REAL-TIME FACIAL EXPRESSION RECOGNITION SYSTEM (FERS)

THELEHAESWARY APPALASAMY

A thesis submitted in partially fulfilment of the requirements for the award of degree of Bachelor of Computer Science (Software Engineering)

Faculty of Computer System & Software Engineering
Universiti Malaysia Pahang (UMP)

May 2011

ABSTRAK

Real-Time Facial Expression Recognition System (FERS) dibangunkan untuk mengenali ekspresi wajah. Penemuan mendakwa bahawa manusia memaparkan emosi mereka yang paling ekspresif melalui ekspresi wajah dan gerak tubuh. Manusia lebih cenderung untuk mempertimbangkan komputer untuk menjadi seperti manusia di mana komputer memahami dan memaparkan perilaku komunikatif yang sesuai. Jadi, interaksi antara manusia dan komputer akan menjadi lebih neutral jika komputer mampu memahami perilaku manusia dan mengenali keadaan afektif mereka. Justeru itu, projek ini dibangunkan untuk membina satu prototaip yang dapat mengenali ekspresi wajah manusia. Evolusi metodologi dilaksanakan dalam perancangan sistem dengan menggunakan teknik pemprosesan imej termasuk teknik dapatan imej, penambahbaikan imej (atau dikenali sebagai tahap pra-pemprosesan) dan pengeluaran cirri-ciri imej. Sebanyak 154 sampel wajah manusia yang berumur di antara 18 hingga 26 tahun telah diuji. Pada peringkat awal, berlaku beberapa peringkat pra-pemprosesan untuk penambahbaikan imej. Kemudiannya, muka akan dikesan. Bahagian muka (iaitu mulut) akan ditentukan, di mana, ciri-ciri muka akan didapatkan. Akhirnya, wajah akan diklasifikasikan ke dalam salah satu daripada tiga kelas yang berbeza menggunakan kaedah K-Mean berdasarkan pada ciri-ciri yang diekstraksi. Sistem ini Berjaya mencapai objektif yang disasarkan berdasarkan keputusan ujian yang menunjukkan 94.80% daripada 154 wajah menunjukkan ekpresi yang tepat.

ABSTRACT

Real-Time Facial Expression Recognition System (FERS) is developed to recognize facial expressions. Findings claiming that humans display their emotions most expressively through face expressions and body gestures. Humans are more likely to consider computers to be human-like when those computers understand and display appropriate nonverbal communicative behaviour. So, the interaction between humans and computers will be more natural if computers are able to understand the nonverbal behaviour of their human counterparts and recognize their affective state. Therefore, this project is carried out to build a prototype to recognize facial expressions. Evolutionary methodology was implemented in this system design by using several image processing techniques include image acquisition, image enhancement (or known as pre-processing stages) and feature extraction. Hundred and fifty-four sample data of human faces in the range of 18 to 26 years old is tested. The system first applies some pre-processing stages to enhance the input image and reduce the noise. The face boundary will then be detected. The region of interest (i.e. mouth and eyes) will be determined, from which, features will be extracted. Finally, the face will be classified into one of three different classes using the K-means method based on the features extracted. The method was applied and tested on a dataset of 154 images of faces and the success rate obtained was 94.80%.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	DECLARATION	ii
	SUPERVISOR'S DECLARATION	iii
	DEDICATION	iv
	ACKNOWLEDGEMENT	v
	ABSTRAK	vi
	ABSTRACT	vii
	TABLE OF CONTENTS	viii
	LIST OF TABLES	xi
	LIST OF FIGURES	xii
	LIST OF APPENDICES	xiv
1	**INTRODUCTION**	1
	1.1 Introduction	1
	1.2 Problem Statements	2
	1.3 Objectives	2
	1.4 Scopes	3
	1.5 Thesis Organization	3
2	**LITERATURE REVIEW**	4
	2.1 Facial Expressions	4
	2.2 Case Study on Existing System	5
	2.2.1 Vision-based Facial Expression Recognition	5
	2.2.2 Systems that Recognize Prototypic Facial Expression	6
	2.2.3 System that Recognize Facial Action	7

		2.2.3.1 Previous Work on Recognizing Facial Action	7
	2.3	Techniques Used in Facial Expression Recognition System (FERS)	14
		2.3.1 Components of an Image Processing System	17
		2.3.2 Image Acquisition	17
		2.3.3 Image Enhancement	18
		2.3.4 Feature Extraction	18
		2.3.4.1 Face Boundary Detection	19
		2.3.4.2 Segmentation	25
	2.4	Software Approach	28
3	**METHODOLOGY**		**29**
	3.1	Introduction	29
	3.2	Software Methodology	29
		3.2.1 Requirements Planning	32
		3.2.1.1 Image Acquisition	33
		3.2.2 User Design	33
		3.2.2.1 Image Enhancement	34
		3.2.2.2 Facial Feature Extraction	34
		3.2.2.3 K-Means Clustering	39
		3.2.3 Construction	41
		3.2.3.1 Testing	41
		3.2.3.2 Black-Box Testing	42
		3.2.4 Cutover	42
	3.3	Software and Hardware Requirements	42
		3.3.1 Software Requirements	43
		3.3.2 Hardware Requirements	44
4	**IMPLEMENTATION**		**45**
	4.1	Image Acquisition	45
	4.2	Image Enhancement	47
		4.2.1 Noise Removal	47

	4.3	Feature Extraction	47
		4.3.1 Face Boundary Detection	48
		4.3.2 Segmentation	50
		4.3.3 Feature Extraction	50
	4.4	Facial Expression Recognition	51
5		**RESULTS AND DISCUSSION**	**52**
	5.1	Introduction	52
	5.2	Image Acquisition	52
	5.3	Image Enhancement	53
		5.3.1 Noise Removal Filtering	54
	5.4	Feature Extraction	55
		5.4.1 Face Boundary Detection	55
		5.4.2 Segmentation	56
	5.5	Facial Expression Detection	56
	5.6	Test Results	57
	5.7	Advantages and Disadvantages	58
		5.7.1 Advantages	58
		5.7.2 Disadvantages	58
	5.8	Constraints	59
	5.9	Assumptions and Further Research	59
		5.9.1 Assumptions	60
		5.9.2 Further Research	60
6		**CONCLUSION**	**61**
	6.1	Conclusion of the Project	61
		REFERENCES	**62**
		APPENDICES	**68**

LIST OF TABLES

TABLE NO.	TITLE	PAGE
2.1	FACS Action Units (AU)	9
3.1	List of software requirements used in the system	43
3.2	List of hardware requirements used in the system	44
5.1	Results from the testing process	57
5.2	Range of density of pixels and ratio of feature	57

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
2.1	The six principle emotions: anger, fear, disgust, surprise, happiness and sad	5
2.2	Sample AU's and the appearance changes they describe	8
2.3	Temporal segmentation of facial gestures	11
2.4	a) Original lip image b) Magnitude of gradient c) Canny edge detector with T=0.2	13
2.5	Points used in training AAM	13
2.6	Viewing AAM generation using training data	14
2.7	Image enhancements and restoration	15
2.8	Adjusting the image histogram to improve image contrast	15
2.9	Overview of Facial Expression Recognition System	17
2.10	Partial rotation invariant versions (PRIVs) of a face	20
2.11	Region growing method	21
2.12	Using region growing method	21
2.13	Boundary tracking method	22
2.14	Using LSM	23
2.15	Skin detection method applied on the original image	23
2.16	SMQT method applied on an image	24
2.17	Comparison between 6 edge detection methods	27
3.1	Martins Life Cycle	31
3.2	Illustration of system modules	32
3.3	Face image	33
3.4	Noise removal filtering applied on the grayscale image	34
3.5	SMQT method applied on an image	36

3.6	Comparison between 6 edge detection methods	37
3.7	Ratio range for smile, sad and neutral expressions	38
3.8	Sample of the dataset values	39
3.9	Basic concepts of K-means clustering	40
3.10	The K-means clustering for 154 data sets	41
3.11	Happy, Neutral and Sad icons	42
4.1	Setup for image acquisition	45
4.2	Source code for image acquisition	46
4.3	Source code for image enhancement	47
4.4	Source code for face detection	49
4.5	Source code for segmentation	50
4.6	Source code for feature extraction	50
4.7	Source code for facial expression recognition	51
5.1	Result of captured human face	53
5.2	RGB image converted to grayscale	54
5.3	Result after image enhancement	54
5.4	Face boundary detected image	55
5.5	Results of cropped image and the detected edges	56
5.6	FERS user interface with result	56

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	Gantt Chart	68
B	Testing Results	70
C	User Manual	113

CHAPTER 1

INTRODUCTION

1.1 Introduction

Emotions can be communicated by various modalities, including speech and language, gesture and head movement, body movement and posture, as well as face expression. According to Mehrabian [1], in human-human interaction (HHI) spoken words only account for 7% of what a listener comprehends; the remaining 93% consist of the speaker's nonverbal communicative behavior (i.e. body language and intonation). There exist other findings claiming that humans display their emotions most expressively through face expressions and body gestures [2, 3]. Moreover, research shows that humans are more likely to consider computers to be human-like when those computers understand and display appropriate nonverbal communicative behavior [4]. Therefore, the interaction between humans and computers will be more natural if computers are able to understand the nonverbal behavior of their human counterparts and recognize their affective state.

Facial expression plays an important role in smooth communication among individuals. Facial Expression Recognition (FER) is a technology used for identification of human behavior based on facial expressions. The extraction and recognition of facial expressions has been the topic of choice of many researchers in the recent years. FER is mainly applied to enable smooth interaction between computers and their users as an attempt to give computers the ability to recognize the emotion of the user and offer advices in response to the mood of the user.

1.2 Problem Statement

It has often been said that the eyes are the "window to the soul." This statement may be carried to a logical assumption that not only the eyes but the entire face may reflect the "hidden" emotions of the individual. Darwin's research on facial expressions has had a major impact on the field in many areas; foremost, his belief that the primary emotions conveyed by the face are universal [5]. Darwin placed considerable emphasis on the analysis of the action of different muscle groups in assessing expression.

The research on the statement of Darwin was done by Ekman and Friesen. They hypothesized that the universals of facial expression are to be found in the relationship between distinctive patterns of the facial muscles and particular emotions (happiness, sadness, anger, fear, surprise, disgust and interest) [6]. They suggested that cultural differences would be seen in some of the stimuli, which through learning become established as elicitors of particular emotions, in the rules for controlling facial behavior in particular social settings, and in many of the consequences of emotional arousal [7].

Many factors impinge upon the ability of an individual to identify emotional expression. Social factors, such as deception, and display rules, affect one's perception of another's emotional state. Therefore, there is a need to develop Face Expression Recognition System (FERS).

1.3 Objective

There are two objectives to be achieved in this project:

i. To develop a prototype to determine facial expressions corresponding to the basic emotions.

ii. To identify the suitable image processing techniques for facial expression recognition.

1.4 Scope

There are some restrictions in this project:

i. Limited to the detection of three main basic facial expressions (happy, sad and neutral).

ii. 154 samples of Universiti Malaysia Pahang (UMP) students.

iii. The range of age of the samples is 18-26 years old.

iv. Based on physical expression

v. The lighting condition is always consistent with 220-240V 50Hz fluorescent lamp.

vi. Images taken had length range from 50cm- 200cm between the camera and the samples.

vii. Only one sample is being captured during image acquisition phase for each testing conducted.

viii. This system is not a fully automated system.

1.5 Thesis Organization

This thesis consists of 6 chapters ranging from Chapter 1 until Chapter 6. Chapter 1 gives an overview of the study conducted. It also supply with the problem statement, objective and the scope of the study. Meanwhile, Chapter 2 reviews the previous research works that was conducted by other researches. All the relevant technical paper, journals, and books taken from those researches will be discussed in detail. Chapter 3 reveals the techniques and the algorithms that will be used in performing this study. It will discuss about the process flow in detail of this research. Details of the implementation of the study will be discussed in Chapter 4. Results of the testing are to be expounding in Chapter 5. Lastly, Chapter 6 concludes the entire thesis.

CHAPTER 2

LITERATURE REVIEW

2.1 Facial Expressions

This project concerns about recognition system through identifying the facial expression of human being. It has often been said that the eyes are the "window to the soul". This statement may be carried to a logical assumption that not only the eyes but the entire face may reflect the "hidden" emotions of the individual. The human face is the most complex and versatile of all species. For humans, the face is a rich and versatile instrument serving many different functions. It serves as a window to display one's own motivational state. This makes one's behavior more predictable and understandable to others and improves communication. A quick facial display can reveal the speaker's attitude about the information being conveyed [5].

As Facial expressions can indicate emotion and pain, regulate social behavior, and reveal brain function, Research in psychology has indicated that at least six emotions are universally associated with distinct facial expressions. The six principal emotions are: happiness, sadness, surprise, fear, anger, and disgust. Several other emotions and many combinations of emotions have been studied but remain unconfirmed as universally distinguishable [8].

Figure 2.1 The six principal emotions: anger, fear, disgust, surprise, happiness and sad

2.2 Case Study on Existing System

In the field of Facial Expressions, was pioneered by Darwin's work and has been extensively studied in psychology during the last twenty years [9]. The studies will be discussed in this section.

2.2.1 Vision-based Facial Expression Recognition

Within the past decade, analysis of human facial expression has attracted interest in machine vision and artificial intelligence areas to build systems that understand and use this non-verbal form of human communication. Most of the systems that automatically analyze the facial expressions can be broadly classified into two categories:

i. systems that recognize prototypic facial expressions corresponding to basic emotions (happy, angry etc)

ii. systems that recognize facial actions (eyebrow raise, frown etc).

There has been a significant amount of research on creating systems that recognize a small set of prototypic emotional expressions, i.e., joy, surprise, anger, sadness, fear, and disgust from static images or image sequences. This focus on emotion-specified expressions follows from the work of Ekman [10, 11] who proposed that basic emotions have corresponding prototypic facial expressions.

2.2.2 Systems that Recognize Prototypic Facial Expressions

Automatic facial expression analysis is done in two different ways: from static images or from video frames. The studies based on facial expression recognition from static images are performed by presenting subjects with photographs of facial expressions and then analyzing the relationship between components of the expressions and judgments made by the observers. These judgment studies rely on static representations of facial expressions with two facial images: a neutral face and an expressive face. Facial expression recognition from image sequences is based on categorizing 5-7 classes of prototypic facial expressions by tracking facial features and measuring the amount of facial movement.

There are various approaches that have been explored. Some of those include analysis of facial motion [12,13,14] measurements of the shapes and facial features and their spatial arrangements [15] holistic spatial pattern analysis using techniques based on principal components analysis [16,17] and methods for relating face images to physical models of the facial skin and musculature [12,14]. All these methods are similar in that they first extract some features from the images or video, then these features are used as inputs into a classification system, and the outcome is one of the pre selected emotion categories. They differ mainly in the features extracted and in the classifiers used to distinguish between the different emotions.

2.2.3 System that Recognize Facial Action

The evidence for six universal facial expressions does not imply that these emotion categories are sufficient to describe all facial expressions [18]. Although prototypic expressions, like happy, surprise and fear, are natural, they occur infrequently in everyday life and provide an incomplete description of facial expression. Emotion is communicated by changes in one or two discrete facial features, such as tightening the lips in anger or obliquely lowering the lip corners in sadness [14]. Further, there are emotions like confusion, boredom and frustration for which any prototypic expression might not exist. To capture the subtlety of human emotion and paralinguistic communication, automated recognition of fine-grained changes in facial expression is needed. Hence, vision-based systems that recognize facial actions were introduced. Generally, the approaches that attempt to recognize action units (AUs) are motivated by Paul Ekman's Facial Action Coding System (FACS) [15].

2.2.3.1 Previous Work on Recognizing Facial Action

Some of the previous work to directly recognize action units has used optical flow across the entire face or facial feature measurement. Patterns of optical flow corresponded to several AUs, but did not attempt to recognize them [19, 20]. Methods to analyze expressions into elementary movements using an animation style coding system inspired by FACS [20, 21] proposed. Then, local principal component analysis was superior to full face Eigen faces for expression recognition founded [13].

The Facial Action Coding System (FACS) is the leading method for measuring facial movement in behavioural science. FACS has been successfully applied, but not limited to, identifying the differences between simulated and genuine pain, differences between when people are telling the truth versus lying, and differences between suicidal and non-suicidal patients [22]. In 1978, Ekman and

Friesen designed the facial action coding system (FACS) for characterizing facial expressions by AUs. This system is a human observed system developed to explain the subtle changes of facial expressions. This system describes how various parts of the face are positioned for various emotions. These parts include the brows, the forehead, the eyes, the eyelids, and the root of the nose, the mouth, the cheeks, and the chin. There are totally 44 AUs. Among them, 30 are related to facial muscle contraction including 12 for upper faces and 18 for lower faces [23]. The anatomic basis of the remaining 14 is unspecified. These 14 are referred to in FACS as miscellaneous actions. A FACS coder "dissects" an expression, decomposing it into specific AUs that produced the motion.

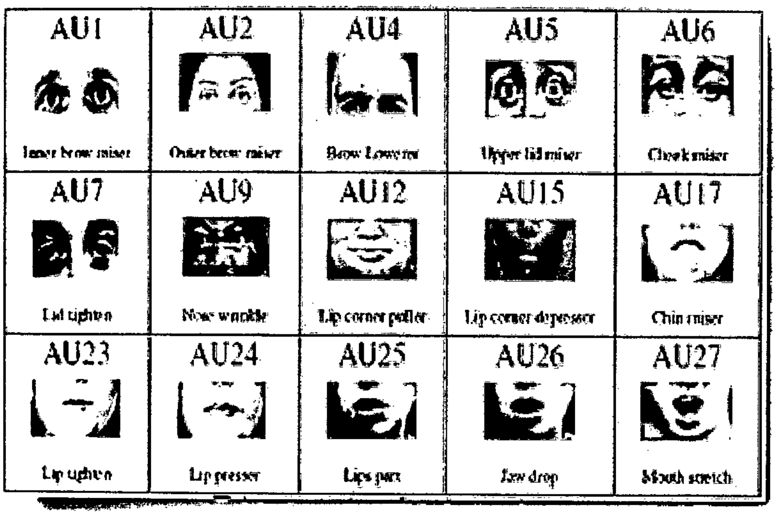

Figure 2.2 Sample AUs and the appearance changes they describe

Table 2.1 FACS Action Units (AU)

Action Unit	Description
Brows	
AU 1+2	Inner and outer portions of the brows are raised
AU 1+4	Medial portion of the eyebrows is raised and pulled together
AU 4	Brows are lowered and drawn Together
Eyes	
AU 5	Upper eyelids are raised, which produces a widening of the eyes
AU 6	The lower-eye and infra-orbital furrows are raised and deepened and the eye opening is narrowed
AU 7	Lower eyelids are tightened, which narrows the eye opening
Mouth	
AU 27	Mouth is stretched open and mandible extended
AU 26	Lips are relaxed and parted; mandible Lowered
AU 25	Lips are relaxed and parted; mandible not lowered
AU 12	Lip corners are pulled up and backward
AU 12+25	AU 12 with mouth opening
AU 20+25	Lips are parted, pulled back laterally, and may be slightly raised or pulled down.
AU 15+17	Lip corners are pulled down and stretched laterally (AU 15), and chin boss is raised, which pushes up the lower lip (AU 17).
AU 17+23+24	AU 17 and the lips are tightened, narrowed, and pressed together (AU 23+24)
AU 9+17±25	The infra-orbital triangle and centre of the upper lip are pulled upwards (AU 9) with AU 17. In 25% of cases, AU 9+17 occurred with AU 25.

A system to recognize facial expressions by identifying Facial Animation Parameter Units (FAPUs) defined in MPEG-4 standard by feature tracking of Facial Definition Parameter (FDP) points, also defined in MPEG-4 framework described.

The CMU/Pittsburgh and UCSD groups are among the most important research groups that have focused on automatic FACS recognition as a tool for behavioural research. From the CMU/Pittsburgh group, [18] developed an automatic AU (Action units) analysis system using facial features to recognize 16 action units and any combination of those. The shape of facial features like eyes, eyebrow, mouth and cheeks are described by multistate templates. The parameters of these multistate templates are used by a Neural Network based classifier to recognize the action units. The degree of manual pre-processing is reduced by using automatic face detection. However, the system requires that the templates be initialized manually in the first frame of the sequence, which prevents it from being fully automatic. The system has achieved average recognition rates of 96.4 percent for upper face AUs and 96.7 percent for lower face AUs [17] and [24] from the UCSD group compared holistic spatial analysis, explicit measurement of features (local feature analysis) such as wrinkles, and estimation of motion flow fields and combined them in a hybrid system which classified 6 upper facial actions but no AUs occurring in combinations. The system achieved 91 percent accuracy. However, only results on manually pre-processed image sequences were reported. The facial gesture recognition system [25] analyzes subtle changes in facial expressions based on profile contour fiducial points in a profile-view video [26]. For tracking the profile face, a profile contour and 10 profile contour fiducial points are extracted. 20 individual Aus occurring alone or in a combination are recognized by a rule-based method and the recognition rate of 85 percent is achieved. In addition, Pantic also proposed a self adaptive facial-expression analyzer that classifies detected facial muscle activity into multiple, quantified, user-defined interpretation categories.

A fully automatic framework that requires no manual intervention to analyze facial activity is focused on recognizing upper action units [27]. The system detects the pupils using an infrared sensitive camera equipped with infrared LEDs. For each frame, the pupil positions are used to localize and normalize eye and eyebrow regions, which are analyzed using PCA (Principal Component Analysis) to recover parameters that relate to the shape of the facial features. These parameters are used as input to classifiers based on Support Vector Machines to recognize upper facial action units and all their possible combinations. The system achieved a recognition accuracy of 62.5 percent for all possible AU combinations. However, the system breaks when the subjects are wearing glasses. Since the system uses infrared LEDs, it can be confused by the presence of strong direct sunlight. It also needs to be extended to recognize lower facial action units.

Temporal segmentation of facial gestures in spontaneous facial behaviour recorded in real-world settings is an important, unsolved, and relatively unexplored problem in facial image analysis. These include non-frontal pose, moderate to large out-of-plane head motion, large variability in the temporal scale of facial gestures, and the exponential nature of possible facial action combinations. Evaluated this method in facial behaviour recorded during face-to face interactions. The method achieved moderate convergent validity with manual FACS (Facial Action Coding System) annotation. Further, when used to pre-process video for manual FACS annotation, the method significantly improves productivity, thus addressing the need for ground-truth data for facial image analysis [28].

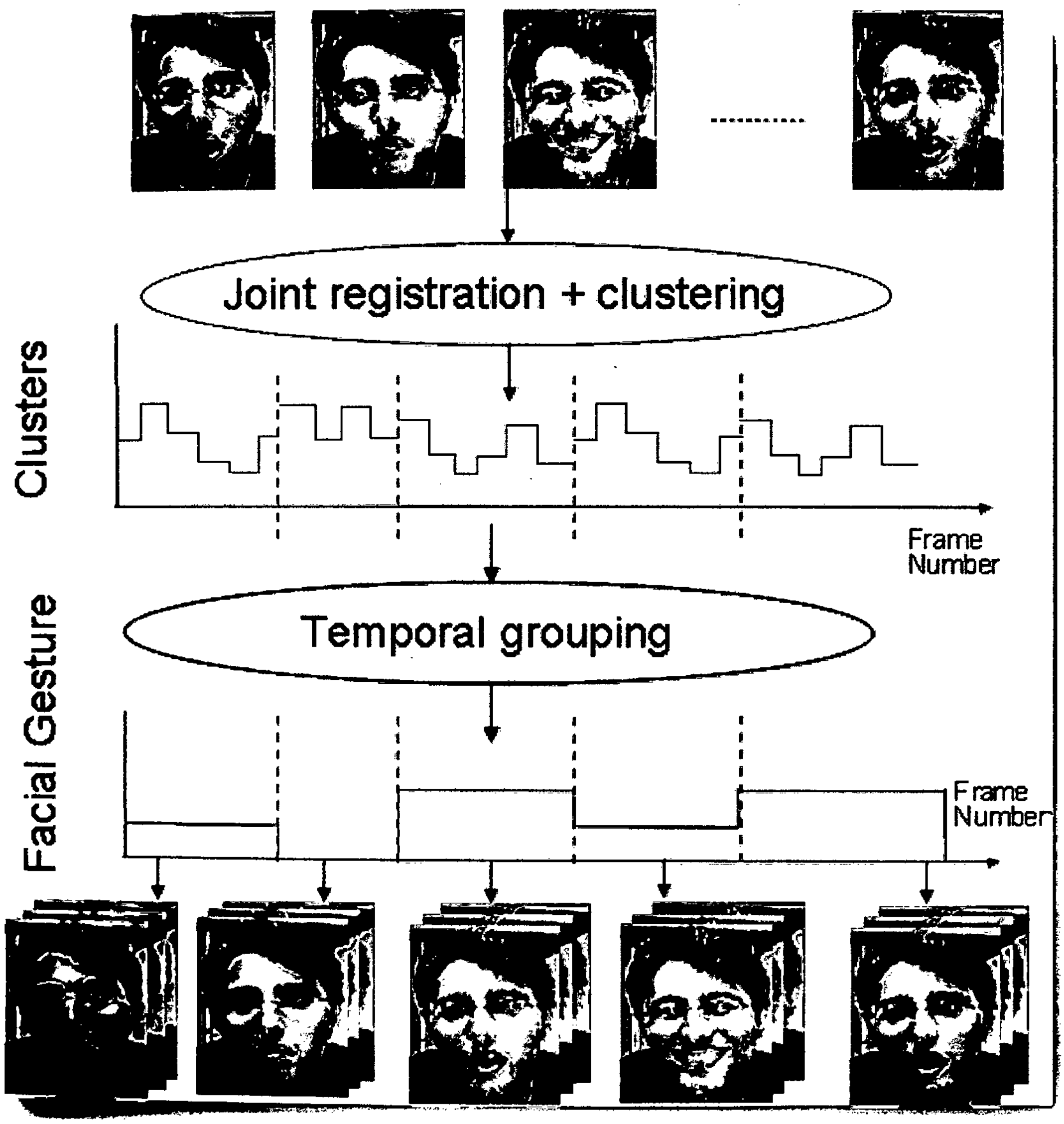

Figure 2.3 Temporal segmentation of facial gestures.

In addition, an emotion recognition system for assessing human emotional behaviour from communication by a speaker includes a processing system configured to receive signals representative of the verbal and/or non-verbal communication. The processing system derives signal features from the received signals. The processing system is further configured to implement at least one intermediate mapping between the signal features and one or more elements of an emotional ontology in order to perform an emotion recognition decision. The emotional ontology provides a gradient representation of the human emotional behaviour [29].

The most of the human emotions are communicated by changes in one or two of discrete facial features. These changes are coded as Action Units (AUs). In this paper, the authors develop a lip shape extraction and lip motion tracking system both in static and dynamic facial images, based on a novel two step active contours model. A knowledge based system is used for estimating initial position of mouth. The first step active contour locks onto stronger upper lip edges by using both high thresholds Canny edge detector and balloon energy for contour deflation. An oval shaped initial active contour is considered inside the estimated mouth region. Four energy terms are used to control motion of control points. At the first step active contour locks onto upper lip. Then using lower threshold image gradient snake inflates and locks onto lower lip edges. Then using lower threshold image gradient as well as balloon energy for inflation, snake inflates and locks onto weaker lower lip edges. Extracted lip feature points are used to extract some geometric features to form a feature vector which is used to classify lip images into AUs, using Probabilistic Neural Networks (PNN). Experimental results show robust edge detection and reasonable classification where an average AUs recognition rate is 85.98% in image sequences and 77.44% in static images [30].

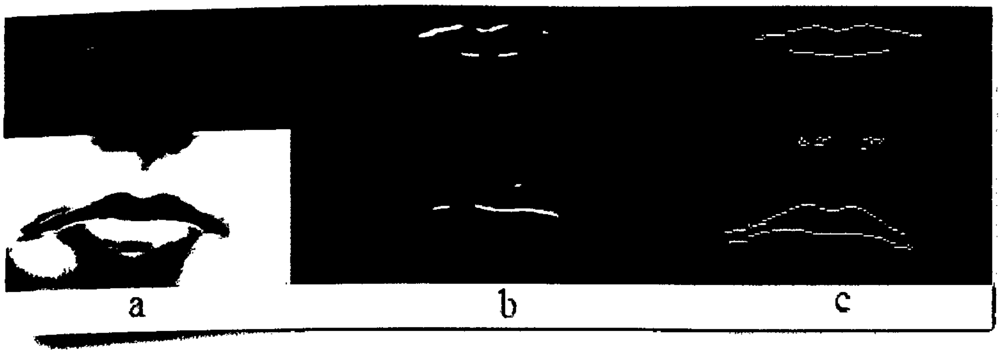

Figure 2.4 a) Original lip image b) Magnitude of gradient c) Canny edge detector with T=0.2

The study of the facial muscle movements classified by FACS in creating certain expressions was used to inform the choice of landmarks for active appearance model (AAM) shape parameter. AAM focused on developing a framework for emotion recognition based on facial expressions. Facial images representing the six universal emotions mentioned previously as well as a neutral expression were labelled in a manner to capture expressions. An AAM was built using training data and tested on a separate dataset. Test face images were then classified as one of the six emotion-based expressions or a neutral expression using the AAM parameters as classification features. The technique achieved a high level of performance in classifying these different facial expressions based on still images.

Figure 2.5 Points used in training AAM

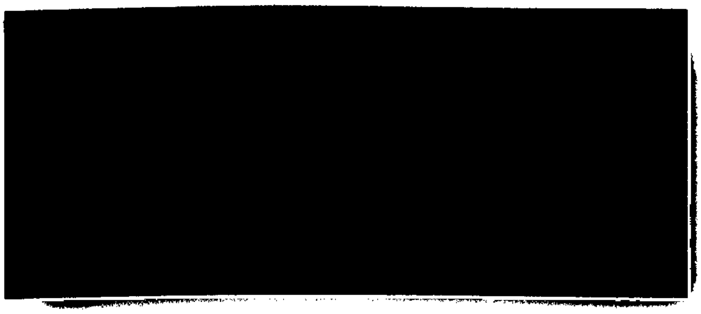

Figure 2.6 Viewing AAM generation using training data

Overall, though AAM parameters achieved significant success using only a Euclidean distance measure and produced results that compare well with other techniques where the percentage of correctness is Fear 90%, Joy 93.3%, Surprise 79.7%, Anger 63.9%, Disgust 93.3%, Sadness 63.9% and Neutral 93.3% [31].

2.3 Techniques Used in Facial Expression Recognition System (FERS)

We are in the midst of a visually enchanting world, which manifest itself with a variety of forms and shapes, colours and textures, motion and tranquillity. The human perception has the capability to a machine in order to interpret the visual information embedded in still images, graphics, and video or moving images in our sensory world. It is thus important to understand the techniques of storage, processing, transmission, recognition, and finally interpretation of such visual scenes. Image processing is a many-steps process. Several steps must be performed one after another until the data of interest could be extracted from the observed scene.

The first stage of any vision system is the image acquisition stage. After the image has been obtained, various methods of processing can be applied to the image to perform the many different vision tasks required today. However, if the image has not been acquired satisfactorily then the intended tasks may not be achievable, even with the aid of some form of image enhancement.