Computational model based emotional state recognition to assist autistic children.

Teoh, Teik Toe.

2011

Teoh, T. T. (2011). Computational model based emotional state recognition to assist autistic children. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/48659>

<https://doi.org/10.32657/10356/48659>

Computational Model based Emotional State Recognition to Assist Autistic Children

**TEOH TEIK TOE
SCHOOL OF COMPUTER ENGINEERING
2011**

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

Acknowledgements

I wish to record the deepest gratitude to my advisor, Dr. David Cho Siu-Yeung who offered prompt, wise, constructive feedback and advice despite his busy schedules. His thoughtful guidance, invaluable suggestions and constant encouragement has been and will continue to be the source of inspiration to me. I have learned a lot about how to be a researcher from his continual passion and serious attitude towards research for which I am truly grateful.

I would like to thank Nanyang Technological University and the School of Computer Engineering for granting me the opportunity to pursue my PhD. I am also grateful to the staff and peers at the Centre of Computational Intelligence (C2I) who have made the centre a very good environment to do research and have helped me in different ways throughout the studies.

Lastly, I would like to thank my family for their continuous support, and encouragement.

Abstract

Children diagnosed with autism, which affects one in every 165, are thought to lack or have impairment in some representational sets of abilities. As a result, they have difficulties operating in our highly complex social environment, and are for the most part, unable to understand other people's emotions. People express their emotion states all the time, even when interacting with machines. These emotion states shape the decisions that we make, govern how we communicate with others, and affect our performance. The ability to attribute emotion states to others from their behaviour, and to use that knowledge to guide one's own actions and predict those of others is known as emotion-recognition.

To allow children with autism to read and respond to the emotions of people, we propose a computer-based device called Cognitive Assistive Computational-based Emotional State Recognition to assist autistic children to understand, interpret and react to emotions of people they interact with. The system is real-time so that computation time is of vital importance to enable real-time interactive training for the system to learn and analyse the human emotions for autistic children. The principal contribution of this thesis is the real time inference of a wide range of emotion states from head and facial displays in a video stream, both pre-recorded (Mind Reading DVD) and live camera. In particular, the focus is on the inference of complex emotion states (agreeing, disagreeing, encouraging, discouraging and unsure): the affective and cognitive states of mind that are not part of the set of basic emotions (in our case is neutral, joy, sad and surprise). The automated emotion state inference system is inspired by and draws on the fundamental role of emotion-recognition in communication and decision-making. The thesis describes the design, implementation and validation of a computational model of emotion-recognition. The design is based on the results of a number of experiments that we have undertaken to analyse the facial signals and dynamics of complex emotion states. In this research, a device will be developed with

camera connected to a computer or a mobile PC. The software, which runs on the computer or mobile PC upon received the images from camera, will recognize the emotional states and pronounce the states through earpiece to children with autism including advises. The whole recognition process is real-time and the training is interactive, such that the knowledge of the system is updated continuously.

This research is believed to be the first attempts based on the combination of pattern recognition and machine learning together with a neuroscience understanding of cognitive and visual signal interplay in solving the above mentioned problems. This work proceeds with some background of Autism Spectrum Disorders (ASD). The state-of-the-arts of existing facial expression recognitions solutions are studied. We then introduce the algorithms of the proposed model in which the model is capable of recognizing four types of facial expression, namely, neutral, joy, sad and surprise. The system is composed of four major blocks, the face locator, the fiducial point locator, the feature extractor, and the classifier. The face locator undergoes some image processing stages to find the face edges. The fiducially point locator finds crucial fiducially points for subsequent feature extraction processing. We adopted Gabor features and Gini extractions that will be discussed in the following chapters. Finally the meaningful features are classified into the corresponding class. A probabilistic based approach using the Naïve Bayesian is adopted to recognize these four types of facial expressions. Emotion indexing is also used to act as a higher-level analysis to interpret the emotional states that include agreeing, encouraging, disagreeing, discouraging and unsure. Generic Algorithm is also used to predict emotion to allow anticipation. With facial expression recognition, emotion index and prediction, we will advise autistic children what to react.

The experiments carried out show that we have achieved the important features of mobile application: speed and efficiency. The system successfully classifies and generalizes to new examples of these classes with a reasonable accuracy (about 75%) and speed (3 frames/sec) that

are comparable to that of human recognition, both results are achieved using PC-based notebook. The research we present here significantly advances the nascent ability of machines to infer cognitive-affective emotion states in real time from nonverbal expressions of people. By developing a real time system for the inference of a wide range of emotion states beyond the basic emotions, we have widened the scope of human-computer interaction scenarios in which this technology can be integrated. This is an important step towards building socially and emotionally intelligent machines.

Table of Contents

Chapter 1. Introduction.....	14
1.1 Overview	14
1.2 Motivation	16
1.3 Contribution	19
1.4 Thesis Outline.....	24
Chapter 2. Literature Review	26
2.1 Introduction	26
2.2 Assistive Tools for Autism	29
2.2.1 Marketable Technologies of Assistive Tools for Autism.....	32
2.3 Research in Emotion and Facial Expression Recognitions.....	42
2.3.1 Emotion Recognition.....	42
2.3.2 Facial Expression Recognition	46
2.3.3 Human Emotional States Recognition for Mind Reading.....	55
2.3.4 Other potential applications of emotion recognition	58
Chapter 3. A Portable Facial Expression Recognition	60
3.1 Introductio	60
3.2 Derivative-based Face Locator	62
3.3 Fiducial Point Locator.....	64
3.4 Feature Extractor	69
3.5 Feature Selection	71
3.5.1 T-test Feature Selection	71
3.6 Classification	75
3.7 Experimental Results	82
3.7.1 Results of Face Localization.....	82
3.7.2 Results of Feature Selection.....	89
3.7.3 Results of Facial Expression Classification.....	92
3.8 Conclusion.....	94

Chapter 4. Hidden Markov Model (HMM) based Emotion Indexing 96

4.1	Introduction	96
4.2	Emotion Indexing in Autism	103
4.3	Hidden Markov Models.....	105
4.3.1	Notation and Definitions.....	108
4.3.2	Likelihood	110
4.3.3	Topology	111
4.3.4	Formulation of State Markov Chain	113
4.4	The proposed Emotion Index Classification Framework using HMM	115
4.5	Proposed Hidden Markov Expert Rule Model (HMER)	120
4.6	Experimental Results - Emotion Indexing with HMER model	122
4.7	Conclusion.....	130

Chapter 5. Genetic Algorithm based Emotion Prediction Model..... 132

5.1	Introduction	132
5.2	Genetic Algorithm for Prediction	136
5.3	Predicting Models	139
5.4	Regression Model	140
5.4.1	Time Series Multiple Regression Model.....	140
5.5	Our Model - GA based Emotion Prediction Model	142
5.6	Experiment Result - Emotion Prediction with Genetic Algorithm.....	149
5.7	Conclusion.....	157

Chapter 6. Fuzzy Rules based Emotion Advisor Model..... 158

6.1	Introduction	158
6.2	Fuzzy System	160
6.2.1	Introduction	160
6.2.1	Fuzzification	162
6.3	Emotion Advisor.....	166
6.3.1	Our Fuzzy Model	168

6.3.2 Fuzzy IF-THEN Rules	170
6.4 Experimental Results and Analysis.....	171
6.4.1 First Phase Results (5 advises/classes).....	171
6.4.2 Final Results (10 advises/classes).....	175
6.6 Conclusion.....	180
Chapter 7. Conclusion and Further Developments	181
7.1 Contribution.....	181
7.2 Conclusion and Future Works	184
Appendix A - Abbreviation	188
Appendix B - Publications.....	190
Appendix C - References	194

List of Figures

Figure 1-1 Examples of autism children	15
Figure 1-2 The basic idea of the proposed emotional aid system to assist autism children	19
Figure 1-3 An overview of the emotional aiding model for an autistic child. There are 4 basic modules, real-time facial expression, emotion indexer, emotion prediction and finally the output is the advisor	21
Figure 2-1 Parts of the brain affected by Autism (Kimmei 2010)	27
Figure 2-2 Chess games for Autistic Children.....	37
Figure 2-3 Example of pictures used for detection of Autism.....	38
Figure 2-4 Video used to play back to monitor activities of autistic children	39
Figure 2-6 Light variations problem: face images are taken from different illumination conditions (source: Yale Face Database B http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html)	47
Figure 2-7 Light variations problem: face images are taken from different illumination conditions (source: NTU Asian Emotion Database.....	48
Figure 2-8 Occlusion problems: facial components are occluded by some artifact objects.(Source: NTU Asian Emotion Database	48
Figure 2-9 The upper facial Action Units (AUs) in FACS (source: Tian et al. 2001).	49
Figure 2-10 The lower facial Action Units (AUs) in FACS (source: Tian et al. 2001).	49
Figure 2-11 Outline of the real-time expression recognition system of Littlewort <i>et al.</i> (2006).....	55
Figure 2-12 User interface of Mind Reading DVD	56
Figure 2-13 Emotion fusion in tree structure (Partially).....	57
Figure 3-1 System diagram of emotion recognition highlighting the Real-Time Expression Recognizer.	61
Figure 3-2 System flow diagram of facial expression recognizer	62
Figure 3-3 Derivative operations for face image	63
Figure 3-4 The AUs from head movement. The examples are quoted from Mind Reading DVD (BGWH04).....	65

Figure 3-5 The AUs related to facial expression. The examples are FACS-coded image sequences from the Cohn-Kanade database (KCT00) of facial expressions.....	66
Figure 3-6 Fiduciary points on the face, basic location is around the eyes, mouth and nose areas.	67
Figure 3-7 Asymmetrical characteristic of a human face. Our experiment is base on both the left and right side of a human face due to their asymmetrical characteristic.....	68
Figure 3-8 Gabor wavelets representation and the overview of optimal excitatory and inhibitory stimuli (S+ res. S-).	71
Figure 3-9 Examples of Gabor Wavelets and corresponding convoluted images.....	71
Figure 3-10 Samples of face images.....	84
Figure 3-11 Samples of non-face images	84
Figure 3-12 Intermediate images (a) Human face, (b) Non-face images.....	85
Figure 3-13 The 3D data distribution by Gini Indexing. (a) Face input, (b) Non face input.	85
Figure 3-14 Scatter Plots (a) Left Eye versus Right Eye, (b) Mouth versus Nose, (c) Left Face versus Right.....	86
Figure 3-15 Statistics of Seven Face Sub-Regions.....	87
Figure 3-16 Result of facial emotion recognition (Using GINI feature selection).....	90
Figure 3-17 Result of facial emotion recognition (Using T-test feature selection)	91
Figure 3-18 Result of facial emotion recognition (Using Euclidean Distance feature selection)	91
Figure 3-19 Result of facial emotion recognition (Using KNN classification, k=5).....	92
Figure 3-20 Emotion Recognition Result (Bar Form)	93
Figure 4-1 Block Diagram for Emotion Indexer (circled in black) which can deduce higher level emotion from basic emotion states	97
Figure 4-2 Four basic emotions together with five emotion/emotional states from emotional indexer	101
Figure 4-3 The proposed emotion indexing framework.	102
Figure 4-4 Illustration of Hidden Markov Model.....	108
Figure 4-5 Choice of Hidden Markov models and topology (a) 4 state ergodic HMM (b) 4-state left- right HMM with a maximum of 2 jumps between states.....	112

Figure 4-6 Three state Markov chain.....	113
Figure 4-7 The displays present in a model of a emotion state are determined by a feature selection mechanism (partially). For simplicity, the display only shows two emotion states.....	116
Figure 4-8 Our Markov Model	118
Figure 4-9 Our Markov Model (Simplified)	119
Figure 4-11An interaction scenario of emotion recognition through the system	124
Figure 4-12 Results of emotion indexer (average).....	125
Figure 4-13 Results of emotion indexer (frames)	125
Figure 4-14 Classification results presented by 3D bar charts and confusion matrices achieved by the HMER model under (a) ergodic Topology, (b) left right topology.	126
Figure 4-15 Emotional state model by the HMER model, (a) From “Unsure” to “Interest” emotional state via a “Joy” transition, (b) From “Interest” to Encouraging” emotional state via a “Surprise” transition.....	128
Figure 4-16 GUI of Emotion Indexer(a) Sadness expression with Disagreeing state (b) Surprise expression with Encouraging state.....	129
Figure 5-1 Block Diagram for Emotion Predictor, predicting emotion from emotion states, also act as one of the three input to Reaction Advisor	136
Figure 5-2 Flow chart for GA population	139
Figure 5-3 Flow chart for GA programming.....	148
Figure 5-4 Flow chart for GA system	149
Figure 5-5 Results of Experiment (Number of Independent Variable)	151
Figure 5-6 Results of Experimental (Number of Generations).....	152
Figure 5-7 Results of “Unsure”	153
Figure 5-8 Results of “Interest”	153
Figure 5-9 Results of “Encouraging”	154
Figure 5-10 Results of “Discouraging”	155
Figure 5-11 Results of “Disagreeing”	156

Figure 6-1 Block Diagram for Emotion Advisor, which takes facial expression, emotion index and emotion prediction as input and generate advises to user	159
Figure 6-2 The aggregations of the three emotional inputs (emotion recognizer, emotion indexer and emotion predictor) to emotion advisor output.....	168
Figure 6-3 Membership function for input 1 (emotion recognizer)	169
Figure 6-4 Membership function for input 2 (emotion indexer)	169
Figure 6-5 Membership function for input 3 (emotion predictor)	169
Figure 6-6 The results obtained by classifiers for data-set 1 & 2 in percentage	172
Figure 6-7 The results obtained by classifiers for data-set 3, 4 & 5 in percentage	173
Figure 6-8 The results obtained by classifiers for data-set 6, 7 & 8 in percentage	174
Figure 6-9 The test results obtained by classifiers for data-set A & B in percentage.....	176
Figure 6-10 The test results obtained by classifiers for data-set C & D in percentage	177
Figure 6-11 The results obtained by classifiers for the eight data-sets in percentage	179

List of Tables

Table 2-1 Summary of some examples of Assistive Tools for Autism.....	33
Table 2-2 Strength and weakness analysis of the marketable technologies of assistive tools for autistic children	41
Table 3-1 Benchmarking results with different classifiers used for face locator	88
Table 3-2 Accuracy of different combinations of feature selection methods and classification methods (the values are in percentages of average and maximum).....	92
Table 3-3 Benchmarking results of different classifiers used for our proposed portable facial expression recognition	94
Table 4-1 Results comparing our model against other real-time facial expression recognition engines performing onto public domain databases.....	127
Table 5-1 Results for the three criterion	151
Table 6-1 Examples of fuzzy IF-THEN designed for the emotional advisor	170

Chapter 1. Introduction

1.1 Overview

Symptoms of autistic children include avoiding eye contact, lacking of interest in playing and social interaction (see Figure 1-1 as examples). Mind-reading is the terminology used in psychology to describe people's ability to attribute emotional states to others from their behaviour and to use that knowledge to guide one's own actions and predict those of others (Hart 2005). Those who lack the ability to do so have difficulties operating in the complex social world in which we live and are sometimes referred to as mind-blind. People diagnosed with Autism Spectrum Disorders belong to this group. Besides social competence, emotion-recognition has a central role in the processes underlying decision-making, perception and memory. Recent findings in neuroscience show that emotions regulate and bias these processes in a way that contributes positively to intelligent functioning. In decision-making, our own intentions, the social context and our appraisal of other people's emotional states all affect the choices we make every day (Baron-Cohen *et al.* 2004). Today, rapid developments in technology can help create a new future for people with autism spectrum.

Figure 1-1 Examples of autism children

Visual cues like facial expression, body posture and gesture are an important carrier of emotional information. Facial expression is used widely in all cultures and civilizations to express and perceive emotion (Whitehill 2006). In an emotion recognition system, the basic task is to extract those features from a face that are most indicative of a person's emotional state. These included eyes, eyebrows, mouth and nose-bridge. The extracted features can be analysed statistically or dynamically by monitoring and measuring the variation of these features within specified duration of time. In conventional computational studies, face detection or tracking is the first task in any facial expression recognition approach. This task can be based on color as a clue to detect face from background. Pose estimation may be involved in the face tracking as different viewing angles cause substantial change in the appearance of the face. After face tracking, mapping of facial features to emotion is essential for the system. One of the approaches is to conduct statistically at the apex or extremity of the expression (mug-shot) with the aim of detecting the presence of static cues such as wrinkles as well as the positions and shapes of facial features. Another approach is to require a set of successive frames of a facial expression, as it involves the measurements of image properties on a frame-by-frame basis so that gradients and variances can be extracted.

The major objective of this research is to develop an assistive aided device for helping children with autism to improve their skills in emotional and social communication. In summary, although most of the facial expression analysers developed so far target on human facial affect analysis and attempt to recognize a small set of prototypic emotional facial expressions like happiness and anger; some progress has been made in addressing a number of other scientific challenges that are considered essential for realization of machine understanding of human facial behaviour. Existing methods for machine analysis of facial expressions discussed thus far assume that the input data are near frontal- or profile-view face image sequences showing facial displays that always begin with a neutral state. In reality, such assumption cannot be made. The existing facial expression analysers were tested on spontaneously occurring facial behaviour, and extract information about facial behaviour in less constrained conditions such as an interview setting. However deployment of existing methods in fully unconstrained environments is still in the relatively distant future. Development of robust face detectors, head and facial component trackers, which are robust to variations in both face orientation relative to the camera, occlusions, and scene complexity like the presence of other people and dynamic background, forms the first step in the realization of facial expression analysers capable of handling unconstrained environments.

1.2 Motivation

Human employs a variety of nonverbal cues consciously or subconsciously, such as vocal nuances, posture and facial expressions to express their emotions and mental states. The automation for recognition of these cues is an open research problem, making the development of a comprehensive emotion-recognition machine an ambitious undertaking.

This thesis addresses the problem of detecting and decoding of emotion from visual expression. Human face possesses excellent expressive ability and provides one of the most powerful, versatile and natural means of conveying a different emotional states. Facial expressions communicate feelings, behavioral intentions, show empathy and acknowledge the actions of other people (Ekman 1969). The possibility of enabling human computer interfaces to recognize and make use of the information conferred by facial expressions has gained significant research interest over the last few years. This has given rise to a number of automated systems that recognize facial expressions in images or video. The starting point of this thesis is the observation of automated facial expression analysis systems which encompasses three problems. The first is the problem of facial action analysis or identifying the basic units of facial activity such as an eyebrow raise. This is essentially a perceptual task, which is a necessary but an insufficient component of emotion-recognition. The second problem is the recognition of basic emotions—happy, sad, neutral, surprise. The third is the inference of high-level emotion, namely the emotion indexing, in which the secondary emotional states may consist of agreeing, disagreeing, encouraging, discouraging and unsure. Recognizing the basic emotions from facial expressions is of limited use in understanding user's cognitive state of mind and intentions. These cognitive states and intentions are more relevant and frequent in Human Computer Interface (HCI) context where the user is typically performing some task (Turk 2005). For example, even though mild forms of fear towards computers are common among inexperienced users or learners (fear is a basic emotion), they are less frequent in an HCI context than cognitive emotion states like thinking or concentrating. The result is that the application of automated facial expression analysis to

human computer interaction is limited to primitive scenarios where the system responds with simple positive or negative reactions depending on which basic emotion the user is in. The range of emotion states that people express and identify extend beyond basic emotions, to include a range of affective and cognitive emotional states which are collectively referred to as complex emotion states. These states encompass many affective and cognitive states of emotion, such as agreeing, confused, disagreeing, interested and thinking. Part of the objectives is to use the ability of machines to infer complex emotion states from a video stream of facial expressions of people. In addition, recognizing complex emotion states widens the scope of applications in which automated facial expressions analysis can be integrated, since these emotion states are indicators of the user's goals and intentions. Hence, the automated inference of complex emotion states serves as an important step towards building socially and emotionally intelligent machines that improve task performance and goal achievement. Our final objective is to develop a working prototype of an automated emotion state inference system. The automated inference of complex emotion states from observed behavior in the face involves a number of challenges. Emotion state inference involves a great deal of uncertainty since a person's emotional state is hidden from the observer, and can only be inferred indirectly by analyzing the behavior of that person. In addition, the automated analysis of the face in video is an open machine-vision problem that is the concern of many research groups around the world. Lack of knowledge of how facial expressions are mapped into corresponding emotion states is another challenge.

This research is targeted towards the development of emotional aid system for autistic children. The proposed system consists of a pair of glasses with miniature camera

connect to a PDA or a mobile PC. The software running on the portable device can recognize facial emotion and pronounce the emotions through earpiece to autistic children upon receiving the image from camera. The whole recognition process is real-time and the training is interactive, meaning the knowledge (database) of the system is continuously updated. Figure 1-2 shows the conceptual idea of this system. The proposed system consists of a pair of spectacles, a camera, earpiece, and a PDA. The miniature camera is mounted beside the frame of the spectacles. The earpiece is meant for providing response to the autism children. The miniature camera and earpiece are hooked up with the PDA where the system resides. The image is captured from camera and is sent directly to the system; the system then performs facial emotion recognition and pronounces the emotion through the ear piece. The whole training and recognition process is live and interactive.

Figure 1-2 The basic idea of the proposed emotional aid system to assist autism children

1.3 Contribution

This work was initially targeted towards the development of an efficient and intelligent facial expression recognition system. The system is capable of locating the face region using derivative-based filtering, and classifying human face through the use of

probabilistic based classifier. The system is tested using PC-based notebook except the emotion recognizer module is light-weighted and is deployed under Windows Mobile environment with short computation time. The motivation behind this work is that we aim to develop a robust model that can help to locate a face for a portable face recognition application that assists children with autism to learn how to communicate with others. The underlying algorithm adopted in this work uses Boosting Naïve Bayesian (BNB) approach for recognition. We examined the structure of training data and the effect of attributes on the class probabilities through the use of Naïve Bayesian classifier (NBC). The experiments carried out show that we have achieved the important features of mobile application: speed and efficiency. This work seeks to provide a launching point for a sound and portable mobile application that is capable of recognizing different facial expressions.

The proposed prototype of this assistive device for children with autism is shown in Figure 1-3. This system can identify a facial event in real time, extracts the dynamic features from facial expressions and infer the underlying emotional state conveyed by the video segment. More works worth looking into includes a reaction advisor that utilizes a GUI to display current emotion state inference and a recommended action, both textually and graphically. They will include the implementation of the emotional indexer using partially observed Markov decision processes, so that the utility of the actions is

Figure 1-3 An overview of the emotional aiding model for an autistic child. There are 4 basic modules, real-time facial expression, emotion indexer, emotion prediction and finally the output is the advisor

also learnt from data rather than hard-coded as in the current rule-based implementations, hence utilizing the information to suggest an appropriate reaction. Representative frames

of that event along with the inferred emotion states label are sent to the emotional indexer to be archived. The emotional indexer module is responsible for keeping an archive of past events; every event is stored as a tuple in the index containing the representative frames of the emotion states with additional parameters and context cues available. The indexed events are made available through the interface layer to the child engaging with his/her tutor/parents for discussion, learning and reviewing purposes. This extends the emotional indexing approach to allow events to be replayed. The archive is also made accessible to the system and the reaction advisor modules to improve inference and suggestions. The reaction advisor appraises the current video input, analysing it within any contextual cues that are available, to suggest appropriate courses of actions to take. Timing issues such as latency and frequency of reactions are key factors in the design of this module. A rule-based version will be implemented in this module. We aim to develop the system to successfully classify and generalize to new examples of other emotion state classes with an accuracy and speed that are comparable to that of human recognition. With this system, an autistic child will make use of this aiding device that will inform them on the facial emotional brings into their towards awareness or attention during the social communication. With successful development of this system, a number of patents will be filed.

Along with this development of the proposed system, the major contributions are listed as entries below:

1. Real-time Facial Expression Recognition: We propose to develop a computational framework that is able to recognize basic facial expressions in real time. Facial features can be extracted using Gabor filtering and the most important features

- can be selected by Gini Index in order to reduce the number of features needed for processing in real time. Boosting Naïve Bayesian classifier can be used to classify the facial expressions (See Chapter 3 in details).
2. Emotion Indexing: We make use of the established facial expression recognition system to further comprehend more involved emotional state like agreeing, disagreeing, interesting, discouraging and unsure from various modalities. The proposed model will mimic human meta-cognitive operation that focuses and reacts to the human emotional states. Hidden Markov Model (HMM) will be used and further improved for inferring and indexing the human secondary emotions (see Chapter 4 in details).
 3. Emotion Prediction: Emotion predictor is developed to predict the human emotional behaviour in order to enhance the accuracy of the system outcome. Genetic Algorithm (GA) based Multi-Regression Time Series model is developed to predict the emotional behaviour based on the primary and secondary emotions capturing from the facial expression recognizer and HMM emotion indexer (see Chapter 5 in details).
 4. A final computational model is developed to assist and advise autistic children to understand emotions. More than ten different types of advices can be produced based on a production rule-based system trained by a fuzzy inference engine. A broad category of affective and cognitive states of mind is able to generate to advise autistic children beyond the basic emotions (see Chapter 6 in details).

1.4 Thesis Outline

This thesis consists of 7 chapters. The thesis begins with Chapter 2 where the literature reviews on assistive technology for autism are presented which forms the foundation of this work. Here, the currently marketable assistive tools are introduced. We then proceed to introduce the research on emotion and facial expression recognition. We provide distinction between emotion and facial expression. The concept of mind reading is then described and the potentials of applying emotion recognition is reviewed.

In Chapter 3, we introduce the model Cognitive Assistive Probabilistic-based Emotional State Recognition (CAPESR). CAPESR is a portable facial expression recognition system.

Chapter 4 investigates emotion indexing based on Hidden Markov Model (HMM). This chapter starts with emotion indexing in autism. The workings of Hidden Markov Model are then illustrated with notations, likelihood, topology and its formulation.

Chapter 5 illustrates a genetic algorithm model for emotion prediction. The genetic algorithm model is a prediction model that can be used to predict the next emotion based on its current and historical emotions. The regression model is then presented. The chapter is then preceded with the GA model in emotion prediction, supported with experimental results and conclusion.

Chapter 6 focuses on emotion advisor using fuzzy sets and rules. The important IF-THEN rule in fuzzy domain is illustrated. There are two phases, the exploring phase and final testing phase.

Finally, Chapter 7 concludes the researches described in this thesis and some further studies are described.

Chapter 2. Literature Review

2.1 Introduction

In psychology, the term ‘emotion-recognition’ is defined as the ability to attribute emotion states to others based on their behaviour, and to apply this knowledge in guiding one’s actions and predicting those of others. Emotional states that can be expressed and attributed include affective states or cognitive states, intentions, beliefs and desires (Koyama 2009). Emotion-recognition is instrumental in ascertaining the intent of an interaction, take account of others’ interests during conversations, empathize with them and persuade them to take certain course of actions. Most people can perform emotion-recognition subconsciously all the time, and with little difficulty. Individuals who do not possess the ability to do so are sometimes referred to as mind-blind. People diagnosed with Autism Spectrum Disorders (ASD) falls into this category. These individuals face difficulties in interacting with the complex social world on a daily basis. Besides being socially-competent, emotion-recognition plays an important role in processes underlying decision-making, perception and memory. Researches in neuroscience reveal that emotions regulate and bias these processes, contributing positively to intelligent functioning in humans. In decision-making, factors such as our intentions, the social context and our appraisal of other people’s emotion states influences the choices we

make each day. The lack of emotional aptitude in people who suffered from traumatic brain injuries has resulted in impaired reasoning. Consequently, this advocates the notion that emotions should be embedded within models of human reasoning. The ability to perform emotion-reading has been displayed since childhood. Children who are 18 months old are able to refer to a variety of emotional states such as emotions, desires, beliefs, thoughts, dreams and pretence. Most of them can attribute many emotional states to other people by the age of five. Some children can even use this knowledge derived to predict and manipulate these people's actions. Figure 2-1 shows parts of the brain affected by Autism.

Figure 2-1 Parts of the brain affected by Autism (Kimmei 2010)

The deficit of, or impairment in, the ability to rationalize emotional states is known as mind-blindness. This condition is seen to be the key inhibitor of social and

emotional intelligence for autistic people. Characterized by abnormalities in various domains, Autism is a spectrum of neuro-developmental conditions which affects one's social functioning, communication and is often accompanied with repetitive behaviours and obsessive interests. Inabilities resulting from mind-blindness include gauging the interest of other parties during conversations, withdrawal from social contact, oblivion to social cues, indifference to people's opinions and incomprehensible non-verbal communication.

Despite that, the nature of autistic intelligence seems to find its place in the area of identifying and characterizing patterns. This observation is supported by the performance of autistic participants on the Ravens Progressive Matrices test of fluid intelligence; the scores were substantially higher as compared to the Wechsler IQ Test (WISC-III Full -Scale for Children and WAISIII Full-Scale for Adults). On the other hand, the scores attained by the non-autism groups for both tests were statistically similar. These findings suggests that certain aspects of autistic intelligence may have been severely undermined (Dawson *et al.* 2007). Baron-Cohen (Baron-Cohen 2006) emphasized autistic people to be strong systemisers. This involves law detection via observations of input-output relationships, and the facilitation of search structure in patterns, rules, regularities and periodicity in data. By acknowledging the power of coupling pattern understanding and systemizing talents with access to unprecedented sets of data, it is probable to expect valuable contributions from individuals with first-hand experience of autism. Fuelled by their fascination for patterns, this special group of people is capable of devoting their natural talents and time to support research efforts in such specialized field of interests.

2.2 Assistive Tools for Autism

Autism, a pervasive neuro-developmental disorder, is primarily characterized by difficulties in social, language and communicative domains (Stanfield *et al.* 2008). In the recent years, autism has been receiving an increase in recognition by the general public, as well as clinicians and researchers. Based on the social-cognitive theories of autism, the Theory of Mind suggests that one of the major challenges experienced by autistic children is making interactions with the natural social environments (Koyama 2005). Another distinctive feature of having the “Theory of Mind” is to be able to make inference of one’s emotions based on their facial expressions. Studies are carried out to illustrate the difficulties of adult autistic subjects in inferring emotions in the eyes (Baron-Cohen 1997; Kleinman *et al.* 2001). fMRI evidence generated from similar tasks of emotional inference also depict decreased activation of the amygdale and increased activation in the superior temporal gyrus (Baron-Cohen 1999), areas classically implicated in emotional perception and language comprehension. One common characteristics of autistic behaviour is the lack of eye contact and this is confirmed by eye-tracking studies which exhibited decreased attention on the eyes and increased attention on the mouth or the surroundings. Given this observation, our research interest extends computer-based technologies that can help autistic children by therapy or providing assistance to them. The basis of the research in this proposal is largely influenced by the impairment of empathizing abilities, in addition to the above-average systemizing skills of people diagnosed with autism.

The emotional hearing aid is a portable, computerized device developed to assist children with Asperger syndrome read and respond to the people whom they interact with

by the facial expressions displayed. The device aims to replicate two important traits that enable one to empathize with others. The first trait is the ability to identify and attribute an emotion state to someone, which is known as emotion-recognition. The other trait is sympathizing, the ability to produce an appropriate response to that emotion information.

This thesis highlights the progress in the development of the emotional hearing aid on two fronts. Firstly, the introduction begins with the description of the reaction advisor, and subsequently documents its implementation, presenting how the persistence, intensity and degree of confidence of an emotion state inferred is taken into account. An automated emotion-recognition system is employed. This real-time system associates an emotion state to a person by observing his or her behaviour. After that, the reaction advisor generates a recommended reaction based on the suggested emotion identified by the emotion-recognition system. Secondly, an experimental evaluation of the emotion-recognition system on five different classes of complex emotion states is conducted. The thesis concludes with a discussion of the challenges that is required to be addressed in the development and validation of the emotional hearing aid.

The research in this thesis draws inspiration from various disciplines. We present the different theories on how humans perceive and interpret emotions and emotional states of others, and examine existing researches done on how to enable computers to mimic some of these functions. We first present the literature for basic emotions, which have received most of the attention to date, and then continue to explore in researches that consists of other emotion states. We conclude the chapter by highlighting the shortcomings of automated facial analysis systems in dealing with a range of emotion states.

According to the Technology-Related Assistance for Individuals with Disabilities Act of 1988 (Public Law 100-407), an assistive technology means any item, piece of equipment, or product system, whether acquired commercially, off-the-shelf, modified or customized, that is used to increase, maintain, or improve functional capabilities of individuals with disabilities. Assistive technology service refers to any service that provides direct assistance to a person with certain disability in the selection, acquisition, or usage of devices equipped with assistive technology.

Autistic children are more apt in processing visual information as compared to auditory information. When assistive technology devices are used with these children, the devices are essentially feeding them information via their strongest processing channel, which is by visual means. Hence, in order to improve the functional capabilities of autistic children, various technologies ranging from low-end to high-end should be integrated into different aspects of daily life.

The following outlines the various skill areas that are commonly associated with children diagnosed with autism and the corresponding supportive technologies strategies defined.

Low-end Technology is usually visual support strategies that excludes the uses of any electronic or battery-operated devices. These are typically low-cost and simple-to-use equipment. Some examples include dry erase boards, clipboards, ring binders, file folders and photo albums.

Middle-range Technology involves battery-operated devices or simplistic electronic devices that do not require advanced technologies. Examples of such devices include tape recorder, overhead projector, timers and simple voice output devices.

High-level Technology is associated with complicated technology-related support strategies that make use of expensive equipment. Some examples include video cameras, computers, and complex voice output devices.

2.2.1 Marketable Technologies of Assistive Tools for Autism

Studies also have showed that computers serve as great assistive technological tool because autistic children are driven by predictability and consistency (Strickland 1996). This is also the reason why computers are the most suitable assistive tool for autistic learning. The ease of using computers is brought to autistic users by having adaptive hardware devices. The child is placed in the driver's seat for learning purposes and also to encourage independent functioning. It is revealed that autistic students who are computer users have increased attention spans, are able to remain in their seats for a longer duration, develop improved fine motor skills, and show greater ability to generalize skills across environments, which includes repeating a wanted behavior at home that was learned previously in school (Dautenhahn 2002). Incorporating adaptive hardware for computer usage has also been instrumental in reducing behavioral patterns such as agitation, perseveration (uncontrollable repetition) and self-stimulatory behavior. Due to the massive benefits computers can provide, they should be an integral part of a special education student's daily curriculum instead of merely labeled as kids' play toy. Some autistic children can stand to benefit from adapting a standard computer with certain devices that can make it easier for the child to use a computer. For instance, a touch screen allows the child to touch the computer screen instead of using a mouse. With this, the child's actions produce changes on the monitor, thus increases the child's awareness of cause and effect. Table 2-1 shows some examples of assistive tools for

autism with the summarized descriptions. Their details are shown in the following sub-sections.

Assistive Technology	Electronic Devices Motivate and Improve Communication Skills in Children with Autism.
AtKidSystems	Education software for color recognition, spatial awareness, directional awareness, numbers, and shapes.
Byonetics	An "In Home" program proving successful in enhancing speech, behavioral, and occupational therapies for autistic children.
Cognitive Enhancement System	A subscription website providing online cognitive enhancement applications for all children ranging from gifted to those with learning disability, attention deficit, developmental disability and autism.
Computer and electronic games	Develop computer games can help children with autism in educational areas such as learning new vocabulary, practicing math skills or improving eye-hand coordination.
Digital Voice Technologies	Augmentative and Alternative Communication (AAC) Devices for children with autism.
Early Autism Detector	Using eye tracker technology that measures eye direction while the babies look at faces, eyes, and bouncing balls on a computer screen to detect autism.
eHow Contributor	Develop computer parts for an autistic student include a trackball mouse is a larger mouse navigated by a large roller ball.
IntelliKeys® keyboard	Provide a versatile enlarged keyboard that plugs into any Macintosh or Windows computer with a simple cable change.
Laureate Learning Systems	The system was specially designed for individuals with Autism. It offers over 50 research-based software programs that train cause and effect, basic vocabulary, grammatical forms, and language concepts.
TEXTHELP Software	A text to speech software that has put to use its specialized knowledge to develop a range of software products designed to assist individuals to improve their reading and writing abilities.
Videotaping Interaction	Provide videotaping interactions allows teachers or parents to replay situations and evaluate the cause of particularly good or bad behavior.

Table 2-1 Summary of some examples of Assistive Tools for Autism.

(A) Assistive Technology: Electronic Devices Motivate and Improve Communication Skills in Children with Autism

When a child experiences difficulty in communicating, this means that their self-esteem, ability to grasp concepts and socializing skills are directly affected. The ability to

communicate is something that people often take for granted. Children diagnosed with autism are presented with communication problems ranging from preverbal to verbal difficulties and they struggle with mundane tasks such as socializing with peers. However, they can now hear their voice with the aid of a user-friendly device called a Voice Output Communication Aide (VOCA). A VOCA is an electronic device designed to help children to overcome the challenges in communication and improve their relationships with people around them. The child communicates by pressing a button that activates pre-recorded messages customized for each child. The touch pads on the device contain visual representation in words, line drawings or pictures that describes the contents of the recording. For instance, the recorded message would play, “Hug Me”, when the picture shows two people with arms around each other. The VOCA falls under the category of soliciting middle-range technology. With this technology, autistic children can improve their communication, understanding and social skills. Devices like this are attractive to children and enable them to focus in class and be more active in class participations.

(B) AtKidSystems: Education software for color recognition, spatial awareness, directional awareness, numbers, and shapes.

AtKidSystems is used to help preschool children with autism, developmental delays and some other learning problems. It guides them in areas like color recognition, directional awareness, spatial awareness, and in identifying numbers and shapes. Cosmo’s Learning Systems (CLS) is a unique and ingenious system that features Mission Control, a computerized device that employs aFFx™ activator technology which senses duration and pressure, increasing interactivity and adaptability of the access device. Mission

Control consists of four aFFx activators, a built-in microphone, and gives user the option of plugging into any individual switch. Using the Mission Control, the child can explore Cosmo's Play and Learn Playground Discovery software, which guides the child through the basics of words and numbers. This is a fun and educational software that autistic children can enjoy and learn from. The software allows customization for each child, complete with data tracking, matching magnetic theme board and curriculum guide. MC Commander, an application that permits the interfacing of Mission Control using off-the-shelf software, provides the user the option of designating Mission Control to function as a mouse or keystrokes from keyboards. These software options are available to further enhance the accessibility of the software for children who are limited by the access device.

(C) Byonetics: an "In Home" program proving successful in enhancing speech, behavioral, and occupational therapies for autistic children

Byonetics, created by Jean Genet, is a software developed to assist autistic children. He conquered his autism and used his own research in brain management to help other autistic people. Byonetics is based on the concept that the brain is akin to a computer that utilizes brain wave frequencies to activate development switches. These switches serve as a connection to our mental, emotional and physical firmware that triggers our innate ability to speak, to maintain emotional balance and help us to stay focused. Cranial Dynamics™ technology is applied for the creation of harmonic frequency codes that the brain uses to repair these switches. Harmonic frequencies are encoded into digitally-mastered CDs and replayed to the children before sleeping. By repairing more of these switches, the connection between the brain and its firmware improves. The effects of

having repaired switches can be shown by better speech ability, emotional balance and concentration skills. Byonetics proves to harness success in enhancing speech, behavioral and occupational therapies. Part of its campaign also includes providing the autistic child with daily support so as to resolve conflicts with autism. Ending autism is a completely different playing field as curing autism.

(D) Cognitive Enhancement Systems: COM is a subscription website providing online cognitive enhancement applications for all children ranging from gifted to those with learning disability, attention deficit, developmental disability and autism.

Challenging Our Minds (COM) is a cognitive enhancement system constructed to enhance attention, memory, visuo-spatial, problem solving, communication and psychosocial skills of children from all walks of life. This cognitive system has proven to work effectively on children ranging from those who suffered traumatic brain injury to those who are undertaking gifted, educational programs. COM is an online service that executes within the web browser. The weblog houses information and guides on the use of the system. Once the subscriber registers and receives the login details, he is able to access the system from anywhere that has Internet connection. Many educational institutions have registered for this service.

(E) Computer and electronics games

Computer games can be assistive tools to autistic children in educational learning such as picking up new vocabulary, practicing mathematical skills or improving eye-hand coordination. Carefully selected video games for children with autism can be very beneficial. The FaceSay software that aims to improve the social interactions of students with Autism has proved to help kids on the spectrum to recognize emotions, facial

expressions and facial features, according to the University of Alabama study. Figure 2-2 shows how a chess game could help autistic children.

Figure 2-2 Chess games for Autistic Children

(F) Digital Voice Technologies: Augmentative and Alternative Communication (AAC) Devices for children with autism.

Digital Voice Technologies produces augmentative and alternative communication devices like MV-1000 Handheld Touchscreen Solution that are specially catered to help children with conditions such as autism, traumatic head injuries, developmental delay, cerebral palsy and other language problems. These devices allow children who are unable to speak, to be able to communicate with others and express their thoughts and feelings through touch technology. Their speech device solutions are of premium quality, and these speech devices are being used in assisting their own autistic children as well.

(G) Early Autism Detector:

The Early Autism Study, spearheaded by Mel Rutherford, uses eye tracker technology to measure the eye direction of babies looking at faces, eyes, and the bouncing balls on a computer screen. Figure 2-3 is an example of pictures used for detection of Autism.

Figure 2-3 Example of pictures used for detection of Autism

The crux of this study brings down to the fact that a group of siblings with autism can be distinguished from a group with no autism -- at 9 months and 12 months. This diagnosis can be accomplished in minutes, and it is objective, meaning that the only measurement is eye direction; it's not influenced by a clinician's report or by intuition. Currently, the earliest diagnostic test for autism is reliable around the age of two, and most children in Ontario are diagnosed around age three or four. Early diagnosis of autism can improve the overall prognosis. Autism requires a quick, reliable and objective screening tool to aid in diagnosing autism much earlier than is presently possible. Figure 2-4 shows how a video is used to play back to monitor activities of autistic children.

Figure 2-4 Video used to play back to monitor activities of autistic children

(H) eHow Contributor: Assistive Computer Hardware

As compared to convention computer peripherals, assistive computer hardware allows autistic users to use a computer with little difficulty. For example, autistic children find a trackball mouse that is navigated by a large roller ball easier to use than a standard computer mouse. Some assistive technological devices include computer touch screen, Voice Outpoint Communication Aide (VOCA), the Language Master, Board Maker and Picture Exchange Communication Systems (PECS). Touch screen computers are implemented in many computer programs, allowing autistic children to select correct answers for questions or problems just by touching the screen. The VOCA assists autistic children in learning to speak via auditory feedback. Upon hearing the exact words, the child starts learning by mimicking them. Approximately the size of a tape recorder, the Language Master reads programmed cards and can be connected to a computer by cables. The Board Maker produces visual pictures that can be used by autistic children for communication means. This program is also capable of creating worksheets using images. PECS has the ability to create pictures and cards with appropriate labels for them.

A summarized table of the some assistive tools technologies for autism is tabulated in Table 2-2. Both strength and weakness are presented to analyze those technologies. We could see from Table 2-2 that basically there are 3 types of assistive tools, first is detection, second is curing and lastly is to assist autistic patient. Our system should be benchmark with assisting autistic patient like IntelliKeys. However, our approach is very niche and currently there is no identical product to compare with.

Assistive Tools for Autism	Strength	Weakness
Assistive Technology	Help to improve the communication skill	Need to customized
AtKidSystems	Create a lot of fun for the children	Lack of final result, how to improve in their social life or career
Byonetics	Enhances speech and behavioral	Cannot cure autism
Cognitive Enhancement System	Enhance children's attention, problem solving, communication and psychosocial skills	Web base, could not use without Internet
Computer and electronic games	Learning new vocabulary, practicing math skills or improving eye-hand coordination	Need supervision, could not standalone
Digital Voice Technologies	Allow children who cannot speak to communicate; giving them the ability to express themselves through touch technology.	Do not cure or improve their skill or knowledge
Early Autism Detector	Can distinguish between a group of siblings with autism from a group with no autism – from 9 to 12 months in 10min	Only detection, no cure or solution
eHow Contributor	Provide computer parts for an autistic student include a trackball mouse is a larger mouse navigated by a large roller ball.	Did not help in improving their soft skill like communication.
IntelliKeys® keyboard	Enabling the user to easily type, enter numbers, navigate on-screen displays, and execute menu commands.	Did not help in improving their soft skill like communication, same as eHow.
Laureate Learning Systems	It enables individuals with special needs to work independently to reach their full potential and experience success through training them the cause and effect, basic vocabulary, grammatical forms, and language concepts.	Does not cure Autism
TEXTHELP Software	Provide a range of software products designed to assist individuals to improve their reading and writing abilities.	Did not cover the communication problem.
Videotaping Interaction	Videotaping interactions allows teachers or parents to replay situations and evaluate the cause of particularly good or bad behavior.	Need supervision, not automatic

Table 2-2 Strength and weakness analysis of the marketable technologies of assistive tools for autistic children

2.3 Research in Emotion and Facial Expression Recognitions

2.3.1 Emotion Recognition

Emotion-recognition is defined by the set of representational abilities that enables one to draw inferences about other people's emotions. Following the works of Baron-Cohen *et al.*(1995), this thesis takes a scientific approach in the application of emotion-recognition, denoting the set of abilities that a person utilizes for inferring the emotional states of others based on nonverbal cues and observable behavioural patterns. From the observer's point-of-view who performs emotion-reading, the input is essentially an array of observations, such as visual, auditory, tactile stimuli and contextual cues. On the other hand, the output is a set of emotion states attributed to other people, determined by the analysis of the details previously mention. People often exhibit and attribute mental states to each others. Some of these mental states include emotions, beliefs, desires, intentions, cognitive states, and focus of attention. In literatures of developmental psychology, emotion-recognition is referred to as a specific faculty that can be separable from generic cognitive abilities like general intelligence and executive function (Montse Pard`as 2002). In recent years, the interest generated in the mechanisms and functions of emotion-recognition has evolved to become a noteworthy and emergent problem for cognitive scientists. With the findings, one can understand how cognitive skills that enable high-level social cognition are compartmentalized in the human brain, and the roles they play in everyday functioning. These findings also form the crux of the computational model of emotion-recognition in the later chapter.

Emotion recognition may seem vague and subtle but it is vital for social functioning which many of us may have taken for granted. It is an important component

of a more general skill set known as social intelligence. It is credit to emotion recognition that we are capable of making sense of other people's behaviour and thus predict their subsequent actions. Besides giving us the ability to effectively communicate with other people (Lacava *et al.* 2007), emotion recognition is also termed as a cognitive component of empathy. A good empathizer can instantaneously sense a change of emotion in someone else, the causes underlying the change, and what may help that person to feel better. Being proficient in recognising emotions also brings about having a greater capacity in persuasion and negotiation. Having the awareness that people's thoughts and beliefs are moulded by the information they possess can in turn influence people's decision to change their thinking or their action plan. Emotion recognition is also a major contributing factor in processes such as perception, attention, learning, memory and decision-making. Studies reveal that brain areas being responsible for such functionalities are interconnected to other brain structures which are involved in the initiation and selection of future behaviours (Wong and Cho 2006). These results emphasize the intricate relationship of emotion and cognition, and thus guide us to a new level of understanding of the human brain, such that the brain is more than a cognitive information processing system. Instead, it is a hybrid system in which affective and cognitive tasks are conjugated in an inexplicable way. Emotion recognition capabilities of humans are displayed since childhood. Children are able to exhibit mental states such as emotions, thoughts, dreams, desires, beliefs and pretence at a tender age of 18 months (Howlin 1999). By the age of five, most of them can attribute a wide range of emotion states to people they interact with, and use this information to predict or manipulate the actions of these people. Mind-blindness is the disability of reasoning with emotional state

and is deemed as the key inhibitor of social and emotional intelligence in people diagnosed with Autism Spectrum Disorder (ASD). Autism Spectrum Disorder (ASD) is a complex neuro-developmental disorder that is characterized by impairments in social interaction such as language skills, in particular, social communication. People suffering from autism exhibit abnormalities in social functioning, communication and display compulsive repetitive behaviours and obsessive interests (Association 1994). Mind-blindness results in the failure to gauge the interest of others during conversations, oblivion to social cues, withdrawing from social contacts, incomprehensive nonverbal communication and being apathetic to other people's feelings or opinions.

There are two components that affects emotion recognition in people. They occur in different parts of the brain and are developed at a certain age. Different people may experience impairments of one or both components. The first component consists of the social-perceptual aspect of emotion recognition (Sullivan 2000), namely detecting or decoding the emotional states of others based on observable information that are available. For instance, one could judge or distinguish the emotional state of another by perceiving their facial expressions or listening to the tone of his or her voice. As its name suggests, this component involves bottom-up processing of facial or other stimuli. It also encompasses cognitive abilities, or top-down processing of abstract models that defines how one's behaviour can be associated with certain emotion states (Aggarwal, 2004). The next component to be highlighted is the social-cognitive aspect of emotion recognition. This component encompasses the reasoning of emotional states with the objective of explaining or foretelling a person's actions. Cognitive tasks such as prediction of people's behaviours based on false beliefs and differentiation of jokes from lies are

undertaken by this component. Tasks pertaining to false beliefs can be used to examine a person's understanding that one's thoughts can be different from others and from reality, and they classically measure the social-cognitive aspect of emotion recognition (Frith 2001). It is imperative to remember that these two components of emotion recognition cannot be explicitly established, as it is impossible to be one hundred per cent certain of one's emotion state. This is due to the fact that the emotional state of a person and the details of it are not directly available to an onlooker. Instead, these information are inferred from observable traits and contextual knowledge derived from the person, both with varying degrees of certainty.

The research presented in this thesis attempts to automate these two key components of emotion recognition. To obtain a deeper understanding of this subject, various tasks that have been devised to facilitate the social-perceptual understanding of people are reviewed. These tasks aim to examine one's ability to identify emotional or other people-oriented information such as personality, or perceptual stimuli like facial expressions, physical actions or vocal expressions. Among these, facial expressions have garnered the most attention. Regarded as an important channel of nonverbal communication, facial expressions help to communicate a variety of emotional states. Besides conveying emotions, facial expressions also serve as signals that improve conversations and regulate turn-taking. A face comprises of permanent facial features such as the eyes, mouth and eyebrows, and transient features such as dimples, furrows and wrinkles. Facial muscles produce transient features and create the movement and appearance of permanent facial features. The effect produced by the facial muscles is what we perceive as facial expressions. On top of that, the gesture and orientation of the

head and the gaze of the eyes are also major indications of emotion recognition in social-perceptual understanding (Bassili 1978).

2.3.2 Facial Expression Recognition

According to the facial feedback theory, emotion is the experience of changes in our facial muscles. In other words, when we smile, we experience pleasure or happiness. When we frown, we experience sadness. It is the changes in our facial muscles that direct our brains and provide the basis for our emotions. Facial expressions communicate feelings, behavioral intentions, show empathy and acknowledge the actions of other people.

Facial expression has hence gained significant research interest over the last few years. This has given rise to a number of automated systems that recognize facial expressions in images or video (Fogel 2000). Recognizing the basic emotions from facial expressions is of limited utility in understanding the user's cognitive state of mind and intentions. The range of emotional states that people express and identify extends beyond the classic basic emotions, to include a range of affective and cognitive emotional states which are collectively referred to as complex emotional states. These states encompass many affective and cognitive states of emotion, such as agreeing, confused, disagreeing, interested and thinking.

As there are many possibilities of muscle configurations in our face, there is seemingly unlimited number of emotions. Facial expression recognition is a challenging task. The challenges of such system are light variation, direction of subject face, the quality of image acquisition device, and occlusion problem. Facial Expression Analysis is a challenging task. A facial expression is formed by contracting or relaxing different

facial muscles on human face that results in temporally deformed facial features like wide-open mouth, raising eyebrows or etc. The challenges of such system related to:

- a. Light variation. Lighting conditions is a very difficult problem to constrain and regulate. The strength of the light depends on the light source (see Figure 2-5).
- b. The direction of the subjects face is not always ideal which may pose difficulties when the system is implemented live that captures moving subjects' facial expression (see Figure 2-6).
- c. Another difficulty is the way image is acquired by the image acquisition system. The characteristics of the image acquisition system can affect the quality of the images or videos captured.
- d. Occlusion of subject face may significantly degrade the hit rate of many established approaches. The experiments being carried out by most researchers do not take occlusion into account (see Figure 2-7).

Figure 2-5 Light variations problem: face images are taken from different illumination conditions
(source: Yale Face Database B <http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html>)

Figure 2-6 Light variations problem: face images are taken from different illumination conditions
(source: NTU Asian Emotion Database
<http://www3.ntu.edu.sg/SCE/labs/forse/Asian%20Emotion%20Database.htm>)

Figure 2-7 Occlusion problems: facial components are occluded by some artifact objects.(Source: NTU Asian Emotion Database
<http://www3.ntu.edu.sg/SCE/labs/forse/Asian%20Emotion%20Database.htm>)

The first known facial expression analysis was presented by Darwin in 1872 (Darwin 1872). He presented the universality of human face expressions and the continuity in man and animals. He pointed out that there are specific inborn emotions, which originated in serviceable associated habits. After about a century, Ekman and Friesen (Ekman and Friesen 1971) postulated six primary emotions that possess each a distinctive content together with a unique facial expression. These prototypic emotional displays are also referred to as basic emotions in many of the later literature. They seem to be universal across human cultures and are; namely happiness, sadness, fear, disgust, surprise and anger. They developed the Facial Action Coding System (FACS) for describing facial expressions. It is appearance-based. FACS uses 44 action units (AUs) for the description of facial actions with regard to their location as well as their intensity (see Figure 2-8) and

NEUTRAL	AU 1	AU 2	AU 4	AU 5

Eyes, brow, and cheek are relaxed.	Inner portion of the brows is raised.	Outer portion of the brows is raised.	Brows lowered and drawn together	Upper eyelids are raised.
AU 6	AU 7	AU 1+2	AU 1+4	AU 4+5

Cheeks are raised.	Lower eyelids are raised.	Inner and outer portions of the brows are raised.	Medial portion of the brows is raised and pulled together.	Brows lowered and drawn together and upper eyelids are raised.
AU 1+2+4	AU 1+2+5	AU 1+6	AU 6+7	AU 1+2+5+6+7

Brows are pulled together and upward.	Brows and upper eyelids are raised.	Inner portion of brows and cheeks are raised.	Lower eyelids cheeks are raised.	Brows, eyelids, and cheeks are raised.

Figure 2-8 The upper facial Action Units (AUs) in FACS (source: Tian et al. 2001).

NEUTRAL	AU 9	AU 10	AU 12	AU 20

Lips relaxed and closed.	The infraorbital triangle and center of the upper lip are pulled upwards. Nasal root wrinkling is present.	The infraorbital triangle is pushed upwards. Upper lip is raised. Causes angular bend in shape of upper lip. Nasal root wrinkle is absent.	Lip corners are pulled obliquely.	The lips and the lower portion of the nasolabial furrow are pulled pulled back laterally. The mouth is elongated.
AU15	AU 17	AU 25	AU 26	AU 27

The corners of the lips are pulled down.	The chin boss is pushed upwards.	Lips are relaxed and parted.	Lips are relaxed and parted; mandible is lowered.	Mouth stretched open and the mandible pulled downwards.
AU 23+24	AU 9+17	AU9+25	AU9+17+23+24	AU10+17

Lips tightened, narrowed, and pressed together.				
AU 10+25	AU 10+15+17	AU 12+25	AU12+26	AU 15+17

AU 17+23+24	AU 20+25

Figure 2-9 The lower facial Action Units (AUs) in FACS (source: Tian et al. 2001).

Figure 2-9). Individual expressions may be modelled by single action units or action unit combinations. FACS codes expression from static pictures.

More variances of recognition approaches emerge after the year 2000, and some are tested with video sequences rather than images (Bourel 2002; Pardas 2002; Bartlett et al. 2003; Shin *et al.* 2000); extracting pleasure/displeasure and arousal dimension emotion features using hybrid approach. The hybrid approach used feature clustering and dynamic linking to extracting sparse local features from edges on expression images. The expressions are happiness, surprise, sadness, disgust, fear, satisfaction, comfort, distress, tiredness and worry. They concluded that the arousal-sleep dimensions may depend on the personal internal state more than pleasure-displeasure dimensions, that is to say, the relative importance of dimension can have an effect on facial expression recognition on the two dimensional structure of emotion (Fasel and Lüttin 2000) described a system that adopts on holistic approach that recognizes asymmetric FACS Action Unit activities and intensities without the use of markers. Facial expression extraction is achieved by difference images that are projected into a sub-space using either PCA or ICA, followed by the nearest neighbour classification. Recognition rates are between 74~83%. The system proposed in (Bartlett *et al.* 2003) detects frontal faces into the video stream and classifies them in seven classes in real time: neutral, anger, disgust, fear, joy, sadness, and surprise. An expression recognizer receives image regions produced by a face detector and then a Gabor representation of the facial image region is formed to be later processed by a bank of SVMs classifiers.

There are six significant works in the year 2004, they are; namely: (Kobayashi 2004), (Ma 2004), (Kobayashi 2004), (Saxena *et al.* 2004), (Ye *et al.* 2004), (Abboud

2004) and (Sebe *et al.*2004). Kobayashi (Kobayashi 2004) proposed an expression learning system by imitating the process of a baby's learning process. The system consists of a self-learning part and a recognition part. They implemented the system in a penguin robot that integrates vision system and other accessories that allows it to responds differently to human action and facial expression. Ma (Ma 2004) applied an adaptive constructive one-hidden layer feed forward neural network with reduced number hidden units and input-side weights to facial expression recognition. Pruning was used to reduce network size while maintaining the performance of the network. Local image-based approach for the extraction of in-transient facial features and recognition of four facial expressions were used in (Saxena *et al.* 2004). The algorithm uses edge projection analysis for feature extraction and creates a dynamic spatio-temporal representation of the face followed by feed-forward neural networks. The recognition rate is 90% on greyscale image sequences. A lip-enhancement transform was used for better segmentation of lip region in colour images. Ye *et al.*(2004) extracted intransient features of sub-regions such as mouth, eyes and eyebrows and applied Gabor wavelets transformation to form elastic graph for expression. Then the features of six basic expressions are extracted and compared with each other. Abboud (Abboud 2004) presented a bilinear factorization based representation for facial recognition and synthesis. A face representation based on appearance model is used and two bilinear factorization models are proposed to separate expression and identity factors from the global appearance parameters. Facial expression recognition is performed through expression factors classification. Facial synthesis is performed using linear regression over a training set. The recognition rate was found to be 83.3%. (Sebe *et al.* 2004) created

an authentic facial expression database based on spontaneous emotions derived from the experiment. This is contrast to most existing database that collects emotion data by asking the subjects to perform a series of facial expressions. Experiments were carried out using classifiers like Bayesian Networks, decision trees, SVM, kNN, etc. The bagging and boosting voting classifiers were used to improve the classification results.

Ji (Ji 2005) based on FACS has developed a system that adopted a dynamic and probabilistic framework based on combining Dynamic Bayesian Networks (DBM) with FACS for modelling the dynamic and stochastic behaviours of spontaneous facial expressions. The three major components of the system are facial motion measurement, facial expression representation and facial expression recognition. Wu *et al.* (Wu *et al.* 2005) modelled uncertainty in facial expressions space for facial expression recognition using fuzzy integral. The fuzzy measure is constructed in each of the facial expression space. They adopted Active Appearance Models (AAM) to extract facial key points and classify based on shape feature vector. Fuzzy C-means (FCM) was used to build a set of classifiers. The recognition rates were found to be 83.2% and 91.6% on JAFFE and FGnet databases respectively. Yeasin *et al.* (Yeasin and Bulot 2005) compared the performances of linear and non-linear data projection techniques in classifying six universal facial expressions. The three data projection techniques are Principal Component Analysis (PCA), Non-negative Matrix Factorization (NMF) and Local Linear Embedding (LLE). The system developed by (Anderson and McOwan 2006) characterized monochrome frontal views of facial expression with the ability to operate in cluttered and dynamic scenes, recognizing six emotions universally associated with unique facial expressions; namely: happiness, sadness, disgust, surprise, fear, and anger.

Faces are located using a spatial ratio template tracker algorithm. Optical flow of a face is subsequently determined using a real-time implementation of gradient model. The expression recognition system then averages facial velocity information. The motion signatures produced are then classified using Support Vector Machines. The best recognition rate is 81.82%. Zeng et al. (Zeng *et al.* 2006) classified emotional and non emotional facial expressions. Piecewise Bezier Volume Deformation (PBVD) was used to track face. They applied kernel whitening to map the data to a spherical symmetrical cluster. Then Support Vector Data Description (SVDD) was applied to directly fit a boundary with minimal volume around the target data. Experimental results suggested the system generalize better than using PCA and single Gaussian approaches. Xiang *et al.* (2007) utilized fourier transform, fuzzy C means to generate a spatio-temporal model for each expression type. Unknown input expressions are matched to the models using Hausdorff distance to compute dissimilarity values for classification. The recognition rate was found to be 88.8% with expression sequences.

In general, a facial expression recognizer comprises of 3 stages; namely: image pre-processing, features selection, and classification. Image pre-processing involves general manipulation of the image. The raw image is processed to provide a region of interest (human face without hairs and background) for the second stage to select meaningful features. Some noise reduction, clustering, labelling or cropping may be done in this stage. For some ready data that are taken off the shelf from online database, the first stage is unnecessary.

Features selection is an important module. Without good features, the effort made in the classification stage would be in vain. Fasel and Luetttin (Fasel and Luetttin 2003)

provides a detailed survey on facial expression, they classified the feature extraction methods into several groups. Some of the methods were highlighted like Gabor wavelets, Active Appearance Model, Dense Flow Fields, Motion and Deformable Models, Principal Component Analysis, High Gradient Components and etc. They group the facial features into 2 types: intransient facial features and transient facial features. Intransient features are like eyes, eyebrow and mouth that are always present in the face. Transient features include wrinkles and bulges.

Figure 2-10 shows an outline of the real-time expression recognition system developed by Bartlett and colleagues (Bartlett *et al.* 2003; Littlewort *et al.* 2006). The system automatically detects frontal faces in the video stream and codes each frame with respect to 7 dimensions: neutral, anger, disgust, fear, joy, sadness, surprise. The system first performs automatic face and eye detection using the appearance-based method of Fasel *et al.* (2005). Faces are then aligned based on the automatically detected eye positions, and passed to a bank of appearance-based features. A feature selection stage extracts subsets of the features and passes them to an ensemble of classifiers which make a binary decision about each of the six basic emotions plus neutral. According to their results, they found that Gabor wavelets and ICA (Independent Component Analysis) gave better performance than PCA (Principal Component Analysis), LFA (Local Feature Analysis), Fisher's linear discriminants, and also outperformed motion flow field templates. More recent comparisons included comparisons of Gabor filters, integral image filters, and edge-oriented histogram (Whitehill and Omlin 2006), using SVMs and AdaBoost as the classifiers. They found an interaction between feature-type and classifier, where AdaBoost performs better with integral image filters, while SVMs

perform better with Gabors. The difference may be attributable to the fact that the pool of integral image filters was much larger. AdaBoost performs feature selection and does well with redundancy; whereas SVMs were calculated on the full set of filters and do not do well with redundancy.

Figure 2-10 Outline of the real-time expression recognition system of Littlewort *et al.* (2006)

2.3.3 Human Emotional States Recognition for Mind Reading

Many autistic people are capable of identifying the basic emotions but not the more complex ones. This is evident when facial signals are vague and the boundaries between each emotional state are ambiguous (Simon Baron-Cohen 2002). Existing corpora of nonverbal expressions are relevant to autistic therapy to some extent. This applies to the context of our thesis as well. This is because only the basic emotions are incorporated. The Mind Reading DVD is an interactive computer-based guide to emotions (Figure 2-11). This project was collaboration between a team of psychologists led by Professor Simon Baron-Cohen at the Autism Research Centre of University of Cambridge and a London multimedia production company. The objective was to design a program that

would assist autistic individuals in recognizing emotions from facial expressions. It stems from taxonomy of emotions, encompassing a diversity of affective and cognitive emotional states. This feature marks the software as a valuable resource in constructing an automated inference system for computer user interfaces.

Figure 2-11 User interface of Mind Reading DVD

The taxonomy by Baron-Cohen *et al.* consists of 412 emotion concepts. They are classified into 24 distinct emotion groups: Afraid, angry, bored, bothered, disbelieving, disgusted, excited, found, happy, hurt, interested, kind, liked, romantic, sad, sneaky, sorry, sure, surprised, thinking, touched, unfriendly, unsure, wanting. The list is partially shown in Figure 2-12. The 24 groups were chosen in an attempt to preserve the semantic uniqueness of the different emotion concepts within each group. This signifies that each group contains the fine shades of the given emotion state. For example, brooding, calculating and fantasizing falls into the “thinking” emotion category, whereas baffled,

confused, likewise and puzzled belong to different classes within the “unsure” emotion category. The ability to identify different shades of the same group reflects one’s empathizing ability (Baron-Cohen 2003). There is a possibility that emotion state groups possessing distinctive semantics can occur concurrently. For instance, one can be both thinking and confused at the same time. The co-occurrence of the emotion state groups within the taxonomy remains to an interesting and open research question.

Figure 2-12 Emotion fusion in tree structure (Partially)

Emotion states constitute the top level of the computational model of emotion-recognition, and denote the affective and cognitive states of the mind. A person’s emotion state is not directly available to an observer. Instead, it is communicated via nonverbal cues of which the face is arguably the most important. This thesis outlines a system designed for the recognition of head gestures and facial expressions from a continuous video stream in real-time. The process of recognizing the emotion displayed by facial expressions is inherently uncertain. People can express the same emotion state using different facial expressions, at varying intensities and durations. In addition, the recognition of head and facial displays is a noisy process. It is very useful and effective

when represented as a graph structure, resulting in a probabilistic graphical model (PGM).

2.3.4 Other potential applications of emotion recognition

Since the inception of the emotion-recognition field almost a decade ago, numerous researchers have spearheaded building machines with affective abilities. In the automobile industry, the automated inference of driver vigilance from facial expressions is gaining a lot of research attention and commercial interest. Despite this significant progress, the vision of an automated, robust, affect-aware system remains elusive. This continues to be a challenging endeavour. The aim of this thesis is to deliver a plausible solution and this is discussed in the following sections.

Despite these important functions of emotion-recognition, existing human-computer interfaces are mostly mind-blind. They are oblivious to the emotion states of their users, fail to reason about their actions and fail to take into account what they seem to know or not. Such interfaces have no user-awareness. They lack understanding of interruption and context of their uses and the ability to adapt to new circumstances. As Matthew Turk notes (Turk 2005), a computer may wait indefinitely for inputs from a user who is no longer there or decide to do irrelevant, computationally intensive tasks while a user is frantically working on a fast approaching deadline. Existing human-computer interfaces rarely take the initiative and lack persuasive power. Most of them are limited by a command-and-control interaction paradigm. This interaction paradigm is especially restrictive as human-computer interaction (HCI) becomes more complex and new methods of computing emerge. Computing is no longer bounded to a desktop setup and is far more than a task that is only executed during certain times of the day. Instead,

computing is becoming ubiquitous, extending to mobile, embedded and wearable devices used by people in different interaction scenarios to perform an assortment of tasks. To unleash the full potential of these new technologies, user-aware interfaces that complement existing interaction methods are necessary. People assume certain emotion states at any time. This includes their interaction with a computer. Emotion states can be affective, as in expressions of emotions, or cognitive as in revealing emotion processes. Both shape the decisions we make and affect our performance. The omnipresence of computers, along with the central role of emotion-recognition in communication and in decision-making serves as the motivation for this thesis: building Emotion-Recognition Machines. We define these computing technologies as ones that are aware of the user's state of emotion and are capable of adapting accordingly to their responses. Our goal is to enhance human-computer interaction through empathic responses, to improve the productivity of the user and to empower applications to initiate interactions with and on behalf of the user, without waiting for inputs from the user explicitly.

Chapter 3. A Portable Facial Expression Recognition

3.1 Introduction

Figure 3-1 System diagram of emotion recognition highlighting the Real-Time Expression Recognizer.

The work of this thesis is mainly focused on developing Cognitive Assistive Probabilistic-based Emotional State Recognition module which forms part of the whole emotion aid system. The whole system diagram is shown in Figure 3-1. The system consists of four modules: Portable Emotion Recognizer, Reaction Advisor, Emotion Indexer and Emotion Predictor. These four modules work hand-in-hand to provide a complete system to aid user who needs advice on emotion. The focus of this chapter is on the first module, i.e., a Portable Emotion Recognizer.

The Emotion Recognizer is capable of recognizing four types of facial expression; namely: neutral, joy, sad and surprise. The module itself is composed of four major blocks, the face locator, the fiducial points locator, the feature extractor, and the classifier which are shown in Figure 3-2. The face locator undergoes image processing stages to find the face edges. The fiducial points locator finds crucial fiducial points for subsequent feature extraction processing. We adopted Gabor and Gini features which are useful for emotion recognition. Finally, the meaningful features are classified into the corresponding class using the Naïve Bayes Classifier (NBC). A probabilistic based approach using the Naïve Bayesian Boost is adopted to recognize these four types of facial expressions. We adopted the boosting Naïve Bayes Classifier (NBC) to classify facial expressions. The Bayesian classifier is the most popular classifier among the probabilistic classifiers used in the machine learning community. Furthermore, Naive Bayesian classifier is perhaps one of the simplest yet surprisingly powerful techniques to construct predictive models from labelled training sets when comparing against other

supervised machine learning methods. NBC can also provide a valuable insight to the training data by exposing the relations between attribute values and classes besides good predictive accuracy with very small computational time.

Figure 3-2 System flow diagram of facial expression recognizer

3.2 Derivative-based Face Locator

We use the Laplacian of Gaussian Derivations filter to locate face area. Figure 3-3 shows the transformation of this stage. Schiele (Schiele 1997) experimentally compares the invariant properties for a number of receptive field functions, including Gabor filter and local derivative operators. Those experiments showed that Gaussian derivatives provide

the most robust recognition results. Accordingly, we use filters which are based on Gaussian derivatives which can be represented by:

$$D_x(x, y) = -\frac{x}{\sigma^2} G(x, y), \quad (3.1)$$

$$D_y(x, y) = -\frac{y}{\sigma^2} G(x, y), \quad (3.2)$$

where $G(x, y)$ denotes an input image of (x, y) domain and σ is Gaussian standard deviation.

Ros and Kak (Rosenfeld and Kak 1982) showed that the simplest isotropic derivative operator is the Laplacian operator. The 2-D Laplacian of Gaussian function centered on zero and with a Gaussian standard deviation with σ has the form:

$$LoG(x, y) = -\frac{1}{\pi\sigma^4} \left[1 - \frac{x^2 + y^2}{2\sigma^2} \right] e^{-\frac{x^2 + y^2}{2\sigma^2}}. \quad (3.3)$$

Or a simplified form of Laplacian of Gaussian is:

$$Lap(x, y) = G_{xx}(x, y) + G_{yy}(x, y), \quad (3.4)$$

where G_{xx} and G_{yy} are the Gaussian derivatives with respect to x and y respectively.

Figure 3-3 Derivative operations for face image

After the Derivative-Based filter, we use Naïve Bayesian to classified into 2 classes, face and non-face.

3.3 Fiducial Point Locator

Head and facial actions are important clues in facial expression recognition. They encode the basic spatial and motion characteristics of the head and facial features in video input. This section discusses the extraction of head and facial actions from a video stream, and shows how the adopted approach is suited for an automated, real time, and user-independent system. A number of studies showed that the visual properties of facial expressions can be described by the movement of points belonging to facial features (Pantic and Rothkrantz 2000). These feature points are typically located on the eyes and eyebrows for the upper face and the lips and nose for the lower face. Figure 3-4 illustrates the 2D face model of the 25 feature points relevant to this work. By tracking these feature points over an image sequence and analyzing their displacements over multiple frames, a characteristic motion pattern for various action units (AUs) can be established. Cohn et al. (Baron-Cohen 1999) have shown that the results of the automated extraction of AUs using methods based on feature point tracking that comparable to that of manual FACS coding (Ekman 1978). Figure 3-4 describes how the relevant head AUs are measured. These are divided into different “sensing” channels based on rotation axis. Figure 3-5 describes the facial AUs, which are grouped into lip, mouth and eyebrow sensors. Only selected AUs from the complete list of AUs in FACS were chosen that are straightforward to identify and we consider them relevant to the emotion states on which we are focused on. AUs are more precise and robust to rigid head motion compared to

similar measurements that also use feature-point tracking. The table includes both additive and non-additive facial AU combinations. In a non-additive combination, the resulting facial action has a different appearance altogether than the individual AUs, and hence requires an analysis rule of its

AU	Description	Example
53	Head up
54	Head down
55	Head tilt left
56	Head tilt right

Figure 3-4 The AUs from head movement. The examples are quoted from Mind Reading DVD (BGWH04).

AU	Description	Example
12	Lip corner pull
27	Mouth stretch
12+27	Lip corner pull + Mouth stretch

Figure 3-5 The AUs related to facial expression. The examples are FACS-coded image sequences from the Cohn-Kanade database (KCT00) of facial expressions.

own. For instance, the combination of a lip corner pull (AU12) with the lips spread (AU25) has a different appearance from each AUs when they are measured independently.

The fiducial points are located at the eyes, nose, and mouth as shown in Figure 3-6. The fiducial points define the extended feature components. The component-based feature detector has two levels: The Micro SVM based independent component detector and The Macro SVM based independent component detector. The micro level uses linear SVM based independent component detection. Each component classifier was trained on a set of extracted facial components (the 4 key fiducial components) and on a set of randomly selected non-face patterns. The macro level uses the maximum outputs of the component classifiers within rectangular search regions as inputs to a combination SVM classifier. The macro SVM performs the final detection of the face component regions.

Figure 3-6 Fiduciary points on the face, basic location is around the eyes, mouth and nose areas.

The detection of asymmetry can be useful in the inference of complex emotion states since asymmetric facial actions frequently occur in expressions of cognitive emotion states as shown in Figure 3-7. In particular, an asymmetric eyebrow raised and asymmetric lip pull is frequent in confusion, while asymmetric lip bites are frequent in expressions of worry. Recently, Mita and Liu (2004) have shown that facial asymmetry has sufficient discriminating power to significantly improve the performance of an expression classifier. With a few exceptions, very little research has gone into the automated recognition of asymmetric facial actions. Fasel and Luetttin (Luetttin 2000) used image differences to identify asymmetry between the left and right facial regions.

Their approach however, has two limitations: it is not robust to rigid head motion, and it does not provide sufficient details on the specific features where asymmetry occurs. The shape and colour-based methods that have been described so far can be extended to describe facial action asymmetry. Instead of the one action symbol output per feature, two sets of symbols are output per feature: one for the left and one for the right region of the face. This is straightforward for the eyebrows, and equally so for the lips because the use of the anchor point to measure shape deformations means that two different measurements are obtained, one for each half of the face. For the mouth, the line passing through the upper and lower lip centers defines the left and right polygons. All the shape and colour parameters are normalized against head turns to remove the effect of one region falsely appearing smaller than the other. The variance in motion, shape and colour between the two regions indicates asymmetry.

Figure 3-7 Asymmetrical characteristic of a human face. Our experiment is base on both the left and right side of a human face due to their asymmetrical characteristic.

3.4 Feature Extractor

After the presence of a face has been detected in the observed scene, the next step is to extract the information about the displayed facial signals. Most of the existing facial expression analyzers are directed towards 2D spatiotemporal facial feature extraction. The feature extractor then adopts Gabor wavelet feature extraction. Gabor wavelet is a popular choice because of its capability to approximate mammals' visual cortex. The primary cortex of human brain interprets visual signals. It consists of neurons, which respond differently to different stimuli attributes. The receptive field of cortical cell consists of a central ON region and is surrounded by 2 OFF regions, each region elongates along a preferred orientation (Daugman 1985). According to Jones and Palmer (John and Palmer 1995), these receptive fields can be reproduced fairly well using Daugman's Gabor function.

The Gabor wavelet function can be represented by:

$$g(x, y) = g_1(x, y) \exp(j2\pi Wx) \quad (3.5)$$

where

$$g_1(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y} \right) \exp \left(-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right) \quad (3.6)$$

We consider the receptive field (RF) of each cortical cell consists of a central ON region (a region excited by light) surrounded by two lateral OFF regions (excited by darkness) (G.E. La Cara 2003). Spatial frequency (W) determines the width of the ON and OFF regions. σ_x^2 and σ_y^2 are spatial variances which establish the dimension of the RF in the preferred and non-preferred orientations. As shown in Figure 3-8, the Gabor wavelets are represented with different orientations and frequencies. These Gabor

wavelets act as stimuli to the system. Figure 3-9 gives an overview of the optimal stimuli for the first 48 units resulting from one typical simulation. Since the processing at the retina and LGN (Lateral Geniculate Nucleus) cortical layers are used by Gaussian kernels applying to some spectral bands to simulate local inhibition, most stimuli resemble Gabor wavelets. The input image is in HSI domain and is projected into different Gabor wavelets to generate the output signals that resemble electrical signals in visual cortex. Different orientations and special frequencies produce different wavelets. After the convolution of input and wavelets, a set of feature vectors is formed that acts like the 'hypercolumns' as described by (Hubel and Wiesel 1962).

In our work, a facial image is convoluted with all these filters so as to extract the facial features. We have chosen 6 orientations and 4 spatial frequencies; generating a total of 24 Gabor filtered images. The lower bound frequency is chosen as 0.05 while the upper bound frequency is chosen to be 0.4. Orientations are in multiples of $\frac{\pi}{6}$ from 0 to π . Figure 3-9 shows the responses of two different facial images for four of the selected Gabor filters. All the convoluted images model the data received by the primary visual cortex (area V1). The mean Gabor image is then generated and form the major feature vector for the next stage to process.

Figure 3-8 Gabor wavelets representation and the overview of optimal excitatory and inhibitory stimuli (S+ res. S-).

Figure 3-9 Examples of Gabor Wavelets and corresponding convoluted images.

3.5 Feature Selection

3.5.1 T-test Feature Selection

In many existing Feature Selection algorithms, feature ranking is often used to show which input features are more important (Guyon and Elisseeff, 2003; Wang and Fu, 2005), especially when datasets are very large. T-test (Devore and Peck, 1997) is a common type of feature ranking measures that is often used to assess whether the means

of the two classes are statistically different from each other by calculating a ratio between the difference of two classes means and the variability of the two classes. The T-test has been used to rank features for microarray data (Jaeger *et al.*, 2003; Su *et al.*, 2003) and for mass spectrometry data (Wu *et al.*, 2003; Levner 2005). These uses of T-test are limited to two class problem. For multi-class problems, a T-statistic value can be calculated as eqn. (3.7) for each feature of each class by evaluating the difference between the mean of one class and the mean of all the classes, where the difference is standardized by the within-class standard deviation.

$$t_{ic} = \frac{\bar{x}_{ic} - \bar{x}_i}{M_c \cdot (S_i + S_0)} \quad (3.7)$$

$$S_i^2 = \frac{1}{N - C} \sum_{c=1}^C \sum_{j \in c} (x_{ij} - \bar{x}_{ic})^2 \quad (3.8)$$

$$M_c = \sqrt{1/n_c + 1/N} \quad (3.9)$$

Here t_{ic} is the T-statistics value for the i -th feature of the c -th class; $\bar{x}_{ic}$ is the mean of the i -th feature in the c -th class, and $\bar{x}_i$ is the mean of the i -th feature for all classes; x_{ij} refers to the i -th feature of the j -th sample; N is the number of all the samples in the C classes and n_c is the number of samples in class c ; S_i is the within-class standard deviation and S_0 is set to be the median value of S_i for all the features.

The steps to compute T-test are given as follow: Firstly, compute the mean for each of the two classes. Secondly, compute the variance for each of the two classes; thirdly, determine by the formula. Lastly, after determining the t-test value for every feature, then the attribute can be ranked according to their t-test values to determine the most discriminative ones. The larger the t-test value is, the more discriminative the

feature is, also imply the ranking of the features to be selected. T-statistics is usually used to shrink class means toward the mean of all classes to constitute a nearest shrunken centroid classifier, but do not rank features regard to all the classes. In our work, another feature ranking method called GINI feature ranking for feature selection is used as discussed in the next section.

3.5.2 GINI feature selection

Apart from the statistical based feature selection likes T-test to rank the features for selection, another commonly used feature ranking is based on information theories. This type of feature ranking is normally called correlation based feature selection. It uses a correlation based heuristic to evaluate the worth of features (Hall and Smith 1998). In order to perform the evaluation, a heuristic will be used. This heuristic takes into account the usefulness of individual features for predicting the class label along with the level of inter-correlation among them. The hypothesis on which the heuristic can be stated as: Good feature subsets contain features highly correlated with the class, yet uncorrelated with each other.

Classification tasks in machine learning often involve learning from categorical features, as well those that are continuous or ordinal. A measure based on information theory estimates the degree of association between nominal features. If X and Y are discrete random variables, the entropy of Y before and after observing X are given as:

$$H(Y) = -\sum_{y \in Y} p(y) \log_2 p(y), \quad (3.10)$$

$$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2 p(y|x). \quad (3.11)$$

The amount by which the entropy of Y decreases reflects the additional information about Y provided by X and is called the information gain (Quinlan 1993). Information gain is given by:

$$\begin{aligned}\text{gain} &= H(Y) - H(Y|X) \\ &= H(Y) - H(X|Y) \\ &= H(Y) + H(X) - H(X, Y)\end{aligned}\quad (3.12)$$

Information gain is a symmetrical measure – that is, the amount of information gained about Y after observing X is equal to the amount of information gained about X after observing Y . Unfortunately, information gain is biased in favour of features with more values, that is, attributes with greater numbers of values will appear to gain more information than those with fewer values even if they are actually no more informative. The purpose of feature selection is to decide which of the initial features to include in the final subset and which to exclude. If there are n possible features initially, then there are 2^n possible subsets. The only way to find the best subset would be to try them all – this is clearly prohibitive for all but a small number of initial features.

Gini Index selects features based on information theories (Hall and Smith 1998). It measures the impurity for a group of labels. Gini Index for a given set S of points assigned to, for example, two classes C_1 and C_2 is given below:

$$GINI(s) = 1 - \sum_{j=1,2} \left[p(C_j | s) \right]^2 \quad (3.13)$$

With $p(C_j | s)$ corresponds to the frequency of class C_j in set S . The maximum value as $1 - \frac{1}{nc}$ occurs when features are equally distributed among all classes, which implies less interesting information. On the other hand, the zero value occurs when all points belong

to one class that represents the most interesting information. We then sort the n features over different classes of samples in ascending order based on their best Gini index. Low Gini index corresponds to high ranking discriminative features.

3.6 Classification

Classification is the task of generalizing known structure to apply to new data. For example, an email program might attempt to classify an email as legitimate or spam. Common algorithms include decision tree learning, nearest neighbour, naive Bayesian classification, neural networks and support vector machines.

Neural network is a popular choice for classification. Most of them fall under supervised learning. The essence of this project is to resolve the emotion mapping problem in an unsupervised manner. This research aims at modelling the behaviour of the brain, the way human visual processing model comprehend human emotion. The other classification methods underscored by Pantic and Rothkrantz (Pantic and Rothkrantz 2000) are Expert System Rules, Discriminant functions by Cohn (Cohn *et al.* 1998), Spatio-temporal motion energy templates by Essa (Essa and Pentland 1997), Thresholded motion parameters by Black (Black and Yacoob 1997), and Adaptive Facial Emotion Tree Structures by (Wong and Cho 2006). Neural network was traditionally used to refer to a network or circuit of biological neurons (Liu and Li 2004). The modern usage of the term often refers to artificial neural networks, which are composed of artificial neurons or nodes. Thus the term has two distinct usages: Firstly, biological neural networks are made up of real biological neurons that are connected or functionally related in the peripheral nervous system or the central nervous system. In the field of neuroscience, they are often identified as groups of neurons that perform a specific physiological

function in laboratory analysis. Secondly, artificial neural networks are composed of interconnecting artificial neurons (programming constructs that mimic the properties of biological neurons). Artificial neural networks may either be used to gain an understanding of biological neural networks, or for solving artificial intelligence problems without necessarily creating a model of a real biological system. The real, biological nervous system is highly complex and includes some features that may seem superfluously based on an understanding of artificial networks. We focus on the relationship between the two concepts; for detailed coverage of the two different concepts refers to the separate articles: Biological neural network and artificial neural network.

We adopted a boosting Naïve Bayesian Classifier (NBC) to classify facial expressions. Both theoretical and practical studies have often been carried out to understand the predictive properties on this boosting NBC method. The Bayesian classifier is the most popular classifier among the probabilistic classifiers used in the machine learning community, especially with the boosting of the tuple. The study of probabilistic classification is the study of approximating a joint distribution with a product distribution. Probabilistic classifiers operate on data sets where each example x consists of feature values $\langle a_1, a_2, \dots, a_i \rangle$ and the target function y can take on any value from a pre-defined finite set $V = \{v_1, v_2, \dots, v_j\}$. Bayesian rule is used to estimate the conditional probability of a class label y , and then assumptions are made on the model, to decompose this probability into a product of conditional probabilities. The formula used by the simple Bayesian classifier is: $P(a_i | v_j)$ and $P(v_i)$ it can be calculated based on their frequency in the training data, under the assumption that features values are conditionally independent given the target value. On the other hand, one can model the

component marginal distributions in a wide variety of ways for numerical features. Zero counts are obtained when a given class and feature value never occur together in the training set, and is problematic because the resulting zero probabilities will wipe out the information in all the other probabilities when they are multiplied. Incorporating a small sample correction into all probabilities is one of the solutions to this problem. Its probability is set to $1/N$, where N is the number of examples in the training set if a feature value does not occur given some class. The assumption of independence is always wrong. However, a large scale comparison of simple Bayesian classifier with state-of-the-art algorithms for decision tree induction and instance-based learning on standard benchmark datasets found that simple Bayesian classifier sometimes is superior to each of the other learning schemes even on datasets with substantial feature dependencies. NBC algorithm always had the best accuracy per needed training time and it predicts the class feature in a very short time.

3.6.1 Naïve Bayesian

After the feature components are ranked and selected, the facial expression recognition tasks make use of the feature vectors to discriminate and identify each of the expressions. Various pattern classifiers such as Support Vector Machines (SVM, K-nearest neighbours (K-NN) (Aha and Kibler 1991), Naïve Bayes Algorithm (John and Langley 1995) and Artificial Neural Networks could be used for emotion recognition. SVM is a learning technique which was strongly motivated by the results of statistically learning theory. SVM operates on the principle of induction, known as structural risk minimization, which minimizes the upper bound of the generalization error. K-nearest neighbours method is a

nonparametric technique in pattern recognition, which is used to generate k numbers of nearest neighbours' rules for classification. Naïve Bayes algorithm is based on the Bayesian decision theory, which is a fundamental statistical approach to the problem of pattern classification. This approach is based on quantifying the tradeoffs between various classification decisions using probability and the costs that accompany such decisions.

Naive Bayesian classifier is a simple probabilistic classifier based on applying Bayesian' theorem with strong (Naive) independence assumptions. Independent feature model is a more descriptive term for the underlying probability model. Naive Bayesian classifiers can be trained very efficiently in a supervised learning setting, depending on the precise nature of the probability model. Naive Bayesian classifiers often work much better in many complex real-world situations than one might expect, in spite of their Naive design and apparently over-simplified assumptions. An advantage of the Naive Bayesian classifier is that it requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix. Detailed analysis of the Bayesian classification problem has shown that there are some theoretical reasons for the apparently unreasonable efficacy of Naive Bayesian classifiers recently. One should notice that the independence assumption might lead to some unexpected results in the calculation of posteriori probability. When there is a dependency between observations, the above-mentioned probability may contradict with the second axiom of probability by which any probability must be less than or equal to one. The Naive Bayesian classifier

has several properties that make it surprisingly useful in practice despite the fact that the far-reaching independence assumptions are often inaccurate. For example, the decoupling of the class conditional feature distributions means that each distribution can be independently estimated as a one-dimensional distribution. This will help to alleviate problems stemming from the curse of dimensionality, such as the need for data sets that scale exponentially with the number of features. It arrives at the correct classification as long as the correct class is more probable than any other class; hence class probabilities do not have to be estimated very well. In other words, the classifier is robust enough to ignore serious deficiencies in its underlying Naive probability model.

Given a set of test data X and a posteriori probability of a hypothesis $H, P(H|X)$, it may follow the Bayesian theorem (Rish 2001) as:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}. \quad (3.14)$$

Informally, this can be written as $\text{posteriori} = \text{likelihood} \times \frac{\text{prior}}{\text{evidence}}$. It predicts X belongs to C_2 iff the probability $P(C_2|X)$ is the highest among all the $P(C_k|X)$ for all the k classes. The detailed steps are as follow: Firstly, let D be a training set of tuples and their associated class labels, and a test tuple that is represented by an n -D attribute vector $X = (x_1, x_2, \dots, x_n)$. Suppose there are m classes $C_1, C_2, \dots, C_m$. The classification is to derive the maximum posteriori, i.e., the maximal $P(C_i|X)$. Alternatively, a simplified assumption: attributes are conditionally independent (i.e., no dependence relation between attributes). This greatly reduces the computation cost: Only counts the class distribution. The formula is stated as below:

$$\begin{aligned}
P(X|C_i) &= \prod_{k=1}^n P(X_k|C_i) \\
&= P(X_1|C_i) \times P(X_2|C_i) \times \dots \times P(X_n|C_i)
\end{aligned} \tag{3.15}$$

Bayesian networks can represent joint distributions we use them to compute the posterior probability of a set of labels given the observable features, and then we classify the features with the most probable label. The idea is to use a strategy that can efficiently search through the whole space of possible structures and to extract the ones that give the best classification results.

Consider the problem of classifying facial emotion by features, for example into joy and non-joy. Imagine that image are drawn from a number of classes of facial features which can be modeled as sets of words where the (independent) probability that the i -th feature of a given image occurs in a image from class C can be written as $P(W_i|C)$. For this treatment, we can simplify further by assuming that the probability of a feature in an image is independent of the total number of features, or that all image contain same number of features. Then the probability of a given image E , given a class C is

$$P(E|C) = \prod_i^n P(W_i|C). \tag{3.16}$$

Bayesian' theorem manipulates these into a statement of probability in terms of likelihood given by:

$$P(C|E) = \frac{P(C)}{P(D)} \times P(E|C). \tag{3.17}$$

Assume for the moment that there are only two classes, J and $\neg J$ (e.g. Joy and not Joy). Using the Bayesian result above, we can write respectively as:

$$P(J|E) = \frac{P(J)}{P(E)} \times \prod_i^n P(W_i|J), \quad (3.18)$$

$$P(\neg J|E) = \frac{P(\neg J)}{P(E)} \times \prod_i^n P(W_i|\neg J). \quad (3.19)$$

Dividing one by the other and the probability ratio $\frac{P(J|E)}{P(\neg J|E)}$ can be expressed in terms of a series of *likelihood ratios*. The actual probability $P(J|E)$ can be easily computed from $\log \frac{P(J|E)}{P(\neg J|E)}$ based on the observation that $P(J|E) + P(\neg J|E) = 1$.

Taking the logarithm of all these ratios, i.e., log-likelihood ratios, in case of two mutually exclusive alternatives, such as this example, the conversion of a log-likelihood ratio to a probability takes the form of a sigmoid curve. Finally, the image can be classified as follows:

It is Joy if $\log \frac{P(J|E)}{P(\neg J|E)} > 0$, otherwise it is not Joy.

3.6.2 Weighted Naïve Bayesian and GINI Split

A Naive Bayesian classifier is a simple probabilistic classifier based on Bayesian' theorem with strong (Naive) independence assumptions. The Naive Bayesian classifier has several attractive properties like the decoupling of the class conditional feature distributions that helps to alleviate problems stemming from the curse of dimensionality. The data arrives at the correct classification as long as the correct class is more probable than any other class. In overall, the classifier is robust enough to ignore serious deficiencies in the underlying Naive probability model.

The proposed algorithm adopted Naïve Bayesian with boosting. Each iteration of boosting uses the GINI split method to remove redundant features. The classifier produced is more complicated than the simple Bayesian model. The main reason for using BNB is that the embedded feature selection technique makes the data more suitable for boosting.

The algorithm is summarized below:

BNB Algorithm:

GINI feature selections:

1. Assign each training sample with weight=1.
2. For ten iteration (ten features):
 - Sort features index S.
 - Split S.
 - Break if GINI criterion is satisfied.

BNB classification:

1. Apply simple Bayesian to weighted data set.
 2. Compute error rate.
 3. Iterate the training examples.
 - Multiply the weight by $\frac{e}{1-e}$.
 - Normalize the weight
 4. Add $-\log \frac{e}{1-e}$ to weight of class predicted.
 5. Return class with highest sum.
-

3.7 Experimental Results

3.7.1 Results of Face Localization

This section presents the experimental results of our proposed face locator. A total of 264 samples were used for testing and training with 100 face pictures and 164 non-face pictures. The pictures are selected randomly with sampling. The dataset is divided into three stratified cross-validation sets of training and testing. Training set is made up of one third of the data, whereas testing set is made up of the remaining two third. The class

distribution is maintained. We use images of different nature in the experiments, face (Figure 3-10) and non face (Figure 3-11) samples for training. For non-face samples, we collected different pictures, which include natural scenery pictures, animal's pictures, flower's pictures and etc. For the face samples, we have adopted the Mind Reading DVD (Baron-Cohen *et al.* 2004), a computer-based guide to emotions, developed by a team of psychologists led by Prof. Simon Baron-Cohen at the Autism Research Centre, University of Cambridge. The database contains images of approximately 100 subjects. Facial images are of size 320x240 pixels in PNG format. Subjects' age ranges from 18 to 30. Sixty-five percent were female, 15 percent were African-American, and three percent were Asian or Latino. The subjects are all randomly selected. Subjects were instructed by experimenter to perform a series of facial displays. Subjects began each display from a neutral face. Before performing each display, the experimenter described and modelled the desired display. The input image is first input to the face locator. After derivative filtering, face region will be cropped out as region of interest. We then segregate the region of interest into seven sub-regions (Figure 3-12): whole face, right face, left face, mouth, left eye, right eye, and nose. Figure 3-13(a) and Figure 3-13(b) show the three dimensional data distributions selected by the Gini indexing of a face and bird input images respectively. The three major components (left eye, right eye and mouth) display differences for face and non-face. The plots show distinctively different shapes are formed in the three attributes. This implies that we should be able to classify face images and non-face images easily as there is a clear difference between the data. Our experiments suggest that these three sub-regions crucially determine if there is any face existing in the picture. The individual sub-regions data distributions are shown in Figure

3-14. The scatter plots also suggest that mouth, left and right eyes features best discriminate face from non-face. Figure 3-15 displays the statistics of the intensity of the seven sub-regions. Red bar in the histogram denotes face samples and blue bar denotes non-face samples.

Figure 3-10 Samples of face images

Figure 3-11 Samples of non-face images

Figure 3-12 Intermediate images (a) Human face, (b) Non-face images

Figure 3-13 The 3D data distribution by Gini Indexing. (a) Face input, (b) Non face input.

(a)

(b)

(c)

Figure 3-14 Scatter Plots (a) Left Eye versus Right Eye, (b) Mouth versus Nose, (c) Left Face versus Right

Figure 3-15 Statistics of Seven Face Sub-Regions

We previously described the use of Derivative-Based filters in conjunction with Naïve Bayesian classifier in face locator. Table 3-1 consolidates the results of Derivative-Based (first row) and it is benchmarked against some typical approaches like BayesNet, Adaboost, Multi-Layer Perceptron, Radial Basis Function, Bagging, J48 Decision tree, and oneR methods. The seven sub-regions as described previously are input to the eight classifiers as listed in Table 3-1. The first column shows the accuracy, followed by mean squared error, precision, Recall, and F-Measure (3.20 to 3.24). The result of Naïve Bayes is on the high side of 94.68% accuracy and 0.23. Naïve Bayes is the second highest in terms of F-measure. The higher the F-measure, the better it is in classifying the dataset. Thus, our approach exhibits the superior performance.

In a Receiver Operating Characteristic (ROC) curve the sensitivity (true positive rate) is plotted in function of specificity (false positive rate) for different cut-off points. A test with perfect discrimination (no overlap in the two distributions) has a ROC plot that passes through the upper left corner. Therefore the closer the ROC plot is to the upper left

corner, the higher the overall accuracy of the test. As shown in Figure 3-16 is the ROC curve of face data. It provides good curve for all data as shown.

$$\text{F-Measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}. \quad (3.20)$$

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}. \quad (3.21)$$

$$\text{MSE} = \frac{1}{n-1} (\sum \text{error}^2). \quad (3.22)$$

$$\text{Accuracy} = \frac{\text{True Negative} + \text{True Positive}}{\text{False Positive} + \text{True Negative} + \text{False Negative} + \text{True Positive}}. \quad (3.23)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{False Positive} + \text{True Positive}} \quad (3.24)$$

APPROACH	ACCURACY (%)	MSE	PRECISION	SENSITIVITY/ RECALL	F-MEASURE
Derivative-Base	94.7	0.23	0.94	0.95	0.94
BayesNet	92.8	0.24	0.92	0.94	0.93
Adaboost	82.9	0.37	0.82	0.83	0.82
MLP	94.3	0.21	0.94	0.95	0.94
RBF	93.2	0.24	0.92	0.94	0.93
Bagging	95.1	0.21	0.94	0.96	0.95
J48	95.4	0.21	0.95	0.96	0.95
oneR	79.5	0.45	0.79	0.80	0.79

Table 3-1 Benchmarking results with different classifiers used for face locator

Figure 3-16 ROC of different classifiers used for Face Locator

3.7.2 Results of Feature Selection

For our feature selection, comparisons with other approaches are necessary for us to investigate how the recognition of our approach can be benchmarked against others. The accuracy of facial emotion recognition using Gini Index feature selection is shown in Figure 3-16, T-test feature selection in Figure 3-17, Euclidean Distance feature selection in Figure 3-18, and K-nearest neighbour classification approach in Figure 3-19. According to the results as shown, our approach achieves the most optimal result.

The T-test assesses whether the means of different groups are statistically different from each other. K-nearest neighbour algorithm is a method for classifying objects based on closest training examples in the feature space. These approaches are generally used for bench-marking. Our approach that combines Gini and Boosting Naïve Bayesian achieves average of 75% and highest of 100% outperforms the rest. To put it simply, in the classification part, the system is able to recognize four classes of facial

expression: neutral, joy, sad, and surprise. The experimental result shows that Gini indexing to the features plays a very important role in recognizing different facial emotion. To summarize, the experimental results suggest the combination of Gini and BNB achieves the most optimal result. The average hit rate is around 75%. Figure 3-16, Figure 3-17, Figure 3-18 and Figure 3-19 illustrate the accuracy in % using Euclidean, Gini, T-test feature selections and KNN, Naïve Bayes and Euclidean classifications for both average and maximum result. Table 3-2 shows GINI feature reduction and Naive Bayesian classification combination gives best result.

Figure 3-16 Result of facial emotion recognition (Using GINI feature selection)

Figure 3-17 Result of facial emotion recognition (Using T-test feature selection)

Figure 3-18 Result of facial emotion recognition (Using Euclidean Distance feature selection)

Figure 3-19 Result of facial emotion recognition (Using KNN classification, k=5)

Feature Selection	Classification	Average	Maximum
Gini	Euclidean	63	83
Gini	KNN	57	99
Gini	Naïve Bayes	75	199
T-Test	Euclidean	59	69
T-Test	KNN	59	89
T-Test	Naïve Bayes	57	100
Euclidean	Euclidean	58	69
Euclidean	KNN	57	100
Euclidean	Naïve Bayes	53	99

Table 3-2 Accuracy of different combinations of feature selection methods and classification methods (the values are in percentages of average and maximum)

3.7.3 Results of Facial Expression Classification

Once the face region has been identified, we then assess its classification ability. We use Naïve Bayesian to train and classify into 4 classes (sad, joy, neutral, surprise). The facial expression recognition result is shown in Figure 3-20. The confusion matrix is included in the figure as well where the column of the matrix represents the instances in a

predicted class, in which each row represents the instances in an actual class. The system correctly recognizes 76.3% of neutral, 78.3% of joy, 74.7% of sad and 78.7% of surprise expressions amongst 100 subjects in the database, although some facial expressions do get confused by the wrong class, however at an acceptable range of less than 12%.

In addition, comparisons with other approaches are necessary for us to investigate how the recognition performance of our approach can be benchmarked with others. Table 3-2 shows the recognition results for facial expression recognition using T-test, Euclidean, and K-nearest neighbour approaches. Only the average and maximum hit rates are included in the table. According to the results in the table, our approach achieves the most optimal result. Table 3-3 and Fig 3-20 compares our result with other models from WEKA.

Figure 3-20 Emotion Recognition Result (Bar Form)

Table 3-3 Benchmarking results of different classifiers used for our proposed portable facial expression recognition

Approach	Accuracy (%)
Adaboost	74.2
BayesNet	67.3
Boosting Naïve Bayes (Our Model)	77
MLP	72.4
RBF	68.2
Bagging	62.3
J48	64.6
oneR	69.2

3.8 Conclusion

In summary, although most of the facial expression analyzers developed so far target human facial affect analysis and attempt to recognize a small set of prototypic facial expressions like happiness and anger, some progress has been made in addressing a number of other scientific challenges that are considered essential for the realization of machine understanding of human facial behavior. The discussed facial expression analyzers were tested on spontaneously occurring facial behavior, and extract information about facial behavior in less constrained conditions such as an interview setting. Development of robust face detectors, head and facial component trackers will be robust to variations in both face orientation relative to the camera, occlusions, and scene complexity. Also, the presence of other people and dynamic background forms the first

step in the realization of facial expression analysers capable of handling unconstrained environments.

This chapter presented the experiments carried out to assess the performance of our proposed portable facial expression recognition system. The face locator and face classifier were separately tested. The face locator undergoes some image processing stages to find the face edges. Subsequently the fiducial points are identified. The result of face locator is very encouraging which can achieved accuracy of 94%. In the classification part, the system is able to recognize four classes of facial expression: neutral, joy, sad, and surprise. The experimental result shows that GINI indexing to the features plays a very important role in recognizing different facial expressions. The experimental result also suggests the combination of GINI feature selection and BNB achieves the most optimal result.

Chapter 4. Hidden Markov

Model (HMM) based Emotion

Indexing

4.1 Introduction

This chapter presents the use of Hidden Markov Model in modelling high-level emotions (namely encouraging, interest, unsure, disagreeing and discouraging) through low-level facial expressions (namely happy, sad, surprise and neutral). Human emotions are complex in nature, human instincts comprise of various hidden layers such as the subconscious mind. In addition, Markov state chain property is suitable for modelling human emotion as our emotions are part of the state of mind and it is dependent on our prior emotion-state. Emotion-state at each instant is the resultant of reflection from events that affect us. We propose to develop emotion indexer acting as a higher-level analysis to interpret advanced emotion states from basic emotions as shown in Figure 4-1. Our proposed model can heighten, maintain, and improve emotion recognition functional capabilities of individuals whose ability to recognize emotion are impaired.

Figure 4-1 Block Diagram for Emotion Indexer (circled in black) which can deduce higher level emotion from basic emotion states

Emotion is a state of feeling and involves thoughts and physical changes; These occurrences brings forth an outward expression. Emotions positively affect intelligent functions such as decision making, perception and empathic understanding (Bechara et al. 2000; Isen 2000). There are five theories which attempt to highlight the sequence of the way human experience an emotion. They are the James-Lange theory (Cook 1911), the Cannon-Bard theory (Newman 1936), the Lazarus theory (Lazarus 1982), the Schachter-Singer theory (Reisenzein 1983), and the Facial Feedback theory (Buck 1980). The facial feedback theory states that emotion is the manifestation of changes in the facial muscles. People frown when they experience sadness while smiling shows one's pleasure or happiness. It is the changes in our facial muscles that direct our brains and

provide the basis for our emotions. As there are many possible combinations of muscle configurations in our faces, this suggests the existence of an infinite number of emotions. Most individual are capable of interpreting emotions expressed by others all the time, however there exists a certain group of people who lack this ability. This includes people who are diagnosed along the autism spectrum (Baron-Cohen 1995). Autistic people lack or suffer impairment in several representational sets of abilities. Consequently, they experience difficulties operating in our highly complex social environment. Most autistic people are incapable of deriving emotions of other people based on their facial expressions. A facial expression is defined as the contraction or relaxation of various facial muscles on the human face, resulting in momentarily deformed facial features such as wide-open mouth, raised eyebrows etc. The first known facial expression analysis was presented by Darwin in 1872 (Darwin 1872). The universality of human facial expressions and the continuity in man and animals were also presented. In the analysis, he stated that certain inborn emotions originated from serviceable associated habits. Nearly a century later, Ekman and Friesen (1971) suggested six primary emotions each possessing a distinctive content, coupled with a unique facial expression. These universal emotional displays were collectively termed as the basic emotions in many of the later literature. They are; namely: happiness, sadness, fear, disgust, surprise and anger. The Facial Action Coding System (FACS) was devised subsequently for describing these primary emotions from one's facial expressions (Bauer and Pawelzik 1992).

Apart from the primary emotions stated previously, “secondary” emotions are often labelled as emotions that are more subtle and sophisticated in nature. Damasio (1994) stated that primary emotions are primitive emotions such as startle-based fear,

innate aversions, as well as attractions. On the other hand, “secondary” emotions are relatively more subtle and sophisticated than primary emotions as they involve cognitive processing and require conscious awareness. Thus the primary emotions are equivalent to the basic emotion which are evolutionary crafted in the limbic system. In a nutshell, the secondary emotions are essentially a blend of basic emotions, similar to the way that different colours can be created by the fusion of primary colours red, green and blue. It is important to fully appreciate the division of emotions into primary and secondary since the secondary emotions are the main concerns when it comes to designing human-computer interfaces. Nevertheless, primary emotions such as anger, can surface in the type of interactions that we have discussed, including human-computer interaction.

Emotion indexing is a method used to read and respond to emotions. There are several approaches in teaching emotion understanding. Each of these methods may differ structurally. Highly structured methods implements carefully planned teaching materials and deploy them in a relatively controlled environment. Existing teaching methods also differ in the different teaching setting. This implies that the method utilized can range from hypothetical-based to natural-based, though the teaching environment may or may not be interactive.

In this thesis, we propose an emotion indexing method to understand four primary emotions; namely: neutral, joy, surprise and sad, and we also aim to identify a few additional human emotional states like interested, unsure, disagreeing, encouraging and discouraging. These five classes of emotional states together with the four basic emotions belong to affective and cognitive states of mind. One of the hypothesis models used for inferring and associating lower-level emotions to higher-level emotions is shown in

Figure 4-2 that features the relationship of the five classes of emotional states with the four basic emotions. Parrot who has created 3 levels of emotion inspires this model: Primary, Secondary and Tertiary (Parrot 2001). It is imperative that our proposed emotion indexer will be able to recognize more advanced emotional states beyond the basic emotions. In our approach, the Hidden Markov Model (HMM) is used to model higher-level emotions (such as: encouraging, interest, unsure, disagreeing and discouraging) from low-level facial expressions (such as: happy, sad, surprise and neutral). As mentioned earlier on, the reason for selecting HMM as the model is that HMM simulates how human brain works. In order to ascertain the current emotion state, prior information such as the previous state of the emotion is required.

Our emotion indexer acts as a higher-level analysis to interpret emotional states. The indexer works as a database; events can be retrieved from the database for subsequent replay. The stimuli are prudently segmented to ensure effective one-to-one mapping of emotion to its corresponding facial expression. A smile, for instance, is always used as the face of happiness. This over-simplifies the recognition process since in real-world scenarios a smile can be used to denote different kinds of emotions, such as pride. This also applies to psychological processes that are not particular to certain emotions, like greeting people. It is also likely that different people

Figure 4-2 Four basic emotions together with five emotion/emotional states from emotional indexer

express the feeling of happiness differently. The combination of these factors makes it challenging to predict emotional states from existing literatures. Further studies are needed to quantify human's ability to recognize emotions under natural stimuli. As such, our facial expressions are presented to the analyser using a facial expression language. The analyser fuses these formulas to determine associated emotions and eventually the derive emotional state. The main characteristic of this analyser is that it takes the current emotional state and internal state into consideration during emotional state calculation. In our approach, we propose the use of Hidden Markov Model (HMM) to generalize the visual output from the facial expression recognizer to infer the expert rules for indexing the human emotions. The framework of emotion indexing is shown in Figure 4-3. Initially, facial expression recognizer identifies the head and facial actions for each video frame. These frames are later aggregated to create displays, establishing the observation vector for the inference of the emotion-states. The inputs of the emotion-indexer comprise of representative frames of the mental state, its valence (i.e. positive or negative state), label, intensity and related contextual cues. Events are appended after the last emotional index and the communication between user and other modules are managed by the interface.

Figure 4-3 The proposed emotion indexing framework.

This framework on emotion-state recognition draws on the literature from emotion recognition. Emotion recognition represents the ability to attribute an emotion state of others by observing their behaviours. The theory of emotion recognition describes a coherent framework of how people combine bottom-up perceptual processes with top-down reasoning to map, in our context, low-level observable emotions such as neutral, joy, sad, surprise, to higher-level emotion states such as encouraging, agreeing, discouraging, disagreeing and unsure. The association of such is shown in Figure 4-2. Facial expressions display abundant geometric and dynamic properties adequate for selection of an emotion state in bottom-up processing (Ekman Paul 1980). In top-down reasoning, human utilize emotion models that map observations of particular facial configurations to emotion-state labels. The emotion recognition theory takes into account the intrinsic ambiguity of the mind-reading process. This uncertainty is a by-product of stochastic nature of facial behaviour: people experiencing the same emotion-state may

exhibit different facial expressions with varying intensities and durations. In order to simulate the emotion-reading process in humans, the computational model for emotion-recognition would first have to create or “learn” mappings between emotion state classes and patterns of facial behaviours as observed in video sequences. These mappings are thus used to predict the probability of an incoming video sequence resulting from the emotion-states detected during classification. To account for the variability in expressing these states, the system learns common patterns underlying generic set of emotion-states from data using statistical machine learning instead of rule-based methods. In other words, the model will learn to map from fast continuous video input to any generic set of slowly changing discrete emotion states with the data provided. This is crucial given the absence of expert domain knowledge on the subject. The data-driven process is one of the key features behind designing this emotion recognition model. Other characteristics of the model are presented in details in subsequent sections.

4.2 Emotion Indexing in Autism

Driven by the increasing demand for assistive devices to help autistic people, our project draws inspiration from the emotion indexing method we discussed in section 4.1. Emotion indexing is one method that is helpful for providing guidance to autistic children on how to identify and respond to emotions of the people.

This section continues on the emotion recognition by first performing an abstraction of the video input into two levels. Each level conveys face-based events at different levels of spatial and temporal abstraction. An overview of the automated emotion-recognition system is presented. The system works by applying a hybrid of top-

down predictions of emotion states and bottom-up vision-based processing of the face for deducing complex emotion-state from real-time video.

The real-time emotion-recognition system automatically identifies a facial event, extracts the dynamic head and facial actions, and deduces the underlying emotion state conveyed by the video segment. As such, symbolic frames belonging to a particular event, together with the deduced emotion state label, are delivered to the emotion indexer for archiving. The emotion indexer (Figure 4-3) acts as a mean of storage medium for past events archives.

Each of these events is stored as a tuple in the index that holds the symbolic frames of the emotion-state. The indexed events are made available for use with the Experts' Rules that will be discussed later. The facial actions denoted by the features are determined and classified into weighted emotion labels with the use of a rule-based system. The automatic emotion-recognition system accesses the archive to update and improve inference and recommendations. Essential modules of the system are modelled after the emotion-indexer approach. The working principle of emotional-indexer revolves around a simple concept: the system analyses and indexes facial expressions of people whom the autistic child interacts with, thus inferring their emotion-states and communicates this emotion information back to the child. Facial expression recognition is a pattern recognition problem. While the modelling of dynamic facial feature vector sequences is critical in the analysis of facial expression sequences, it also needs to consider the stochastic nature of human facial expression that consists of both the observable or quantifiable action and the hidden or unquantifiable emotion state. For instance, different people can express the same emotion very differently. Factors

attributing to the disparity can be due to the innate difference between one's controls of facial muscles, and the durations and intensities of their expressions. Despite this disparity, a human observer is still capable of identifying the emotions expressed by others, indicating that there are similarities between elements underlying each emotion. Hence, the objective of facial expression modelling is to unearth the hidden and unknown patterns that link specific expressions to the measured, observable data. A criterion must be defined in order to gauge a particular expression via facial expression modelling, making it desirable to analyse a series of images to capture the dynamics.

Expressions are recognized in the context of an entire image sequence of arbitrary length. We will develop a recognition system based on the stochastic modelling of encoded time series describing facial expressions, which should perform well in the spatio-temporal domain, analogous to the human performance. The system adopts Hidden Markov Models to automate the inference of a wide range of emotion-states from facial expressions. While there has been a great deal of research in facial detection and recognition, there has been very limited work on identifying the expression on face. We focus on the problem of classifying low-level feature sequences into emotional events. Unlike most classical pattern classification problems, the data to be classified is time series data. The next section discusses the details of using HMM to understand sequential emotion data.

4.3 Hidden Markov Models

Markov chain is a mathematical system characterized by its chain-like style of transition from one state to the next possible state. The assignment of the next state is a random process that depends solely on the current state and not the past states. This is a unique

characteristic of the Markov property. Many real-world application processes implement Markov chains as statistical models (Pankin 2007). A Markov process is defined as a random procedure such that the result generated by the current experiment can influence the result of the succeeding experiment. The current state of process is the sole variable influencing its following state. In nearly all fields of modern applied mathematics, Markov processes serve as a useful tool for statistical modelling and for analyzing random-dependent events, whose likelihood of occurrence is determined by the preceding event. They do not depend on independent events such as flipping a coin which has no memories of the previous event. A Markov analysis studies a sequence of events and analyzes the likelihood of one event preceding another event. Based on the observations from the analysis, a fresh sequence of random but related events resembling the original can be obtained.

Leonard E. Baum and few other authors were among the first to discuss about Hidden Markov Models in statistical papers during the late 1960s. One of the first applications of HMMs was speech recognition, originated from the mid-1970s (Rabiner 1989). In the second half of the 1980s, HMM was put to practice in the analysis of biological sequences, especially in DNA (Xuedong Huang 1990). The use of HMMs have since become universal in the field of bioinformatics (Richard Durbin 1999). The state is directly visible to the user in a typical Markov model. Consequently, the state transition probabilities are marked as the only parameters. On the contrary, in a Hidden Markov Model (HMM), two kinds of states exist. They are the observable state and the hidden state. The processes underlying both states are Markov chain processes. Typically, a HMM can be considered as the simplest dynamic Bayesian network. Every state carries a

probability distribution of the possible output tokens. Thus, these sequence of tokens generated by the HMM produces certain information regarding the sequence of states. The term ‘hidden’ in our context describes that the state sequence is passed via the model and not by its parameters. Even though the parameters of the model are accurately identified, the model is still ‘hidden’. Any system modelled after the Hidden Markov Model is presumed to be a Markov process with unobserved state. The subsequent implementation of Hidden Markov Models is well known in the area of temporal pattern recognition. This includes applications such as bioinformatics, gesture recognition, speech, handwriting, part-of-speech tagging, partial discharges and musical score following.

Figure 4-4 illustrates the general structure of an instantiated HMM. Each oval shape in the diagram represents a random variable that can undertake any set of values. The arrows indicate conditional dependencies. The random variable $x(t)$ refers to the hidden state at time t whereas the random variable $y(t)$ corresponds to the observation at time t . As illustrated in the diagram, given the values of the hidden variable x at any time instance, the conditional probability distribution of the hidden variable $x(t)$ at time t is based solely on the value of the hidden variable $x(t - 1)$. As opposed to that, the values at time $t - 2$ and earlier does not induce any effect or influence. This is denoted as the Markov property. Likewise, the value of the observed variable $y(t)$ is also based solely on the value of the hidden variable $x(t)$, both at time t .

Figure 4-4 Illustration of Hidden Markov Model

4.3.1 Notation and Definitions

The observed sequence of length τ is denoted by, $X_0 = X_0, \dots, X_{\tau-1} = X_{\tau-1}$ by $X_0^{\tau-1} = x_0^{\tau-1}$.

The same convention is used for representing state sequence S_t . The entire collection of parameters in the model is denoted by θ . A Hidden Semi-Markov Model (HSMM) comprises of a pair of discrete-time stochastic processes $\{S_t\}$ and $\{X_t\}$. Similar to HMMs, the observed process $\{X_t\}$ is associated with the unobserved state process $\{S_t\}$ by the conditional distributions. The state process itself is a finite-state semi-Markov chain that is constructed as follows: the transitions between states are modelled using a homogeneous Markov chain with J states, which we label $0, 1, \dots, J-1$. The chain specifies the initial probabilities

$$\pi_j := P(S_0 = j) \text{ with } \sum \pi_j = 1 \quad (4.1)$$

and the transition probabilities for the state i : for each $j \neq i$

$$p_j := P(S_{t+1} = j | S_{t+1} \neq i, S_t = i) \text{ with } \sum P_{ij} = 1 \text{ and } p_{ii} = 0 \quad (4.2)$$

Associated with each state is a sojourn time distribution, which describes the number of time periods the chain remains in the state:

$$d_j(u) := P(S_{t+u+1} \neq j, S_{t+u-v} = j, v = 0, \dots, u-2 | S_{t+1} = j, S_t \neq j) \quad (4.3)$$

In contrast to HMMs, the diagonal elements of TPM for each HSMM are required to be zero, assuming that the states are non-absorbing. The sojourn time in the last visited state is subject to right-censoring and is depicted by the survivor function.

$$D_j(u) := \sum_{v \geq u} d_j(v) \quad (4.4)$$

The survivor function is significant in extending Ferguson's original algorithms (Ferguson 1980), which is to abandon the assumption that a change in state occurs immediately after the last observation. Thus the only difference between HMMs and HSMMs lies in the way that their state processes $\{S_t\}$ are defined. The Markov property is absent in the semi-Markov chain related to HSMMs at each time t ; this property is shifted to the level of the embedded first-order Markov chain. The observed process $\{X_t\}$ at time t is linked to the state process $\{S_t\}$ by the component distributions,

$$b_j(x_t) = P(X_t = x_t | S_t = j) \quad (4.5)$$

For the observation module, the conditional independence property obtained from the HMM is satisfied as follows:

$$P(X_t = x_t | S_t = j) = P(X_t = x_t | X_0^{\tau-1}, S_0^{\tau-1} = s_0^{\tau-1}, S_t = j, S_{t+1}^{\tau-1} = s_{t+1}^{\tau-1}), \quad (4.6)$$

implying that the output process at time t depends only on the value of S_t .

A typical HMM has N states and transitions are available among these states. At different times, the system is in one of the states; each transition between the states carries an associated probability, and each of these states carries an associated observation output (symbol). A HMM is characterized by the following notations:

1. N , the number of states present in the model. The collection of individual states is denoted as $S = \{S_1, S_2, \dots, S_N\}$, and the state at time t as q_t .
2. T , the number of observations. A typical observation sequence is denoted as:

$$O = \{O_1, O_2, \dots, O_T\}. \quad (4.7)$$

3. $A = \{a_{ij}\}$, the state transition probability distribution, of size $N \cdot N$, defines the probability of transition from state i at time t , to state j at time $t + 1$

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i), 1 \leq i, j \leq N \quad (4.8)$$

4. $B = \{b_j(k)\}$, probability distribution of observation symbols for each state. b_j refers to the observation symbol of probability in state j .

5. The initial state distribution, $\pi = \{\pi_i\}$, indicates the probability of any given state, where

$$\pi = P(q_1 = S_j), 1 \leq i \leq N \quad (4.9)$$

6. M , the number of observation symbols in the alphabet.

The complete specification of a HMM requires definition of model parameters, as well as the three probability measures A , B , π . HMM parameters use the following set:

$$\Lambda = \{A, B, \pi\} \quad (4.10)$$

4.3.2 Likelihood

The likelihood of a HMM is obtainable in a tractable form. Unfortunately, such a convenient representation of the formula is non-existent for modelling the likelihood of HMM. We first have to introduce the complete-data likelihood that plays a critical role for the parameter estimation. The likelihood of the complete data, i.e., the observations $x_0^{\tau-1}$ as well as the unobserved sequence $s_0^{\tau-1+u}$, is given by:

$$L_c(s_0^{\tau-1+u}, x_0^{\tau-1} | \theta) = P(S_0^{\tau-1}, S_{\tau-1+u} = s_{\tau-1}, v = 1, \dots, u-1, S_{\tau-1+u} \neq S_{\tau-1}, X_0^{\tau-1} = x_0^{\tau-1} | \theta) \quad (4.11)$$

where θ is the model parameter. The last visited state is marked at time $\tau-1+u$ and therefore the completed state sequence stops at this time, instead of $\tau-1$ for algorithms

without right-censoring (Guédon, 2003). The association of the state sequence to the complete-data likelihood is given by:

$$\pi_{s_0} d_{s_0}(u_0) \prod_{r=1}^R P_{s_{r-1}s_r} d_{s_r}(u_r) I\left(\sum_{r=0}^{R-1} u_r < \tau \leq \sum_{r=0}^R u_r\right) \quad (4.12)$$

where $s_0, \dots, s_R$ denote the $R + 1$ states visited by $s_0^{\tau-1+u}$. Combining equations (4.11) and (4.12), the likelihood of the observed sequence can be calculated by using the summation of complete-data likelihood of all permissible paths. In comparison with the classical likelihood by Ferguson (1980), the likelihood function described here includes an additional summation of all possible continuations of the state sequence $s_0^{\tau-1}$ up to the departure from the last visited state:

$$L(\theta) = \sum_{s_0, \dots, s_{\tau-1}} \sum_{u_{\tau+}} L_c(s_0^{\tau-1+u}, x_0^{\tau-1} | \theta) \quad (4.13)$$

where $\sum_{s_0, \dots, s_{\tau-1}}$ denotes the summation of all possible state sequences of length τ and $\sum_{u_{\tau+}}$ denotes the sum of every supplementary duration from time τ spent in the state that initially occurred at time $\tau - 1$.

4.3.3 Topology

The Ergodic model and the Bakis model are two of the most popular types of HMMs. In the Ergodic, or fully connected HMM, every state is reachable in one single step from every other state in the model as shown in Figure 4-5(a). In the left-right or Bakis model, the state sequence must commence from the left at state number 1 and end at the right that is the final state N. As time progresses, the observable symbols in each sequence either remains at the same state or increase in a progressive manner. An example of a 4-state

left-right model is shown in Figure 4-5(b). Ergodic HMMs subsume Bakis models. However, Bakis models are generally more efficient as fewer parameters are required. For small HMMs, Ergodic models are chosen instead as efficiency is not the main issue (Rabiner 1989). Selection of a feasible HMM topology is necessary for a specific model.

Figure 4-5 Choice of Hidden Markov models and topology (a) 4 state ergodic HMM (b) 4-state left-right HMM with a maximum of 2 jumps between states

There is no simple, theoretical methodology in ascertaining the best topology for HMM. By performing trial-and-error methods, the number of states is approximated by taking into account the complexity of patterns that the HMMs are required to differentiate. The observable symbols or features of an HMM should be as simple as possible to enable fast computation. Such symbols or features must also contain adequate details so as to be able to specify differences between patterns. The number of symbols is established by the number of possible actions the action extraction level is capable of identifying for each display.

4.3.4 Formulation of State Markov Chain

Consider a first-order N-state Markov chain for N=3 as illustrated in Figure 4-6. The system can be described as bearing one of the N distinct states 1, 2, . . . , N at any discrete

Figure 4-6 Three state Markov chain

time instance t . We use the state variable q as the state of the system at discrete time t . The Markov chain is then characterized by a state transition probability matrix $A = [a_{ij}]$, where:

$$a_{ij} = \Pr(q_t = j \mid q_{t-1} = i), \quad 1 \leq i, j \leq N \quad (4.14)$$

with the following axiomatic constraints:

$$a_{ij} \geq 0 \quad (4.15)$$

and

$$\sum_{j=1}^N a_{ij} = 1 \quad \text{for all } i \quad (4.16)$$

Note that in (4.16) the assumption of a homogenous Markov chain was made so that the transition probabilities do not rely on time. Assume that at $t = 0$ the state of the system q_0

is given by an initial state probability $\pi_i = P_r(q_0 = i)$. Then, for any state sequence $q = (q_0, q_1, q_2, \dots, q_t)$, the probability of q being generated by the Markov chain is:

$$\Pr(q | A, \pi) = \pi_{q_0} a_{q_0 q_1} a_{q_1 q_2} \dots a_{q_{t-1} q_t} \quad (4.17)$$

Supposing that the state sequence q is unavailable for observation, each observation O_t , for instance, a cepstral vector, is envisioned as a portion of the result produced by the system in state q_t where $q_t \in \{1, 2, \dots, N\}$. The creation of O_t in any possible state i ($i = 1, 2, \dots, N$) is assumed to be a stochastic process and is characterized by a collection of observation probability measures $B = \{b_t(O_t)\}$ where

$$b_t(o_t) = \Pr(o_t | q_t = i). \quad (4.18)$$

On the other hand, if the state sequence q leading to the observation sequence $o = (o_1, o_2, \dots, o_t)$ is available, then the probability of o being generated by the system is assumed to be:

$$\Pr(O | q, B) = b_{q_1}(O_1) b_{q_2}(O_2) \dots b_{q_t}(O_t). \quad (4.19)$$

The joint probability of O and q obtained by the system is the product of (4.18) and (4.19), written as:

$$\Pr(o, q | \pi, A, B) = \pi_{q_0} \prod_{t=1}^T a_{q_{t-1} q_t} b_{q_t}(o_t), \quad (4.20)$$

It then proceeds to execute the stochastic process, represented by the observation sequence O , which is denoted by:

$$\Pr(o, q | \pi, A, B) = \sum_q \Pr(o, q | \pi, A, B) = \sum_q \pi_{q_0} \prod_{t=1}^T a_{q_{t-1} q_t} b_{q_t}(o_t), \quad (4.21)$$

This renders the probability of O generated by the system without any assumptions of the information pertaining to the state sequence in which it was generated. The triple $\lambda = (\pi,$

A. B) thus outlines an HMM. The following terminology will be used instead; ‘*i*’ will be used to refer to the model and the model parameter set interchangeably to avoid confusion. The construction of (4.18) resembles that of the incomplete data statistical problem. Having certain interpretations can prove to be useful in gaining initial understanding of a problem. For instance, in the case of speech signal processing, a state represents an abstract speech code such as a phoneme, which can be found in a succession of spectral observations. As speech is usually shaped in a continuous routine, it is often challenging and at times, unnecessary, to figure out how and when a state transition (i.e. from one abstract speech code to the next), is carried out. No assumptions of clear and absolute observations of the state sequence were made, despite the strict implication of the Markovian structure for the state sequence. Hence this is known as a “Hidden” Markov Model.

4.4 The proposed Emotion Index Classification Framework using HMM

Facial expressions are an important channel of nonverbal communication. They communicate a wide range of emotional states. We use Hidden Markov Models (HMMs) to segment and recognize facial expressions from video sequences. In our work, the emotion recognizer with primary emotion (joy, neutral, surprise and sad) is used to infer a secondary emotional state (encouraging, interesting, unsure, disagreeing, discouraging). While many systems are dynamic, they at most consider the progression of facial motion within a single expression. The transition between one facial expression and another is not considered. In addition, facial events are represented at a single time- scale, which is

often close to the capture rate of the raw video stream. A more useful model of facial events and emotions requires the abstraction of larger, more meaningful, elements at temporal scales that are progressively greater than the frame rate sampled. Figure 4-7 presents a model of a emotion state are determined by a feature selection mechanism. To account for, and indeed exploit, the dynamics of periodic and episodic displays, our system employs HMMs for the classification of temporal sequences of actions into a corresponding head or facial display. Displays are described as periodic or

Figure 4-7 The displays present in a model of a emotion state are determined by a feature selection mechanism (partially). For simplicity, the display only shows two emotion states.

episodic depending on the dynamics of the input action sequences. Each display is represented as an HMM that is trained as a classifier for that display. HMMs are particularly suited for representing and classifying displays because they incorporate the dynamics of the actions that constitute these displays, while accounting for variations in these dynamics. Action sequences from a video of arbitrary length are analysed spatio-temporally. Classification that is implemented by a sliding window of observations recognizes a display well within its total duration. The experiments demonstrated reliable recognition of displays that were sampled from a range of complex emotion-states. The

output of the HMM classifier is concatenated to form the input to the topmost level of the computational model of emotion-recognition: emotion-state inference.

In summary, there is a possibility of developing a system for the automated recognition of emotional states beyond the basic emotions. That system would have to account for asynchronous facial cues that occur within a video stream and for the uncertainty in the relationship between facial expressions and underlying emotional states. Finally, it would have to consider the transition between facial expressions. The facial displays are implemented as Hidden Markov Models (HMMs). These are essentially a quantization of a system's configuration space into a small number of discrete states, together with probabilities for the transitions between these states. The training phase at this level involves defining and training an HMM for each of the supported head and facial displays. HMMs are a good choice of classifiers for representing and classifying displays because they encode the temporal regularity inherent in the head/facial actions that constitute these displays. Once the parameters are estimated for each HMM (a one time off-line process), the HMMs can act as classifiers for the online recognition of displays. The problem of classification can be stated as follows: given an HMM model $\lambda = (\pi, A, B)$ and a running sequence of head and facial actions $Z[1 : t]$, the objective is to find the probability that the observations are generated by the model $P(Z[1 : t] | \lambda)$. This is computed using the forward-backward algorithm. The output of each classifier is a probability that generated the vector HMM. To use HMMs, several things such as topology, observation vectors and statistical parameters of HMM, have to be determined. We construct a HMM topology shown in Figure 4-8. In Figure 4-8, a four-state circular HMM model (A, B, π) is created. The states model one of the

emotional events such as Surprise, Sadness, Joy and Neutral state. For example, in Figure 4-8, “N” represents Neutral state. The “SU” represents Surprise emotional state. “SA” and “H” represent the Sadness emotional state and Joy emotional state, respectively. The arrowed lines indicate possible transitions between states. The parameter A, B, and π are determined during the training process. Here

$$A = \{a_{ij} \mid a_{ij} = P(s_{t+1} \mid s_t = q_i)\} \quad (4.22)$$

a_{ij} : the probability of transiting from state q_i to the state q_j .

Figure 4-8 Our Markov Model

$$B = \{b_j(k) \mid b_j(k) = P(v_k \mid s_t = q_j)\} \quad (4.23)$$

$b(k)$: the probability of output symbol v at state q .

$$\pi = \{\pi \mid \pi = P(s = q)\} \quad (4.24)$$

π_i = initial state probability.

Figure 4-9 Our Markov Model (Simplified)

To detect emotional events through HMM like Figure 4-8 (Figure 4-9 is the simplified version), we compute the observation vector sequences first. The observation vector or token v_k is computed from three low level features. Then, we decode the observation vector sequences into the most likely sequence of hidden states by dynamic programming. Emotional events like Joy are detected by identifying sequences of hidden states “N-Hs”. To detect emotional events through HMM like Figure 4-8, we decode state trajectory by dynamic programming. We determine whether emotional events exist or not by counting the number of emotional states. Our system first do the mappings between emotion-state classes and patterns of facial behaviour as observed which are then used during classification to infer the probability of the facial behaviour in an incoming video sequence being "caused" by each of the states. The system combines top-down predictions of emotion-state models with bottom-up vision-based processing of the face to recognize complex emotion-states in real time communication. The output of the system is a number of inference instances, where each instance depicts the probability of

each of the emotion states. Categorisation of emotion-states has to be early enough after their onset to ensure that the resulting knowledge is relevant and actionable. The sequence of actions is presented to a corresponding hidden Markov model (HMM) classifier. The goal is to estimate the most likely emotion-state model which has given rise to the observed facial displays. We have shown the importance of considering inter-expression as well as within-expression facial dynamics in emotion-state recognition. This chapter describes how consecutive actions are analysed spatio-temporally to recognize high-level, communicative, head and facial displays. Displays differ in total duration and intensity. Such variations often signify different user intents. For instance, a long laughter indicates more joy than a weaker or shorter smile. Despite these variations, displays follow a pattern of temporal regularity that can be exploited when analyzing these displays. By modelling the temporal progression of actions across an image sequence, one can infer the underlying display. Displays are interpreted as head or facial events that are meaningful in the contexts of communication. They are the logical unit that people use to describe facial expressions and to link these expressions to emotion-states.

4.5 Proposed Hidden Markov Expert Rule Model (HMER)

We realized that HMM classifiers of facial displays are imperfect: displays may be misclassified or undetected by the system. Both cases result in incorrect evidence being presented to the emotion-state classifier. We want to analyse the effect of recognizing current expression given knowledge about prior emotions. This will enhance the value of our study to detect real effects, avoiding phantom effects that may be caused by random variation in the emotion-recognition abilities of the participants. The emotion-indexer

keeps an archive of past events: every event is stored as a tuple in the index containing the representative frames of the emotion-state, a label, any additional parameters and context cues available. The archive will be able to improve inference and suggestions. The application is coded as a sliding window to incorporate event-history in an efficient way. Timing issues such as latency and frequency of reactions are key factors in the design of this module. A rule-based version of this module has been implemented. Finally, at the topmost level, the model represents $x = 6$ emotion state events $\{X_1, \dots, X_x\}$. For example, X_1 may represent the emotion state agreeing; $P(X_1[t])$ is the probability that agreeing was detected at time t . The probability of a emotion-state event is conditioned on the most recently observed displays and previous inferences of the emotion state: $P(X_i[t] | Y[1 : t], P(X_i[1 : t - 1])$. The classification framework is implemented as a sliding window of evidence. The evidence size is w displays, and it progresses dw actions at a time. At any instant t , the observation vector that is input to the HMM is a vector of w most-recent displays $Y[t - w : t]$, and the corresponding most-recent emotion state inferences $P(X[t - w : t - 1])$. The output is a probability that the observation vector was generated by each of the HMM. The window size or evidence length w describes the number of most recent observations to include as evidence on each HMM invocation. In physical terms, it denotes the amount of inter-expression dynamics or the amount of head and facial display history to consider when making an emotion-state inference. The criterion for choosing a value for w is as follows: a small value of w may result in inaccurate results, as it may not incorporate sufficient inter-expression dynamics during inference. For larger values of w , the system becomes more resilient to noise but the results are produced much later. The system is configured to use the five most recent

head and facial displays for emotion state inference.

Our system is currently implemented as a rule-based system that acknowledges the currently observed facial expression, storing it in an array and suggests the resulting emotion state based on the window of stored facial expression in sequence. The reactions are invoked according to the following criteria. Firstly, persistence that is the number of inference instances of a particular emotion-state has to meet a persistence-threshold to warrant a reaction. This differs from one emotion-state to another. Next, the intensity of an emotion state invokes different reactions. For instance the emotion glad is a mild expression of happiness would result in a different reaction from an enthusiastic emotion-state, even though both belong to the happy group of emotions. The confidence of an emotion state inference has to meet a particular level to invoke a reaction (this is used in combination with the persistence threshold) is another criteria. Lastly, time elapsed since last inference is another factor to consider. There is a minimum threshold that is imposed between recommendations; otherwise it would cognitively overload (and indeed frustrate) the child if a resulting emotion-state is suggested with every incoming facial recognition even if it was a noise.

4.6 Experimental Results - Emotion Indexing with HMER model

Complex emotion states are often expressed through multiple head gestures and facial expressions, which may occur asynchronously. On their own, these displays are weak classifiers of the corresponding emotional states. The rules that govern how people read other people's emotion from nonverbal cues continue to challenge researchers in the domain of behavioural sciences. From an engineering point of view, this means that there

is no "rule-book" to follow when automating these processes. Instead, statistical machine language and data-driven approaches have to be combined with the limited domain knowledge that is available to encode the automated system's functions.

Figure 4-10 illustrates an interaction scenario of emotional recognition through the system, both the physical setup and the user interface. Our implementation of HMER state diagram (subject-independent) has achieved results as shown in Figure 4-11. With our implementation of HMM and Experts' Rules, we have been able to improve inference and avoid to a certain extent phantom effects that may be caused by random variation in the emotion-recognition abilities of the system. The findings from the studies have shown that complex emotion-states are often expressed through multiple head gestures and facial expressions, which may occur asynchronously. On their own, these displays are weak classifiers of the corresponding emotion-states. The studies have also shown that the incorporation of complex emotion-states, and that only a relatively small amount of facial event history accounts for most of the improvement. The rules that govern how people read other people's emotion from nonverbal cues continue to challenge researchers in the behavioural sciences.

In a rule-based system, domain knowledge is encapsulated in rules that have the form of IF A THEN B, where A is an assertion or group of assertions, and B may be an assertion or action. The main problem with using rule-based systems for emotion recognition, as in the system described in Pantic and Rothkrantz (Pantic and Rothkrantz 2000), is that often the input available may be inadequate or insufficiently reliable to enable a conclusion to be reached, or the rules themselves may not be logically certain. This section shows the results achieved by subject-independent of the HMER state

classification. Figure 4-11, Figure 4-12 and Figure 4-13 shows the live chart of the recognition rates corresponding to the different video frames. In overall, the recognition rate is about 75% to 85% and we could deliver less than 0.2sec per frame.

Figure 4-10An interaction scenario of emotion recognition through the system

Figure 4-11 Results of emotion indexer (average)

Figure 4-12 Results of emotion indexer (frames)

(a)

(b)

Figure 4-13 Classification results presented by 3D bar charts and confusion matrices achieved by the HMER model under (a) ergodic Topology, (b) left right topology.

The proposed HMER model aims at estimating the most likely emotion-state model that has given rise to the observed facial expressions. The emotional state models are the HMER models that represent secondary emotional states in which each emotional states class is modelled as a separate HMER where the hidden emotional state of each HMER is a binary transition. By having a HMER classifier per class, it is possible that the output of more than one class is true. Hence, emotional states that are not mutually exclusive or may co-occur, such as “Interest” and “Encouraging”, can be represented by the system.

Examples of the emotional state models are shown in Figure 4-14. It can be shown that the emotional state is inferred from an initial state (primary emotion) to a next state (secondary emotion) by the HMER model.

Moreover, Figure 4-15 demonstrates the screenshots from the proposed emotion indexer in which the primary emotions are recognized by the facial expression recognizer. The HMER model infers the secondary emotions. Figure 4-15(a) shows a subject displaying ‘sad’ as the primary emotion from her facial expression and the emotion displaying “disagreeing” as the secondary emotion that is detected by the HMER model. Similarly, Figure 4-15(b) shows a subject displaying surprise as the primary emotion and then the HMER model detects and infers the secondary emotion is “encouraging” that the corresponding “emotion-icon” has been displayed. Table 4-1 shows our result is as comparative as other models.

Table 4-1 Results comparing our model against other real-time facial expression recognition engines performing onto public domain databases

Approach	Accuracy (%)
Hidden Markov Model	83.2
BayesNet	72.3
Adaboost	74.6
MLP	69.2
RBF	64.3
Bagging	72.4
J48	76.3
oneR	77.1

Figure 4-14 Emotional state model by the HMER model, (a) From "Unsure" to "Interest" emotional state via a "Joy" transition, (b) From "Interest" to Encouraging" emotional state via a "Surprise" transition.

(a)

(b)

Figure 4-15 GUI of Emotion Indexer(a) Sadness expression with Disagreeing state (b) Surprise expression with Encouraging state

A closer analysis of the results explains why our undetected and false positive displays occurred though. The specific reasons for mis-classifications for each display are discussed in the sections that follow. These reasons fall into one of the following three cases:

1. Error at the action recognition level: An incorrect action sequence, which also persists, results in an incorrect output by the corresponding HMM classifier.
2. Error at the HMM level: Failure at the HMM level is mainly due to under- represented patterns of a particular display in the training examples.
3. Noise in coding: Some noise is introduced when coding the videos because of difficulties in determining the precise time of onset and offset of a display.

As a result, Experts' Rules was used to minimize the noise and avoid phantom effects that may be caused by random variation in the emotion recognition abilities of the system. We realized that HMM classifiers for facial displays are imperfect: displays may be mis-classified or undetected by the system. Both cases result in incorrect evidence being presented to the emotion state classifier. We want to analyse the effect of recognizing a current expression given knowledge about previous ones. This will

increase the power of the study to detect real effects, avoiding phantom effects that may be caused by random variation in the emotion-recognition abilities of the participants.

4.7 Conclusion

This chapter presents the suitability of a Hidden Markov Expert Rule (HMER) model in representing the dynamics of displays and a classification framework is demonstrated that enables their fast recognition, so that there is no significant delay. Experimental results showed reliable and real time recognition of displays in a context of complex emotional states.

In contrast to static classifiers that classify single frames into an emotion class, dynamic classifiers model the temporal information inherent from facial events. HMM are one of the basic (and perhaps the best known) probabilistic tools used for time series modelling. HMM has been successfully implemented in a number of applications including speech recognition, handwriting recognition, head gesture recognition, and automated facial expression recognition to classify facial action units and basic emotions. HMER model is particularly suited for the recognition of facial displays from action sequences. It provides a sound probabilistic framework for modelling time-varying sequences.

Currently, we are working on the improvement of the recognition model in both structure and algorithm to obtain better accuracy and reduce the computation time for training. We are looking into developing a computer-based system that can interact with humans that would inspire us to build a real-time system for helping people to read and respond to the people from facial expressions.

Furthering on the emotion indexing, the next chapter presents the framework on emotion prediction through the use of Genetic Algorithm. The emotion indexer and emotion predictor work together to provide intelligent information and act as part of the overall assistive technology.

Chapter 5. Genetic Algorithm

based Emotion Prediction Model

5.1 Introduction

Earlier chapters we had looked at emotion-state recognition and emotion indexing which form the building blocks of overall emotion system. This chapter underlines the role of emotion prediction as part of the human interactive process in order to reinstate the importance of nonverbal communication in this domain. The main theme here is concerned with the aspects from the emotion prediction approach to the importance of nonverbal as an instrument of usability evaluation and the role of predicting emotions in human interaction. We make use of four basic human facial expressions that are commonly presented, i.e., happy, surprise, sad and neutral, to predict the five common secondary emotions namely: encouraging, interest, unsure, disagreeing, and discouraging. We adopt Genetic Algorithm (GA) to model emotion prediction in a Time Series Multiple Regression manner. The model is capable of estimating optimal parameters through learning from historical data. Multiple-Regressive Integrated Moving Average (MRIMA) models are integrated with Genetic Algorithm in order to overcome the linear and data limitations of MRIMA models, thus obtaining more accurate results. The

experiments were conducted to confirm these hypotheses by evaluating the predictive capability of the developed ensemble of models in the domain of emotion prediction.

The role of predicting emotions is essential in understanding user's behaviours while interacting with others. Human interaction suffers from several challenges like re-orientation beyond efficiency to better understand the way people perceive emotion. Without neglecting the constructs of informatics, we cannot disregard those of affects, predicting emotions, satisfaction when relating to human interaction domain (Ekman 1999). Recent research has shown that it is not necessary to set up a stiff delimitation between cognition and emotions when we study human interaction. The importance of the psychological aspect of predicting emotions in the domain of human interaction can lead to solving some important aspects concerning human interaction. Recent research works pointed out the primordial role of affects over the rationality concerning both the evolution of brain and also its function in real time (Lavie 2004). In addition, some laboratory experiments have shown that information processing in the neocortex is slower than processing it with the emotional brain. Also, the number of connections from limbic system to neocortex is larger than the number of connections that start from the rational brain (neocortex) and head towards the emotional one (the limbic system) (Kaliouby *et al.* 2005). Therefore, we should not jump to the conclusion that the functionality of the emotional brain is independent from the rational one, or vice versa. Instead of judging in terms of opposition between rational and emotional, we should point out that predicting emotions are indispensable for human knowledge. However, explaining the human behaviour and their interaction exclusively in terms of rational or

emotional elements is very difficult. Therefore, we have to find a middle way to tackle them.

The basic emotions, included happiness, surprise, fear, anger, disgust and sadness, are in some circumstances associated with facial expressions (Kaliouby 2005). The first known facial expression analysis was presented by Darwin in 1872 (Darwin 1872). He presented the universality of human face expressions and the continuity in man and animals. He pointed out that there were specific inborn emotions, which originated in serviceable associated habits. After about a century, Ekman and Friesen (1971) postulated six primary emotions that possessed each a distinctive content together with a unique facial expression. In 2002, Jeremy R. Gray (Gray 2002) showed that emotions and cognition play a balanced role in controlling thoughts and behaviors. Therefore, emotions and cognition are mutually correlated: emotions contribute to behavior and thought adjustment and cognition contributes to adjusting emotions. The domain of artificial intelligence concurs with this paradigm of integrating emotions with cognition. So, studying emotions by measuring the facial expressions in human interaction offers the chance of identifying the difficulties that users stumble upon (Hassenzahl 2006).

Many models were introduced to recognize human emotion in the past three decades (Kaliouby *et al.* 2003). However, none of these models has proven to perform well in predicting emotion. In this chapter, we propose to use Genetic Algorithms (GA) as an alternative approach to derive these models. GA is a powerful search and optimization techniques to estimate the parameters of well known reliably growth models. The main motivation to choose GA for this task is its capability of estimating optimal parameters through learning from historical data. Moreover, machine learning

algorithms, proposed the solution to overcome the uncertainties in modelling by combining multiple models aiming at a more accurate prediction at the expense of increased uncertainty. Some experiments in the literature were conducted to confirm these hypotheses by evaluating the predictive capability of the developed ensemble of models and the results were compared with traditional models.

In this thesis, we present a method to deal with the prediction problems which can lead to a higher accuracy rate than the existing models, such as, fuzzy logic or neural networks. Most likely, Genetic-based approach is one of the computational intelligence techniques that could be used for prediction using historical data. Therefore, we propose a genetic-based approach used in predicting emotion based on a sequence of facial images. Detailed results are provided to explore the advantages of using GA in solving this problem.

This chapter covers the emotion prediction as shown in Figure 5-1 which takes emotion states as an input and use Genetic Algorithm to predict the emotion and feed into Reaction Advisor as one of the three inputs. The existing work on emotion prediction was based on a traditional and cognitive approach. We will present a new method to deal with the forecasting problems based on high-order fuzzy time series and genetic algorithms, where the length of each interval in the universe of discourse is tuned by using genetic algorithms and the historical enrolments of the University of Alabama are used to illustrate the forecasting process of the proposed method. The proposed method can achieve a higher forecasting accuracy rate than the existing methods.

Figure 5-1 Block Diagram for Emotion Predictor, predicting emotion from emotion states, also act as one of the three input to Reaction Advisor

5.2 Genetic Algorithm for Prediction

Genetic Algorithm (GA) is a type of machine learning algorithm that possesses great machine learning capabilities and optimization techniques for estimating parameters. Such machine learning algorithms offer a sophisticated framework for dealing with uncertainty in modeling by aggregating various models targeting at higher prediction accuracy, but at the expense of increased uncertainty (Fogel 1998). Our system adopts GA as the model for emotion prediction in a Time Series Auto Regression manner. The emotions include encouraging, interest, unsure, disagreeing, and discouraging. GA has

the ability to perform estimation and optimization of parameters by learning from past data. Recently, numerous researchers have proposed various methods of forecasting based on fuzzy time series in order to overcome forecasting problems. To tackle forecasting problems, it is vital to determine the length of each interval in the universe of discourse as this directly affects the forecasting accuracy rate.

Similar to neural networks, genetic algorithms are machine learning and optimization schemes. Yet, genetic algorithm is capable of avoiding local minima, which neural networks do not do as well. Genetic algorithms are established upon the model of evolution, where a population survives and achieve overall fitness, despite the fact that individuals from the population perish. According to the theory of evolution, superior individuals are more likely to reproduce than inferior individuals, thus they stand a better chance in passing their above-average traits on to the following generation. Akin to the survival-of-the-fittest, this criterion was firstly introduced to an optimization algorithm by Holland in year 1975 and has evolved to become a key optimization technique for complicated, nonlinear problems in the present-day (Markoff, 1990).

Within a genetic algorithm, every single individual of a population represents a potential solution to an optimization problem, encoded as a binary string known as a chromosome. A collection of such individuals is generated to vie for the reproduction privileges or they may even advance to the next generation of the population. Competition is executed by applying a fitness function to all the individuals in the population; the individuals yielding the best results are the fittest. The following generation is formed by passing over several of the best individuals in order to perform reproduction and mutation. Reproduction is executed by a “crossover” operation. This is

analogous to what takes place in the embryo of an animal. A new pair of individuals is formed by allowing two chromosomes to exchange segments of their codes. The simplest version of crossover is executed by selecting a crossover point on the pair of chromosomes in an arbitrary manner. After that, the chromosomes exchange all remaining data from that selected point onwards, while maintaining their own data up to the selected point. A mutation operator randomly alters a bit or more in some chromosome(s) to introduce greater degree of diversity in the population. In general, the mutation rate is kept minimal so as to retain the stability of good solutions in the population. The representation of solutions and the fitness function are the two most crucial elements of a genetic algorithm, and both of them are dependent on the problem. The latter is the doorway to bringing the entire population of solutions towards a global best. The coding of the solution must cater for the representation of an implicit, intricate idea or a sequence of procedures. As for the fitness function, it is mandatory to perform interpretation of the encoded solutions, as well as derivation of ranking among the different solutions.

We propose a genetic-based approach to predict emotion based on historical data. Detailed results are documented to explore the advantages of implementing GA in the system. Figure 5-2 illustrates the basic steps in the canonical genetic algorithms. The algorithm starts with populating the genes followed by calculating their fitness values. For each of the gene, the corresponding mating pairs are selected and attempt to produce offspring with the matching. The gene is then mutated, and the newly produced offspring is inserted to the population database. A set of criteria is pre-defined and it is testing to see if this whole process needs to be repeated again.

Figure 5-2 Flow chart for GA population

5.3 Predicting Models

Hundreds of models were utilized for predictive purposes in the past few decades. Many research works were geared towards the problem of constructing growth models that can assist in prediction (Minerva T 2002). In prediction research, the application of Non-Homogeneous Poisson process (NHPP) models, time series analysis and Bayesian inference are the three most common trends. An NHPP is a Poisson method consisting of a time-varying mean value function. For Bayesian inference, it involves managing the parameters of the prediction model as random variables rather than constants for approximation. After assuming certain prior and realistic distributions for these parameters, Bayes' theorem is invoked for ascertaining subsequent distributions using reliable data. On the other hand, time series analysis makes use of an auto-regressive

procedure and an auto-regressive integrated moving average (ARIMA) model. Apart from the massive trends mentioned earlier on, there are numerous software reliability models that offer seemingly unique features. An auto-regression model is adopted in this work that is introduced in the following section.

5.4 Regression Model

A time series is defined as a time-ordered sequence of observation values of a physical or fiscal variable occurring at equally-spaced time intervals Δt , denoted as a set of discrete values. Time series analysis deals with the problems related to the identification of basic characteristics/features of time series. By analysing the observation data from which the time series is constructed, it uncovers the internal time series structure for the prediction of time series data values, influencing the subsequent actions to be taken. The auto-regression model is commonly used among the different time series models. Much of its appeal stems from the simplicity and accessibility offered by many popular statistical packages (Baragona 2001). The AR model can be described using the following equation:

$$y_j = \omega_0 + \sum_{i=1}^n \omega_i y_{j-i} \quad (5.1)$$

where y_{j-i} is the number of faults previously observed and $i = 1, 2, K, n$. The value of n is termed as the "order" of the model, ω_0 and ω_i , $i = 1, 2, K, n$ are the model parameters.

5.4.1 Time Series Multiple Regression Model

Multiple-Regressive Integrated Moving Average (MRIMA) models are one of the most important time series models used in financial markets. They have been implemented for

forecasting purpose for the past three decades and time series forecasting has drawn considerable interest for research and applications of various fields (Minerva 2001). Two fundamental limitations have been discovered through recent research activities in time series forecasting. (a) MRIMA models may be inadequate for complex nonlinear problems as the models assume that futuristic values of a time series shares linear relationships with current and past values as well as with white noise and (b) in order to deliver accurate result, MRIMA models require a large quantity of historical data. Theoretical and empirical findings have suggested the integration of various models can lead to the enhancement of their predictive abilities. This is particularly true when the models involved are relatively different. Improved forecasting accuracy by the use of different models has been proven by empirical results of financial markets. Hence, this model was proposed as an alternative for financial market forecasting tools. To achieve greater accuracy, MRIMA models are fused with Artificial Neural Networks (ANNs) and fuzzy logic so as to overcome the linearity and data limitations of MRIMA models.

An error occurs when reality is compromised in a model, reflecting the disparities between the real-world aspects of interest and its corresponding model representation. Error measuring can be used to quantify how effective a model is in representing reality. Stochastic uncertainty should be categorized as one of the characteristics of reality as faults occur during the software testing process can manifest and behave in countless, erratic ways. Such errors can have a correlation with the model's structure due to approximation, assumptions, generalization or uncertainties in arbitrary values assigned to the model parameters. Certainly, errors can occur during the measurement process itself. Machine learning algorithms aim to attain greater predictive accuracy at the

expense of increased uncertainty. The fusion approach, which is an amalgamation of the average predictions of multiple models, is taken here. This can be described by the following equation:

$$y_j = M_j(\Omega_j, S_j) \quad (5.2)$$

where the model's prediction about a real-world aspect of interest represents the model's structure reflecting a collection of assumptions and generalizations that is encrypted into the mathematical model. $Y_j S_j M_j \Omega_j = (\omega_0, \omega_1, \omega_2, K)$ is a finite set of model parameters.

In a general case of a discrete set of n models Ψ , each model $M_j(\Omega_j, S_j), j = 1, 2, K, n$ symbolizes an alternate form with the given set of parameters. Every model in the set $S_j \Omega_j \Psi$ provides estimation about the magnitude of interest in the form of a predictive probability distribution. The literature on the combination of methods is abundant and diverse; among the methods, the simple averaging (equal weights) and the weighted average (Hashem 1993) are the most discussed one. In this study, the combination function v is implemented in both schemes. Equation (5.3) represents the average predictions of multiple models whereas equation (5.4) represents the weighted average predictions of multiple models

$$y = \frac{1}{n} \sum_{j=1}^n y_j \quad (5.3)$$

$$y = \frac{1}{n} \sum_{j=1}^n w_j y_j \quad (5.4)$$

5.5 Our Model - GA based Emotion Prediction Model

Based on the principles of evolution and natural genetics, generic algorithms are

randomized search-and-optimization techniques, and works by maintaining a population of individual solutions. This is according to the theory of “survival-of-the-fittest” evolution, where “stronger” individuals will survive and reproduce offspring while “weaker” ones will be eliminated. In the field of pattern recognition, there are many tasks that require analysing or identifying patterns. These can only be executed with the selection of suitable parameters and having efficient searches in complex spaces so that optimum solutions can be reached (Li, 1999). The application of GA in solving problems associated with pattern-recognition is appropriate and acceptable. This is especially pertinent for tasks that requires optimization of computation, and demands quick, stable and close-approximate solution. In our model, we are trying to use GA to model our regression by finding out what is the optimum number of independent variable (time series regression). Let $X \equiv \{X_1, X_2, ..., X_m\}$ be the set of m independent variables which representing the historical emotional results, where n represents the total number of observations, and Y be the dependent variable in a multivariate regression model. $X_1=X_{t-1}$ which represent a value of 0 to 1 where 0 to 0.20 is encouraging, 0.21 to 0.40 is agreeing, 0.41 to 0.6 is unsure, 0.61 to 0.8 is disagreeing and 0.81 to 1 is discouraging.

$$Y = \beta_0 + \beta_1 \bar{X} + \beta_2 \bar{X}_2 + + \beta_p \bar{X}_p + \varepsilon \quad (5.5)$$

The model (above) is our fitness function and it depicts the relationship shared between the dependent and independent variables, where $\bar{X} \subseteq X$ is the set of the $p \leq m$ independent variables chosen as regressors and $B \equiv \{\beta_0, \beta_1, ..., \beta_p\}$ is the parameters set. If B is estimated as $\hat{B} \equiv \{\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p\}$ by using the Ordinary Least Square method (OLS), the main task that remains is to pick the independent variables that should

be included in $\bar{X} \subseteq X$. This task is an optimization problem and the objective is to select $\bar{X} \subseteq X$ such that the estimated models:

$$Y = \beta_0 + \beta_1 \bar{X}_1 + \beta_2 \bar{X}_2 + \dots + \beta_p \bar{X}_p \quad (5.6)$$

is optimal and competent to satisfy certain statistical criteria. In addition to the selection of the top subset $\bar{X} \subseteq X$ of independent variables, the next task could be to find the type of mathematical transformations that can be applied to the independent variables so that the overall adequacy of the model can be further enhanced. Examples of functions that can be tested include exponential and logarithmic transformation. In this situation, the challenge lies in the selection of subset $\bar{X} \subseteq X$ and the mathematical transformations $f: \bar{X}_i \rightarrow T_i(\bar{X})$ such that the model:

$$Y = \beta_0 + \bar{\beta}_1 T_1 \bar{X}_1 + \bar{\beta}_2 T_2 \bar{X}_2 + \dots + \bar{\beta}_p T_p \bar{X}_p \quad (5.7)$$

obtains optimal results after being accessed by the statistical criterion. P is the total of frames used for forecasting the subsequent emotion.

Different standards could be adopted in quantifying the degree of optimality of a regression model. Three statistical criteria are elected to stop the iteration for GA. Firstly is the Asymptotic Information Criterion (AIC) (Akaike 1969):

$$AIC(p) = n \log(s_p^2) + 2p \quad (5.8)$$

where n is the number of observations, p is the number of independent variables in the

regression model, $S_p^2 = \frac{\sum_{i=1}^n (E_i)^2}{n-p-1}$ is the variance of the residual when p independent

variables is considered, and E_i are the residuals. The AIC offers an approximation of the

“distance” between the estimated model and the unknown mechanism behind the known data. The model with the lowest AIC is the best. The amount of parameters that needs to be taken into consideration by the optimal model is often greater than the actual number.

The Bayesian Information Criterion (BIC) (Akaike 1969) is characterized by:

$$BIC(p) = (n - p) \log\left(\frac{nS_p^2}{n - p}\right) + p \log\left(n \frac{S_0^2 - S_p^2}{p}\right) \quad (5.9)$$

where p represents the number of independent variables in the model, n refers to the number of observations, S_p^2 is the variance of the residual when p independent variables is considered and S_0^2 is the variance of all the observations of the dependent variable. In this case, the model with the lowest BIC value is the best. On the contrary, BIC’s aim is to consider the reduction of variance. This is done by estimation of the model with p covariates.

As for the Schwarz Information Criterion (SIC) (Schwartz G, 1978):

$$SIC(p) = \log(s_p^2) + \frac{p}{n} \log(n) \quad (5.10)$$

where p represents the number of independent variables in the model, n refers to the number of observations, S_p^2 is the variance of the residual when p independent variables is considered. The model with the lowest SIC is the best.

Expert-based selection, stochastic search algorithms and classical stepwise approaches are approaches proposed for the selection of regression model. For expert-based selection, it is usually a synthesis of trial-and-error and the expert’s past experience, where he or she attempts to manually identify the variables to be

incorporated. Different models are then investigated and the one that is deemed to be the most suitable model is then chosen. However, the method of selection is subjective and cannot promise a regression model with optimal adequacy.

Classical stepwise regression applies techniques for finding a model that possess optimal adequacy that satisfies the statistical criterion by taking alternative models into consideration. All the models are tested at every step, unlike the best, current model which includes/excludes one independent variable. If any of these models yield better result than the current best model, it replaces and becomes the current best model and the search is resumed from that point onwards. The execution of the algorithm ends when no better model differing in one variable is found. The backward stepwise regression method begins with a complete model; with the number of variables decrease at each iteration. In contrast, the forward stepwise regression method begins by taking all the models with only one independent variable into account. Defined as a local search process, stepwise regression iteratively attempts to improve the existing solution by advancing to the best neighbour to determine if it is better. Otherwise the search terminates. Eventually, every single run of the process congregates to a local optimum.

Population-based stochastic search heuristics is probably the most favourable method in handling multiple local optima in non-linear optimization problems. One example will be genetic algorithms; they explore the search space concurrently by a population of candidate solution which solutions compete and recombine. While they are capable of selecting the most suitable variables in a regression model, GAs can also discern the most fitting transformations of the independent variables. Images are converted to integer values that are used to decide if a variable should be included and

which transformation will be applied. The algorithm commences by random generation of the population, after which every GA individual is evaluated according to the AIC, BIC or SIC criterion which in our case, we are selecting AIC as the criteria to stop the algorithm.

The approximation of regression parameters $\bar{B} = \{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p\}$ were carried out using the training set whereas the output values were calculated for the validation set with increasing competency of the variable selection mechanism. The calculations of R^2 and the Residual Mean Squared Error (*RMSE*) statistics were tabulated with reference to the test set. The R^2 statistics refers to a simplistic linear regression model with intercept, where the dependent variable is the Y on the test set and the independent variable is the estimated value:

$$Y = \beta_0 + \beta_1 \bar{X}_1 + \beta_2 \bar{X}_2 + \dots + \beta_p \bar{X}_p \quad (5.11)$$

on the test set with $\bar{B}$ estimated on the training set. The Residual Mean Squared Error is calculated as

$$RMSE = \sqrt{\sum_{i=1}^k (Y_i - \bar{Y}_i)^2 / k} \quad (5.12)$$

where k is the length of the test set.

In this thesis, the GA is trained once for each emotion (total 5 trainings). We define the number of generation as (G) and population as (P). We found out experimentally that the bigger the values of parameter G and P , the better the prediction results. This program only runs when there is a need to train or re-train the system. The third program uses the

coefficient found out earlier to predict the next emotion. This program runs continuously. Three programs are utilized to predict emotions. The first program captures the facial expression from the live program (VB) and records into 5 files (binary), namely encouraging.txt, interest.txt, unsure.txt, disagreeing.txt, and discouraging.txt. This program runs continuously. The second program trains the coefficient (β_0 to β_s) using GA. The coefficient is our time series regression coefficient. The third program is used to predict emotions using test data. An overview of the software flow and design incorporating the use of GA is shown as follows in

Figure 5-3 and Figure 5-4.

1. input.txt– main.vbp write the emotion to this file as neutral sad surprise happy
2. neutral.txt, happy.txt, surprise.txt, sad.txt – They are the binary files generated by GA_live.m base input.txt
3. co_eff.txt – This file is generated by GA_Train.m and use in Prediction.m

Figure 5-3 Flow chart for GA programming

Figure 5-4 Flow chart for GA system

5.6 Experiment Result - Emotion Prediction with Genetic Algorithm

Through GA, a certain portion of the population is chosen as parents for reproduction. The probability of the rule being chosen is proportional to the fitness of the rule based on our fitness function (equation 5.5) base on AIC. Our maximum iteration time is 5min. The fitness function measures the strength of rules not only based on their individual performance in terms on classification accuracy and coverage, but also based on their competitive contribution in the population.

The experiments have shown an encouraging result of over 60% accuracy and in some cases, over 70%. However, more research is needed to further improve the algorithm and to extend it to a more general situation. It is known that elitist model of GAs provide the optimal string as the number of iterations goes to infinity when the probability of going from any population to the one containing the optimal string is greater than zero.

The system has been tested using video sequence data and live data. We have adopted the Mind Reading DVD video sequence data (refer to page 85). The significant population is chosen as parents for reproduction. The probability of the rule chosen is proportional to the fitness of the rule based on our fitness function. The fitness function measures the strength of rules not only based on their individual performance in terms on classification accuracy and coverage, but also based on their competitive contribution in the population. Based on our fitness function (equation 1) and implementation (subject-independent), we achieved the following results for our emotional prediction with generic algorithm as follows in Figure 5-5, Figure 5-6 till Figure 5-11. Figure 5-5 explains the independent variable (1st independent variable is t_1 , 2nd independent variable is t_2 and so on) selection. We could observe that about 30 independent variables give the optimum results. Figure 5-5(c) shows that about 30 independent variable give best result for criterion value as well. Number of generation results is show in Figure 5-6. In terms of generation, an optimum vale is about 800 generations. Figure 5-7 till Figure 5-11 illustrates the overall results and error for emotions Unsure, Interesting, Encouraging, Discouraging and Disagreeing.

(a) RMSE versus independent variable which shows minimum error ~ 30

(b) R Square Determinant shows best result around 30 independent variable, similar to RMSE

Figure 5-5 Results of Experiment (Number of Independent Variable)

Table 5-1 Results for the three criterion

Number of Independent Variable: 10		R Square	RMSE	AIC	BIC	SIC
Unsure		0.755	0.231	0.870	0.925	0.831
Interest		0.762	0.228	0.887	0.903	0.805
Encouraging		0.758	0.226	0.858	0.912	0.804
Disagreeing		0.753	0.234	0.860	0.917	0.806
Discouraging		0.748	0.238	0.880	0.949	0.802
Number of Independent Variable: 20		R Square	RMSE	AIC	BIC	SIC
Unsure		0.759	0.227	0.820	0.881	0.78
Interest		0.767	0.219	0.833	0.895	0.789
Encouraging		0.768	0.217	0.834	0.902	0.81
Disagreeing		0.758	0.224	0.826	0.892	0.815
Discouraging		0.756	0.222	0.839	0.872	0.812
Number of Independent Variable: 30		R Square	RMSE	AIC	BIC	SIC
Unsure		0.768	0.220	0.794	0.86	0.774
Interest		0.779	0.209	0.807	0.846	0.769
Encouraging		0.776	0.211	0.793	0.854	0.755
Disagreeing		0.762	0.219	0.838	0.844	0.785
Discouraging		0.76	0.223	0.837	0.836	0.768
Number of Independent Variable: 40		R Square	RMSE	AIC	BIC	SIC
Unsure		0.758	0.234	0.860	0.883	0.787
Interest		0.769	0.23	0.842	0.897	0.785
Encouraging		0.77	0.229	0.842	0.899	0.797
Disagreeing		0.763	0.236	0.856	0.905	0.791
Discouraging		0.757	0.230	0.832	0.871	0.802

(a) RMSE base on number of generation. This shows that optimum generation is about 700

(b) Accuracy base on number of generation. This shows that optimum generation is about 700

Figure 5-6 Results of Experimental (Number of Generations)

(a) Result for “Unsure”

(b) Error for “Unsure”

Figure 5-7 Results of “Unsure”

(a) Result for “Interest”

(b) Error for “Interest”

Figure 5-8 Results of “Interest”

(a) Result for “Encouraging”

(b) Error for “Encouraging”

Figure 5-9 Results of “Encouraging”

(c) Result for “Discouraging”

(d) Error for “Discouraging”

Figure 5-10 Results of “Discouraging”

(e) Result for “Disagreeing”

(f) Error for “Disagreeing”

Figure 5-11 Results of “Disagreeing”

5.7 Conclusion

A real-time facial expression predictor system is presented in this paper. The system adopts a genetic-based approach in predicting emotion based on historical data. GA works according to theory of survival of the fittest evolution; relatively “strong” individuals will live and reproduce offspring, while “weak” ones will die off. In the area of pattern recognition, there are many tasks involved in the process of analysing or identifying a pattern that need appropriate parameter selection and efficient search in complex spaces in order to obtain optimum solutions. Therefore, the application of GAs for solving certain problems of pattern recognition (which need optimization of computation requirements, and robust, fast and close approximate solution) appears to be appropriate and natural. The described system makes use of four basic human facial expressions which are commonly presented, i.e., happy, surprise, sad and neutral, to predict the six common secondary emotions namely, Encouraging, Interest, Unsure, Disagreeing, and Discouraging. The experiments have shown an encouraging result of over 60% accuracy and in some cases, over 70%. However, more research is needed to further improve the algorithm and to extend it to a more general situation.

This emotion predictor lays ground for the next chapter on emotion advisor. The emotion advisor takes facial expression, emotion index and emotion prediction as input and generate advises to user.

Chapter 6. Fuzzy Rules based Emotion Advisor Model

6.1 Introduction

This research is based on the combination of fuzzy logic, learning and pattern recognition together with a neuroscience understanding of cognitive and visual signal interplay in bridging the communication chasm between autistic children and the world. The state-of-the-art research on fuzzy system implementations is studied. After which, the computational framework of the proposed fuzzy system is introduced. The methodology of fuzzy sets and membership, reaction-to-emotion association by fuzzification and defuzzification and fuzzy IF-THEN rules are discussed. This research focus is on examining the feasibility of applying fuzzy principles into the emotional advisor, the functionality and practicality of the features of emotional advisor. Results from the experiments conducted have shown that the emotional advisor, which forms the last module of the intelligent emotion system as shown in Figure 6.1, has fulfilled the essential criteria of mobile application: speed and efficiency.

Figure 6-1 Block Diagram for Emotion Advisor, which takes facial expression, emotion index and emotion prediction as input and generate advises to user

This unique characteristic of fuzzy logic is essential in tackling real-world scenarios because the world is full of uncertainty (Biacino 2002). However, controversies still remain among control engineers whose preferences sway towards two-valued logic and statisticians who only accept Bayesian logic. Nevertheless, fuzzy logic has been successfully incorporated in many of the specialized fields today and it has also been a research topic that is extensively studied over the past few decades. Because of its ability to tame uncertainty, fuzzy logic is a logic theory that suits the nature of our project and can be adopted to fashion the framework of our emotion advisor.

In 1965, Lotfi Zadeh (Zadeh, 1965) came up with the basic notion of fuzzy logic, which is formulated to tackle fuzzy or quantities that are not describable in terms of probability distributions. Fuzzy logic, as the name imply, is not logic that is fuzzy, but logic that describes and handles fuzziness. Fuzzy logic works on the idea that all things are subjected to degrees of truth. For instance, Tom is 170cm, considering the average height of guys in the room is 160cm, we may say that Tom is tall. However, if we add a few more guys into the room and now the average height of guys in 180cm, is he still tall? Using conventional logic, Tom is said to be tall in the first case, but after the change in average height, he becomes short. If this is the case, then when did he start to become short? Conventional logic cannot give a solution for this, but fuzzy logic can. In fuzzy logic, we can say that Tom is tall to 0.7 extents, where 1 means he is definitely considered tall. Basically, fuzzy logic is a multi-valued logic that permits intermediate values to be defined, instead of conventional ‘crisp’ logic, for example, binary sets that have two-value logic. By means of fuzzy logic, variables such as *very cold* or *a little warm* can be defined and mathematically computed. Fuzzy logic has been applied in many specialized fields like artificial intelligence and control theory. Fuzzy logic’s approach of controlling problems is akin to how a real person makes decisions.

6.2 Fuzzy System

6.2.1 Introduction

Fuzzification is the process of transforming crisp values into grades of membership for fuzzy sets. The membership function associates a grade to each term defined in the sets

(Kosko 1993). Emotions are complex and vague; hence there is a need to associate each emotional input to a fuzzy set, to accurately pinpoint the overall, dominant emotion that the user is experiencing. Individual may display different reactions towards certain emotions. Therefore, this emotion-to-reaction association is not a one-to-one function. For instance, when one is happy, one may start to sing. Alternatively, others may express happiness by buying an ice cream for themselves. Both are logical and subjected to the individual's preferences.

The fuzzy inference engine consists of several membership functions mapping the input variables to the fuzzy sets defined, and a set of fuzzy IF-THEN rules (Novak 2005). The input variables in this fuzzy system are the three inputs derived from the facial expression recognizer, and they are mapped into sets of membership functions known as fuzzy sets. Each of the five emotional states represents a membership function as shown above. The figure depicts that each state has an overlapping region with another state. This highlights the distinct feature of fuzzy logic; instead of setting an arbitrary threshold of *happy* and *sad*, incorporating the overlapping regions allows the states to change gradually from one to the next. Consequently, the emotional states no longer abruptly change from one state to the next. On the contrary, as the emotional input changes, it loses degree of belongingness in one membership function while gaining entry to the next function. This accurately portrays emotions as complex and constituting of a mixture of different emotions. Given the input variables and the truth-values produced, the system then decides on the type of advice to generate for the end user based on a set of fuzzy IF-THEN rules. This combination of fuzzy operations and rule-based inference system creates a fuzzy system.

Many practical applications use a relatively restricted yet important part of fuzzy logic that centres on the use of IF-THEN rules (Gerla 2005). This aspect of fuzzy logic comprises of collection of concepts and methods for handling a diversity of knowledge which can be represented in the form of a set of fuzzy IF-THEN rules whereby the antecedents, consequences, or both, are fuzzy rather than crisp values. The term ‘crisp’ refers to exactness of an entity. Fuzzy values are defined as fuzzy because they partially belong to one or more sets. A set that consists of fuzzy values is known as fuzzy sets. In essence, the IF-THEN rules convert inputs to outputs, one fuzzy set into another. Fuzzy logic allows the conversion of linguistic control strategy based on expert knowledge, into an automated control strategy. The main beauty is that fuzziness of the antecedents eliminates the need for an exact match with the input; hence giving room for ambiguity that is omnipresent in almost every situation (Cordon 1999).

6.2.1 Fuzzification

Fuzzy set theory deals with a subset A of the universe of discourse X , where the transition between full membership and no membership is gradual rather than sudden. The fuzzy subset A is defined as a membership function $\mu_A(x)$ which maps the degree to which an element x belongs to a fuzzy subset A from domain X to the range of $[0,1]$.

Fuzzification is the conversion of a crisp entity into a fuzzy value that has grades of memberships to fuzzy sets. In this section, we discuss the underlying reasons for emotional association and the derivation of membership functions for each emotional class. Fuzzification of a real-valued variable is done with intuition, experience and analysis of the set of rules and conditions associated with the input data variables. There

is no fixed set of procedures for fuzzification. In the emotional advisor, the fuzzification procedures implemented makes use of the Triangular Fuzzifier. Mamdani-style inference is applied alongside with the membership functions to determine the truth-value of the consequent membership function (Mamdani 1974). Defuzzification of the resultant truth-value will be done by using the Max-membership Defuzzification method. In our case, the truth-value essentially determines the resultant emotional state. By aggregating the three emotional inputs (emotion recognizer using Naïve Bayesian, emotion indexer using Hidden Markov and emotion predictor using Genetic Algorithm) and using the designed membership functions and Mamdani-style inference method, the association between the inputs and overall output emotion can be established.

The three popular Fuzzifiers are listed as follows:

a) Singleton Fuzzifier:

$$\mu_A(x) = \begin{cases} 1 & \text{if } x = x^* \\ 0 & \text{otherwise} \end{cases} \quad (6.1)$$

b) Gaussian Fuzzifier:

$$\mu_A(x) = e^{-\left(\frac{x_1 - x_1^*}{\partial_1}\right)^2} \times \dots \times e^{-\left(\frac{x_n - x_n^*}{\partial_1}\right)^2} \quad (6.2)$$

where ∂_i is a positive constant

c) Triangular Fuzzifier:

$$\mu_A(x) = \begin{cases} \left(1 - \frac{|x_1 - x_1^*|}{b_1}\right) \times \dots \times \left(1 - \frac{|x_n - x_n^*|}{b_1}\right) & \text{if } |x_i - x_i^*| \leq b_i, i = 1, \dots, n \\ 0 & \text{otherwise} \end{cases} \quad (6.3)$$

where b_i is a positive constant

In our case, we are using triangular fuzzifier as it is very fast in computation. After the aggregation process, there is a fuzzy set for each output variable that needs defuzzification.

The Mamdani-style fuzzy inference process is performed in four steps:

- i) Fuzzification of input variables
- ii) Rule evaluation (inference)
- iii) Aggregation of the rule outputs (composition)
- iv) Defuzzification

After fuzzifying the inputs and applying the membership functions to them, the resultant output is fuzzy. These fuzzy outputs need to be converted into a crisp output quantity so that the nature of the action to be performed can be determined by the fuzzy system. This process, which converts the fuzzy output into a crisp entity is called defuzzification. Before an output is defuzzified, all the fuzzy outputs of the system are aggregated with union operator. The union is the maximum of the set of given membership function and can be expressed as $\mu_A = \bigcup_i (\mu_i(x))$

There are three defuzzification techniques that are commonly used:

a) Maximum Defuzzification Technique:

This technique is given by algebraic expression $\mu_A(x^*) \geq \mu_A(x)$ for all $x \in X$ where x^* is the defuzzified value. It gives the output with the highest membership function. This defuzzification technique is very fast but is only accurate for peaked output. Based on the overall truth-values derived, the maximum defuzzification technique is applied to obtain the crisp entity, which comes in the form of an emotion class.

b) Centroid Defuzzification Technique:

This method is also called ‘centre of gravity’ or ‘centre of area’ defuzzification. This is the most commonly used technique and is very accurate. The centroid defuzzification technique can be expressed as:

$$x^* = \frac{\int \mu_i(x) \times x \times dx}{\int \mu_i(x) \times dx} \quad (6.4)$$

where x^* is the defuzzified output, $\mu_i(x)$ is the aggregated membership function and x is the output variable. The only disadvantage of this method is that it is computationally difficult for complex membership functions.

c) Weighted Average Defuzzification Technique

In this method the output is obtained by calculating the weighted average of each output of the set of rules stored in the knowledge base of the system. The weighted average defuzzification technique can be expressed as:

$$x^* = \frac{\sum_{i=1}^n m^i w_i}{\sum_{i=1}^n m^i} \quad (6.5)$$

where x^* is the defuzzified output, m^i is the membership of the output of each rule, and w_i is the weight associated with each rule. This method is computationally faster and easier and gives fairly accurate result. This defuzzification technique is applied in fuzzy application of signal validation and fuzzy application on power.

6.3 Emotion Advisor

We examined the possibility and feasibility of using fuzzy logic to create the emotion advisor. However, the test instances and the number of possible outputs are limited in creating a meaningful and practical emotion advisor. In order to accomplish this feat, a wider range of emotional inputs is being introduced. With the increase in the possible permutations, the variety of outputs also multiplies. The aggregations of emotion-to-reaction association become slightly more complex with the influx of emotional inputs but the methodology of creating the entire fuzzy system remains unchanged.

The three inputs derived from the facial expression recognizer are fed to the fuzzy emotion advisors in the form of three plain text file namely *output.txt* (*emotion recognizer*), *ei.txt* (*emotion indexer*), and *predict.txt* (*emotion predictor*). Four possible inputs can be reached for *output.txt* and 5 possible inputs from *ei.txt* and *predict.txt*. A total of 100 unique combinations can be formed based on the three different inputs (4 from emotion recognizer x 5 from emotion indexer x 5 emotion predictor).

We started our test on the feasibility of implementing fuzzy logic in the emotion advisor. To achieve this, different values are deployed for each emotional input. It is to simplify and reduce the magnitude of the prototyping process. The testing is executed in real-time using MATLAB programs. The performance of using fuzzy logic in the emotion advisor is compared and benchmarked against other popular classifiers. Detailed analysis of the output results are discussed and summarized in the later section.

As highlighted earlier, there are 100 unique combinations that can be formed by the 3 different emotion inputs (emotion recognizer, emotion indexer and emotion predictor). To determine the number of classes (i.e. the number of possible emotion

advices that can be generated) and the number of data combination that the system should deploy, experiments are conducted with different number of classes and data-sets using several classifiers provided by WEKA. From the given results, the number of classes and data sets to be deployed in the actual test set is determined. This test set will be used for the testing of the fuzzy rules and used for the comparison between all the different classifiers. The results are tabulated and presented in the later section.

In order to carry out various tests, there is a need to attribute the different emotional inputs to a single, specific emotional state. This emotion-to-reaction association is constructed based on the acceptability, and could be understood by the other party during communication. This association may differ slightly based on different individuals.

The following Figure 6-2 illustrations outline the association between various emotional inputs. Emotion recognizer consists of joy, sadness, neutral and surprise. As for emotion indexer and emotion predictor, there are five classes: encouraging, agreeing, unsure, discouraging and disagreeing.

Figure 6-2 The aggregations of the three emotional inputs (emotion recognizer, emotion indexer and emotion predictor) to emotion advisor output

6.3.1 Our Fuzzy Model

Firstly, in order to aggregate the three different kinds of inputs (emotion recognizer, emotion-indexer and emotion-predictor), we need to perform fuzzification so that we can see the degree of membership each of these emotions has. Membership functions (triangular) are represented in a graph that defines how each point in the input space is mapped to membership value between 0 and 1. Input space is often referred to as the universe of discourse or universal set (μ), which contains all the possible elements of concern in each particular application. Several membership functions are deployed for the emotional aggregation. Emotion can interpreted as positive emotion and negative emotion (Gaulin 2003). In our case, we translate emotion as continuous from positive emotion like Joy and encouraging, to negative emotion like sad and discouraging.

Membership function of the input using triangular fuzzifier is shown as follows (graphical representation Figure 6-3 to Fig 6-5):

Figure 6-3 Membership function for input 1 (emotion recognizer)

Figure 6-4 Membership function for input 2 (emotion indexer)

Figure 6-5 Membership function for input 3 (emotion predictor)

The fuzzy semantic is using maximum value. For example, a 0.22 for emotion recognition will imply 0.8 for Joy, 0.2 for neutral, 0 for surprise and 0 for Sad, we would take the maximum which is Joy. The resultant value derived from the membership functions determines the degree to which the aggregation of the three inputs belongs to the fuzzy set of the dominant emotion. The membership value only ranges from 0 to 1,

hence the maximum threshold of 1 is taken. Using Mamdani's Rule, the three inputs (emotion recognizer, emotion indexer and emotion predictor) are fuzzified. Results from these fuzzy sets are aggregated to form the resultant, composite fuzzy value.

6.3.2 Fuzzy IF-THEN Rules

Based on the 3 emotional inputs (4 from emotion recognizer, 5 from emotion indexer and 5 from emotion predictor), there are 100 (4x5x5) possible combinations to generate emotion 10 advises (classes). 30 IF-THEN rules are selected base on the training data. These rules are applied in parallel. When the input values correctly match the rules, the fuzzy system generates an output in the form of a reaction advice. The nature of advice generated by the system defines the resultant overall emotional state of the user. This process of combining fuzzy values and converting back to one crisp entity is known as defuzzification.

The subsequent tables 6-1 presented in the following list out part of the 30 fuzzy IF-THEN rules implemented in the system, the detail of the rules is in appendix B. We are making use of fuzzy rules, Multiple-Input Single-Output (MISO), which in our case is Three-Input Single-Output. The rules are represented as such:

R_i : If w is A_i and x is B_i and y is C_i , then z is D_i

where A_i , B_i and C_i are fuzzy numbers, $i=1, \dots, n$ (in our case, $n=30$)

w = emotion recognizer

x = emotion indexer

y = emotion predictor

and z = emotion advisor

Table 6-1 Examples of fuzzy IF-THEN designed for the emotional advisor

Rule		Conditions		Consequences
1	IF	emotion recognizer is neutral	AND	
		emotion indexer is unsure	AND	
		emotion predictor is unsure	THEN	Generate neutral advise
2	IF	emotion recognizer is joy	AND	
		emotion indexer is encouraging	AND	
		emotion predictor is interesting	THEN	Generate happy advise

These rules are actualized and simulated in the MATLAB program to test the feasibility of real-time generation of reaction-advice. For faster computation, decision-tree structure is administered. By feeding in the sample input data, the experiments show that the program can generate advice smoothly using one-second intervals.

6.4 Experimental Results and Analysis

6.4.1 First Phase Results (5 advises/classes)

Because the number of classes (Advices) used in the testing sets differs, representations of each emotional state are adjusted according to the number of classes that were made available. For each test set, the emotion-to-reaction associations are refined so as the classes made available can adequately differentiate and articulate the emotional advice. The emotional advice is the truth-value obtained by aggregation of the three emotional inputs (emotion recognizer, emotion indexer and emotion predictor).

Fifty instances were randomly selected and extracted from the entire available combination and formed 8 different sets of data sets to test the predictive performances of fuzzy rules and other popular classifiers. The results are summarized and shown below (Figure 6-6 to 6-8, note that classes imply advises).

Figure 6-6 The results obtained by classifiers for data-set 1 & 2 in percentage

Figure 6-7 The results obtained by classifiers for data-set 3, 4 & 5 in percentage

Figure 6-8 The results obtained by classifiers for data-set 6, 7 & 8 in percentage

6.4.2 Final Results (10 advises/classes)

Initially, data sets with 60 test instances (3 inputs from emotion recognizer, emotion indexer and emotion predictor with 1 output of emotion advisor) are created. These data sets are tested using classifiers provided by WEKA for benchmarking purpose. However, the results are of lower standard because the test instances are insufficient to accurately predict the 10 classes that we have arbitrary defined. Results are shown in Figure 6-9 and Figure 6-10. Data sets are modified and test instances are increased from 60 to 80 which is shown in the next paragraph. Note that classes imply advises.

Figure 6-9 The test results obtained by classifiers for data-set A & B in percentage

Figure 6-10 The test results obtained by classifiers for data-set C & D in percentage

Using 8 different sets of test data, results using various classifiers and our own fuzzy model are tabulated as shown in the following. From the Figure 6-11, we can see that the fuzzy system achieved a hit rate of 66.25% in new data-set A, and the second-in-place is a tie between Naïve Bayes, SMO, LBR, which obtained an accuracy of 55%, followed by VFI, with an accuracy of 50%. The subsequent results generated by the data-sets are similar to that of data-set A.

New Data-set B(80 data, 10 classes)

New Data-set C(80 data, 10 classes)

New Data-set D(80 data, 10 classes)

New Data-set E(80 data, 10 classes)

Figure 6-11 The results obtained by classifiers for the eight data-sets in percentage

Results (Figure 6-11) obtained from the experiments of all 8 data-sets show that the fuzzy classifier achieved at least 62.5% accuracy, which is slightly higher than the other classifiers (like Decision Trees) used. The performance of fuzzy logic is fairly constant, having an accuracy that ranges from 62.5% to 71.25% for the 8 sets of test data. Other popular classifiers such as Naïve Bayes and LBR managed to achieve an acceptable accuracy rate of 61.25% and 68.75% in new data-set H and E.

6.6 Conclusion

We can conclude base on the experimental results that incorporating fuzzy logic in the emotional advisor is a feasible method that can yield a relatively high accuracy in predicting the correct emotional state and generating appropriate advice for the end user. As fuzzy expert systems are modelled empirically, they have the potential to catalyse better performance and are receptive to changes and improvements. The research we present here significantly advances the nascent ability of machines to infer cognitive-affective emotional states in real time from non-verbal expressions of people. By using fuzzy logic in developing a real-time system for the inference of a wide range of emotion states beyond the basic emotions, we have widened the scope of human-computer interaction scenarios in which this technology can be integrated. This is an important step towards building socially and emotionally intelligent machines. In order to implement the emotional advisor in the real-world situation, the standards of the fuzzy rules should be determined by experts in the field of psychology, emotions and autism spectrum disorders to correctly represent and establish the correlation between various emotions, and also to supply suitable reaction-responses which autistic children is able to understand and perform when generated by the emotional advisor.

Chapter 7. Conclusion and Further Developments

This last chapter of the thesis describes the principle contributions of this research, followed by several directions for future work. It ends by summing up the progress it has made towards the development of socially and emotionally intelligent interfaces.

7.1 Contribution

This thesis started with addressing the problem of automated inference of complex emotional states, the group of affective and cognitive states of the mind. These emotional states are not part of the basic emotions set which are easily observable from facial expression. This is a challenging endeavour due to several reasons. Firstly, there is uncertainty inherent from the inference of hidden emotional states. Secondly, the automated analysis of the face is an open machine-vision problem. Thirdly, there is a lack of knowledge about the facial expression composition of complex emotion states. In the light of these challenges, this thesis makes four principal contributions in addressing the research problem. Firstly, the thesis describes a computational model of emotion-recognition as a novel framework for machine perception and emotion state inference. The second contribution is that the research undertaken here particularly focuses on complex emotion states, which advances the state-of-the-art in affective computing beyond basic emotions. Thirdly, this thesis has emphasized on a working prototype of an

emotion state inference system that runs in real time and is thus suited for direct application to human-computer interaction. Finally, we present a body of knowledge about the configurations and dynamics of facial expressions of complex emotional states, which have the potential to inform future work in face perception and emotions.

The computational model of emotion-recognition is a novel framework for machine perception and emotional state recognition. The model describes a coherent framework for fusing low-level behaviour in order to recognize high-level emotion state concepts. It is defined as a multi-level probabilistic graphical model that represents a raw video stream at three levels of abstraction: actions, displays and emotion states. Each level of the model captures a different degree of spatial and temporal details that are determined by the physical properties of the facial event at that level. The framework works well due to its hierarchical, probabilistic architecture and it is a good match to the attributes of complex emotion states. The multi-level representation mimics the way with which people perceive facial behaviour in a hierarchically structured manner. It also accounts for the inter-expression dynamics that, through several studies, are found to improve human's recognition of complex emotional states. The top-most level of the model is implemented using HMM. These classifiers allow multiple asynchronous observations of head and facial displays to be combined within a coherent framework, and provide a principled approach to handle the uncertainty inherent in emotion state inference. The output probabilities and their development over time represent a rich modality analogous to the information humans receive in our everyday interaction through emotion-recognition. The application of automated facial expression analysis to human-computer interaction is limited to primitive scenarios where the system responds

with simple positive or negative reactions depending on which basic emotion the user is expressing. This is because basic emotions are of limited utility in understanding the user's cognitive state of emotion and intentions. The automated inference of complex emotion states is a significant step forward from existing facial analysis systems that only address the basic emotions. Recognizing mental states beyond the basic emotions widens the scope of applications in which automated facial expressions analysis can be integrated, since complex emotion states are indicators of the user's goals and intentions. The automated emotion-recognition system combines bottom-up vision-based processing of the face with top-down predictions of emotion state models to interpret the meaning underlying head and facial signals. The system executes in real time, does not require any manual pre-processing. It is user independent, and supports natural rigid head motion. These characteristics make it suitable for application to HCI. The classification accuracy, generalization ability, and real time performance of the system were evaluated for five groups of complex emotion states—agreeing, disagreeing, interested, encouraging and unsure. The videos of these emotion states were sampled from two different sources—the Mind Reading DVD and real-time capturing. The results show that the automated emotion-recognition system successfully classifies the five emotional state groups, generalizes well to new examples of these classes, and executes automatically with an accuracy and speed that are comparable to that of human recognition. The research described throughout this thesis provides insight to the configuration and dynamics of facial expressions in complex emotion states, which is lacking in the literature. The findings from the studies have shown that complex emotion states are often expressed through multiple head gestures and facial expressions, which may occur asynchronously.

On their own, these displays are weak classifiers of the corresponding emotional states. The studies have also shown that the incorporation of inter-expression dynamics, or previous facial events, boosts the recognition results of complex emotional states, and that only a relatively small amount of facial event history accounts for most of the improvement.

7.2 Conclusion and Future Works

In summary, although most of the facial expression analyzers developed so far target on human facial affect analysis and attempt to recognize a small set of prototypic emotional facial expressions like happiness and anger, some progress has been made in addressing a number of other scientific challenges that are considered essential for realization of machine understanding in human facial behaviour. Existing methods for machine analysis of facial expressions discussed thus far assume that the input data are near frontal- or profile-view face image sequences showing facial displays that always begin with a neutral state. In reality, such assumption is not warrant. The discussed facial expression analysers were tested on spontaneously occurring facial behaviour, and extract information about facial behaviour in less constrained conditions such as in an interview setting. However, the deployment of existing methods in fully unconstrained environments is still in the relatively distant future. Development of robust face detectors, head and facial component trackers, which will be robust to variations in both face orientation relative to the camera, occlusions, and scene complexity like the presence of other people and dynamic background, forms the first step in the realization of facial expression analyzers capable of handling unconstrained environments.

There are two aspects still unsolved. The first issue is how the grammar of facial behavior can be learned and how this information can be properly represented and used to handle ambiguities in the observation data. Another issue is how to include information about the context in which the observed expressive behaviour was displayed so that a context-sensitive analysis of facial behaviour can be achieved. These aspects of machine analysis of facial expressions form the main focus of the current and future research in the field. Yet, since the complexity of these issues concerned with the interpretation of human behaviour at a deeper level is tremendous and spans several different disciplines in computer and social sciences, we believe that a large, focused, interdisciplinary, international program directed towards computer understanding of human behavioural patterns (as shown by means of facial expressions and other modes of social interaction) should be established if we are to experience true breakthroughs in this and the related research fields.

Currently, we are working on the improvement of the recognition model structurally and algorithmically to obtain better accuracy and reduce the computation time for training. We are looking into developing a computer-based system that can interact with humans. Also, with about 80 percent recognition rate, we are currently going to build a real-time system for autistic children to read and respond to the emotions of people from facial expressions. This system will identify a facial event in real time, extracts the dynamic features from facial expressions and infers the underlying emotion state conveyed by the video segment.

Future work worth looking into includes a reaction advisor in the form of a GUI to display current emotion state inference and a recommended action, both textually and

graphically. Future directions will include the re-implementation of the reaction advisor using partially observed Markov decision processes, so that the utility of the actions is also learnt from data rather than hard-coded as in the current rule-based implementations, hence utilizing the information to suggest an appropriate reaction. Representative frames of that event along with the inferred emotion states label are sent to the emotional indexer to be archived. The emotional indexer module functions like a database; every event is stored as a tuple in the index containing the representative frames of the emotion state, a label, any additional parameters and context cues available. The indexed events are made available (through the interface layer) to the child and carer for discussion, learning and reviewing purposes. This extends the emotional indexing approach to allow events to be replayed. The archive is also made accessible to the system and the reaction advisor modules to improve inference and suggestions. The reaction advisor appraises the current video input, analyzing it within any context cues that are available, to suggest appropriate courses to actions to take. Timing issues such as latency and frequency of reactions are key factors in the design of this module. A rule-based version will be implemented in this module. We aim to develop the system to successfully classify and generalize to new examples of other emotion state classes with an accuracy and speed that are comparable to that of human recognition. We also aim to improve our prediction of the next emotion to allow anticipation.

We also aim to make considerable use of the contexts in which facial expressions occur to assist interpretation, including situational context. A simple implementation of location-context would entail defining several profiles which reflect the various situations the child can be in, such as “in school” or “in playground”. The contexts would have to

be explicitly selected, but as the tool gets more sophisticated, more detailed profile information would be deduced automatically by the system.

In theory, the users would benefit from the integration of context and other modalities in the system since many people diagnosed with Asperger syndrome have problems integrating emotion state concepts from facial expression into wider contexts (e.g. previous encounters with a person or an environment). The integration of context and a study of its effect on recognition accuracy is a research direction worth pursuing. Future works regarding the emotional indexer having the ability to review, define emotion classes and update the Index through the GUI is another good direction we steering towards. With this system in the future, an autistic child will make use of this aiding device which will inform them in what the facial emotion brings into their attention during the social communication.

Appendix A - Abbreviation

AAI	Adult Attachment Interview
AAM	Active Appearance Models
ASD	Autism Spectrum Disorders
AU	Action Unit
AVI	Audio Video Interleave
BNB	Boosted Naïve Bayesian
BNT	Bayes Net Toolbox
CVPR	Computer Vision and Pattern Recognition
DBN	Dynamic Bayesian Network
DVD	Digital Video Disc
FACS	Facial Action Coding System
FCM	Fuzzy C-means
FER	Facial Expression Recognition
FPS	Frames Per Second
GA	Genetic Algorithm
HCI	Human-Computer Interaction
HMM	Hidden Markov Model
HMER	Hidden Markov Expert Rules
ICA	Independent Component Analysis
LFA	Local Feature Analysis

LLE	Local Linear Embedding
MI	Mutual Information
MLE	Maximum Likelihood Estimation
MLP	Multilayer Perceptron
NB	Naïve Bayesian
NMF	Non-negative Matrix Factorization
PBVD	Piecewise Bezier Volume Deformation
PCA	Principal Component Analysis
PGM	Probabilistic Graphical Model
PNL	Intel's Probabilistic Networks Library
ROC	Receiver Operator Characteristic
SD	Standard Deviation
SVDD	Support Vector Data Description
SVM	Support Vector Machines

Appendix B - Publications

Journal Papers:

1. Teik-Toe Teoh, Yok-Yen Nguwi and Siu-Yeung Cho, "Towards a Portable Intelligent Facial Expression Recognizer", *Intelligent Decision Technologies, An international Journal*, vol. 3:3, pp. 181-191, Published and Featured July 2009. (featured by IOS press, CNET and Science Daily on 19 Oct 2009)
2. Yok-Yen Nguwi, Teik-Toe Teoh and Siu-Yeung Cho, "Developing an Assistive Technology to Help Children with Autism for Recognising Human Emotion ", *HKIE Transactions*, 17(4), pp.61-68 2010 (shortlisted paper for The HKIE Outstanding Paper Award for Young Engineers/Researchers 2010).
3. Yok-Yen Nguwi, Siu-Yeung Cho and Teik-Toe Teoh, "Prototype Ranking for Solving Imbalanced Conditions of Facial Emotion Recognition", *International Journal of Innovative Computing, Information and Control* (ISI impact factor: 2.932), May 2010.
4. Yok-Yen Nguwi, Teik-Toe Teoh, Insu Song, Siu-Yeung Cho, "Real-Time Road Sign Recognition", *Australian Journal of Intelligent Information Systems*, Vol 12, No 4 (2010): Machine Learning Applications (Part II).
5. Teik-Toe Teoh, Kiat-Han Kok, Yok-Yen Nguwi, Oruganti Venkata Ramana Murthy, Siu-Yeung Cho, "Cognitive-based Online Face Matching System", *Journal of on Computer Sciences and Convergence Information Technology Korea*, Nov 2011

6. Teik-Toe Teoh, Shi-Min Lim, David Siu-Yeung Cho, Yok-Yen Nguwi, "Advances in Assisted Tool for Autism", Journal of Computer Sciences and Convergence Information Technology Korea, Nov 2011
7. Teik-Toe Teoh, Siu-Yeung Cho, "Expert Rules based Emotion Indexer Using Hidden Markov Model", *Expert Systems and Application*, under review, Oct 2010.
8. Teik-Toe Teoh; David Siu-Yeung Cho, Yok-yen Nguwi "Emotional Prediction using Time Series Auto-Regression Genetic Algorithm (TSARGA) for children with autism", in preparation

Conference Papers:

1. Teik-Toe Teoh, Kiat-Han Kok, Yok-Yen Nguwi, Oruganti Venkata Ramana Murthy, Siu-Yeung Cho, "Cognitive-based Online Face Matching System", The 6th International Conference on Computer Sciences and Convergence Information Technology Korea, 2011
2. Teik-Toe Teoh; Shi-Min Lim; David Siu-Yeung Cho, "Emotional Advisor to help Children with Autism in Social Communication", The 6th International Conference on Computer Sciences and Convergence Information Technology Korea, 2011
3. Teik-Toe Teoh, Siu-Yeung Cho and Yok Yen Nguwi, "Cognitive Assistive Probabilistic-based Emotional State Recognition (CAPESR) for Assisting Children with Autism, Decade of the Mind VI 2010
4. Teik-Toe Teoh, Yok-Yen Nguwi, "Emotion Indexing using Hidden Markov Expert Rule Model (HMER) for autism children", International Conference on Control, Automation, Robotics and Vision Dec 2010. Singapore. pp 668 - 672

5. Teik-Toe Teoh, Yok-Yen Nguwi and Siu-Yeung Cho, “Intelligent Face Locator for Smartphone”, in IEEE International Symposium on Industrial Electronics 2009, July 2009, Seoul, Korea. pp 1662 – 1667.
6. Teik-Toe Teoh, Siu-yeung Cho, Yok-yen Nguwi “Hidden Markov Model for Detecting Hard-Drive Failure”, The 7th International Conference on Computer Science and Education (ICCSE 2012), Melbourne, Australia, Jan 2012
7. Teik-Toe Teoh, Siu-Yeung Cho and Yok-Yen Nguwi. “Emotional Prediction using Time Series Multiple-Regression Genetic Algorithm for Autistic Syndrome Disorder”, The 7th International Conference on Computer Science and and education (ICCSE 2012), Melbourne, Australia, 2012
8. Siu-Yeung Cho and Teik-Toe Teoh, "A Real-time Facial Emotion Recognition System Using Fuzzy Neural Networks", *International Conference on Artificial Intelligence (ICAI-08)*, Las Vegas, USA, March 2008.
9. Siu-Yeung Cho, Teik-Toe Teoh and Yok-Yen Nguwi, "Development of an Intelligent Facial Expression Recognizer for Mobile Applications". KES International Symposium on *Intelligent Decision Technologies*, Himeji, Japan, 2009.
10. Yok Yen Nguwi and Teik-Toe Teoh, “Two-tier Emergent Self-Organizing (TtEsom) Approach of Understanding Emotions”, International Conference on Control, Automation, Robotics and Vision 2010. Singapore. pp 654 – 658

Book Chapters:

1. Siu-Yeung Cho, Teik-Toe Teoh and Yok-Yen Nguwi, “Development of an Intelligent Facial Expression Recognizer for Mobile Applications”, Studies in

Computational Intelligence: New Advances in *Intelligent Decision Technologies*, Springer Berlin / Heidelberg, pp. 21-29, 2009

2. Siu-Yeung Cho, Teik-Toe Teoh and Yok-Yen Nguwi, "Advanced Feature Selection and Classification methods", In Yu-Jin Zhang (eds.), *Advances in Face Image Analysis: Techniques and Technologies*: Tsing Hua University, IGI Global Publishing, pp. 239-258, 2009.

Appendix C - References

- Abboud, B. (2004). Appearance factorization based facial expression recognition and synthesis. International Conference on Pattern Recognition.
- Aggarwal, S. (2004). "Semantic-level Understanding of Human Actions and Interactions using Event Hierarchy." In IEEE Workshop on Articulated and Non Rigid Motion at the IEEE Conference on Computer Vision and Pattern Recognition.
- Aha, D. and D. Kibler (1991). "Instance based learning algorithms. Machine Learning." Machine Learning 6: 37-66.
- Akaike, A. (1969). "Statistical predictor identification." Annal of the Institute of Statistical Mathematics 22: 203-217.
- Alexandrova, A. (2005). "Search for the Lowest-Energy Structures Using the ab Initio Gradient Embedded Genetic Algorithm (GEGA). Elucidation of the Chemical Bonding in the Lithium Clusters." Journal of Chemical Theory and Computation: 556-580.
- Anderberg, M. R. (1973). Cluster analysis for applications: probability and mathematical statistics, New York : Academic Press, .
- Anderson, K. and McOwan P.W. (2006). "A real-time automated system for the recognition of human facial expressions." IEEE Transactions on Systems Man and Cybernetics Part B 36(1): 96-105.
- Ang, K. K. and Quek C. (2005). "RSPOP: Rough Set-Based Pseudo Outer-Product Fuzzy Rule Identification Algorithm." Neural Computation 17(1): 205-243.
- Ashish Kapoor, R. and Yuri Ivanov (2004). "Probabilistic Combination of Multiple Modalities to Detect Interest." International Conference on Pattern Recognition (ICPR) 3: 969-972.
- Association, A. P. (1994). "Diagnostic and Statistical Manual of Mental Disorders." 4th Edition. Washington, DC: American Psychiatric Association, DSMIV.

- Baragona, R. and Calzini C. (2001). "Genetic Algorithms for the identification of additive and innovation outliers in time series." Computational Statistics and Data Analysis 37: 1-12.
- Baron-Cohen, S. (1995). Mindblindness: An essay on autism and theory of mind. Cambridge, MIT Press.
- Baron-Cohen, S. (2003). "The Essential Difference: The Truth about the Male and Female Brain." New York: Basic Books.
- Baron-Cohen, S. (2006). "The hyper-systemizing, assortative mating theory of autism." Progress in Neuro-Psychopharmacology & Biological Psychiatry(30): 865–872.
- Baron-Cohen, S. and 1994. (1994). "How to Build a Baby That Can Read Minds: Cognitive Mechanisms in Mindreading." Current Psychology of Cognition13(5): 513-552.
- Baron-Cohen, S., O. Golan, et al. (2004). Mind Reading: The Interactive Guide to Emotions. Lndon, Jessica Kingsley Publishers.
- Bartlett, M. S., G. Littlewort, I. Fasel and J. R. Movellan (2003). Real time face detection and facial expression recognition: Development and applications to human computer interaction. CVPR, Madison.
- Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. J. Experimental Psychology, Vol. 4, No. 3: 373-379.
- Baudry, B. F. F., Jean-Marc Jézéquel, and Yves Le Traon (2005). "Automatic Test Case Optimization: A Bacteriologic Algorithm." IEEE Software (IEEE Computer Society) 22: 76-82.
- Biacino, L. and Gerla, G. (2002). "Fuzzy logic, continuity and effectiveness". *Archive for Mathematical Logic* **41** (7): 643–667.
- Black, M. J. and Y. Yacoob (1997). "Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion." Internationall Journal of Computer Vision 25(1): 23-48.
- Bourel, F., Chibelushi, C.C. and Low, A.A. (2002). Robust facial expression recognition using a state-based model of spatially-localised facial dynamics. IEEE International Conference on Automatic Face and Gesture Recognition.

- Bremner D, D. E., Erickson J, Iacono J, Langerman S, Morin P, Toussaint G (2005). "Output-sensitive algorithms for computing nearest-neighbor decision boundaries." Discrete and Computational Geometry 33(4): 593-604.
- Casti, J. L. (1989). Alternate Realities: Mathematical Models of Nature and Man Wiley-Interscience
- Chandrasiri, N.P., Min Chul Park, Naemura, T. and Harashima, H. (1999). Personal facial expression space based on multidimensional scaling for the recognition improvement. Proceedings of the Fifth International Symposium Signal Processing and Its Applications.
- Cheng L., and Orlich F. (2003). "KidTalk: online therapy for Asperger's syndrome." Technical Report, Social Computing Group, Microsoft Research.
- Ching, W., Ng, M., & Fung, E. (2003). "Higher-order hidden Markov models with applications to DNA sequences." IDEAL, Lecture notes in computer science, Springer.
- Chow, T. and D. Huang (2005). "Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information." Neural Networks, IEEE Transactions on 16(1): 213-224.
- Cohen, P. (2003). "High Frequency of Facial Expressions Corresponding to Confusion, Concentration, and Worry in an Analysis of Naturally Occurring Facial Expressions of Americans." Emotion 3: 68-75.
- Cohn, J. F. and J. Zlochow (1998). Feature-Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression. International Conference on Face & Gesture Recognition, IEEE Computer Society.
- Cohn., J. F. (2004). "What the face reveals (2nd edition): Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System (FACS), chapter Automated Analysis of the Configuration and Timing of Facial Expressions." New York: Oxford University Press Series in Affective Science.
- Cordon O., M. J. del Jesus & F. Herrera (1999). "A proposal on reasoning methods in fuzzy rule-based classification systems." International Journal of Approximate Reasoning, 20 (1), 1999, pp.22-45

- Cottrell, G. W. and M. K. Fleming (1990). Categorisation of Faces Using Unsupervised Feature Extraction. Int'l Conf. Neural Networks, San Diego.
- Cross, S. B. (1992). "Reading the Eyes: Evidence for the Role of Perception in the Development of a Theory of Mind." Mind and Language6(173).
- Darwin, C. (1872). The Expression of the Emotions in Man and Animals, J. Murray, London.
- Dash, M. and H. Liu (1997). "Feature selection for classification." Intelligent Data Analysis1(3): 131–156.
- Daugman, J. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J. Opt. Soc. Amer. 2: 1160-1169.
- Dautenhahn K. (2002). "Games children with autism can play with Robota, a Humanoid Robotic Doll. In: Keates S, Clarkson PJ, Langdon PJ, Robinson P (eds) Proceedings of the 1st Cambridge workshop on Universal Access and Assistive Technology (CWUAAT)." Springer, London: 179-190.
- Dawson, M., Soulieres, I., Gernsbacher, M.A. and Mottron, L. (2007). "The level and nature of autistic intelligence." Psychological Science (18): 657-662.
- Deza, E. D. (2009). "Encyclopedia of Distances." Springer, London: 94.
- Edwards, G. J. and T. F. Cootes, et al. (1998). Face Recognition Using Active Appearance Models. Proc. European Conf. Computer Vision, Vol. 2: 581-695.
- Ekman, P. (1969). "The Repertoire of Non-verbal Behaviour: Categories, Origins, Usage, and Coding." Semiotica1: 49-98.
- Ekman, P. (1979). "Human Ethology, chapter About Brows: Emotional and Conversational Signals." London: Cambridge University Press: 169-200.
- Ekman, P. (1992). "An Argument for Basic Emotions." Cognition and Emotion 6: 169-200.
- Ekman, P., Darwin T. and Dalglish, T.. (1999). "Deception and Facial Expression." Handbook of Cognition and Emotion 301-320.
- Ekman, Paul, Darwin, Deception and Facial Expression. In T. Dalglish, T., Power, M., Handbook of Cognition and Emotion (pp. 301-320). Sussex, UK: JohnWiley and Sons Ltd, 1999.

- Ekman, P. and W. V. Friesen (1971). "Constants across cultures in the face and emotion." J. Personality Social Psychol 17(2): 124-129.
- Ekman Paul, and W. V. Friesen (1978). "Facial Action Coding System: A Technique for the Measurement of Facial Movement." Palo Alto, CA: Consulting Psychologists Press.
- Ekman Paul, W. V. Friesen, and Sonia Ancoli (1980). "Facial Signs of Emotional Experience." Journal of Personality and Social Psychology 39: 1125-1134.
- Essa, I. A. and A. P. Pentland (1997). "Coding, analysis, interpretation, and recognition of facial expressions." IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7): 757-763.
- Evgeniou, T. and M. Pontil (2003). "Image Representations and Feature Selection for Multimedia Database Search." IEEE Trans. on Knowl. and Data Eng. 15(4): 911-920.
- Fasel, B. and J. Luetten (2003). "Automatic facial expression analysis: a survey." Pattern Recognition 36(1): 259-275.
- Fasel, B. and J. Lüttin (2000). Recognition of Asymmetric Facial Action Unit Activities and Intensities. Proceedings of the International Conference on Pattern Recognition, Barcelona, Spain.
- Fasel, I. R. (2006). Learning Real-Time Object Detectors: Probabilistic Generative Approaches. PhD thesis, Department of Cognitive Science. University of California, San Diego, USA.
- Fasel, I. R., B. Fortenberry, et al. (2005). A generative framework for real time object detection and classification. Int'l J Computer Vision and Image Understanding. 98: 181-210.
- J. D. Ferguson, "Variable duration models for speech," in Symp. Application of Hidden Markov Models to Text and Speech, Oct. 1980, pp.143–179
- Fogel, D. B. (1998). "Evolutionary Computation: The Fossil Record." New York: IEEE Press.
- Fogel, D. B. (2000). "Evolutionary Computation: Towards a New Philosophy of Machine Intelligence." New York: IEEE Press: 140.

- Forney, G. (1973). "The Viterbi algorithm." IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 61(3): 268-278.
- Frith, U. (2001). "Mind Blindness and the Brain in Autism. *Neuron*, 32(6):969–979, 2001." Neuron 32(6): 969-979.
- La Cara G.E., M. Ursino and M. Bettini (2003). Extraction of Salient Contours in Primary Visual Cortex: A Neural Network Model Based on Physiological Knowledge. Engineering in Medicine and Biology Society. 3: 2242-2245.
- Gaulin, Steven J. C. and Donald H. McBurney (2003). *Evolutionary Psychology*. Prentice Hall. 2003. Chapter 6, p 121-142
- Gerla G. (2005), Fuzzy Logic Programming and fuzzy control, *Studia Logica*, 79 (2005) 231-254.
- Goldberg, D. E. (1989). "Genetic Algorithms in Search Optimization and Machine Learning." Addison Wesley: 41.
- Gray, J. R., Braver, Todd S. and Raichle, Marcus E (2002). "Integration of emotion and cognition in the lateral prefrontal cortex." New York.
- Guyon, I. and A. Elisseeff (2003). An introduction to variable and feature selection. J. Mach. Learn. Res. 3: 1157-1182.
- Gwen Littlewort, Ian Fasel, Joel Chenu, Takayuki Kanda, Hiroshi Ishiguro, and Javier R. Movellan. n. (2004). "Towards Social Robots: Automatic Evaluation of Human-Robot Interaction by Face Detection and Expression Classification." Advances in Neural Information Processing Systems 16.
- Hall, M. A. and L. A. Smith (1998). Practical Feature Subset Selection For Machine Learning. In Proceedings of the 21st Australasian Computer Science Conference. Springer: 181-191.
- Han and Yang (2004). "Screening important design variables for building a usability model: genetic algorithm-based partial least-squares approach." International Journal of Industrial Ergonomics 33(2): 159-171.
- Hart, M. (2005). Autism/Excel study. ASSETS 2005 - The Seventh International ACM SIGACCESS Conference on Computers and Accessibility.
- Hashem S. and Yih Y. (1993). "Optimal Linear Combinations of Neural Networks: An Overview." IEEE International Conference in Neural Networks.

- Hassenzahl, M. (2006). "User experience - a research agenda." Behaviour & Information Technology 25.
- Haykin, S. S. (1999). Neural networks : a comprehensive foundation. Upper Saddle River, N.J., Prentice Hall.
- Hindley, J. R. (2008). Basic Simple Type Theory, Cambridge University Press.
- Hirose M., Shirakawa K. and Nihei K. (1997). "Development of a virtual sand box: an application of virtual environment for psychological treatment. In: Riva G (ed) Virtual reality in neuro-psycho-physiology: cognitive, clinical and methodological issues in assessment and treatment." IOS Press, Amsterdam.
- Hoffman, P., G. Grinstein, et al. (1997). DNA visual and analytic data mining. Visualization '97., Proceedings.
- Hsu, C. W. and C. J. Lin (2002). "A comparison of methods for multiclass support vector machines." Neural Networks, IEEE Transactions on 13(2): 415-425.
- Hubel, D. and T. Wiesel (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. J. Physiol: 106-154.
- Ira Cohen, N. S., Lawrence S. Chen, Ashutosh Garg, and Thomas S. Huang (2003). "Facial Expression Recognition from Video Sequences: Temporal and Static Modeling." Computer Vision and Image Understanding (CVIU) Special Issue on Face recognition 91: 160-187.
- Ji, H. (2004). "Facial Event Classification with Task Oriented Dynamic Bayesian Network." IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2: 870-875.
- Ji, Y. (2005). "Active and dynamic information fusion for facial expression understanding from image sequences." IEEE Transactions on Pattern Analysis and Machine Intelligence 27(5): 699-714.
- John, G. H. and P. Langley (1995). Estimating Continuous Distributions in Bayesian Classifiers. The Eleventh Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Mateo.
- John, G. H. (1995). "Estimating Continuous Distributions in Bayesian Classifiers." Morgan Kaufmann, San Mateo: 338-345.

- Kaliouby, P. M. (2003). "Real Time Facial Expression Recognition in Video using Support Vector Machines." IEEE International Conference on Multimodal Interfaces (ICMI): 258-264.
- Kaliouby, R. (2005). Mind-reading machines: automated inference of complex mental states, University of Cambridge, Computer Laboratory.
- Kaliouby, R. and P. Robinson (2005). "The emotional hearing aid: an assistive tool for children with Asperger syndrome." Universal Access in the Information Society 4(2).
- Kanndel, E. R. and T. Martinetz (2000). Principles of Neural Science. New York, McGraw Hill.
- Kim, D. J. and Z. Bien(2003). Fuzzy neural networks(FNN)-based approach for personalized facial expression recognition with novel feature selection method. IEEE International Conference on Fuzzy Systems.
- Kimmei Cowaill. (2010). "Autism in the Media." WordPress.com
- Kleinman, J. and L. Marciano. (2001). "Advanced theory of mind in high functioning adults with autism." J. Autism Dev Disord 31(1): 29-36.
- Kobayashi, T., Ogawa, Y., Kato, K. and Yamamoto, K. (2004). Learning system of human facial expression for a family robot. IEEE International Conference on Automatic Face and Gesture Recognition.
- Kohavi, R. (1995). "A Study of Cross-validation and Bootstrap for Accuracy estimation and Model Selection." International Joint Conference on Artificial Intelligence (IJCAI) 2: 1137-1143.
- Kosko, Bart and Isaka, Satoru (July 1993). "Fuzzy Logic". *Scientific American* **269** (1): 76–81
- Koyama, A. (2005). "A Review on the Cognitive Neuroscience of Autism."
- Koyama, A. (2009). "A review on the cognitive neuroscience of autism." Activitas Nervosa Su-perior 51(5): 125-139.
- L. E. Baum, T. P., G. Soules, and N. Weiss (1970). "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains." Ann. Math. Statist 41(1): 164-171.

- Lacava, Paul G., Golan Ofer, Baron-Cohen, Simon Myles and Brenda Smith (2007). "Using Assistive Technology to Teach Emotion Recognition to Students With Asperger Syndrome: A Pilot Study." Remedial & Special Education 28(3): 174-181.
- Lanitis, A. T. and Cootes, T.F. (1997). "Automatic interpretation and coding of face images using flexible models." Pattern Analysis and Machine Intelligence, IEEE Transactions on 19(7): 743-756.
- Lavie, T. (2004). "Assessing dimensions of perceived visual aesthetics of web sites." International Journal of Human-Computer Studies 60: 269-298.
- Ledoux, J. (1994). "The Emotional Brain: The Mysterious Underpinnings of Emotional Life." New York: Rockefeller Center.
- Littlewort, G., Bartlett, M.S., Fasel, I., Susskind, J. and Movellan, J. (2006). Dynamics of facial expression. J. Image & Vision Computing, 24(6): 615-625.
- Liu, P. and H. Li (2004). Fuzzy Neural Network Theory and Application, World Scientific Publishing Company.
- Liu, S. M. (2004). "Local Facial Asymmetry for Expression Classification." IEEE Conference on Computer Vision and Pattern Recognition (CVPR)2: 889-894.
- Luetin, B. F. a. J. (2000). "Recognition of Asymmetric Facial Action Unit Activities and Intensities." International Conference on Pattern Recognition (ICPR)1: 1100-1103.
- Ma, L. and Khorasani, K. (2004). "Facial expression recognition using constructive feedforward neural networks." IEEE Transactions on Systems, Man and Cybernetics, Part B. 34(3): 1588-1595.
- MacDonald, I. L. and Zucchini, W. (1997). "Hidden Markov and Other Models for Discrete-valued Time Series." Monographs on Statistics and Applied Probability 70.
- Mamdani, E. H. (1974), Application of fuzzy algorithms for the control of a simple dynamic plant. In *Proc IEEE*, 121-159.
- Mase, K. and A. Pentland (1991). "Recognition of facial expression from optical flow." IEICE Trans.74(10): 3474-3483.

- Matsuno, K. and C.W. Iee (1994). Recognition of Human Facial Expressions Without Feature Extraction. ECCV.
- Minerva T. (2001). "Building ARMA Models with Genetic Algorithms." EvoWorksho335-343.
- Minerva T. (2002). "Evolutionary Approaches for Statistical Modelling. In: Fogel DB." IEEE Press, Piscataway NJ (2002) Proc.of the Fourth Congress on Evolutionary Computation 2, : 2023-2038.
- Mitchell, M. (1996). "An Introduction to Genetic Algorithms." MIT Press, Cambridge, Massachusetts.
- Montse Pard` and Jos`e Luis Landabaso (2002). "Emotion Recognition based on MPEG4 Facial Animation Parameters." IEEE Internatoinal Conference on Acoustics, Speech and Signal Proccsing 4: 3624-3627.
- Moore D., M. P. and Thorpe J. (2000). "Computer-aided learning for people with autism- a framework for research and development." Innov. Educ. Training Int 37(3): 218-228.
- Naiwala P. Chandrasiri and Hiroshi Harashima (2001). "Real Time Facial Expression Recognition with Applications to Facial Animation in MPEG-4." IEICE Trans. INF. and SYST 8: 1007-1017.
- Novák, V. (2005). "Are fuzzy sets a reasonable tool for modeling vague phenomena?". Fuzzy Sets and Systems 156: 341—348.
- Ning, Z. and S. Andrzej (1999). New Directions in Rough Sets, Data Mining, and Granular Soft Computing. 7th International Workshop, RSFDGrC'99, Yamaguchi, Japan.
- O'Connor, J. J. R. and Edmund F. (1987). "Student's t-test." MacTutor History of Mathematics archive, University of St Andrews.
- Otsuka, T. and J. Ohya (1998). Spotting segments displaying facial expression from image sequences using HMM. IEEE Proceedings ofthe Second International Conference on Automatic Face and Gesture Recognition, Japan.
- Pankin, M.D. (2007). "MARKOV CHAIN MODELS: THEORETICAL BACKGROUND." <http://www.pankin.com/markov/theory.htm>.

- Pantic, M. and L. J. M. Rothkrantz (2000). "Automatic analysis of facial expressions: the state of the art." Pattern Analysis and Machine Intelligence, IEEE Transactions on 22(12): 1424-1445.
- Pardas, M.B. and Landabaso, J.L. (2002). Emotion recognition based on mpeg4 facial animation parameters. IEEE International Conference on Acoustics, Speech, and Signal Processing.
- Parrott, W. (2001), Emotions in Social Psychology, Psychology Press, Philadelphia
- Patricia Howlin and Julie Hadwin (1999). "Teaching Children with Autism to Mind-read: A Practical Guide for Teachers and Parents." New York: John Wiley and Sons.
- Pavlovic, B. K. (2005). "Real-Time Vision for Human-Computer Interaction, chapter Real-Time Algorithms: From Signal Processing to Computer Vision." Springer-Verlag.
- Quinlan and J. R. (1993). C4.5: Program for Machine Learning. Morgan Kaufmann.
- Rabiner, L. R. (1989). "A tutorial on Hidden Markov Models and selected applications in speech recognition." Proceedings of the IEEE 77(2): 257-286.
- Raju, T. (2005). "William Sealy Gosset and William A. Silverman: two "students" of science"." Pediatrics 116(3): 732-735.
- Richard Durbin, Anders Krogh and Graeme Mitchison (1999). "Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids." Cambridge University Press. ISBN 0-521-62971-3.
- Rosenblum, M., Yacoob, Y. and Davis, L. S. (1996). "Human Expression Recognition from Motion Using a Radial Basis Function Network Architecture." IEEE Transactions on Neural Networks 7(5): 1121-1138.
- Rosenfeld, A. and A. C. Kak (1982). Digital picture processing. Academic Press. 1.
- S. Baron-Cohen (1997). "Another advanced test of theory of mind : evidence from very high functioning adults with autism or asperger syndrome." J Child Psychol Psychiatry 38(7): 813-822.
- S. Baron-Cohen (1999). "Social intelligence in the normal and autistic brain: an fMRI study." Eur J. Neurosci 11(6): 1891-1898.

- Saxena, A., Anand, A. and Mukerjee, A. (2004). Robust Facial Expression Recognition using Spatially Localized Geometric Model. International Conference on Systemics, Cybernetics and Informatics.
- Schiele, B. (1997). Object Recognition using Multidimensional Receptive Field Histograms. PhD thesis, I.N.P.Grenoble.
- Sebe, N., Lew, M.S., Cohen, I., Yafei Sun, Gevers, T. and Huang, T.S. (2004). Authentic facial expression analysis. IEEE International Conference on Automatic Face and Gesture Recognition.
- Shah-Hosseini, H. (2009). "The intelligent water drops algorithm: a nature-inspired swarm-based optimization algorithm." International Journal of Bio-Inspired Computation (IJBIC)1.
- Shieh, M. D. and C. C. Yang (2008). "Multiclass SVM-RFE for product form feature selection." Expert Systems with Applications 35(1-2): 531-541.
- Youngsuk Shin, Shin Sooyong Lee, Chansup Chung and Yillbyung Lee (2000). Facial expression recognition based on two-dimensional structure of emotion. International Conference on Signal Processing Proceedings.
- Shinghal, R. (1979). "Experiments in text recognition with the modified Viterbi algorithm." IEEE Transactions on Pattern Analysis and Machine IntelligencePAMI-1: 184-193.
- Simon Baron-Cohen, A. R., Masato Fukushima, Davina French, Julie Hadwin, Pippa Cross, Catherine Bryant, and Maria Sotillo (1996). "Reading the Mind in the Face: A Cross-cultural and Developmental Study." Visual Cognition 3: 39-59.
- Simon Baron-Cohen, S. W., John Lawson, Rick Griffin, and Jacqueline Hill. (2002). "Handbook of Childhood Cognitive Development, chapter The Exact Mind: Empathising and Systemising in Autism Spectrum Conditions." Oxford: Blackwell Publishers: 491-508.
- Stanfield, A. C. and. M. McIntosh. (2008). "Towards a neuroanatomy of autism: A systematic review and meta-analysis of structural magnetic resonance imaging studies." European Psychiatry23(4): 289-299.
- Strickland, D. (1996). "A virtual reality application with autistic children presence." Teleoper Virtual Environ5(3): 319-329.

- Sullivan, H. (2000). "A Componential View of Theory of Mind: Evidence from Williams Syndrome." Cognition76(59-89).
- Syswerda, G. (1989). "Uniform crossover in genetic algorithms." Third International Conference on Genetic Algorithms.
- T. Xiang, M.K.H. Leung and S.Y. Cho (2007). "Expression recognition using fuzzy spatio-temporal modeling." Pattern Recognition41(1): 204-216.
- Takeo Kanade and Ying-Li Tian (2000). "Comprehensive Database for Facial Expression Analysis." International Conference on Automatic Face and Gesture Recognition: 46-53.
- Takeo Kanade and Ying-Li Tian.. (2000). "Comprehensive Database for Facial Expression Analysis." International Conference on Automatic Face and Gesture Recognition: 46-53.
- Turk, M. (2005). "Real-Time Vision for Human-Computer Interaction." A Historical Overview.
- Viola, P. and M. Jones (2004). Robust real-time face detection. J. Computer Vision. 57: 137-154.
- Vukadinovic, D. and M. Pantic (2005). Fully automatic facial feature point detection using Gabor feature based boosted classifiers. IEEE Int'l Conf. Systems, Man and Cybernetics: 1692-1698.
- Welch, L. (2003). "Hidden Markov Models and the Baum-Welch Algorithm." IEEE Information Theory Society Newsletter53(4).
- Werry, I., Ogden B. and Harwin W. (2001). "Can social interaction skills be taught by a social agent? the role of a robotic mediator in autism therapy. Proceedings CT2001, the fourth international conference on cognitive technology." Lecture Notes in Computer Science, sub series Lecture Notes in Artificial Intelligence, Springer, Berlin Heidelberg New York.
- Whitehill, J and C. Omlin (2006). "Haar Features for FACS AU Recognition." Int'l Conf Face and Gesture Recognition.
- Whitehill, J. and C. Omlin (2006). Haar Features for FACS AU Recognition. Int'l Conf. Face and Gesture Recognition.

- Wolpert, D. H. and Macready, W.G. (1995). "No Free Lunch Theorems for Optimisation." Santa Fe Institute.
- Wong, J.J. and S.Y. Cho (2006). A Brain-Inspired Framework for Emotion Recognition. Neural Information Processing—Letters and Reviews. 10: 169-179.
- Wong, J.J. and S.Y. Cho (2006). Facial emotion recognition by adaptive processing of tree structures. Proceedings of the 2006 ACM symposium on Applied computing, Dijon, France, ACM Press.
- Wu, Y., Liu, H. and Zha, H. (2005). Modeling facial expression space for recognition. IEEE/RSJ International Conference on Intelligent Robots and Systems.
- Xuedong Huang and Y. Ariki (1990). "Hidden Markov Models for Speech Recognition." Edinburgh University Press.
- Yacoob, Y. and L. S. Davis (1996). "Recognizing Human Facial Expressions from Long Image Sequences using Optical Flow." IEEE Transactions on Pattern Analysis and Machine Intelligence 18(6): 636-642.
- Ye, Jingfu, Yongzhao Zhan and Shunlin Song (2004). Facial expression features extraction based on Gabor wavelet transformation. IEEE International Conference on System, Man and Cybernetics.
- Yeasin, M. and B. Bulot (2005). Comparison of linear and non-linear data projection techniques in recognizing universal facial expressions. IJCNN.
- Ying-Li Tian, Arun Hampapur, Sharat Pankanti, Andrew W. Senior and Ruud M. Bolle (2003). "RealWorld Real-Time Automatic Recognition of Facial Expressions." IEEE workshop on Performance Evaluation of Tracking and Surveillance.
- Ying-Li Tian and Jeffrey Cohn. (2001). "Recognizing Action Units for Facial Expression Analysis." IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI) 23: 97-115.
- Zadeh, L. A. (1965). "Fuzzy Sets." Information and Control 8(3): 338-353.
- Zadeh, L. A. (1996). "Fuzzy Sets, Fuzzy Logic, Fuzzy Systems." World Scientific Press.
- Zeng, Z., Fu Y., Roisman G.I., Wen Z., Hu Y. and Huang, T.S. (2006). One-class classification for spontaneous facial expression analysis. International Conference on Automatic Face and Gesture Recognition.

Zhang. J (2007). "Clustering-Based Adaptive Crossover and Mutation Probabilities for Genetic Algorithms." IEEE Transactions on Evolutionary Computation 11(3): 326-335.