B. Onaral, and P. A. Shewokis, “Towards a hybrid P300-based BCI using simultaneous fNIR and EEG,” *In Foundations of Augmented Cognition, Springer Berlin Heidelberg*, vol. 8027, pp. 335-344, Jul. 2013.

[64] M. Balconi, E. Grippa, and M. E. Vanutelli, “What hemodynamic (fNIRS), electrophysiological (EEG) and autonomic integrated measures can tell us about emotional processing,” *Brain and cognition*, no. 95, pp. 67-76, Apr. 2015.

[65] M. Tamietto, L. Castelli, S. Vighetti, P. Perozzo, G. Geminiani, L. Weiskrantz, and B. De Gelder, “Unseen facial and bodily expressions trigger fast emotional reactions,” *Proc. of the National Academy of Sciences*, vol. 106, no. 42, pp. 17661-17666, Jun. 2009.

[66] S. A. Trautmann, T. Fehr, and M. Herrmann, “Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations,” *Brain research*, no. 1284, pp. 100-115, Aug. 2009.

[67] L. R. Rabiner and B. Juang. “An introduction to hidden Markov models,” *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4-16, Jan. 1996.

[68] A. N. Akansu and R.A. Haddad, *Multiresolution signal decomposition techniques: transforms, subbands, and wavelets*, Academic Press, San Diego, CA, USA, 1992.

[69] F. Samaria, and S. Young, “HMM based architecture for face identification”, *Image and Vision Computing*, vol. 12, pp. 537-543, Oct. 1994.

[70] S. P. Lloyd, “Least squares quantization in PCM’s”, *Bell Telephone Laboratories Paper*, Murray Hill, NJ, 1957.

[71] S. Das and K. Yamada. “A hidden Markov model based approach to identify emotion from facial expression using a combination of emotion probability classifier and facial action unit intensity classifier,” *International Symposium on Advanced Intelligent Systems*, pp. 446-449, Sep. 2011.

[72] A. Sanchez, J. V. Ruiz, A. B. Moreno, A. S. Montemayor, J. Hernández, and J. J. Pantrigo, “Differential optical flow applied to automatic facial expression recognition,” *Neurocomputing*, vol. 74, no. 8, pp. 1272-1282, Mar. 2011.

[73] J. M. Buenaposada, E. Munoz, and L. Baumela, “Recognising facial expressions in video sequences,” *Pattern Analysis and Applications*, vol. 11, no. 1, pp. 101-116, Jan. 2008.

[74] I. Kotsia and I. Pitas, “Facial expression recognition in image sequences using geometric deformation features and support vector machines,” *IEEE Trans. on Image Processing*, vol. 16, no. 1, pp. 172-187, Jan. 2007.

[75] T. Wu, M. S. Bartlett, and J. R. Movellan, “Facial expression recognition using Gabor motion energy filters,” *IEEE Computer Vision and Pattern Recognition Workshops*, pp. 42-47, Jun. 2010.

[76] J. Saragih, S. Lucey, and J. Cohn, “Deformable model fitting by regularized landmark mean-shifts,” *International Journal of Computer Vision*, vol. 91, no. 2, pp. 200-215, Jan. 2011.

[77] T. Cootes, G. Edwards, and C. Taylor, “Active appearance models,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681-685, Jun. 2001.

[78] Y. Wang, S. Lucey, and J. Cohn, “Enforcing convexity for improved alignment with constrained local models,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, Jun. 2008.

[79] L. Gu and T. Kanade, “A generative shape regularization model for robust face alignment,” *European Conference on Computer Vision*, pp. 413-426, Oct. 2008.

[80] T. Kanade, J. F. Cohn, and Y. Tian, “Comprehensive database for facial expression analysis,” *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 46-53, Mar. 2000.

[81] S. V. Stehman, “Selecting and interpreting measures of thematic classification accuracy,” *Remote Sensing of Environment*, vol. 62, no. 1, pp. 77-89, Oct. 1997.

[82] J. Wong and S. Cho, “A brain-Inspired framework for emotion recognition,” *Neural Information Processing - Letter and Review*, vol. 10, no. 7, Jul. 2006.

[83] C. Shan, S. Gong, and P. McOwan, “Facial expression recognition based on local binary patterns: a comprehensive study,” *Image and Vision Computing*, vol. 27, no. 6, pp. 803-816, May 2009.

[84] P. T. Trzepacz and R. W. Baker, *The psychiatric mental status examination*, Oxford University Press, New York, New York, USA, 1993.

[85] D. Watson, L. Clark, and A. Tellegen, “Development and validation of brief measures of positive and negative affect: the PANAS scales,” *Journal of Personality and Social Psychology*, pp.1063-1070, Jun. 1988.

[86] Y. Tian, L. Brown, A. Hampapur, S. Pankanti, A. Senior, and R. Bolle, “Real world real-time automatic recognition of facial expression,” *IEEE workshop on Performance Evaluation of Tracking and Surveillance*, Mar. 2003.

[87] P. Ekman, W. V. Friesen, M. O’Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Helder, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti, K. Scherer, M. Tomita, A. Tzavaras “Universals and cultural differences in facial expressions of emotion,” *Journal of Personality and Social Psychology*, vol. 53, no. 4, pp. 712-717, Mar. 1987.

[88] Y. Sun and A. N. Akansu, “Automatic inference of mental states from spontaneous facial expressions,” *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 695-702, May 2014.

[89] A. S. Grove, *Only the paranoid survive: how to exploit the crisis points that challenge every company*, Random House Inc., New York, New York, USA, 2010.

[90] L. A. Bebchuk, “Why firms adopt antitakeover arrangements,” *University of Pennsylvania Law Review*, vol. 152, pp. 713-753, Dec. 2003.

[91] J. W. Barnard, “Narcissism, over-optimism, fear, anger, and expression: the interior lives of corporate leaders,” *University of Cincinnati Law Review, William & Mary Law School Research Paper No. 08-10*, May 2008.

[92] A. Mehrabian, *Silent messages*, Wadsworth Publishing, Boston, MA, USA, 1971.

[93] http://emotions.njit.edu.

[94] Y. Huang, Y. Tie, A. Venetsanopoulos, and L. Guan, “Human emotion recognition using the adaptive sub-layer-compensation based facial edge detection,” *IEEE International Symposium on Circuits and Systems*, pp. 2876-2879, May 2013.

[95] B. K. Boyd, “CEO duality and firm performance: a contingency model,” *Strategic Management Journal*, vol. 16, no. 4, pp. 301-312, May 1995.

[96] K. Gramann, J. T. Gwin, D. P. Ferris, K. Oie, T. P. Jung, C.-T. Lin, L. D. Liao, and S. Makeig, “Cognition in action: imaging brain/body dynamics in mobile humans,” *Reviews in the Neurosciences*, vol. 22, no. 6, pp. 593-608, Dec. 2011.

[97] P. Gruneberg and K. Suzuki, “An approach to subjective computing: a robot that learns from interaction with humans,” *IEEE Trans. on Autonomous Mental Development*, vol.6, no.1, pp.5-18, Mar. 2014.

[98] A. Marchewka, L. Zurawski, K. Jednorog, and A. Grabowska, "The Nencki affective picture system (NAPS): introduction to a novel, standardized, wide-range, high-quality, realistic picture database," *Behavior research methods*, vol. 46, no. 2, pp. 596-610, Jun. 2014.

[99] H. Ayaz, P. A. Shewokis, A. Curtin, M. Izzetoglu, K. Izzetoglu, and B. Onaral, "Using maze suite and functional near infrared spectroscopy to study learning in spatial navigation," *Journal of visualized experiments*, vol. 56, no. e3443, Oct. 2011.

[100] Y. Sun, J. E. Cicon, and A. N. Akansu, "The power of fear: facial emotion analysis of CEOs to forecast firm performance," *IEEE International Conference on Information Reuse and Integration*, pp.695-702, Aug. 2014.

[101] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Networks*, vol. 10, no. 3, pp. 626-634, May 1999.

[102] M. A. Lindquist, J. M. Loh, L. Y. Atlas, and T. D. Wager, "Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling," *Neuroimage*, vol. 45, no. 1, Mar. 2009.

[103] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9-21, Mar. 2004.

[104] S. Koelstra and I. Patras, "Fusion of facial expressions and EEG for implicit affective tagging," *Image and Vision Computing*, vol. 31, no. 2, pp. 164-174, Feb. 2013.

[105] S. K. Sutton, and R. J. Davidson, "Prefrontal brain asymmetry: a biological substrate of the behavioral approach and inhibition systems," *Psychological Science*, vol. 8, no. 3, pp. 204-210, May 1997.

[106] S. Fazli, J. Mehnert, J. Steinbrink, G. Curio, A. Villringer, K. R. Muller, and B. Blankertz, "Enhanced performance by a hybrid NIRS-EEG brain computer interface," *Neuroimage*, vol. 59, no. 1, pp. 519-529, Jan. 2012.