COMPARATIVE ANALYSIS OF EMOTION RECOGNITION METHODS

A Thesis

Submitted to the Department of Computer Science and Engineering

of

BRAC University

by

Mustafa Zaman

Student ID: 03101053

Iftekhairul Islam

Student ID: 03201033

In Partial Fulfillment of the

Requirements for the Degree

of

Bachelor of Science in Computer Science and Engineering

April 2009

DECLARATION

We hereby declare that this thesis is based on the results found by ourselves. Materials of work found by other researcher are mentioned by reference. This thesis, neither in whole nor in part, has been previously submitted for any degree.

Signature of Supervisor	Signature of Author

ACKNOWLEDGMENTS

Special thanks to Ahmedul Kabir, Lecturer, BRAC University, Department of Computer Science and Engineering. Without his proper guideline we could never come up with an abstract idea. Moreover some of the researcher Nico Sabe, P Ekman, J Cohn work encourages us.

ABSTRACT

Recent technological advances have enabled human users to interact with computers in ways previously unimaginable. The aim our thesis is to find out a system where computer can communicate with human more efficiently and more friendly. For that purpose we do studies various possible ways for a computer to communicant with human with limited recourses. In this paper we are trying to describe various algorithms and processes that have been studied all around the world. We also discuss the difficult issues of obtaining reliable affective data, obtaining ground truth for emotion recognition, and the use of unlabeled data.

Keywords: Emotional Recognition, FACS, Optical Flow, Rule-based, Multimodal Approaches.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Introduction

There are in multiple ways Emotions can be recognized. One can observe emotions by Looking once face or by observe the voice also by looking someone's body gesture humans emotions can be observe. However As the growth of Technology It is obvious for present word to make more human interaction between human and machine. It is argued that to truly achieve effective human-computer intelligent interaction (HCII), there is a need for the computer to be able to interact naturally with the user, similar to the way human-human interaction takes place. For example, if a machine talks to you but never listens to you, then it is likely to be annoying, analogous to the situation where a human talks to you but never listens. If a machine can understand human emotions and re-act in proper way then communication between human will be more lively and more efficiently. Emotional expression has been a research subject of physiology since Darwin's work on emotional expression in the 19th century. After then Ekman & Friesen develop a system based on Anatomic of face named Facial Action Coding System (FACS) in 1970. After then research have been going on in different fields. In computer vision system there are number of works have been done by individual and by groups. Some of the key persons are Jeffrey F. Cohn, Keqji Mase, Nico Sabe and lots others. However most of the work has been focused on Facial Expression. The aim of the facial expression is to identify by observing a single image or multiple image what is the emotion that the image shows. But nothing comes to make a standard result. However in very recent days other approaches are taken place.

Human face has several components such as eye, nose, mouth, Brow and few others. Based on movement of those components and change of shape and sizes emotions may be extracted in various ways. However those movements are the result of various combinations of contraction and/or relaxation of the facial muscles. Facial expression is a major media for non-verbal language in human to human communicating. As the statistical analysis shows only 7 percentage of message in human to human communication was covered in verbal communication while 55 percentages transmitted by facial expression.

However, human interacts with other mainly through speech, but also through body gesture, to emphasize a certain part of the speech and display of emotions. As a consequence, the new interface technologies are steadily driving toward accommodating information exchanges via the natural sensory modes of sight, sound, and touch. In face-to-face exchange, humans employ these communication paths simultaneously and in combination, using one to complement and enhance another. The exchanged information is largely encapsulated in this natural, multimodal format. Typically, conversational interaction bears a central burden in human communication, with vision, gaze, expression, and manual gesture often contributing critically, as well as frequently embellishing attributes such as emotion, mood, attitude, and attentiveness. But the roles of multiple modalities and their interplay remain to be quantized and scientifically understood. What is needed is a science of human-computer communication that establishes a framework for multimodal “language” and “dialog”, much like the framework we have evolved for spoken exchange.

Not in every system need emotions like auto pilot or automated surgical system. But system like “Instructor”, “helper” or even “companion” which plays a social role need some level of emotions.

1.2 Objective of the work

Emotion recognition is an application area where human computer interaction research is utilized in both military and commercial products. It is a process of identifying or verifying a person from an image, sounds, gesture and comparing the selected features from the input with a given database. In our everyday human to human communication, interpreting we encounter emotional response all the time. Emotion can make life worth living, or sometime ending. So it is not surprising that most of the great classical philosophers—Plato, Aristotle, Spinoza, Descartes, Hobbes, Hume—had recognizable theories of emotion, conceived as responses to certain sorts of events of concern to a subject, triggering bodily changes and typically motivating characteristic behavior. What is surprising is that in much of the twentieth-century philosophers of mind and psychologists tended to neglect them—perhaps because the sheer variety of phenomena covered by the word "emotion" and its closest neighbors tends to discourage tidy theory. In recent years, however, emotions have once again become the focus of vigorous interest in philosophy, as well as in other branches of cognitive science.

When one talk to give emotion to computer systems, it is hard not to imagine the science fiction futures embodied in book and film, that of computers becoming self-aware sentient machines that then suffer the same dreams of self-importance or even megalomania that afflict their human creators. These system tend to go bad and try to take over the world, and have to be brought back down to size by a hero who engages in some daring-do and in doing so reasserts the values of being properly human. However the reality is that most researchers in that field are not trying to produce this future, real or imagined. Adding emotion in computer is not about making new forms of Artificial Intelligence, human like system. According to Rosalind Picard "Computer does not need affective abilities for the fanciful goal of becoming humanoids, which need them for a meeker and more practical goal: to function with intelligence and sensitivity towards human."

Emotion recognition is one of the latest challenges in human-computer intelligent interaction. Most of the work on emotion recognition focuses on extracting emotions from visual or audio information separately. In human to human communication, interpreting the mix of audio-visual signals is essential in communicating. Researcher in many fields recognize this and thanks to advances in the development of unimodal techniques (in speech and audio processing, computer vision, etc), and in hardware technology (inexpensive cameras and sensors), there has been a significant growth in MMHCI research. Multimodal human computer interaction (MMHCI) lays at the crossroads of several research areas including computer vision, psychology, artificial intelligence, and many others. Unlikely in traditional HCI applications (a single user facing a computer and interacting with it via a mouse or a keyboard), in the new applications interactions are not always explicit commands, and often involve multiple users. This is due in part to the remarkable progress in the last few years in computer processor speed memory and storage capabilities, matched by the availability of many new input and output devices that are making ubiquitous computing a reality. Devices include phones, embedded systems, PDA, laptops, wall size displays, and many others. The wide range of computing devices available, with differing computational power and input/output capabilities, means that the future of computing is likely to include novel way of interaction. Some of the methods include gestures, speech, haptics, eye blinks, and many others. As in human-human communications, however effective communication is likely to take place when different input devices are used in combination.

Extensive surveys have been previously published in several areas such as face detection, face recognition, facial expression analysis, vocal emotion, gesture recognition, human motion analysis, and eye tracking. A review of vision-based HCI is presented in with a focus on head tracking, face and facial expression recognition, eye tracking, and gesture recognition. Adaptive and intelligent HCI is discussed in with a review of computer vision for human motion analysis, and a discussion of techniques for lower arm movement detection, face processing, and gaze analysis. Multimodal interfaces are discussed in. Real-time vision for HCI (gestures, object tracking, hand posture, and gaze) is discussed in. Here, we discuss work not included in previous surveys, expand the discussion to areas not covered previously, and discuss new applications in emerging areas while highlighting the main research issues.

1.3 Organization of the Thesis

This thesis consists of five chapters. The first chapter gives an introduction to the thesis. The second and third chapters present the background, theory and related work in this field. Chapter four shows the methods used for human emotion recognition that were studied. Finally, in Chapter Five, conclusions are drawn and a brief explanation of future work is presented.

CHAPTER II

Human emotions

2.1 Human emotions

It is widely accepted from psychological theory that human emotions can be classified into six archetypal emotions: surprise, fear, disgust, anger, happiness, and sadness. Facial motion and the tone of the speech play a major role in expressing these emotions. The muscles of the face can be changed and the tone and the energy in the production of the speech can be intentionally modified to communicate different feelings. Human beings can recognize these signals even if they are subtly displayed, by simultaneously processing information acquired by ears and eyes. Based on psychological studies, which show that visual information modifies the perception of speech, it is possible to assume that human emotion perception follows a similar trend. Motivated by these clues, De Silva et al. conducted experiments, in which 18 people are required to recognize emotion using visual and acoustic information separately from an audio-visual database recorded from two subjects. They concluded that some emotions are better identified with audio such as sadness, and fear, and other with video such as anger and happiness. Moreover Chen at el. showed that these two modalities give complementary information, by arguing that the performance of the system.

Recent studies show that Facial Expression of emotions are hand written in our genes, According to the Journal "Journal of Personality and Social Psychology". The research suggests that facial expressions of emotion are innate rather than a product of cultural learning. The variance of emotions is a vital part of human luxury, or some may say, misery. But with our psyche being so chronically drenched in the stagnant swamp of love, fear, jealousy and anger, day in, day out, our sponge-like mind has become so saturated it no longer absorbs. Researchers dedicated their life to find out human emotions and categorized them. Paul Ekman has dedicated his career to researching emotions, focusing primarily on these six basic emotions.

- Fear — Danger lurks
- Sadness — Impending loss
- Anger — Conspecific threat, trespass, thwarted goals, plea for justice
- Happiness — Impending gain
- Surprise — Unexpected event
- Disgust — Contamination, toxic contact

However in the 90s Ekman expanded his list of basic emotions, including a range of positive and negative emotions not all of which are encoded in facial muscles. The newly included emotions are:

- Amusement
- Contempt
- Contentment
- Embarrassment
- Excitement
- Guilt
- Pride in achievement
- Relief
- Satisfaction
- Sensory pleasure
- Shame

Emotion	Valence	Arousal	Gesture
Anger	Negative	High	Violent descend of hands
Despair	Negative	High	Leave me alone
Interest	Positive	Low	Raise hands
Pleasure	Positive	Low	Open hands
Sadness	Negative	Low	Smooth falling hands
Irritation	Negative	Low	Smooth go away
Happiness	Positive	High	Circular Italianate movement
Pride	Positive	High	Close hands towards chest

Table1: Human Emotions and its classifications.

2.2 Non-Emotions

In his 1991 book, Emotion and Adaptation, Richard Lazarus lists several mental states that may be emotion related, but are not themselves actual emotions. The list includes the complex states of: grief and depression; the ambiguous positive states of: expansiveness, awe, confidence, challenge, determination, satisfaction, and being pleased; the ambiguous negative states of: threat, frustration, disappointment, helplessness, meaningless, and awe; the mental confusion states of bewilderment and confusion; the arousal states of: excitement, upset, distress, nervousness, tension, and agitation; and finally the pre-emotions of: interest, curiosity, amazement, anticipation, alertness, and surprise.

2.3 Universal Emotions:

Scientific research in the 19th century stated that the young and the old of widely different races, both with man and animals, express the same state of mind by the same movements. Still, up to the mid-20th century most anthropologists believed that facial expressions were entirely learned and could therefore differ among cultures. Studies eventually supported Darwin's belief to a large degree, particularly for expressions of anger, sadness, fear, surprise, disgust, contempt, happiness and caring. Recent psychological research has classified six facial expressions which correspond to distinct universal emotions: disgust, sadness, happiness, fear, anger, surprise. These expressions are manifestations of particular emotions, regardless of cultural background, and regardless of whether or not the culture has been isolated or exposed to the mainstream.

2.3.1 Happiness:

Happiness is the most common emotional expression perform by human being. A subject smiles when he/she enjoy the environment. In physiology, a smile is a facial expression formed by flexing those muscles most notably near both ends of the mouth. This expression is universal and easy to recognize.

2.3.2 Sadness

Sad expressions convey messages related to loss, bereavement, discomfort, pain, helplessness, etc. This emotion is often consider as opposite of Happy/joy. The subject who is sad is usually also angry and mad at something.

2.3.3 Anger:

Anger is a natural response to certain threats. As a result, aggression is sometimes the appropriate response to anger. As daily stresses and frustrations underlying anger seem to increase, but the expectation of reprisals decrease with the higher sense of personal security. This expression may very in different ages and sexes.

2.3.4 Fear

Fear expressions convey information about imminent danger, a nearby threat, a disposition to flee, or likelihood of bodily harm.

2.3.5 Surprise

Surprise expressions almost always occur in response to events that are unanticipated, and they convey messages about something being unexpected, sudden, novel, or amazing. The brief surprise expression is often followed by other expressions that reveal emotion in response to the surprise feeling or to the object of surprise, emotions such as happiness or fear. In surprise expressions, the eyebrows are slightly raised straight up, making faint horizontal wrinkles on the forehead, the upper eyelid is raised slightly, the mouth is opened by the jaw drop, and the lips are relaxed

2.3.6 Disgust

Disgust is an emotion that is typically associated with things that are perceived as unclean, inedible, or infectious. Disgust expressions are often part of the body's responses to objects that are revolting and nauseating, such as rotting flesh, fecal matter and insects in food, or other offensive materials that are rejected as suitable to eat. Obnoxious smells are effective in eliciting disgust reactions. Disgust expressions are often displayed as a commentary on many other events and people that generate adverse reactions. Normally a disgust expression has the following properties: wrinkled nose with the eyebrows pulled down and the upper lip drawn up, the lower eyelid is tensed and the eye opening narrowed. The pressing of the lips and rising of the upper eyelids are relevant to an anger expression; whereas the mouth would be open and the upper eyelids relaxed in the typical disgust expression

2.4 Basic and complex emotions

Many theorists define some emotions as basic where others are complex. Basic emotions are claimed to be biologically fixed, innate and as a result universal to all humans and many animals as well. Complex emotions are then either refined versions of basic emotions, culturally specific or idiosyncratic. A major issue is then to define which emotions are basic and which are complex.

One of the problems here is that there is no consensus on the method by which basic emotions can be determined. Theorists can point to universals in facial expression (e.g. Ekman), distinctive physiological symptoms (e.g. the blush of embarrassment), or labels common to different languages. Moreover there should be some plausible developmental story concerning how the various non-basic emotions can be grounded in the basic ones.

The Li Chi: Joy, anger, sadness, fear, love, disliking and liking (1st Century BC Chinese encyclopedia, cited in Russell 1991: 426).
The Stoics: Pleasure/delight, distress, appetite and fear (Cicero, Tusculan Disputations, IV: 13-15).
René Descartes: Wonder, love, hatred, desire, joy and sadness (Passions, 353).
Baruch Spinoza: Pleasure, pain and desire (Ethics, pt. III, prop. 59).
Thomas Hobbes: Appetite, desire, love, aversion, hate, joy and grief (Leviathan, pt. I, ch. 6).
Paul Ekman (1972): Anger, disgust, fear, happiness, sadness and surprise.
Paul Ekman (1999): Amusement, anger, contempt, contentment, disgust, embarrassment, excitement, fear, guilt, pride in achievement, relief, sadness/distress, satisfaction, sensory pleasure and shame.
Jesse Prinz (2004): Frustration, panic, anxiety, physical disgust, separation distress, aversive self-consciousness, satisfaction, stimulation and attachment.

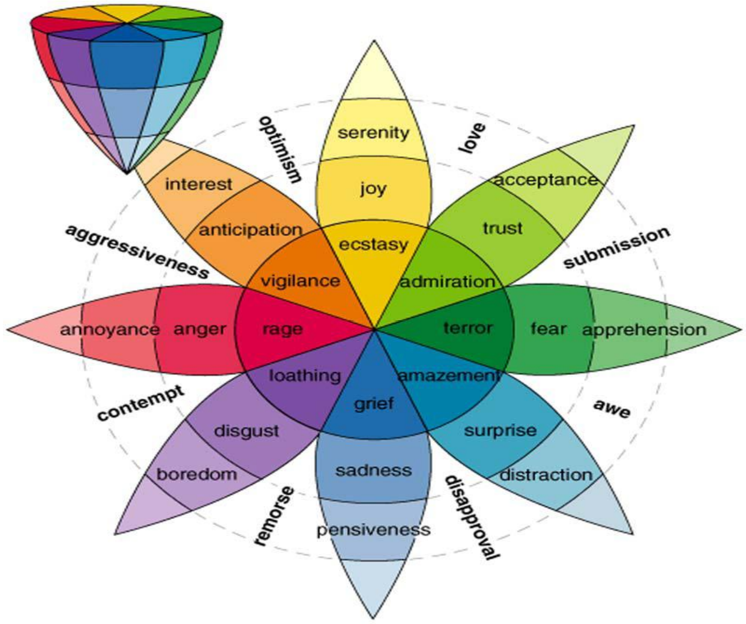


Figure 1: Robert Plutchik's circumplex model

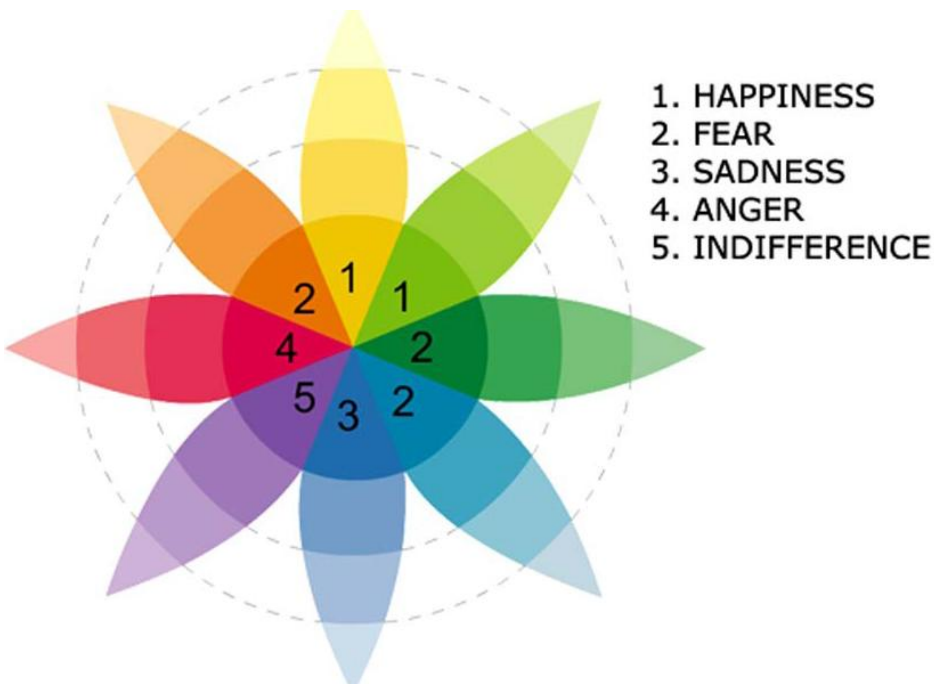


Figure 2: Plutchik's reduced five basic emotions

CHAPTER III

Related Works and Methods

3.1 Introduction:

Emotional Expression has been a vastly discussed topic since P. Ekmane work. In a number of fields this topic was discussed. Not only in Human Computer Interface but in many sectors this topic was discussed. In the following there the accomplishments in the related areas such as:

- Psychological studies,
- Human movement analysis,
- Face detection,
- Face tracking and recognition
- Make the automatic facial expression analysis possible.
- Emotion and paralinguistic communication,
- Clinical psychology,
- Psychiatry,
- Neurology,
- Pain assessment,
- Lie detection,
- Intelligent environments, and
- Multimodal human-computer interface (HCI).

However emotion can be recognized by a computer in multiple ways Emotions can be recognized by

- Emotion Recognition by Facial Expression
- Emotion Recognition from Body Gesture
- Emotion Recognition by Speech
- Multimodal Emotion Recognition
- Also by Body Physiological Signal.

To detect Facial expressions probably P. Ekmane was the first who have done vast work.

3.2 Facial Action Coding System (FACS)

Facial Action Coding System (FACS) is the most widely used and versatile method for measuring and describing facial emotions. Paul Ekman and W.V. Friesen developed the original FACS in the 1970s by determining how the contraction of each facial muscle (singly and in combination with other muscles) changes the appearance of the face. They examined videotapes of facial behavior to identify the specific changes that occurred with muscular contractions and how best to differentiate one from another. They associated the appearance changes with the action of muscles that produced them by studying anatomy, reproducing the appearances, and palpating their faces. Their goal was to create a reliable means for skilled human scorers to determine the category or categories in which to fit each facial behavior. FACS measurement units are Action Units (AUs), not muscles, for two reasons.

First, for a few appearances, more than one muscle was combined into a single AU because the changes in appearance they produced could not be distinguished.

Second, the appearance changes produced by one muscle were sometimes separated into two or more AUs to represent relatively independent actions of different parts of the muscle. (After all, facial muscles were identified and named by anatomists, not behavioral psychologists.)

FACS consists of 44 action units, including those for head and eye positions. AUs are anatomically related to contraction of specific facial muscles. They can occur either singly or in combinations. AU combinations may be additive, in which case combination does not change the appearance of the constituents, or no additive, in which case the appearance of the constituents changes (analogous to co-articulation effects in speech). For action units that vary in intensity, a 5-point ordinal scale is used to measure the degree of muscle contraction. Although the number of atomic action units is small, more than 7,000 combinations of action units have been observed. FACS provides the necessary detail with which to describe facial expression. In 2002, a new version of FACS was finally published, with large contributions by Joseph Hager [13]. On new version of FACS some of the redundant action unites were removed and some are added. The new versions were not renaming the system. It is still simply known as FACS, not as FACS2, FACS 2002 revision or FACS version 2. The website of Paul Ekman's lab refers to it as the "new" FACS.

Their primary goal in developing the FACS was to develop a comprehensive system which could distinguish all possible visually distinguishable facial movement. They chose to derive FACS from an analysis of the anatomical basis of facial movement. Since every facial movement is the result of muscular action, a comprehensive system could be obtained by discovering how each muscle of the face acts to change visible appearance. With that knowledge it would be possible to analyze any facial movement into anatomically based minimal action units.

What are those action Units:

1 Inner Brow Raiser,
2 Outer Brow Raiser
4 Brow Lowered
5 Upper Lid Raiser
6 Cheek Raiser
7 Lid Tightened
9 Nose Wrinkles
And more

3.2.1 How to read AU:

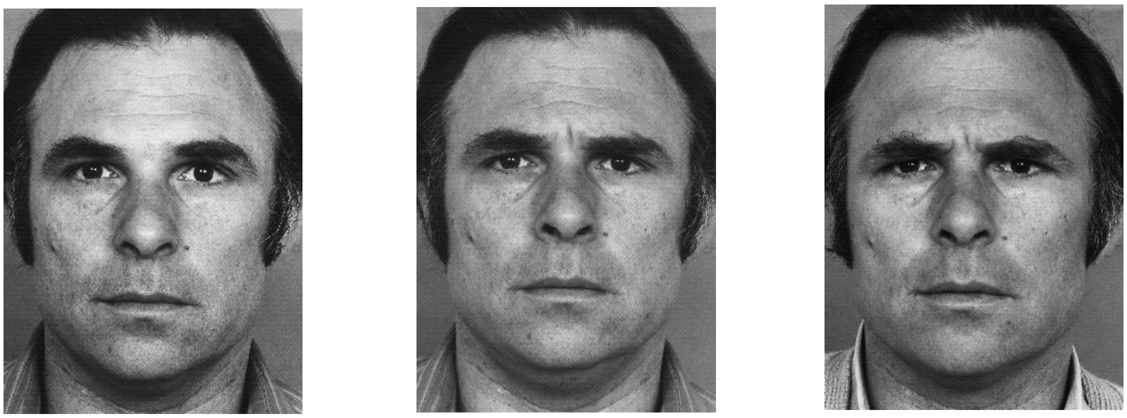
Each AU is identified by a number and name (for example, the first AU explained is "AU 4 - Brow Lowerer"). Names like "Brow Lowerer" are provided as a more meaningful handle than the more arbitrary numbers, and might make it easier for you to relate to the AUs as you begin learning this FACS. Among all other AU. Action Unit – 4 is easier to do and most people do this while they are creating Expression. On Action Unit – 4 the following things are happen.

1. Lowers the eyebrow. In different instances it may be only the inner portion of the eyebrow that is lowered or it may be both inner and central portions that are lowered, or it may appear that the entire eyebrow is lowered.

2. Pushes the eye cover fold downwards and may narrow the eye aperture.

3. Pulls the eyebrows closer together.

4. Produces vertical wrinkles between the eyebrows, which may be deep. In some people the wrinkles between the eyebrows may not be vertical but at a 45 degree angle, or both angled and vertical. May also produce one or more horizontal wrinkles at the root of the nose. If the vertical, angled, or horizontal wrinkles are permanently etched, they deepen.



Normal **AU – 4a** **AU- 4b**

Figure 3: Action Unit 4

This movement is easy for most people to do. Lower your eyebrows and pull them together. Try not to wrinkle your nose (if your nose is wrinkling, you are doing AU 9). By this AU the result might be inferior, surprise or disgust.

Generalized to Different Databases:

		Test Data base		Train Database
		Cohn - Kande	Ekman - Hager	
Recognition Rate	Upper Face	93.2%	96.4%	Ekman –
	Lower Face	96.7%	93.4%	Cohn - Kande

Table 2: Test result of FACS in two different databases

3.3 Hidden Markov Model:

Hidden Markov models (HMMs) are the most popular means of temporal classification. They have found application in areas like speech, handwriting and gesture recognition. Hidden Markov Model has been successfully used for speech recognition where data is essential one dimensional. Extension to a fully connected two dimensional HMM has been shown to be computationally very complex. Kuo and Agazzi have used a pseudo two dimensional HMM for character recognition that was shown to perform reasonably fast for binary images.

For frontal face images, the significant facial regions(hair, forehead, eyes, nose, mouth) come in a natural order from top to bottom, even if the image are taken under small rotations in the image plane and/or rotations in the plane perpendicular to the image plane. Each of these facial regions is assigned to a state in a left to right 1D continuous HMM. The state

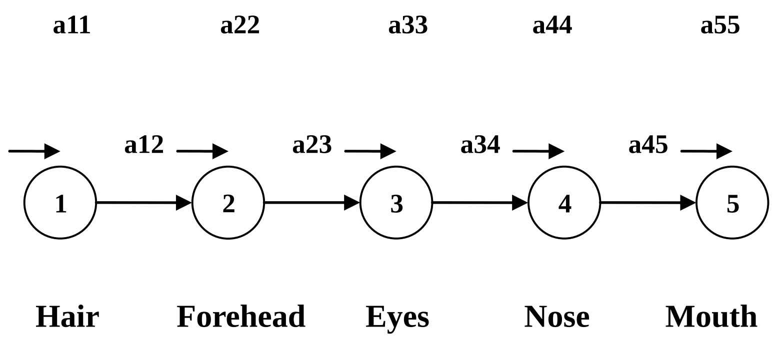


Figure 4: HMM State Structure of Face model

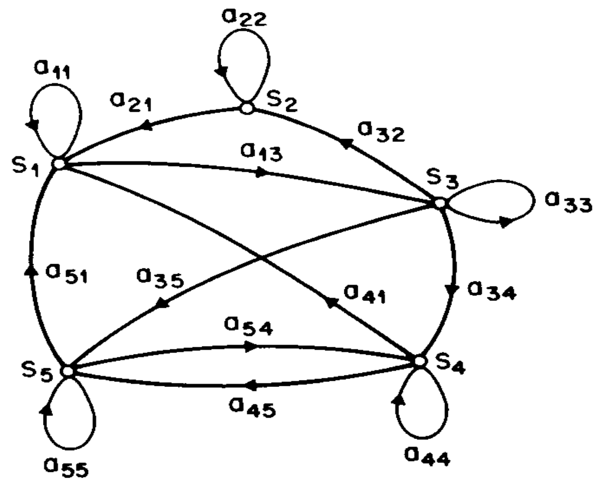


Figure 5: Diagram of HMM

3.4 Eigenfaces

Human face recognition is a very difficult and practical problem in the field of pattern recognition. On the foundation of the analysis of the present methods on human face recognition, a new technique of image feature extraction is presented. And combined with the artificial neural network, a new method on human face recognition is brought up. By extraction the sample pattern's algebraic feature, the human face image's eigenvalues, the neural network classifier is trained for recognition. The Kohonen network we adopted can adaptively modify its bottom up weights in the course of learning. Experimental results show that this method not only utilities the feature aspect of eigenvalues but also has the learning ability of neural network. It has better discriminate ability compared with the nearest classifier. The method this paper focused on has wide application area. The adaptive neural network classifier can be used in other tasks of pattern recognition.

3.5 Neural Networks

Face recognition as an idea comes from the real life term "face recognition". Part of the human "face recognition" is the "memory" of the humans, which means the brain. The connection between this term of a biological organization and the computer is the neural networks. The neuron has a biological meaning even thought the neural networks are just inspired from that biological term of neuron and neuron systems, "network". Even thought that is not that positively true, as long as the aim of neural networking is to create a computer system which actually works in a similar way to how we think the neurons in the human brain works. Neural networks can be defined as:

A neural network is an interconnected group of nodes, akin to the vast network of neurons in the human brain.

A neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing element or nodes. Neural network architecture is inspired by the architecture of biological nervous systems, which use many simple processing elements operating in parallel to obtain high computation rates.

Neural networks are a form of microprocessor computer system with simple processing elements, a high degree of interconnection, simple scalar messages and adaptive interaction between elements.

The neural networks resemble the brain mainly in two respects-

- Knowledge is acquired by the network through a learning process
- Interneuron connection strengths known as synaptic weights are used to store the knowledge.

That means to construct a machine that is able to think. Somehow, not really known yet, the brain is capable to think and perform some operations and computations, much faster sometimes from a computer even the "memory" is much less. How the brain is managed to do that is a hardware parallelism. The computing elements are arranged so that very many of them are working on a problem at the same time. Since there is a huge number of neurons, somehow the weak computing powers of these many slow elements are combined together to form a powerful result.

A Neural Network is an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the inter-unit connection strengths, or weights, obtained by a process of adaptation to, or learning from, a set of training patterns.

In order to see how different this is from the processing done by conventional computers it is worth examining the underlying principles that lie at the heart of all such machines.

Neural network algorithms for face recognition work by applying one or more networks directly to portions of the input image and arbitrating their results. Each network is trained to output the presence or absence of a face. Training now a neural network for face recognition is a challenging task because of the difficulty in characterizing prototypical "non-face" images. The structure of the images that are used is going to be analyzed in the dataset chapter.
Finally we have to mention the why that a neural network system operates. Mainly face recognition systems operate in two stages. The first stage is to apply a set of neural network based filters to an image and then use an arbitrator to combine the outputs. Then the arbitrator merges detection from individual filters and eliminates overlapping detection.

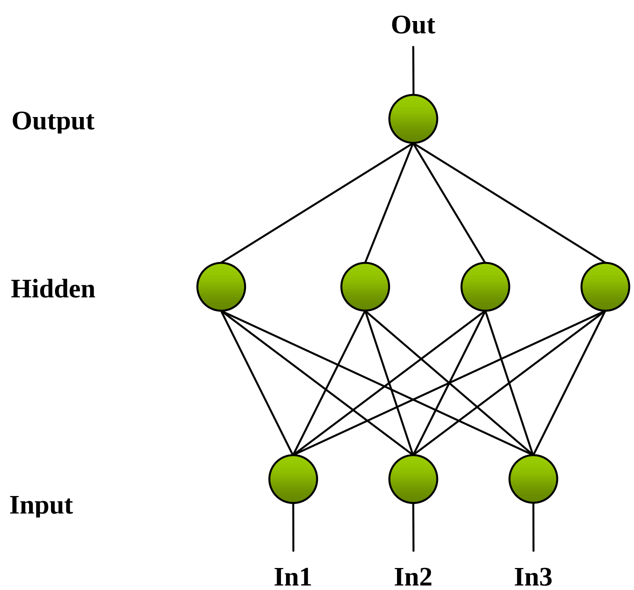


Figure 6: A Simple Neural Network

CHAPTER IV

System Overview

4.1 Introduction

The method of Ekman and Friesen the FACS where movement on the face are describe by a set of Action Units. Each AU has some related muscular basis. Each facial expression may be described by a combination of AUs. This system of coding facial expressions is done manually by following a set prescribed rule. The inputs are still images of facial expressions, often at the peak of the expression. This process is very time-consuming. Ekman's work inspired many researchers to analyze facial expression from images and videotapes. Facial feature were tracked measuring the amount of facial movement, they attempt to categorize different facial expression. However there are many of the research done in the recent years and most of these work are categorize in two different processes.
Region Based.
Feature Based.

4.2 Region Based

May be Mase[12] was one of the first who use image processing techniques to recognize facial expression. He uses Optical Flow to estimate facial muscle movement which could then be recognized. There are many algorithm develop to estimate Optical Flow. Those are categorized in three parts gradient based, correlation based and filtering based. The conventional gradient based algorithm is simple and work well on facial skins deformation expect for the aperture problem and motion discontinuity. Although facial skins are not rigid, we can treat its motion field smooth locally. Tracking a human face is a vast of body work, with applications ranging from motions capture to human-computer interaction. The muscles of expression interact in a very complex way under the facial skins. Keqji Mase use windows which define the major direction of muscle contraction (skin deformation). This allows the effective of each muscle to be determined. They calculate the average length of directional components in the major window axis from the optical flow vectors in each window. In this figure shows recorded image and estimated muscle velocities during one full cycle of a happy expression. The white and the black arrows are computed optical flow and estimated velocities, respectively. The optical flow field is computed for each pixel and is shown in every three pixels. There are several possible ways to recognize expression from the data acquired by the above method. For instance, if the expressions are simple (one expression), we can develop a vector consisting of muscle velocities at the time a t which one arbitrary muscle reaches its maximum velocity.

Since the data consists of velocity, it becomes zero a t maximum extent of the expression. The integral of the velocity is roughly regarded as the displacement. Then the integrals of the velocity until the time the velocity reaches zero are used for pattern matching.
The waveform pattern of each velocity may be used directly from pattern matching.
We related each muscle movement to corresponding AU, and then analyze it using FACS.

4.3 Feature Based

Facial expressions give important clues about emotions. Therefore, several approaches have been proposed to classify human affective states. The features used are typically based on local spatial position or displacement of specific points and regions of the face, unlike the approaches based on audio, which use global statistics of the acoustic features. For a complete review of recent emotion recognition systems based on facial expression the readers are referred to.

Mase proposed an emotion recognition system that uses the major directions of specific facial muscles. With 11 windows manually located in the face, the muscle movements were extracted by the use of optical flow. For classification, K-nearest neighbor rule was used, with an accuracy of 80% with four emotions: happiness, anger, disgust and surprise.

Yacoob et al. proposed a similar method. Instead of using facial muscle actions, they built a dictionary to convert motions associated with edge of the mouth, eyes and eyebrows, into a linguistic, per frame, mid-level representation. They classified the six basic emotions by the used of a rule-based system with 88% of accuracy.

Black et al. used parametric models to extract the shape and movements of the mouse, eye and eyebrows. They also built a mid- and high-level representation of facial actions by using a similar approach employed in, with 89% of accuracy.

Tian et al. attempted to recognize Actions Units (AU), developed by Ekman and Friesen in 1978, using permanent and transient facial features such as lip, nasolabial furrow and wrinkles. Geometrical models were used to locate the shapes and appearances of these features. They achieved a 96% of accuracy. Essa et al. developed a system that quantified facial movements based on parametric models of independent facial muscle groups. They modeled the face by the use of an optical flow method coupled with geometric, physical and motion-based dynamic models. They generated spatial-temporal templates that were used for emotion recognition. Without considering sadness that was not included in their work, a recognition accuracy rate of 98% was achieved.

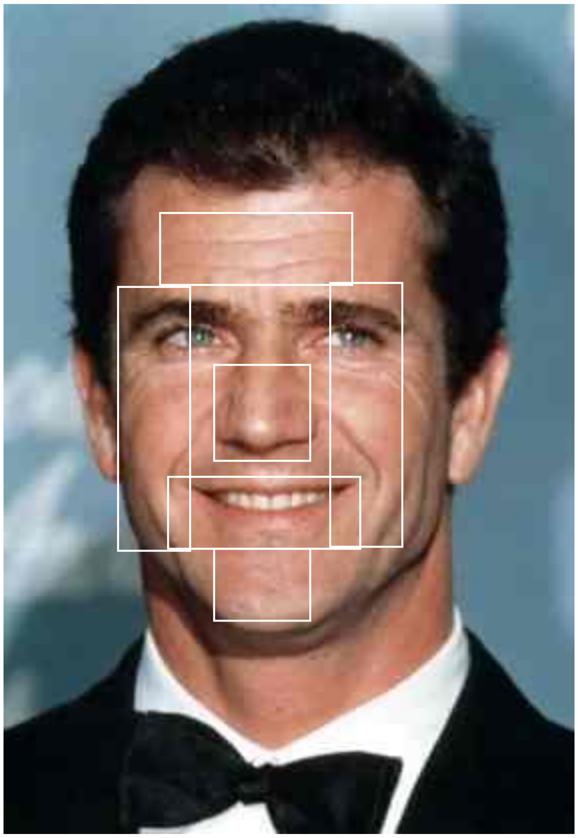


Figure 7: Multiple Features of Face

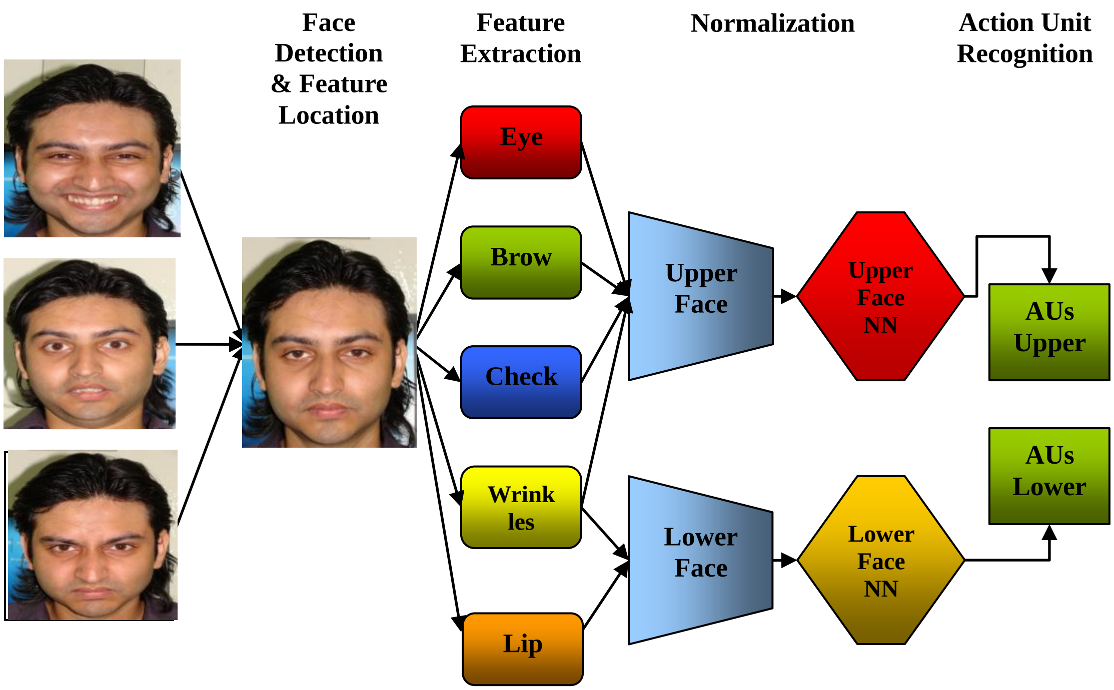


Figure 8: Feature Based Face Recognition System

Author	Processin g	Classification	Number Of Categories	Number Of Subjects	Performance
Mase	optical flow	NN	4	1	86%
Black & Yacoob	parametric model	rule-based	6	40	92%
Yacoob & Davis	optical flow	rule-based	6	32	95%
Rosenblum et al.	optical flow	neural networks	2	32	88%

Table 3: A comparative analysis of different Expression recognition methods

4.3.1 Application of Face Recognition System:

International Police Organization Proposes Worldwide Facial Recognition System. Interpol, the Europe-based international law enforcement group, has proposed an automated face-recognition system for international borders. Such a system could require travellers to undergo face scans, and make the information available to numerous countries. An Interpol face-recognition database would permit Interpol member nations to search records containing travellers' personal biometric information, and could be used in conjunction with travel watch lists. The inaccuracy of facial recognition technology has repeatedly been criticized. Privacy watchdogs have questioned the efficacy and wisdom of government programs that collect ever-more personal information at border crossings. "We need to get our data to the border entry points. There will be such a large role in the future for fingerprints and facial recognition," said Mark Branchflower, head of Interpol's fingerprint unit. (Oct. 20, 2008)

Companies Use Surveillance Cameras for Advertising Studies. Surveillance cameras have long been used as anti-crime devices. However, companies are now seeking to use surveillance cameras to watch people for advertising research. In Germany, developers are placing video cameras into street advertisements and attempting to discern people's emotional reactions to the ads. Dutch researchers are using experimental emotion-recognition software to test individuals' reactions to advertisements and marketing. (July 10, 2007)

Federal Air Marshals to Surreptitiously Photograph Travellers. The US Department of Homeland Security is investing in face recognition technology so that federal marshals can surreptitiously photograph people in airports, bus and train stations, and elsewhere to check whether they are in terrorist databases. The Los Angeles police department already is using handheld facial recognition devices. See EPIC's Video Surveillance page. (May 10, 2007)

British Police Look to Build National Mug shot Database. The Police Information Technology Organisation aims to create a national database of still and video facial images, tattoos, and other imagery linked to criminal biographical information. They are also looking into how they can incorporate facial recognition software into the mug shot database for the police forces of England, Scotland, and Wales. (Jan. 16, 2006)

Google's Picasa took a giant step forward with the implementation of face recognition software. This improvement is brought by Google to users in its web based photo management software application, Google Picasa. Besides photo editing, slide show presentation, printing, etc, Picasa web album users will now enjoy another new face recognition feature. It is a cool feature which allows users to organize their photo collection by people. The feature is smart enough to recognize similar faces or features in users' photo collections and allows them to tag those photos.

4.4 Emotion Recognition by Speech

The importance of automatically recognizing emotions from human speech has grown with the increasing role of spoken language interfaces in human-computer interaction applications. This study addresses the design of an automatic emotion recognition system using spoken language information through signal processing and pattern recognition techniques. The specific focus is on a case study of detecting "negative" and "non-negative" emotions using data obtained from a call centre application. Those two emotional categories are defined in a pragmatic way to follow the need for the given application domain.

	Anger	Happiness	Sadness	Fear	Disgust
Speech Rate	slightly faster	faster or slower	slightly slower	much faster	very much slower
Pitch Average	very much higher	much higher	slightly lower	very much higher	very much lower
Pitch Range	much wider	much wider	slightly narrower	much wider	slightly wider
Intensity	higher	higher	lower	normal	lower
Voice Quality	breathy	blaring	resonant	irregula r	grumbled

Table 3: Differences so speech on different emotional status.

The emotion recognition system by speech consists of two main blocks as shown in Figure 9: feature extraction, and emotion recognizer. A typical interactive spoken dialog system comprises elements enabling dialog management between human and machines, and the emotion recognition system can help the machine manage the interaction in a natural and effective manner. The focus of this work is in the design of an emotion recognizer that will work in conjunction with such a system.

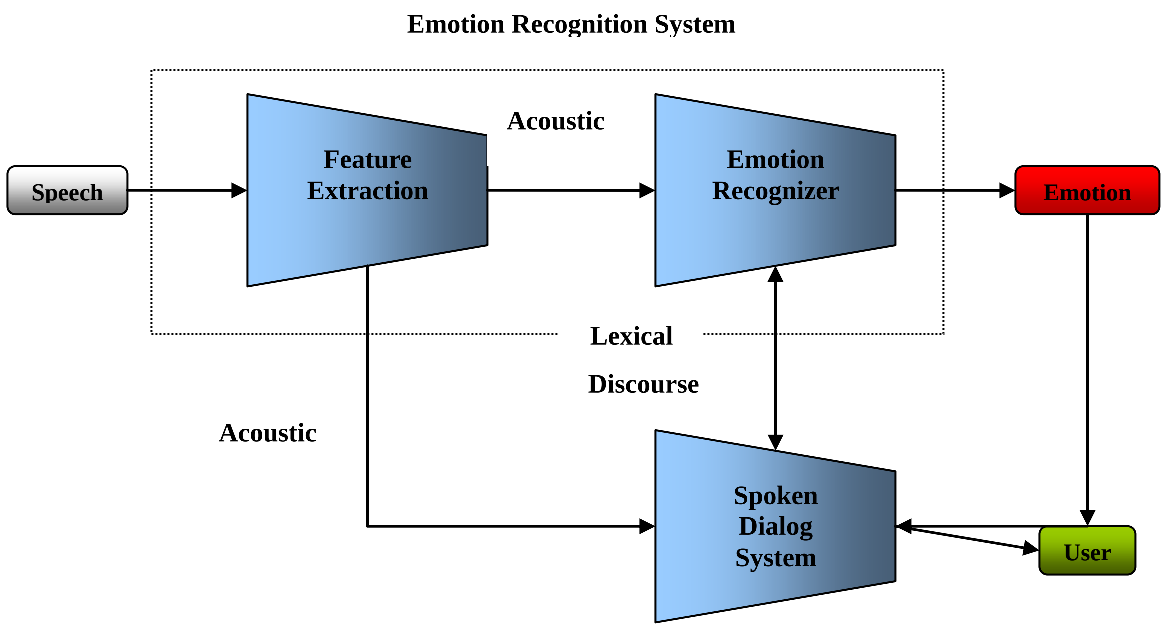


Figure 9: Emotion Recognition System hat Uses Spoken Language

The emotion recognition system in this work utilizes three kinds of information as its input: acoustic information, lexical information, and discourse information. The acoustic information to emotion recognizer includes prosody-based acoustic correlates: fundamental frequency, energy of speech signals, and time duration features of voiced and unvoiced portion of the speech. Many of prosody-related acoustic correlates have been explored in psychology and linguistics, and it is known that they play a significant role in recognizing human emotion. Lexical information to emotion recognizer is “emotionally salient word lexicon”, which are obtained from the data in a data-driven fashion. The words in an utterance, which match those in the “emotionally salient word lexicon”, are used to make a decision. Another important piece of information that would be useful for emotion recognition is discourse information. The history of dialog between the user and the machine carries information regarding the emotional state of the user. For example, if the automated system has rejected or misunderstood the user's response several times, the user's emotional states would likely change. Since each information source varies, different classifiers are used for each information stream in our emotion recognition system. The outputs of these classifiers are combined at the decision level.

There is a basic problem in trying to recognize emotions. In emotion recognition in speech signals, the class boundaries between emotion categories are vague or fuzzy because of the linguistic uncertainties in the definition of the emotional classes in everyday use. To tackle this problem, we developed Fuzzy Inference System (FIS) for emotion recognition. FIS is basically a rule-based system to tackle the language-related uncertainties in the given problems. FIS is used in the classification of emotions using acoustic information, and better performs compared with other statistical pattern classification methods. FIS also improves the interpretability of the results by providing more insight into the classifier structure and decision making systems.

With regards to the problem of speech signal features for emotion classification, most previous research has used suprasegmental/prosodic features as their acoustic cues. Such cues have been known to be an important indicator to emotional states, and thus used in the design of many emotion recognition systems. The spectral information of speech is yet another important feature for representing emotional states, which has been found to be useful for emotion classification. One recent study has also shown that there are variations across emotional states in the spectral features at the phoneme level, especially vowel sounds. Our study explores this notion further in the context of automatic emotion recognition. Specifically, the central hypothesis of the study is that different emotional categories affect different phonemes in distinct ways; hence, automatic emotion classification has to incorporate phoneme dependencies.

4.5 Multimodal Emotion Recognition

Multimodal Human computer interaction (MMHCI) lies at the crossroads of several research areas including computer vision, psychology, artificial intelligence, and many others. As computers become integrated into everyday objects (ubiquitous and pervasive computing), effective natural human-computer interaction becomes critical: in many applications, users need to be able to interact naturally with computers the way face-to-face human-human interaction takes place. We communicate through speech and use body language (posture, gaze], hand motions) to express emotion, mood, attitude, and attention

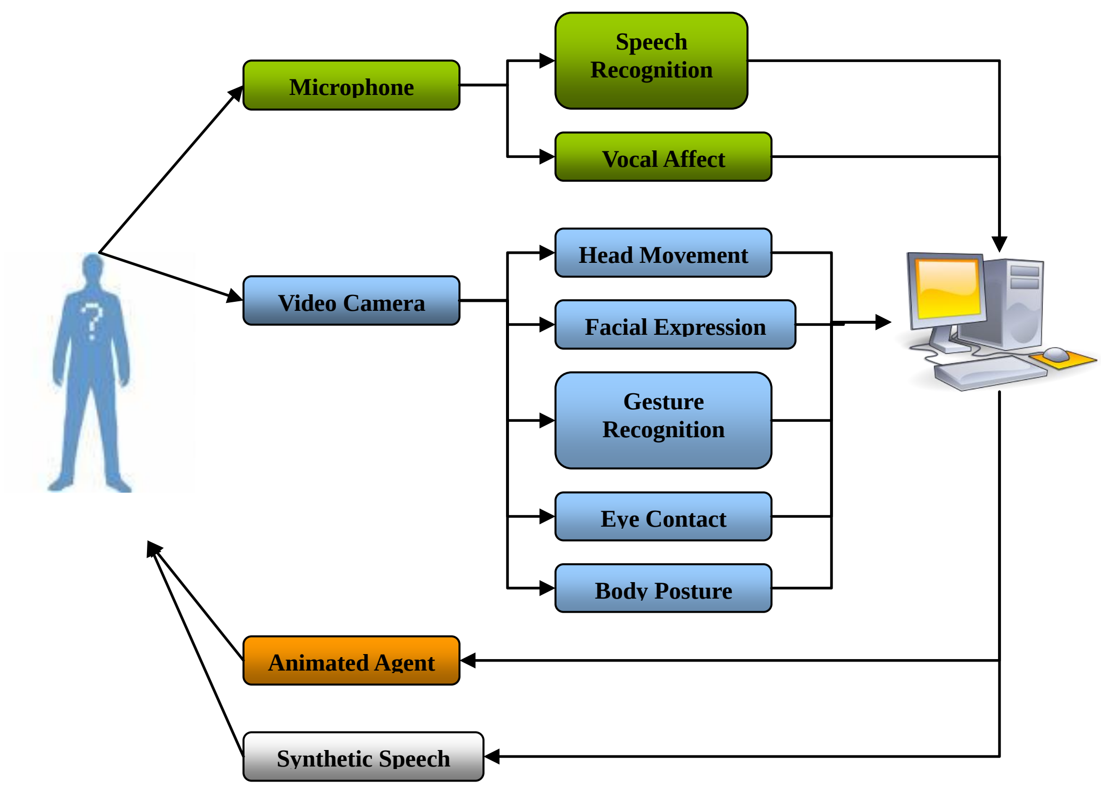


Fig: Multimodal Emotion Recognition System

CHAPTER V

Conclusion

5.1 Future work:

Based on our thesis we come up with a tentative solution. Our future works will be an audio visual system. Our future system will be a multimodal system. Where we consider Images and also voice data. On Images we will process our image/video by using Optical flow and then we will make a rule base model from where we will determine the expressions.
We will also include speech because we want to make the system more accurate.

5.2 Challenges:

The main challenges are to use Optical flow. As optical flow will works where the environment will suitable. If the room environment will dim or the video quality is not good then it will be a problem to determine. Moreover movement of head could be a problem. If there are multiple faces then the system might not work. One of the biggest challenges that we might facing is to create a Rule-based classifiers. Maybe to make a rule-based classifier is one big challenge to overcome. Other thing we need to focus is to think of the video quality and the sound quality. As this could be vary from different conditions. However we will try to overcome our problems while we are facing some barrier. While we are taking inputs sounds as a input then we need to make it clear that we did take input of the right persons. If we get multiple sounds as input then we might get the wrong data. May be we calculate one person’s image and calculate other person’s sounds. May be the image person is happy but the sound person is not feeling happy.

However we strongly believe this is a learning process. We did start with some thing and finally come up with that we can not work with it. And we believe that in near future we can able to make something which will able to communicate with human as a friend. Which can understand humans’ emotions and is the user is sad and able to show sympathy.

6.0 References:

M. Bartlett, J. Hager, P.Ekman, and T. Sejnowski. Measuring facial expressions by computer image analysis.*Psychophysiology*, 36:253–264, 1999

P. Ekman and W. Friesen, Facial Action Coding System: Investigator's Guide, Consulting Psychologists Press, 1978.

Yingli Tian,Takeo Kanade, Jeffrey F. Cohn: Recognizing Action Units for Facial Expression Analysis, CMURITR9940, December, 1999.

Alejandro Jaimes*,1 and Nicu Sebe: Multimodal Human Computer Interaction: A Survey, IDIAP, Switzerland.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G.: Emotion recognition in human-computer interaction, IEEE Signal Processing Magazine, January 2001.

Nicu Sebe, Ira Cohen, and Thomas S. Huang: Multimodal Emotion Recognition, June 18, 2004.

Ara V. Nefian, Monson H. Hayes III: Hidden Markov Models For Face Recognition.

Sebe, N., Cohen, I., Huang, T.S.: Multimodal Emotion Recognition, Handbook of Pattern Recognition and Computer Vision, World Scientific, ISBN 981-256-105-6, January 2005.

Pantic, M., Sebe, N., Cohn, J., Huang, T.S.: Affective Multimodal Human-Computer Interaction, ACM Multimedia, pp. 669 - 676, Singapore, November 2005

Ze-Jing Chuang, Chung-Hsien Wu: Multi-Modal Emotion Recognition from Speech and Text, Computational Linguistics and Chinese Language Processing, Vol. 9, No. 2 , August 2004, pp. 45-62.

Carlos Busso, Zhigang Deng , Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, Shrikanth Narayanan: Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information, published in the Proc. of ACM 6th International Conference on Mutlmodal Interfaces (ICMI 2004), State College, PA, Oct 2004.

Irena Mihalinec, Radomir Stevanović, Karolj Skala: Emotion Recognition and its Aesthetic Interpretation, Rudjer Boskovic Institute, Center for Informatics and Computing, Zagreb, Croatia.

Ginevra Castellano, Loic Kessous, and George Caridakis: Multimodal emotion recognition from expressive faces, body gestures and speech, EU 6th Framework Programme (2004-2007).

Albino Nogueiras, Asunción Moreno, Antonio Bonafonte, José B. Mariño: Speech Emotion Recognition Using Hidden Markov Models, Eurospeech 2001, Scandinavia.

K. Mase, "Recognition of facial expression from optical flow," IEICE Trans. E74(10), pp. 3474–3483, 1991.

Kenji Mase: AN APPLICATION OF OPTICAL FLOW EXTRACTION OF FACIAL EXPRESSION, MVA'9O IAPR Workshop on Machine Vision Applications Nov. 28-30.1990, Tokyo

Caifeng Shan, Shaogang Gong, Peter W. McOwan: Beyond Facial Expressions: Learning Human Emotion from Body Gestures, Department of Computer Science Queen Mary, University of London.

Vogt, T., André, E.: Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. IEEE International Conference on Multimedia & Expo (ICME 2005).