Automatic Facial Expression Recognition using Geometrical Features

A thesis

submitted in partial fulfillment of the requirement for the Degree of

Master of Computer Science and Engineering

of

Jadavpur University

By

Tanmoy Banerjee

Registration No.: 128989 of 2014-15

Examination Roll No.: M4CSE1602

Under the Guidance of

Prof. Debotosh Bhattacharjee

Department of Computer Science and Engineering

Jadavpur University, Kolkata-700032

India

2016

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

Certificate of Recommendation

This is to certify that the dissertation entitled "Automatic Facial Expression Recognition using Geometrical Features " has been carried out by Tanmoy Banerjee (University Registration No.: 128989 of 2014-15, Examination Roll No.: M4CSE1602) under my guidance and supervision and be accepted in partial fulfillment of the requirement for the Degree of Master of Computer Science and Engineering. The research results presented in the thesis have not been included in any other paper submitted for the award of any degree in any other University or Institute.

…………………………………………………

Prof. Debotosh Bhattacharjee (Thesis Supervisor)
Department of Computer Science and Engineering
Jadavpur University, Kolkata-32

Countersigned

…………………………………………………….

Prof. Debesh Kumar Das
Head, Department of Computer Science and Engineering,
Jadavpur University, Kolkata-32.

…………………………………………………….

Prof. Sivaji Bandyopadhyay
Dean, Faculty of Engineering and Technology,
Jadavpur University, Kolkata-32

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

Certificate of Approval*

This is to certify that the thesis entitled “Automatic Facial Expression Recognition using Geometrical Features” is a bonafide record of work carried out by Tanmoy Banerjee in partial fulfillment of the requirements for the award of the degree of Master of Computer Science and Engineering in the Department of Computer Science and Engineering, Jadavpur University during the period of June 2014 to May 2015. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, the opinion expressed or conclusion drawn therein but approve the thesis only for the purpose for which it has been submitted.

……………………………………………………………………………..
Signature of Examiner 1
Date:

……………………………………………………………………………..
Signature of Examiner 2
Date:

*Only in case the thesis is approved

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

Declaration of Originality and Compliance of Academic Ethics

I at this moment declare that this thesis entitled "Automatic Facial Expression Recognition using Geometrical Features" contains literature survey and original research work by the undersigned candidate, as part of his Degree of Master of Computer Science & Engineering.

All information have been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name: Tanmoy Banerjee
Registration No: 128989 of 2013-14
Exam Roll No.: M4CSE1602
Thesis Title: Automatic Facial Expression Recognition using Geometrical Features

…..……………………………………….

Signature with Date

Acknowledgement

I would like to start by thanking the holy trinity for helping me deploy all the right resources and for shaping me into a better human being.

I would like to express my deepest gratitude to my advisor, **Dr. Debotosh Bhattacharjee**, Professor, Department of Computer Science and Engineering, Jadavpur University for his admirable guidance, care, patience and for providing me with an excellent atmosphere for doing research. Our numerous scientific discussions and his many constructive comments have greatly improved this work.

I would like to thank **Prof. Debesh Kumar Das**, Head, Department of Computer Science and Engineering, Jadavpur University also **Prof. Sivaji Bandyopadhyay**, Dean, Faculty of Engineering and Technology, Jadavpur University for providing me with moral support at times of need.

I am highly grateful to Miss Kaushiki Roy, who took the pivotal role in helping me in understanding MATLAB tools and the functions.

Among the seniors, I would like to thank Mr. Sourav Pramanik, Mr. Ratnadeep Dey, and Mr. Koushik Dutta, who have given me valuable suggestions both inside and outside the department in successfully completing this thesis work.

I am also thankful to Mr. Nirmal Das, Miss Reema Sarkar, who were always present to motivate me. Without all of them, it would surely be a very lonely lab.

Most importantly none of this would have been possible without the love and support of my family. I extend my thanks to my mother whose forbearance and wholehearted support helped this endeavor succeed. Finally, I would also like to mention my sister without whose aegis this work could have never been able to flourish.

This thesis would not have been completed without the inspiration and support of a number of wonderful individuals — my thanks and appreciation to all of them for being part of this journey and making this thesis possible.

……………………………………………..

Tanmoy Banerjee

Registration No: 128989 of 2014-15

Exam Roll No.: M4CSE1501

Department of Computer Science & Engineering

Jadavpur University

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Just as a picture is worth a thousand words, our faces can express a wealth of information. Every person express himself or herself with different facial expression. Now a day's facial expression recognition is a popular research topic due to its real life applications in a wide variety of areas. The study of human facial expressions is one of the most challenging domains in pattern research community. Each facial expression is generated by non-rigid object deformations and these deformations are person-dependent. The goal of the project was to design and implement a system for automatic recognition of human facial expression. The results of the project are of a great importance for a broad area of applications that relate to both research and applied topics.

Almost in every automatic expression analysis system, researcher used lots of different approaches for recognizing some small set of expressions: Happiness, Surprise, Anger, Fear and Disgust [1, 2]. In our system we also tried to recognize this set of expression by using some geometrical features. Expressions are changed for every different emotions by changing two or three discrete facial features. Sometimes this emotion change can be determined by changing one discrete feature, such as anger can be represent by showing tightening lips, surprise can be represented by raising eyebrow [3] and so on.

Automatic recognition of facial expressions is a process primarily based on analysis of permanent and transient features of the face, which can be only assessed with errors of some degree. The expression recognition model is oriented on the specification of Facial Action Coding System (FACS) of Ekman and Friesen [4]. The FACS is a human-observer-based system designed to describe subtle changes in facial features. This Facial Action Coding system is discussed later in this thesis.

In a Human Computer Interface, the expression has a great importance as potential input, and this is especially true in voice-activated control systems. Experiments show when people are speaking, 55% of communication happens via expression whereas only 7% happens via spoken words.

In the graphic computer area, a facial expression derived from real images can be used to animate synthetic characters. This technique is useful in video telephony with limited bandwidth. Instead of transmitting the video, we can just send the facial expression sequence, with the help of that the original video can be reconstructed. Another application is in the movie industry where they use this technique to produce high-quality computer animation.

Automatic recognition of expressions very challenging experiment, as there are many underlying factors that can affect the mode of facial expressions. One factor is the presence of subject differences like the texture of the skin, hair style, age, gender, and ethnicity. All these factors have a huge influence on the expression of the face. In addition to the differences in appearance, there might be differences in expressiveness; that is, individuals express their emotion differently from each other. Expressiveness specifically refers to the degree of facial plasticity, morphology, frequency of intense expression, and the overall rate of expression. These individual differences are a very important aspect of identity, and characteristic facial actions may be used as a biometric property to improve the accuracy of the facial recognition algorithms.

1.1 Applications:

Automatically detecting facial expressions has become an increasingly important research area in machine learning, computer vision and behavioral research area. Not only in research area, but also it can be used for many applications such as security [20], driver safety [24], health-care [27] and human-computer-interaction [23].

In a Human Computer Interface, the expression is a great potential input, and this is especially true in voice-activated control systems. Experiments in [32] show when people are speaking, 55% of communication happens via expression whereas only 7% happens via spoken words. This implies an FER module can markedly improve the performance of such systems. Customers' facial expressions can also be collected by service providers as implicit user feedback to improve their service. Compared to a conventional questionnaire-based method, this should be more reliable and furthermore, has virtually no cost. In the graphic computer area, facial expression estimated from real images can be used to animate synthetic characters [13].This technique is useful in video telephony where bandwidth is limited. Instead of transmitting the video, we can just send the facial expression sequence, using which the original video can be reconstructed. Another application of this technique is in the movie industry where they use this to produce high quality computer animation.

There are some other possible applications, including emotion surveillance for employees in high work intensity industry (*e.g.*, long distance drivers), pain assessment, image and video database management and searching, lie detection and so on.

1.2 Goal

The primary goal of this project is to design a fully automatic human emotion recognition system based on the analysis of facial expressions from still pictures. The use of different machine learning techniques is to be investigated for testing against the performance achieved in the current context of the project.

The facial expression recognition system consists of a set of processing components that perform processing in the input video sequence. All the components are organized on distinct processing layers that interact for extracting the subtlety of human emotion and paralinguistic communication at different stages of the analysis.

The main component of the system stands for a framework that handles the communication with the supported processing components. Each component is designed in such a way so as to comply with the communication rules of the framework through well-defined interfaces. For experimental purposes, multiple components having the same processing goal can be managed at the same time and parallel processing are possible through a multithreading framework implementation.

The following steps for building a facial expression recognition system are taken into consideration and detailed in the thesis:

-A throughout literature survey that aims at describing the recent research on the realization of facial expression recognition systems.

- The presentation of the model, including the techniques, algorithms and adaptations made for an efficient determination of an automatic facial expression recognizer.
- The implementation of the models described and the tools, programming languages and strategies used for integrating all the data processing components into a single automatic system.

- The presentation of experimental setups for conducting a comprehensive series of tests on the algorithms detailed in the thesis. The presentation also includes the results of the tests that show the performance achieved by the models designed.
-The discussions on the current approach and the performance achieved by testing the facial expression recognition models.

All the processing steps are described in the current report. The final part contains the experiments run by using different functional approach.

Chapter 2

Background and Survey of Related Work

The significance of facial expression in social interaction and social intelligence is greatly recognized. Facial expression analysis has been an active and interesting research topic since 19th century.

The first automatic facial expression recognition system was introduced in 1978 by Suwa*et al.* [13]. This system attempts to analyze facial expressions by analysing the motion of 20 identified spots on an image sequence. Since then, a lot of work has been done in this domain. Various computer systems have been made to help us understand and use this natural form of human communication. This chapter reviews the state of the art of what has been done in processing and understanding facial expression. When building an FER system, these prime issues must be considered: face detection and alignment, image normalization, feature extraction, and classification. Most of the recent work in FER is based on methods that implement these steps sequentially and independently. Before exploring what has been done in literature for implementing these steps, we will briefly describe the problem space for facial expression analysis.

2.1 Problem Space for Facial Expression Analysis

2.1.1 Level of Description

In general there are two types of method to describe facial expression.

2.1.1.1 Facial Action Coding System (FACS)

The facial action coding system [14] is a human-observer-based system widely applied in psychology to describe subtle changes in facial features. FACS consists of 44 action units which are related to contraction of a specific set of facial muscles (Fig.2.2).Some of the action units are shown in Fig.2.2 [16]. Conventional, FACS code is manually labelled by trained observers while viewing videotaped facial behaviour in slow motion. In recent years, some attempts have been made to do this automatically [19]. The advantage of FACS is its ability to capture the subtlety of facial expression, however FACS itself is purely descriptive and includes no inferential labels. That means in order to get the emotion estimation, the FACS code needs to be converted into the Emotional Facial Action System (EMFACS [18]) or similar systems.

Facial expression described by Action Units can be then analyzed on the semantic level in order to find the meaning of particular actions. According to the Ekman's theory [15], there are six basic emotion expressions that are universal for people of different nations and cultures. Those basic emotions are joy, sadness, anger, fear, disgust and surprise (Fig. 2.1) [17].

Fig 2.1: Six Universal Facial Expression

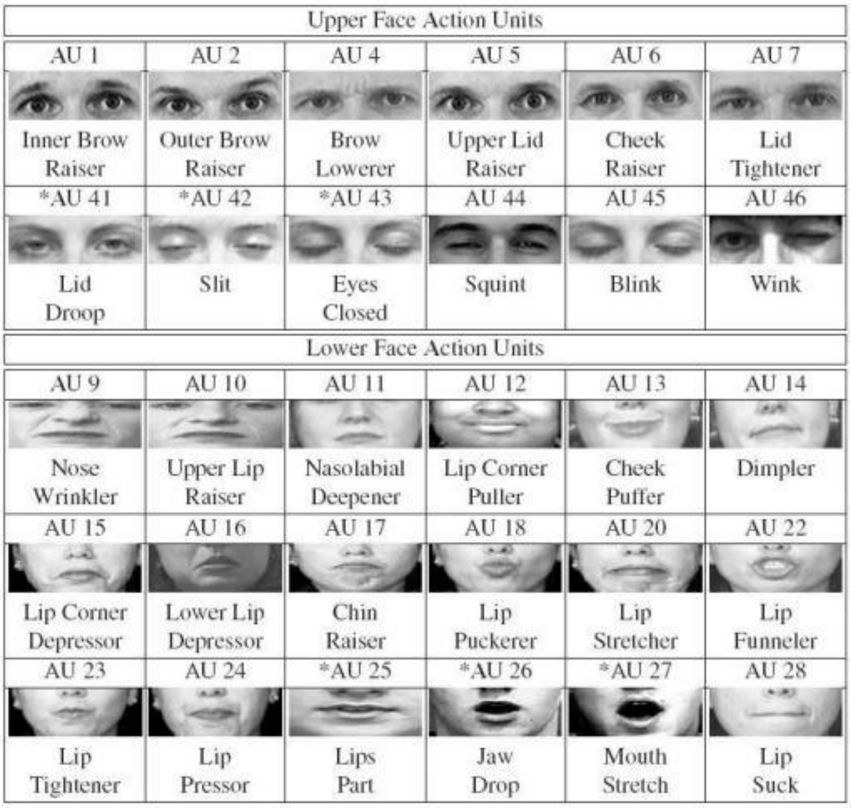


Fig 2.2: FACS Action Unit

2.1.1.2 Prototypic Emotional Expression

Instead of narrating the detailed facial features, most FER systems attempt to identify a small set of prototypic emotional expressions. The most widely-used set is perhaps human universal facial expressions of emotion which consists of six basic expression categories that have been shown to be recognizable across cultures (Fig.2.3).

These expressions, or facial configurations have been recognized in people from widely divergent cultural and social backgrounds [28], and they have been observed even in the faces of individuals born deaf and blind [29].

These 6 basic emotions, *i.e.*, disgust, fear, joy, surprise, sadness and anger plus “neutral” which means no facial expression are considered in this work. Given a facial image, our system either works as a conventional classifier to determine the most likely emotion or estimates the weights (or possibility) of each emotion as a fuzzy classifier does.

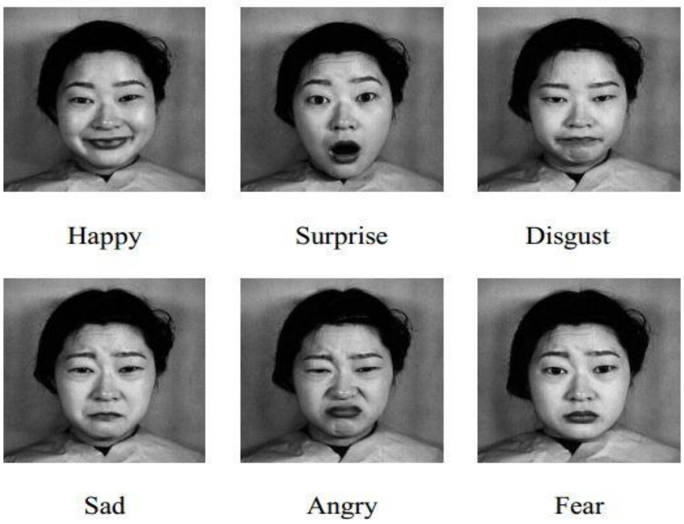


Fig 2.3: Basic Facial Expression (JAFE)

2.2 System Structure

FER can be considered as a special face recognition system or a module of a face recognition system. So it should be instructive to look at the general architecture of a face recognition system. Normally, it consists of four components as depicted in Fig.2.4.

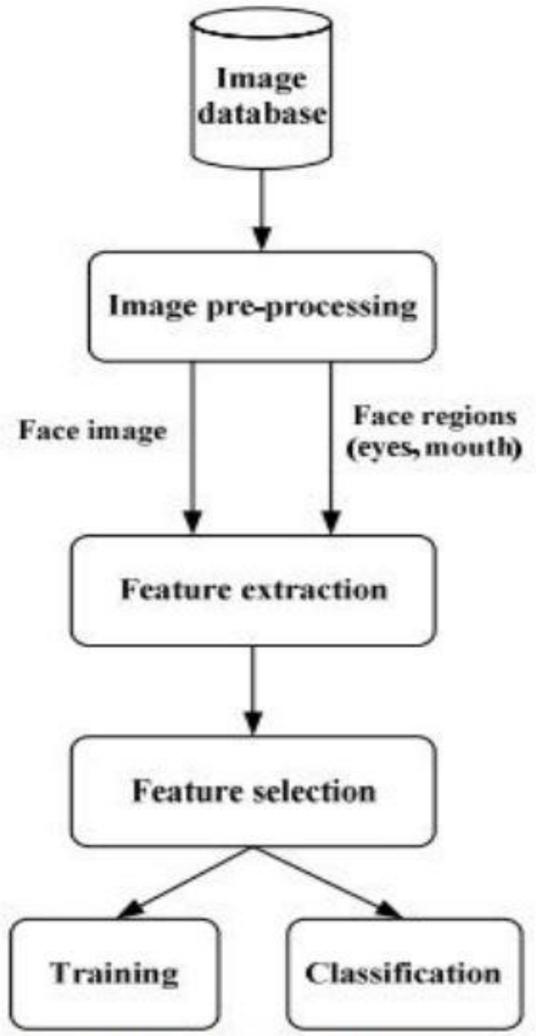


Fig 2.4: Block Diagram of Proposed System

Face detection finds the face areas in the input image. If the input is a video, to be more efficient and also to achieve better robustness, face detection is only performed on key frames and a tracking algorithm is applied on interval frames. Face alignment is very similar to detection, but it is aimed at achieving a more accurate localization. In this step, a set of facial landmarks (facial components), such as eyes, brows and nose, or the facial contour are located; based on that, the face image is rotated, chopped, resized and even warped, this is called geometrical normalization. Usually the face is further normalized with respect to photometrical properties such as illumination and gray scale.

Feature extraction is performed on a normalized face to provide effective information that should be useful for recognizing and classifying labels in which there is interest, such as identity, gender, or expression.

2.3 Face Detection

Face detection is the first step in face recognition. It has a major influence on the performance of the entire system [30]. Several cues can be used for face detection, for example, skin color, motion (for videos), facial/head shape, and facial appearance. Most successful face detection algorithms are based on only appearance [30]. This may be because appearance-based algorithms avoid difficulties in modelling 3D structures of faces. However, the variations of 3D structures due to facial expression and head pose actually heavily affect the facial appearance and make the face/non-face boundary highly complex [31]. To deal with this, a vast arrange of methods have been proposed since the 1990s.

Turk and Pentland [47] describe a detection system based on eigen decomposition which is also known as principal component analysis (PCA). In their method, an image is represented by an average face plus a set of weighted "eigenfaces". Whereas only the face images are considered in eigenface, Sung and Poggio [42] also consider the distribution of non-face images and apply Bayes' rule to obtain a likelihood estimation. Rowley *et al.* [32] use neural networks and Osuna*et al.* [33] trained a Kernel Support Vector Machine to classify face and non-face images. In these systems, a bootstrap algorithm is iteratively used to collect meaningful examples for retraining the detector. Schneiderman and Kanade [35] use AdaBoost learning to construct a classifier based on wavelet representation of the image. This method is computationally expensive because of the wavelet transformation. To overcome this problem, Viola and Jones [36] replace wavelets with Haar features, which can be computed very efficiently [34] [40]. Their system is the first real-time frontal-view face detector [37].

Under Viola's framework, some improvements have been proposed. Lienhart*et al.* [52] use rotated Haar features to deal with in-plane rotation. Li *et al.* [51] [50] [52] propose a multiview face detection system which can also handle out-of-plane rotation using a detector-pyramid. In the following sections, we will describe two face detection algorithms: Eigenface is one of the simplest methods and Viola's framework may be the most successful one. AdaBoost learning is an important component in Viola's framework and this algorithm will also be useful in the feature extraction module, so our presentation focuses on this part.

2.3.1 Eigenface and Template Matching

Eigenface assumes the face image $x = (x1, x2, \ldots, xN)$ is amenable to a multivariate normal distribution from which the training images are identically independently drawn. This distribution can be described by the following probability density function:

$$f(x_1, x_2, x_3 \ldots, x_N) = \left(\frac{1}{2\pi^{N/2}|E|^{1/2}}\right)^{exp(-\frac{1}{2}(x-\mu)^2 E^{-1}(x-\mu))}$$

(2.1)

Where Σ is the covariance matrix and μ is the expectation of *x*.
Eigenface decomposes Σ using Eigen decomposition as

$$E = USU^T \qquad (2.2)$$

where *U* is a unitary matrix and *S* = *diag*(S_1^2, S_2^2, . . . , S_N^2) is a diagonal matrix with allelements non-negative. Each column of *U*, U_i, is called an Eigenface. A face image *x* can be represented by μ and U_i as

$$x = \mu + \sum_i a_i U_i. \qquad (2.3)$$

It can be shown that $\frac{a_i}{s_i}$ are i.i.d. standard normal variables. So the probability density function of *x* is:

$$f(x) = \prod_i - \frac{1}{(2\pi)^{1/2}} \exp(\frac{1}{2} \frac{a_i^2}{s_i^2}) = \frac{1}{(2\pi)^{N/2}} \exp(-\frac{1}{2} \sum_i \frac{a_i^2}{s_i^2}) \qquad (2.4)$$

Equation (2.4) can be used as a probability estimate and we can define a distance measure according to

$$D = \sum_i \frac{a_i^2}{s_i^2} \qquad (2.5)$$

which is called normalized Euclidean distance. A large D implies a small probability of being a face image and vice versa. Based on (2.5) Turk and Pentland [47] built a face detection system. Sung and Poggio's paper [42] used a similar idea, they assume images are produced by a mixture of Gaussian models: a face image Gaussian and a non-face image Gaussian. So they also estimate the probability of x of being a non-face image $f_*(x)$ and the final decision is made using a Bayesian classifier.

If we further assume E is an identity matrix, (2.5) degenerates into Euclidean distance which means the probability density function is controlled by, $|x - \mu|^2$, the variation of the image from the average face μ. This gives the simplest detection algorithm: template matching, *i.e.*, finding a "face template" μ, and then for each x determine whether it is a face image by thresholding $|x - \mu|$.

2.3.2 Viola's Framework

Almost all the state of the art face detection systems are developed upon Viola's Frame-work, based on AdaBoost and Haar features. The Viola–Jones object detection framework is the first object detection framework to provide competitive object detection rates in real-time proposed in 2001 by Paul Viola and Michael Jones. The philosophy of AdaBoost is that if it is hard to find a good (strong) classifier directly, we can construct a lot of poor quality ones (weak classifiers) $h_m(x)$, $m = 1, 2, \ldots \ldots \ldots, M$ and use the combination of them to form a strong classifier [39]:

$$H_M(x) = \frac{\sum_m a_m h_m(x)}{\sum_m a_m} \tag{2.6}$$

where $a_m \geq 0$ are the combining coefficients. In the discrete version $h_m(x)$ gives either -1 or 1 whereas in the real version, the output can be a real number. Because we need a lot of weak classifiers, $h_m(x)$ is normally chose to have a simple form (so it is easy to construct). In Viola and Jones's work [36] [37], they use threshold classifiers, each of which works on one feature selected from an over-complete set of Haar wavelet-like features. We'll first introduce Haar Wavelet, then give the iterative algorithm of AdaBoost.

2.3.2.1 Haar Wavelet

Wavelet analysis is a tool for signal processing which can perform local analysis. It is useful in face detection because we want to focus on a localized area of the image and find whether it contains a face. Haar wavelet is the first known wavelet, and probably the simplest one. A one-dimensional Haar wavelet function is just a step function, as shown in Fig.2.5 (a).

In image processing, two-dimensional wavelets are used which look like the ones in Fig.2.5 (b). The functions in Fig.2.5 (b) don't satisfy some conditions of a wavelet (because we don't need those in our application), so strictly speaking they are "Haar-like features".

An important property of Haar wavelet-like features is that they can be computed efficiently using an integral image [36]. The integral image $II(x, y)$ of image $I(x, y)$ is defined by:

$$II(x, y) = \sum_{x'<x, y'<y} I(x', y') \tag{2.7}$$

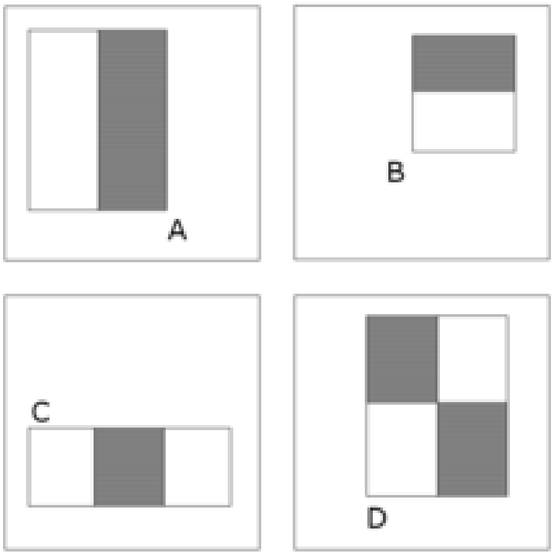


Fig 2.5 Feature used by Viola Jones

This can be computed in one pass over the original image as follows [30]:

$$S(x,y) \quad = \quad S(x,y-1)+I(x,y) \tag{2.8}$$

where $S(x, y)$ is the cumulative sum of the xth row. Apparently $S(x, 0) = 0$ and $II(0, y) = 0$. Any rectangular sum in an image can be expressed in terms of its integral image as illustrated in Fig.2.6.

The sum of the pixels within rectangle D in Fig.2.6 can be computed using the integral image value on points a, b, c, d as follows [36]:

$$II(x2, y2) = \sum_{x<x2, y<y2} I(x, y) = A + B + C + D \tag{2.9}$$

$$II(x2, y1) = \sum_{x<x2,y<y1} I(x, y) = A + C \quad (2.10)$$

$$II(x1, y2) = \sum_{x<x1,y<y2} I(x, y) = A + B \quad (2.11)$$

$$II(x1, y1) = \sum_{x<x2,y<y2} I(x, y) = A \quad (2.12)$$

$$D = II(x2, y2) + II(x1, y1) - II(x2, y1) - II(x1, y2) \quad (2.13)$$

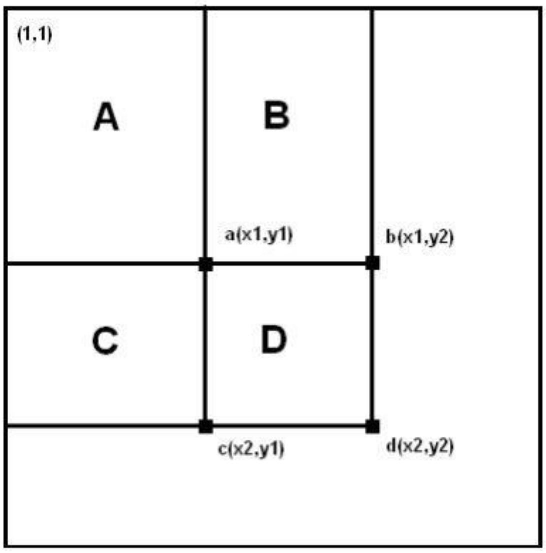


Fig 2.6: Compute rectangular sum using integral image

2.3.2.2 AdaBoost Algorithm

AdaBoost is an algorithm for constructing a "strong" classifier as linear combination of "weak" classifiers ht(x). The ht(x) can be thought of as one feature.

1）Given example images: $(x1, y1), \ldots, (xL\ , yL)$, where $y_i \in \{1, 0\}$ indicates positive or negative examples; $g_j x_i$ is the jth Haar-Like feature of ith example xi.

2）Initialize weights

$$w_{i,j} = \begin{cases} 0.5/_{m} & i \leq m \\ 0.5/_{n} & otherwise \end{cases} \quad (2.14)$$

Where m, n are the number of positive or negative examples respectively. $L = m + n$

3）For $t = 1 \ldots T$

a. Normalize the weights

$$w_{t,j} = w_{t,j} / \sum_{j=1}^{L} w_{t,j} \quad (2.15)$$

b. For each feature j, train a weak classifier h_j, and evaluate its error ε_j with respect to W_t,

$$\varepsilon_j = \sum_{i=1}^{L} w_{i,j} \ \left| h_j(x_i) - y_i \right| \quad (2.16)$$

$$h_j(x) = \begin{cases} 1 & p_j g_j(x) < p_j \theta_j, \\ 0 & otheriwise \end{cases} \quad (2.17)$$

c. .Choose the classifier h_t with the lowest error ε_t

d. Update the weights $w_{t+1,i} = w_{t,i}\, \beta_t^{1-ei}$,where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise and $\beta_t = \varepsilon_t / (1 - \varepsilon_t)$

4）Final Classifier

$$H(x) = \begin{cases} 1, & \sum_{t=1}^{r} a_t h_t(x) \geq 0.5 \sum_{t=1}^{r} a_t \\ 0, & x \geq 0 \end{cases}$$

Where $\alpha_t = \log(1/\beta_t)$

2.3.2.3 Cascading Classifier

The Viola and Jones face detection algorithm eliminates face candidates quickly using a cascade of stages. The cascade eliminates candidates by making stricter requirements in each stage with later stages being much more difficult for a candidate to pass. Candidates exit the cascade if they pass all stages or fail any stage. A face is detected if a candidate passes all stages. This process is shown in “Fig. 2.7”.

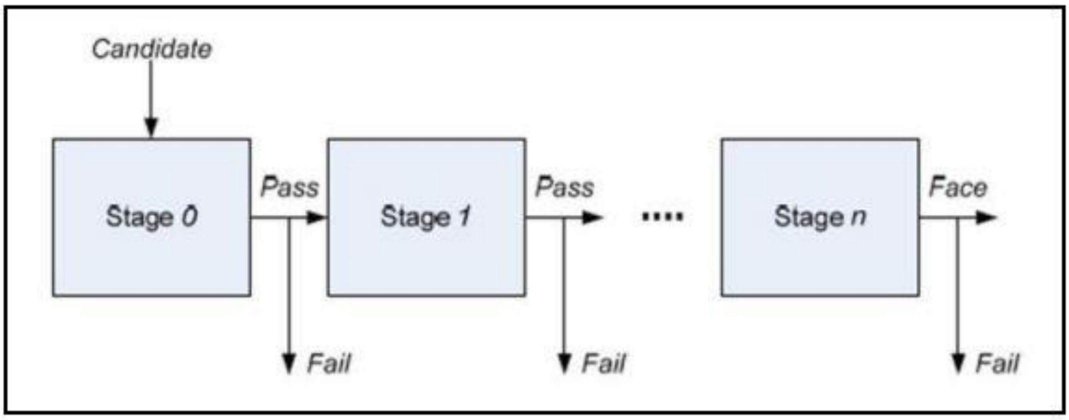


Fig 2.7: Cascade of stages. Candidate must pass all stages in the cascade to be concluded as a face.

2.4 Face Alignment

The aim of face alignment is to achieve more accurate localization of the face, and usually the facial components (feature points). A typical result of face detection and face alignment is compared in Fig.2.7. As we can see, detection considers the images in terms of areas whereas alignment has a precision of pixels.

Various algorithms have been proposed since 1990's. Gu *et al.* [45] use histogram information to localize mouth corners and eye corners. In [44], Ian and Marian implement a preliminary alignment system using Gabor filter to detect pupils and philtrum. Curvefitting algorithms, especially Active Shape Model [43] and its offspring may be the most successful alignment methods nowadays. Cootes *et al.* [43] [45] [41] propose Active Shape Model and apply it to face image. After that, ASM is widely used in face image processing; a great amount of effort has been made to improve its speed, accuracy and robustness. In [54] the Active Shape Model is combined with the Gabor filter; Li *et al.* [55] propose Direct Appearance Model; [56] and [53] enhance ASM by using 2-D local textures feature for local search.

This section will mainly focus on curve-fitting types of method, in particular Active Shape Model and its ancestor Active Contour Model.

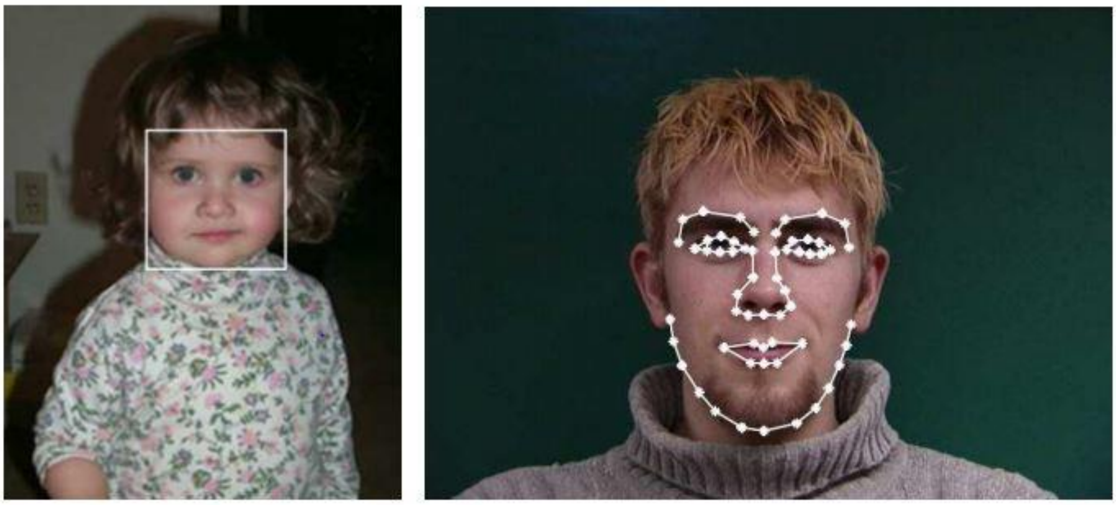


Fig 2.8: Detection vs. Alignment

2.5 Feature Extraction

"Feature extraction converts pixel data into a higher-level representation of shape, motion, color, texture, and spatial configuration of the face or its components. The extracted representation is used for subsequent classification. Feature extraction generally reduces the dimensionality of the input space. The reduction procedure should (ideally) retain essential information possessing high discrimination power and high stability" [58]. In the face recognition area, various features have been used.

The coefficients of Eigenface can be used as features and recently, an a to

$$x = \bar{x} + \sum_{i_1} \dots \dots \dots \sum_{i_N} a_{i_1} \dots \dots \dots a_{i_N} x_{i_1 \dots\dots i_N} \qquad (2.19)$$

Compared to Eigenface which ignores the label of images, Tensorface analyzes a face ensemble with respect to its underlying factors(labels): for example, identities, views, and illuminations. The "principal components" in this multilinear system are referred to as Tensorfaces which are shown in Fig.2.9. The Tensorface coefficients a_{i_k} can be used as features in a recognition task, and because the original image can be reproduced using 2.26, Tensorface coefficients can also be used for image synthesis where we first generate a set of coefficients, then use them to synthesize images.

Since Tensorface is shown to be a promising method in face recognition, some improvements have been proposed. [71] Proposes Multilinear Independent Component Analysis where they try to find the independent directions of variation. In [72] Shashua *et al.* introduce Non-Negative Tensor Factorization which is a generalization of Non-negative Matrix Factorization.

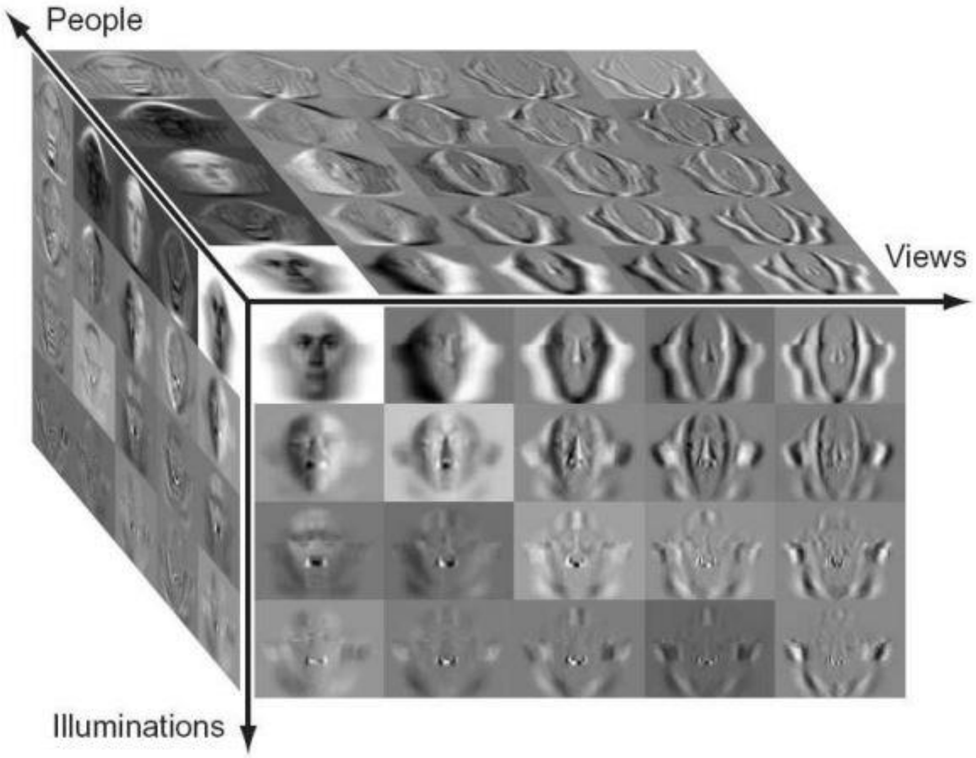


Figure 2.9: A partial visualization of TensorFaces bases for an ensemble of 2,700 facial images spanning 75 people, each imaged under 6 viewing and 6 illumination conditions [72]

2.5.1 Potential Net

Matsuno *et al.* [61] propose Potential Net to extract facial features. As shown in Fig.2.10, Potential Net is a two dimensional mesh of which nodes are connected to their four neighbors with springs, while the most exterior nodes are fixed to the frame of the Net. Similar to curve fitting, Potential Net considers two forces: each node in the mesh is driven by the external forces which come from the image gradient and the elastic forces of springs propagate local deformation throughout the Net. Eventually, equilibrium is reached, and the nodal displacements represent the overall pattern of facial image.

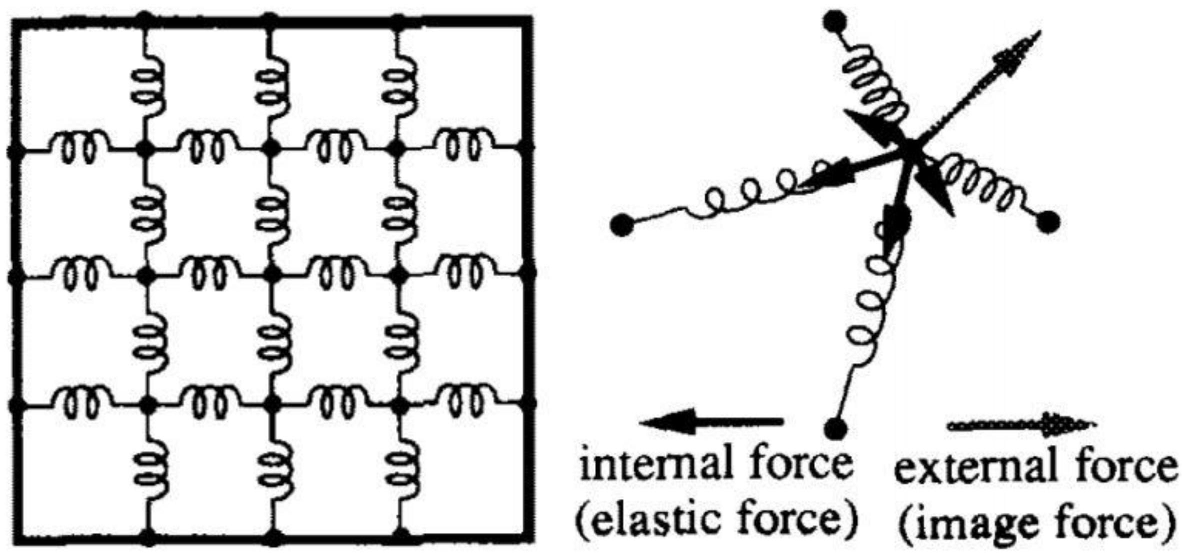


Figure 2.10: Structure of Potential Net and nodal deformation

2.5.2 Active Appearance Model

Cootes *et al.* propose Active Appearance Model [60] to represent objects which have deformable shape as well as diverse textures. In Active Appearance Model, the shape variation, texture variation and their interdependence are modeled using three linear systems respectively as depicted in Fig.2.11. The shape of the facial contour is coded into a shape vector V_{shape} by applying Active Shape Model, then the texture is warped to have identical shape. This “shape-free texture” is converted into a texture vector $V_{texture}$ using PCA, finally $V_{texture}$ is concatenated with V_{shape} and translated into AAM coefficient.

AAM coefficient has fairly small dimensionality, normally 30-40 depending on the training parameters; also it is intuitive and easy to visualize. It has some limitations too, the major one is that AAM is purely linear, which means it is incapable of modeling nonlinear variation of the facial image

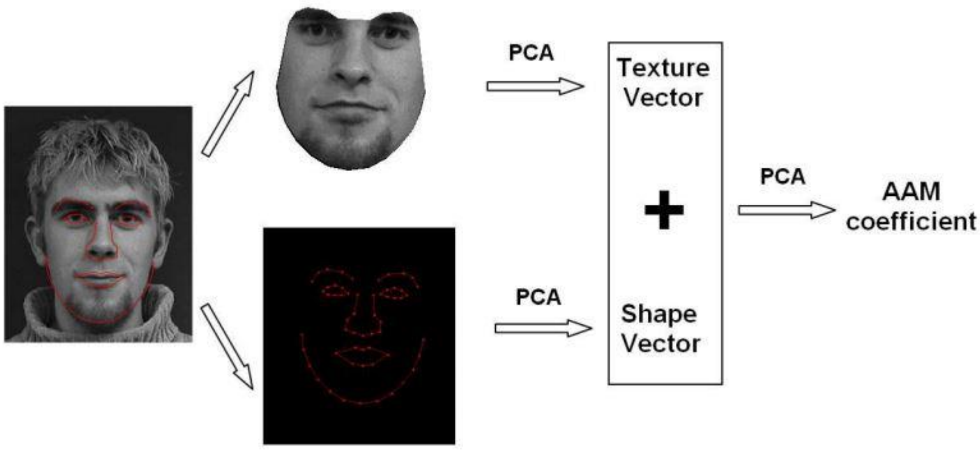


Figure 2.11: Active Appearance Model processing flow

2.6 Survey result on different dataset

The following table-2 list out the result derived from the respective techniques which are used for facial expression recognition. In the section, there are many useful methods such as Neural Network, Independent Component Analysis (ICA), PCA, Gabor Filter Energy and some of them are the combination of 2 or more method for better performance provides 85+% recognition rate in Face Recognition as well as Facial Expression Recognition. Table-2 covered list of most recent techniques and their relevant information like Dataset, Accuracy, Conclusion and Future work.

Sr. no	**Methods/Technique(s)/ Database**	**Result/Accuracy**	**Conclusion**	**Future Work**
1	Neural Network + Rough Contour Estimation Routine (RCER) [53] (Own Database	92.1% recognition Rate	In this paper, they describe radial basis function network (RBFN) and a multilayer perception (MLP) network.	–

2	Principal Component Analysis [18] (FACE94)	35% less computation time and 100% recognition	Useful where larger database and less computational time	They want to repeat their experiment on larger and different databases.
3	PCA + Eigenfaces [19] (CK, JAFFE)	83% Surprise in CK, 83% Happiness in JAFFE, Fear was the most confused expression	Compared with the facial expression recognition method based on the video sequence, the one based on the static image is more difficult due to the lack of temporal information.	Future work is to develop a facial expression recognition system, which combines body gestures of the user with user facial expressions.
4	2D Gabor filter [22] (Random Images)	12 Gabor Filter bank used to locate edge	Multichannel Gabor filtration scheme used for the detection of salient points and the extraction of texture features	They work on adding global and local colour histograms and parameters connected with the

			for image retrieval applications.	shapes of objects within images.
5	Local Gabor Filter + PCA + LDA [23] (JAFFE)	Obtained 97.33% recognition rate with the help of PCA+LDA features	They conclude that PCA+LDA features partially eliminate sensitivity of illumination.	_
6	PCA + AAM [24] (Image sequences from FG-NET consortium)	The performance ratios are 100 % for expression recognition from extracted faces,88% for expression recognition from frames and 88 % for the combined recognition.	The computational time and complexity was also very small. Improve the efficiency	Extend the work to identify the face and its expressions from 3D images
7	Gabor + SVM approach HAAR + Adaboost [25] (Cohn-Kanade database)	99.54% for Mouth AU in G+S and 82.81% with H+A	The Haar+Adaboost method achieved comparable accuracy to the	They are going to establish a large, publicly available AU

			Gabor+SVM method for AUs of the eye and brow regions, but it performed very poorly for AUs of the mouth.	database with singly-occurring AUs to facilitate future research.
8	Dynamic HAAR-like features [26] (CMU expression Database + Own Database)	Experiments on the CMU facial expression database and our own facial AU database showed that the proposed method has a promising performance.	They extracted dynamical HAAR-like features to capture the temporal information of facial AUs and expressions, and then further coded them into binary pattern features.	This method can be extended to video based face recognition.
9	2D appearance-based local approach + Radial Symmetry Transform [27] (JAFFE)	83% for expressions of happy and surprise and an accuracy of about 78% for	In the face, they use the eyebrow and mouth corners as main „anchor"	_

		expressions of anger and sad.	points. The system, based on a local approach, is able to detect partial occlusions.	
10	2D-LDA and SVM [29] (JAFFE)	The recognition rate of this method is 95.71% by using leave-one-out strategy and 94.13% by using cross-validation strategy.	They investigate various feature representation and expression classification schemes to recognize seven different facial expressions on the JAFFE database. Experimental results show that the proposed system using DWT, 2D-LDA and linear one-again-one	_

			SVMs outperforms others.	
11	2-D Gabor Filter [30] (Palm print database)	24 verification tests are carried out for testing the 12 sets of parameters on the two databases and their results in table 1 [30].	They proposed Gabor Filter to extract features in different angle, size and phase offset.	–

Table 2.1: Related work done in past

Chapter 3

System Architecture & Methodology Used

The goal of this project is to design a framework and implement the facial expression recognition system. On a basis of the extensive study of different approaches to the problem of face action representation, appropriate algorithms are selected for each stage of a system.

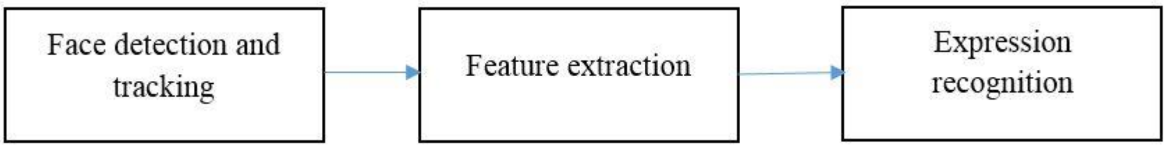


Fig 3.1: Method of Facial Expression Recognition

In our thesis we used 3 basic system to build the algorithm as shown in fig 3.1.

3.1. Face detection and tracking

3.2. Face expression representation

3.3. Expression recognition

System operates on both static images and image sequences. Static images are used in training and testing procedures but the interaction with a system is designed for video analysis.

This chapter includes the description of all three stages of a system. Algorithms used at each stage will be explained from theoretical aspect. Next, the implementation details will be mentioned and the system's behaviour will be illustrated.

3.1 Face detection and tracking

First part of my system is module for face detection and landmark localization in the image. Here we used to two approaches detecting face and its other component: Viola and Jones framework for face detection and Template matching algorithm.

3.1.1 Viola and Jones framework for face detection

The algorithm for face detection is based on work by Viola and Jones [13]. In this approach image is represented by a set of Haar-like features as we see in chapter 2. In previous chapter we see that how four features are described by Viola Jones (Fig 2.5). An important property of Haar wavelet-like features is that they can be computed efficiently using an integral image [35].

We see in 2nd chapter in the article number 2.3.2 how Viola Jones framework extract the feature and then compute the integral image.

Feature value is calculated by subtracting sum of the pixels covered by white rectangle from sum of pixels under grey rectangle. Two rectangular features detect contrast between two vertically or horizontally adjacent regions. Three rectangular features detect contrasted region placed between two similar regions and four rectangular features detect similar regions placed diagonally as shown as Fig. 3.2.

Fig 3.2: Haar-Like Feature detection by Viola and Jones

The method is widely used in area of face detection. However, it can be trained to detect any object. What is more, this algorithm is quick and efficient and could be used in real-time applications. In proposed system, the algorithm is applied for face, eyes and mouth localization with use of already existing functions from MATLAB.

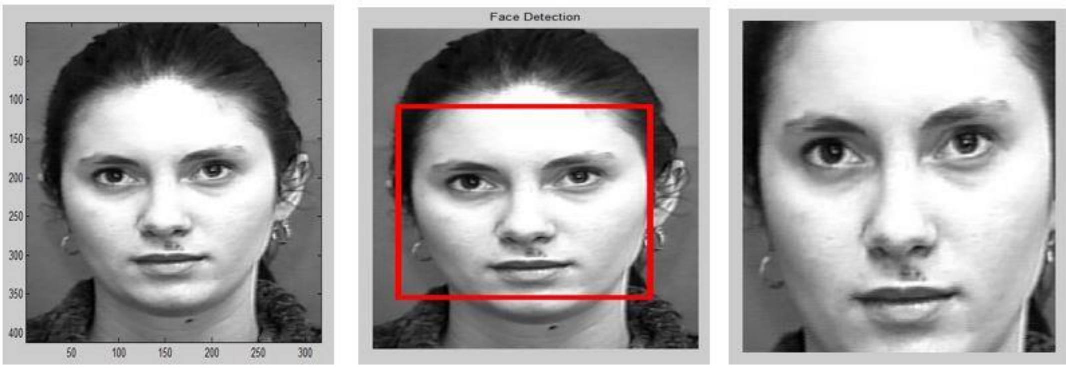


Fig 3.3: Face Detection using Viola Jones

3.1.2 Template Matching

Template matching is a technique in digital image processing for finding small parts of an image which match a template image. It can be used in manufacturing as a part of quality control, a way to navigate a mobile robot, or as a way to detect edges in images.

In chapter 2 article number 2.3.1, we discussed how template matching used in face detection. We applied the same procedure in our proposed system.

In our proposed system we used the template matching algorithm for detect eye detection. At first we took an image of eye (both left eye and right eye) as shown in the Fig 3.2. Then it was used to detect the eye part in face which was detected by Viola Jones algorithm.

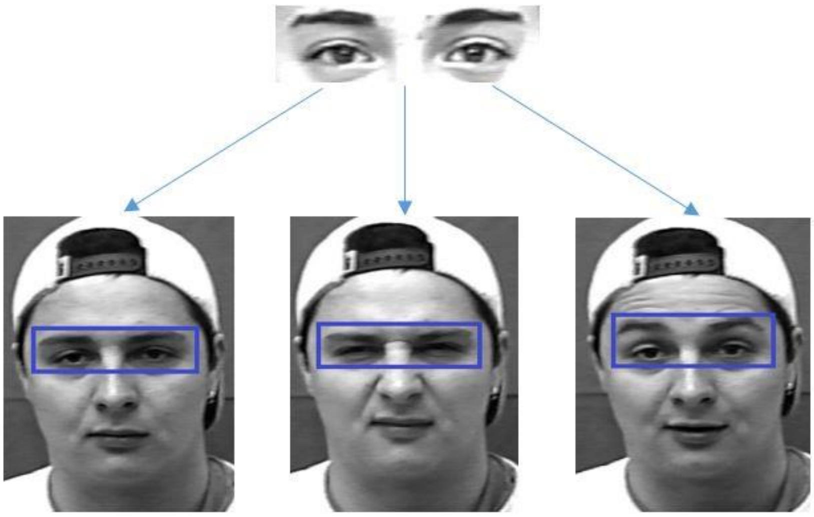


Fig 3.4: Eye detection using Template Matching

3.2 Feature Extraction

As mentioned in the previous chapter, two main approaches are used to describe the expression. Geometric or appearance features can be used either separately or in combination. In geometric feature based systems the face is represented by a set of facial points which are tracked. Deformations between neutral state and current frame are parameters of facial action. This approach requires reliable methods for point's detection and tracking which are difficult to obtain [10]. Appearance based methods measure the appearance changes which are mainly based on texture analysis. Although, Gabor filters are proved to be powerful in face expression analysis, they are time and memory consuming.

Contraction of the facial muscles produces changes in both the direction and magnitude of the motion on the skin surface and in the appearance of permanent and transient facial features. Examples of permanent features are the lips, eyes that have become permanent with age. We assume that the first frame is in a neutral expression. After initializing the templates of the permanent features in the first frame, both permanent and transient features can be tracked and detected in the whole image sequence regardless of the states of facial components. The tracking results show that our method is robust for tracking facial features [9] .

3.2.1 Multistate of Facial Components

After detection of facial components in near frontal images, we developed a multistate facial component models. The models are describe in table 3.1[1], which includes the permanent facial component (i.e., eye, eyebrows, mouth and lips) of frontal images.

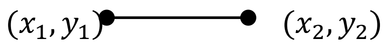
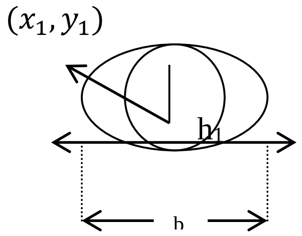
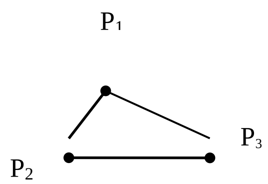
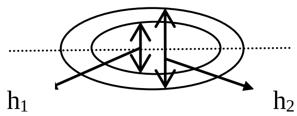
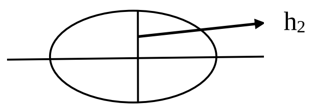
Component	State	Description/Feature
Eye	Closed	
	Open	
Eyebrow	Present	
Lip	Open	
	Close	

Table 3.1: Different component of facial expression

After tracking all the facial component, we implement some method on the components. From this table (3.1) we see that all component have different state for different expression. Now in this session we discussed about this.

3.2.2 Eye and Eyeball detection

Eye is most effective component of face. Most eye trackers developed so far are for open eyes and simply track the eye locations. However, for recognizing facial action units, we need to recognize the state of eyes, whether they are open or closed, and the parameters of an eye model, the location and radius of the iris, and the corners and height of the open eye. As shown in Table 1, the eye model consists of "open" and "closed".

In both cases of eye, some feature are same, but not all. In case of eyes are opened we find the circle of eyeball, but in case of closed eyes, we could not find any circle for eyeball. Closed eyes can only defined as a straight line (as shown in Table 3). In our proposed system when eyes are opened we first try to find the circle for the eye ball using circle detection algorithm as mentioned in below. Then we consider the circle with the maximum radius. As a result we find the eyeball as shown as fig 3.5.

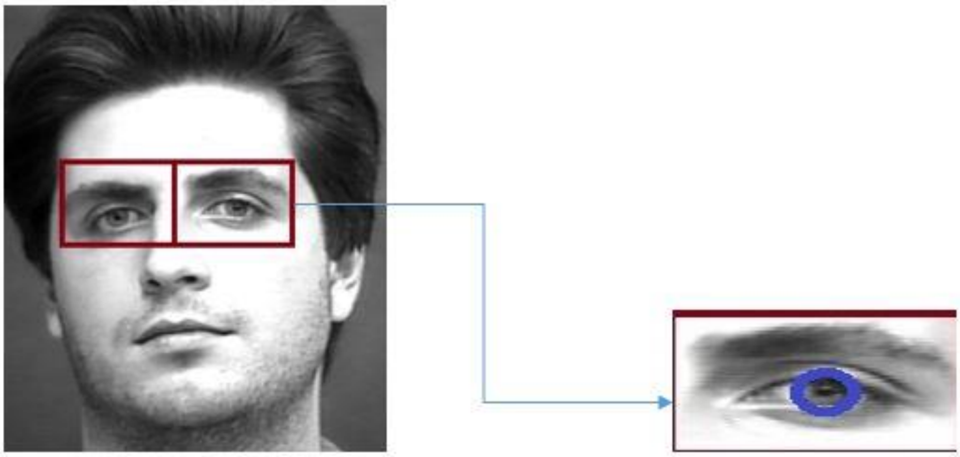


Fig 3.5: Eye and eyeball detection using Template Matching and Circle Hough Transform

3.2.2.1.1 Circle Hough Transform

The Circle Hough Transform (CHT) is a feature extraction technique for detecting circles. It is a specialization of Hough Transform. The purpose of the technique is to find circles in imperfect image inputs. The circle candidates are produced by "voting" in the Hough parameter space and then select the local maxima in a so-called accumulator matrix.

In a two dimensional space, a circle can by described by:

$$(x - a)^2 + (y - b)^2 = r^2 \tag{3.1}$$

Where (a,b) is the center of the circle, and r is the radius. If a 2D point (x,y) is fixed, then the parameters can be found according to (1). The parameter space would be three dimensional, (a, b, r). And all the parameters that satisfying (x, y) would lie on the surface of an inverted right-angled cone whose apex is at (x, y, 0). In the 3D space, the circle parameters can be identified by the intersection of many conic surfaces that are defined by points on the 2D circle. This process can be divided into two stages. The first stage is fixing radius then find the optimal centre of circles in a 2D parameter space. The second stage is to find the optimal radius in a one dimensional parameter space.

Find parameters with Known radius R

If the radius is fixed, then the parameter space would be reduced to 2D (the position of the circle center). For each point (x, y) on the original circle, it can define a circle centered at (x, y) with radius R according to (1). The intersection point of all such circles in the parameter space would be corresponding to the center point of the original circle.

Consider 4 points of circle in the original image (left). The circle Hough transform is shown in the right. Note that the radius is assumed to be known. For each (x,y) of the four points (white points) in the original image, it can define a circle in the Hough parameter space centered at (x, y) with radius r. An accumulator matrix is used for tracking the intersection point. In the parameter space, the voting number of points through which the circle passing would be increased by one. Then the local maxima point (the red point in the center in the right figure) can be found. The position (a, b) of the maxima would be the center of the original circle.

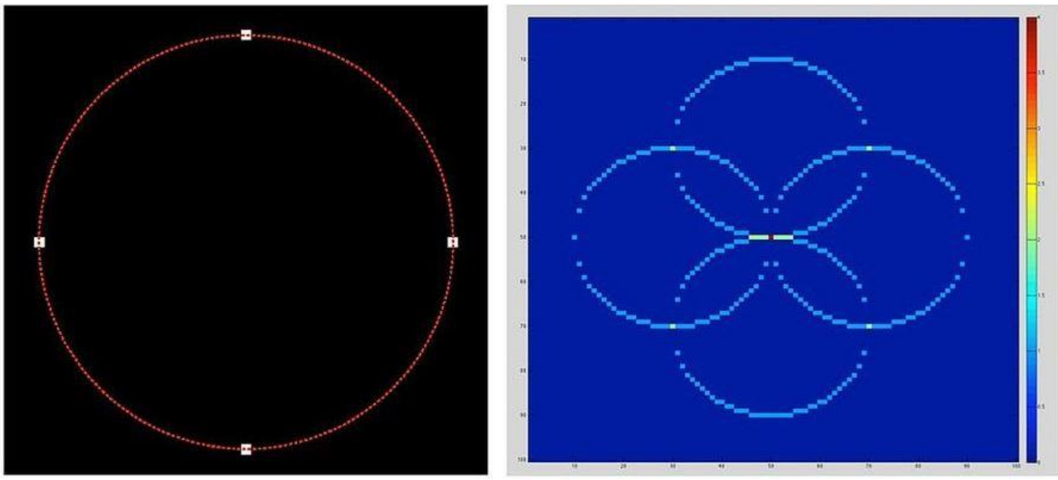


Fig 3.6: finding circle with unknown radius

In our proposed system when eyes are opened we first try to find the circle for the eye ball using circle detection algorithm as mentioned above. Then we consider the circle with the maximum radius. As a result we find the eyeball.

Whenever we get the eyeball we can detect the center easily. This center is used later for calculate the distance between eyebrow and eyeball for the different expression of face.

3.2.3 Eyebrow Detection:

Ding [48], claims the simplicity of the solutions since they are placed above the eye and color of the brows is either darker or lighter that of skin; indeed on color picture, with even illuminations and normal color sources, the eyebrows are holes in the skin map. However, if any of the mentioned conditions is not met, as suggested in [49] the precision of localization is given only by the existing contrast.

But in our system we work with grey-level images where detection of skin is very difficult. Therefore we used different technic for detecting eyebrow. Before this we were already calculate the center (x,y) of eyeball. Now we consider the template of eye, we can easily see that the change of intensity is maximum in the eyeball part. If we check further we see that change in the eyebrow portion. So when we get the center of eyeball, from this point we check the change of intensity vertically upwards. When we get the maximum changes we stop the algorithm. This center point is eyebrow.

This method localization algorithm is based on growing region procedure. The seed for growing algorithm must be point inside the brow.

3.7: Eyebrow detection

Fig 3.8: (I) Normal, (II) Brow Lower and (III) Brow Riser

3.2.3.1 Growing Region Property:

Region growing is a simple region-based image segmentation method. It is also classified as a pixel-based image segmentation method since it involves the selection of initial seed points.

This approach to segmentation examines neighboring pixels of initial seed points and determines whether the pixel neighbors should be added to the region. The process is iterated on, in the same manner as general data clustering algorithms.

The main goal of segmentation is to partition an image into regions. Some segmentation methods such as thresholding achieve this goal by looking for the boundaries between regions based on discontinuities in grayscale or color properties. Region-based segmentation is a technique for determining the region directly. The basic formulation is:

(a) $\bigcup_{i=1}^{n} R_i = R$
(b) R_i is connected region, $i = 1, 2, \dots \dots, n$.
(c) $R_i \cap R_j = \emptyset$ for all $i = 1, 2, \dots \dots, n$.
(d) $P(R_i) = \text{TRUE}$ for $i = 1, 2, \dots \dots, n$.
(e) $P(R_i \cup R_j)$ = FALSE for any adjacent region R_i and R_j.

$P(R_i)$ is a logical predicate defined over the points in set R_i and $\emptyset$ is the null set.

(a) Means that the segmentation must be complete; that is, every pixel must be in a region.
(b) Requires that points in a region must be connected in some predefined sense.
(c) Indicates that the regions must be disjoint.
(d) Deals with the properties that must be satisfied by the pixels in a segmented region. For example $P(R_i) = \text{TRUE}$ if all pixels in R_i have the same grayscale.
(e) Indicates that region R_i and R_j are different in the sense of predicate P.

3.2.4 Lips:

A three-state lip model is used for tracking and modelling lip features. As shown in Table 3.1, we classify the mouth states into open, closed, and tightly closed. Different lip templates are used to obtain the lip contours. Currently, we use the same template for open and closed mouth.

In our system after detecting the face we divided the frame in 3 segment. First one is upper version of face. From this we can detect the upper part of face like eye, eye brow etc. Second segment is detect the nasal part. Third and last one is used to detect the lips.

3.2.4.1 Segmentations:

In computer vision, image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyse. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics.

The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image (see edge detection). Each of the pixels in a region are similar with respect to some characteristic or computed property, such as colour, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristics. When applied to a stack of images, typical in medical imaging, the resulting contours after image segmentation can be used to create 3D reconstructions with the help of interpolation algorithms like marching cubes.

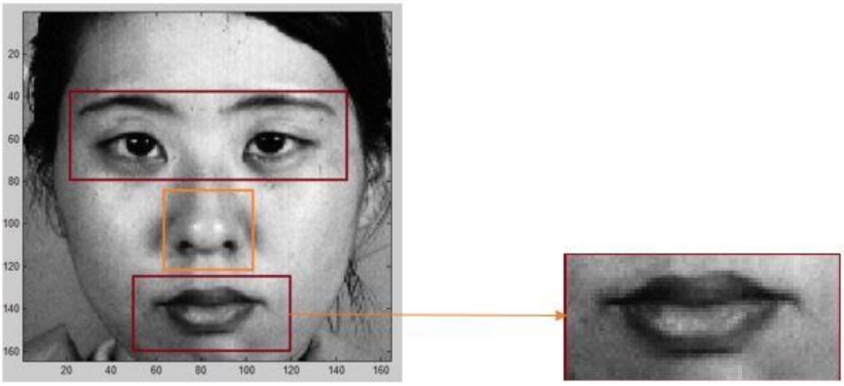


Fig 3.9: Lip extraction by segmenting

3.3 Expression recognition

A facial expression is one or more motions or positions of the muscles beneath the skin of the face. According to one set of controversial theories, these movements convey the emotional state of an individual to observers. Facial expressions are a form of nonverbal communication.

In our system after detecting several part of facial component, we tried to recognize the expression. In previous section we described about six basic facial expressions. Now we tried to recognize this using several methods. When we used the feature extraction method we used Edge detection algorithm and also for further experiment. As example, when we get the eyeball we also detect the edge of eyeball. After that when we check for maximum intensity change vertically for eyebrow detection this Edge detection is more useful for correct result. The basic theory of edge detection is used given below.

3.3.1 Edge Detection Algorithm

Edge detection is the name for a set of mathematical methods which aim at identifying points in a digital image at which the image brightness changes sharply or, more formally, has discontinuities. The points at which image brightness changes sharply are typically organized into a set of curved line segments termed *edges*. The same problem of finding discontinuities in 1D signals is known as step detection and the problem of finding signal discontinuities over time is known as change detection. Edge detection is a fundamental tool in image processing, machine vision and computer vision, particularly in the areas of feature detection and feature extraction.

The purpose of detecting sharp changes in image brightness is to capture important events and changes in properties of the world. It can be shown that under rather general assumptions for an image formation model, discontinuities in image brightness are likely to correspond to:

- discontinuities in depth,
- discontinuities in surface orientation,
- changes in material properties and
- Variations in scene illumination.

In the ideal case, the result of applying an edge detector to an image may lead to a set of connected curves that indicate the boundaries of objects, the boundaries of surface markings as well as curves that correspond to discontinuities in surface orientation. Thus, applying an edge detection algorithm to an image may significantly reduce the amount of data to be processed and may therefore filter out information that may be regarded as less relevant, while preserving the important structural properties of an image. If the edge detection step is successful, the subsequent task of interpreting the information contents in the original image may therefore be substantially simplified. However, it is not always possible to obtain such ideal edges from real life images of moderate complexity.

Edges extracted from non-trivial images are often hampered by *fragmentation*, meaning that the edge curves are not connected, missing edge segments as well as *false edges* not corresponding to interesting phenomena in the image – thus complicating the subsequent task of interpreting the image data.

Edge detection is one of the fundamental steps in image processing, image analysis, image pattern recognition, and computer vision techniques.
There are several technics used for edge detection, like Roberts, Prewitt, Sobel, Canny and so on (shown in table 3.2).

Methods	Descriptions
Canny	Finds edges by looking for local maxima of the gradient of I. The edge function calculates the gradient using the derivative of a Gaussian filter. This method uses two thresholds to detect strong and weak edges, including weak edges in the output if they are connected to strong edges. By using two thresholds, the Canny method is less likely

	than the other methods to be fooled by noise, and more likely to detect true weak edges. Not supported on a GPU.
Sobel	Finds edges using the Sobel approximation to the derivative. It returns edges at those points where the gradient of I is maximum.
Prewitt	Finds edges using the Prewitt approximation to the derivative. It returns edges at those points where the gradient of I is maximum.
Roberts	Finds edges using the Roberts approximation to the derivative. It returns edges at those points where the gradient of I is maximum.
Logs (Laplacian of Gaussian)	Finds edges by looking for zero-crossings after filtering I with a Laplacian of Gaussian filter.

Table 3.2: Technique of edge detection

In our project we used Prewitt and Canny in most of the cases for edge detections. Illustration of an example for this technique is given in fig 3.10.

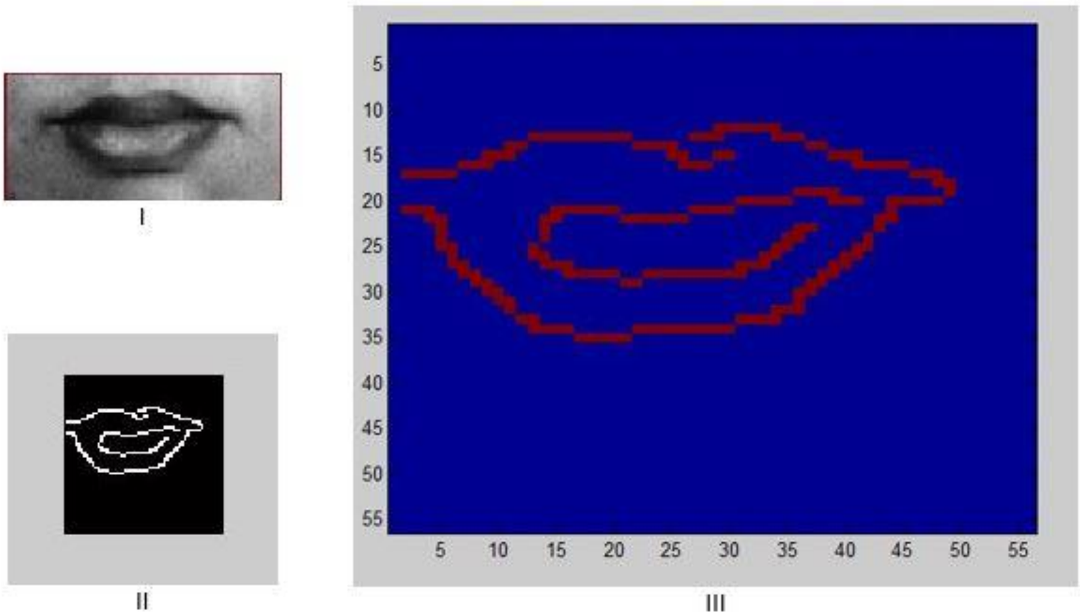


Fig 3.10: Example of edge detection
I. Normal Lips
II & III: Edge detection by Using Canny

After detecting edge we are calculate the distance like from center of eyeball to eyebrow, between upper lip and lower lips etc. For calculating this distance we used Euclidean Distance (discussed below).

3.3.2 Euclidean Distance:

In mathematics, the Euclidean distance or Euclidean metric is the "ordinary" (i.e. straight-line) distance between two points in Euclidean space. With this distance, Euclidean space becomes a metric space. The associated norm is called the Euclidean norm. Older literature refers to the metric as Pythagorean metric. A generalized term for the Euclidean norm is the L^2 norm or L^2 distance.

In the Euclidean plane, if p = (p_1, p_2) and q = (q_1, q_2) then the distance is given by

$$d(p,q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}$$

This is equivalent to the Pythagorean Theorem.

Alternatively, it follows from (*2*) that if the polar coordinates of the point p are (r_1, θ_1) and those of q are (r_2, θ_2), then the distance between the points is

$$\sqrt{r_1^2 + r_2^2 - \cos(\theta_1 + \theta_2)}$$

After calculating the distance for different component of faces, we see that distance are varies for every expressions as shown as figure. Then we set a threshold value for every expressions. This threshold value may differ in different database. Here we used two different database for our system- The Extended Cohn-Kanade (CK+) Dataset and Japanese Female Facial Expression (JAFFE) Database. And get a success full result for three expression like surprise, disgust and neutral. Basic information about this two database and results are shown in fourth chapter.

Chapter 4

Experimental Results & Database Used

Before this we are discussed about the methodology used in our system to recognize different facial expression. Now we see how we implement this method step by step using the mentioned method in chapter 3. In this chapter we see not only the methods but also basic information about the database we used and experimental result and table of experiment. Before discussed about the method used for recognizing facial expression and their experimental result we want to introduce the database used in our system.

4.1 Database Used:

In our system we used two common database for our algorithm-The Extended Cohn-Kanade (CK+) Dataset and Japanese Female Facial Expression (JAFFE) Database.

4.1.1 The Extended Cohn-Kanade (CK+) Dataset:

Facial behaviour of 210 adults was recorded using two hardware synchronized Panasonic AG-7500 cameras. Participants were 18 to 50 years of age, 69% female, 81%, Euro-American, 13% Afro-American, and 6% other groups. Participants were instructed by an experimenter to perform a series of 23 facial displays; these included single action units and combinations of action units. Each display began and ended in a neutral face with any exceptions noted. Image sequences for frontal views and 30-degree views were digitized into either 640x490 or 640x480 pixel arrays with 8- bit gray-scale or 24-bit color values. Full details of this database are given in [35].

Fig 4.1: Different expressions used in CK+ dataset

4.1.2 Japanese Female Facial Expression (JAFFE) Database:

The database contains 213 images of 7 facial expression (6 basic facial expression+1 neutral) posed by 10 Japanese female models. Each image has been rated on 6 emotions adjectives by 60 Japanese subjects. The database was planned and assembled by Michael Lyons, Miyuki Kamachi and Jiro Gyoba.

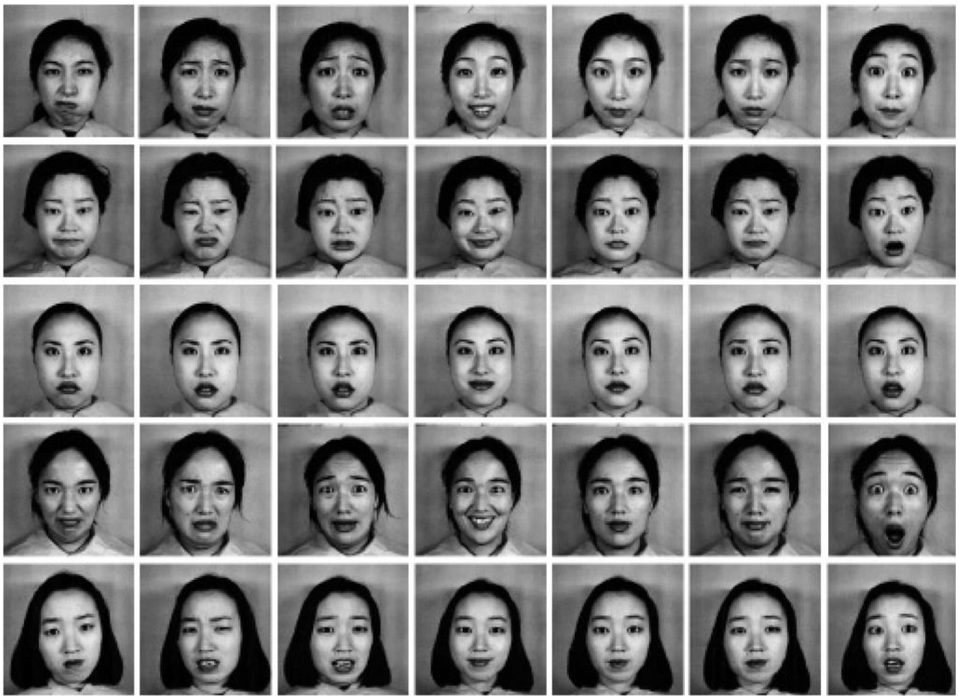


Fig 4.2: Different expression used in JAFFE dataset

4.2 Experimental Procedure:

From previous chapter we came to know that our system is based on three parts: face detection & tracking, feature extraction and recognizing. For this we used several methods. First we observe the images from our database. This images are treated as subjects. When we get our subjects from the database, first we apply to detect the facial part using Viola Jones algorithm. After detecting the face we perform on eye part and mouth or lips part. Our whole thesis are depending on this three basic component of face: eye and eyebrow and lips. After detecting the face, we give our maximum time to work on Eye part and lips part. Using some important methods we detect the eye and the lips part. These methods will be discussed later. After that we calculate the distance of different facial components and we also calculate the action unit for different expressions. Our whole system is depending on these distance calculation.

We used a template of eye for detecting the eye part from the face. When we recognize the eye, we try to observe the circle on this frame by using Hough transformation of circle for detecting the eyeball. It has seen that sometimes lots of unnecessary circle are detected on this frame. Therefore we tried to find the circle with maximum radius. Hence we get the effective result and track the eyeball. Whenever we track the eyeball, we can easily find the mid-point of that circle.

Now our next focus was to detect the eyebrow. So when we select the center of that circle, we used a loop from this mid-point to upward vertically to measure the change of intensity of the pixels. When we observe that the change is maximum we will stop the loop. If we normally watch the face, we can easily tell that after eyeball, the maximum intensity change is in eyebrow portion. We used this technique here. Therefore we observe the eye brow portion. I have described all the methods in this portion to detect the eyeball and eyebrow. Before this we used the edge detection algorithm to find an effective result.

Whenever we get the center of eyeball $(x_{1,}y_1)$ and a point $(x_{2,}y_2)$ of eyebrow where intensity change are maximum, we calculate the Euclidian distance (h_1) between $(x_{1,}y_1)$ and $(x_{2,}y_2)$. This $h_1$is not fixed for every expression. For normal the value of $h_1$and for surprise the value of $h_1$are always different. Therefore we used a threshold value for each of the expressions. For example if $0 < h_1 < 10$, then we set as disgust or if $13 < h_1 < 20$, the set as neutral or normal or if $h_1 \geq 20$, then we set as surprise and so on.

It is not possible for every expression using this h_1 to recognize all the expressions. Therefore we used another feature of face i.e., mouth or lips. To extract this feature, we divide the frame in three part, after detecting the face using Viola Jones face detection algorithm. Last part of the frame (means from $^2/_3$ to end) is used to detect the lips. In this part we saw that there are two parts of lips: upper lips and lower lips. And for different expressions lip's movement is different. The movement of lips means actually the movement of upper and lower lips. Using canny edge detection technique we can detect the edge of the lips clearly. Now using the same method for feature extractions. We calculate the Euclidian distance (h_2) between upper lips $(a_{1,}b_1)$ and lower lips $(a_{2,}b_2)$. Here also we saw that for different expressions $h_2$is different. Sometimes for some expressions when lips are tightly closed we could not measure the distance (h_2). Therefore we had to use curve fitting algorithm. We saw that by using only Euclidian distance (h_1) between eyeball and eyebrow it was difficult to identifying the expression. But now using both distance h_1 and h_2, result is more accurate and easy to get.

To build the system we used some parameters in our algorithm as shown in table 3.1.

	c_1	r_1	m_1	n_1	x_1, y_1	ed
D1	32.9223 51.5931 43.7427 36.3806 **38.5777 40.8236** 29.9632 60.5567	2.0156 2.0510 **4.4057** 2.8335	54	32	32 35	34
D2	33.7684 59.9538 **39.0805 49.3682** 27.5347 68.9011	2.6151 **5.1182** 4.9315	62	33	33 43	42
D3	31.9160 28.7794 **47.2533 40.2171**	1.8502 **3.5344**	32	31	31 25	24
D4	**33.3150 30.2910** 40.3509 49.0393 48.3008 31.3021	**2.0120** 1.3756 1.5295	31	33	33 26	29
D5	38.8763 26.3184 34.4090 61.6893 **37.9370 39.2590** 25.6269 64.2216	1.5275 2.0601 **6.1545** 1.6609	29	38	38 23	22
D6	34.0548 57.1188 35.2479 25.3330 **43.2738 39.1771** 25.0601 62.0862	3.2375 1.7069 **6.6336** 2.9744	60	34	34 22	19

Table 4.1: parameter used to build the system and result

In table 4.1 rows are used for the parameter and column used for data.

D1, D2, ……, D6 : represents the different expression of a single subject.

c_1 : represents the center of the circles detect for the first time.

r_1: represent the corresponding radius of circle.

m_1: represent floor value of y co-ordinates.

n_1: represent floor value of x co-ordinates.

(x_1, y_1) : represent the point where maximum intensity change is detect.

ed: Euclidian distance.

In c_1 and r_1 some data are bold and underlined. This is the data of that circle whose radius is maximum. We store this radius into rs and the center of that circle store into cs. After finding radius we check for maximum intensity change upwards through x co-ordinate.

4.3 Experimental Result:

With the help of previously discussed methods and its results we have given some examples in this part and according this create our whole system. The results are show in this figure shown below:

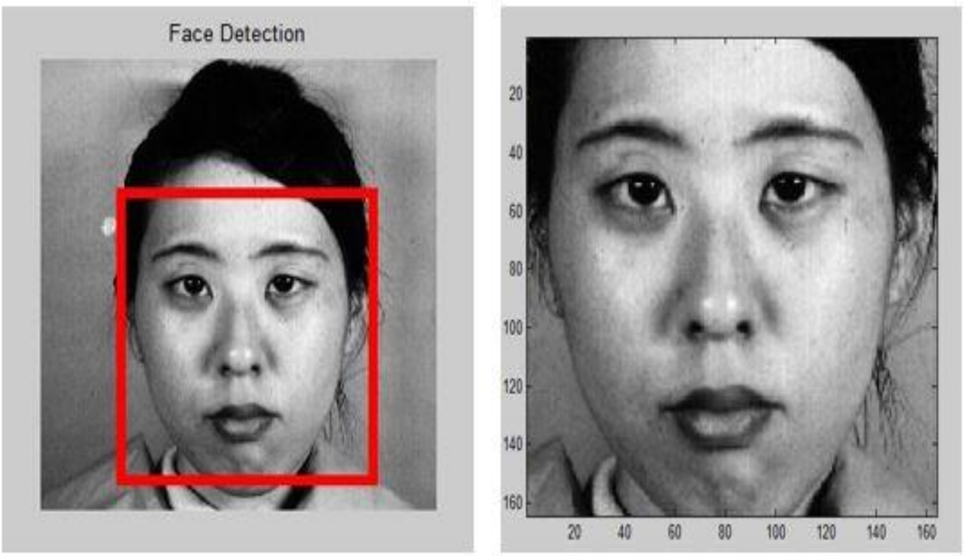


Fig 4.3: Detecting face using Viola Jones

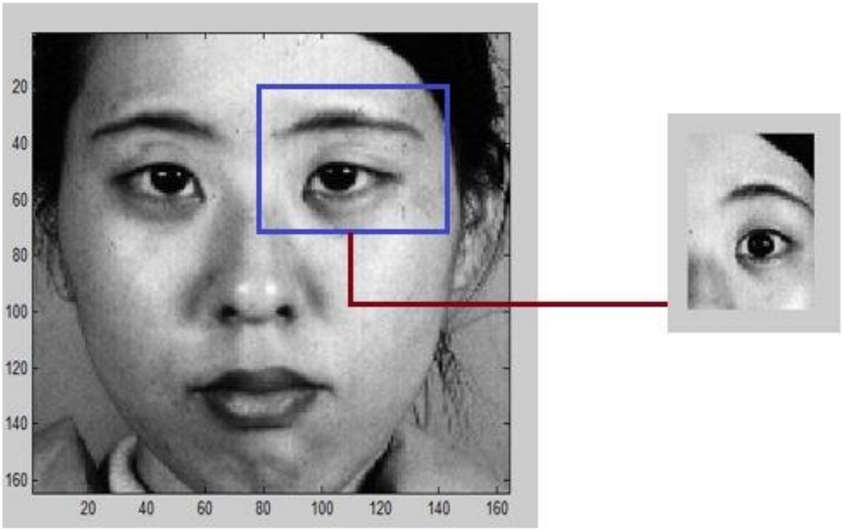


Fig 4.4: Detecting eye using template matching

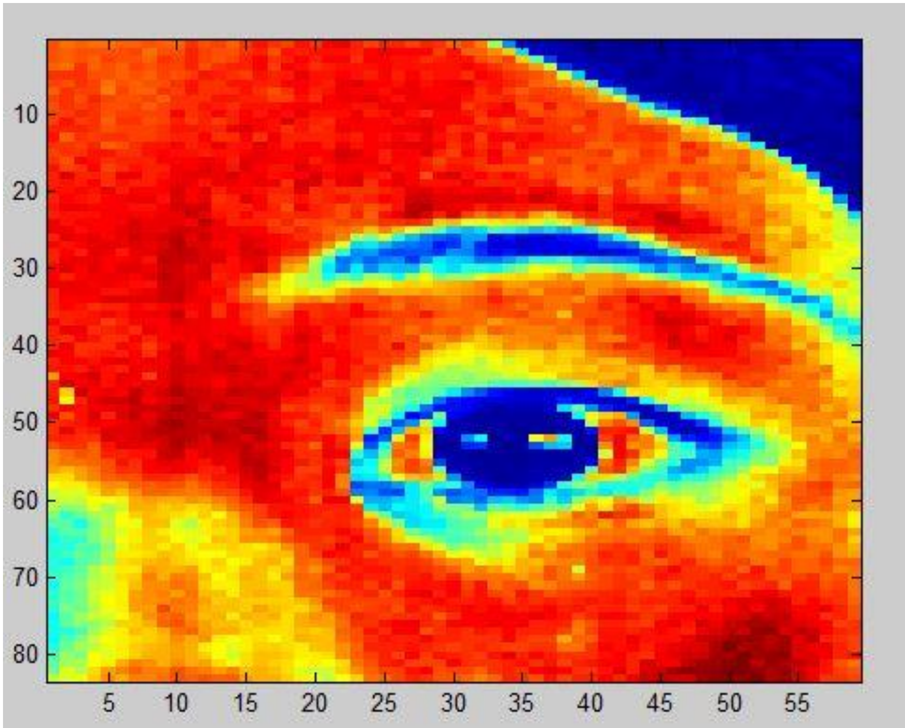


Fig 4.5: Axial representation of eye region

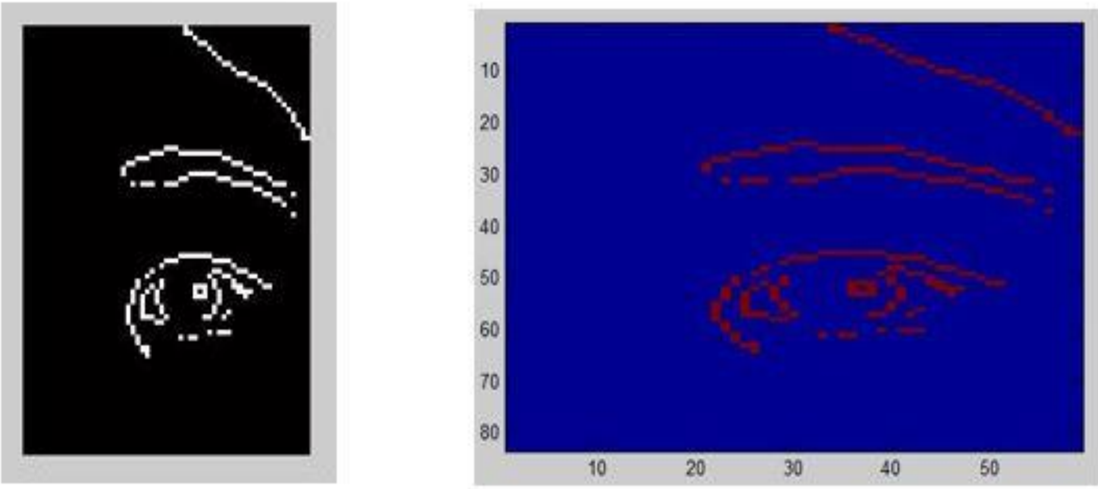


Fig 4.6: Edge of eye are detected using Perwitt

Fig 4.7: Eyeball was detecting using
circle detection algorithm

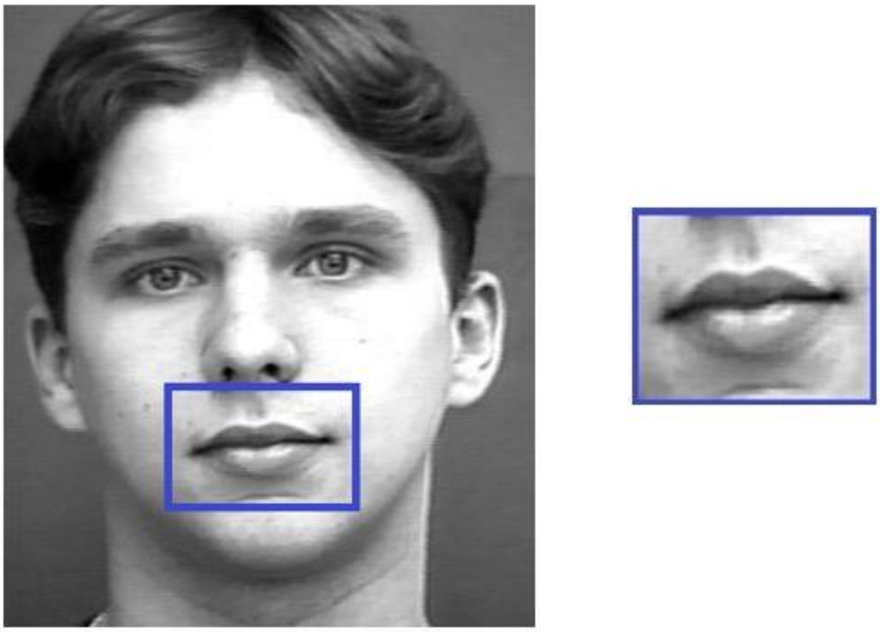


Fig 4.8: Lips detection

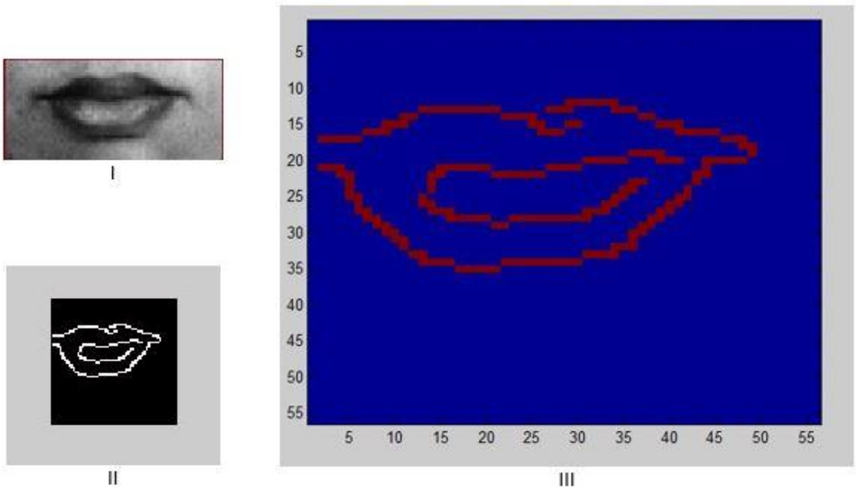
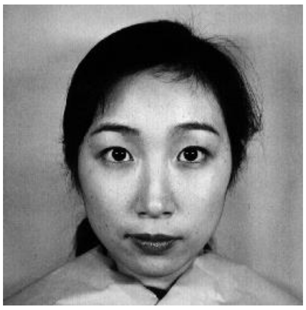
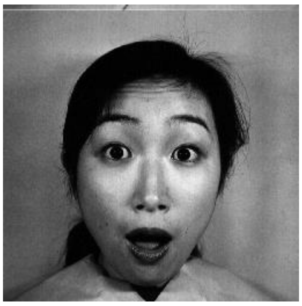


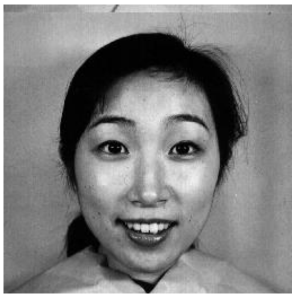
Fig 4.9: Edge of lips



Natural



Surprise



Happy

Disgust

Fear

Angry

Sad

Fig 4.10: Output of different Expressions

The performance of the our system for features for emotion detection is given in Table 4.2 and 4.3 for two different dataset used in our system and it can be seen that Disgust, Neutral and Surprise all perform well compared to the other emotions. This result is intuitive as these are very distinctive emotions causing a lot of deformation within the face. The AUs associated with these emotions so movement of these areas is easily detected by our system as shown in Table 4.1 [ck+]. Conversely, other emotions (i.e. Anger, Sadness and Fear) that do not perform as well. For detecting all the emotions of face we have to use some another features like nose wrinkling, eyebrow distance and so on. Although when we only worked on eye part, the result was too poor for JAFFE dataset. But form this two table we can see that the result is improved from previous. However we combined the two features the result of detection of emotions jumps from just 30% to 70%. An explanation for this can be from that fact that this emotion is quite subtle and it gets easily confused with other, stronger emotions. However, the confusion does not exist in both features sets.

AU	*Name*	N	AU	*Name*	N	AU	*Name*	N
1	*Inner Brow Raiser*	173	13	*Cheek Puller*	2	25	*Lips Part*	287
2	*Outer Brow Raiser*	116	14	*Dimpler*	29	26	*Jaw Drop*	48
4	*Brow Lowerer*	191	15	*Lip Corner Depressor*	89	27	*Mouth Stretch*	81
5	*Upper Lip Raiser*	102	16	*Lower Lip Depressor*	24	28	*Lip Suck*	1
6	*Cheek Raiser*	122	17	*Chin Raiser*	196	29	*Jaw Thrust*	1
7	*Lid Tightener*	119	18	*Lip Puckerer*	9	31	*Jaw Clencher*	3
9	*Nose Wrinkler*	74	20	*Lip Stretcher*	77	34	*Cheek Puff*	1
10	*Upper Lip Raiser*	21	21	*Neck Tightener*	3	38	*Nostril Dilator*	29
11	*Nasolabial Deepener*	33	23	*Lip Tightener*	59	39	*Nostril Compressor*	16
12	*Lip Corner Puller*	111	24	*Lip Pressor*	57	43	*Eyes Closed*	9

Table 4.2: Frequency of the AUs coded by manual FACS coders

Performance of our system are described in table 4.2 and 4.3 by using confusion matrix for two different dataset (CK+ and JAFFE) respectively. In this matrix we used

6 universal pose of face (anger, discussed, surprise, sadness happy and fear) and one is for natural. So in dis matrix 7 rows and 7 column is there for represent the result. From FACS manual we can see that action units are always differ for different expression. By comparing this actin unit we get the result.

Fom this two table 4.2 and 4.3 we see that we get almost the accurate resul for natural, disgust and surprise. For example when we try to find out the expression of surprise for CK + dataset, we get 96 correct result for 100 subject as shown in table 4.1. So accuracy is 96% for that expression.

	Natural	Anger	Disgust	Surprise	Sadness	Happy	Fear
Natural	**92**	2	1	0	2	0	3
Anger	1	**48**	20	2	17	1	11
Disgust	0	6	**89**	0	3	0	2
Surprise	0	0	1	**96**	1	2	0
Sadness	6	21	20	3	**33**	1	16
Happy	1	13	8	17	6	**51**	4
Fear	3	13	14	4	23	3	**40**

Table 4.3: Confusion matrix for facial expression detection in CK+ dataset

	Natural	Anger	Disgust	Surprise	Sadness	Happy	Fear
Natural	**96**	1	1	0	2	0	0
Anger	6	**38**	22	2	19	1	12
Disgust	0	1	**93**	0	2	0	4
Surprise	1	0	0	**97**	0	2	0
Sadness	1	23	17	0	**39**	1	19
Happy	0	19	6	14	11	**47**	3
Fear	7	23	18	2	14	8	**28**

Table 4.4: Confusion matrix for facial expression detection in JAFFE dataset

Chapter 5

Conclusion & Future Work

The aim of this thesis was to explore the area of facial expression recognition. Beginning with the psychological motivation for facial behavior analysis, this field of science has been extensively studied in terms of application and automation. Manual face analysis used by psychologists was quickly replaced by suitable computer software. A wide variety of image processing techniques was developed to meet the facial expression recognition system requirements.

However, there are still many challenges and problems to solve in such systems, especially in the area of their performance and applicability improvement. Apart from theoretical background, this work provides the design and implementation of Facial Expression Recognition System. Proposed system was developed to process the video of facial behavior and recognize displayed actions in terms of six basic emotions. Major strengths of the system are full automation as well as user and environment independence.

Even though by using this approach all expressions were not recognition the all emotions. Before this we see in result part that our system were giving good result only on three expression. Now for better result in future we want use lots of others facial component like nose wrinkling, eyebrow distance and so on. We expecting that using this components we will get more accurate result in future.

References

[1] Ying-li Tian, Takeoa Kanade Jeffrey F. Cohn. Recognizing action units for Facial Expression Analysis. IEEE Transc. On Pattern Analysis and Machine Intelligence, Vol. 23, No. 2, Feb 2001.

[2] Michel F. Valstar, Maja Pantic. Fully Automatic Recognition of the Temporal Phases of Facial Actions.IEEE Transaction on systems, Man and Cybernetics- Part B: Cybernetics. Vol. 42. No. 1. 2012.

[3] Laura Florea and Raluca Boia. Eyebrow Localization for Expression Analysis. Intelligent Computer Communication and Processing (ICCP), 2011 IEEE International Conference Publications, 2011.

[4] P. Ekman and W. V. Friesen. The Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press Inc., San Francisco, CA, 1978.

[5] J. M. Carroll and J. Russell. Facial expression in hollywood's portrayal of emotion. Journal of Personality and Social Psychology. 72:164–176. 1997.

[6] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying facial actions. International Journal of Pattern Analysis and Machine Intelligence, 21(10):974–989, October 1999.

[7] I. A. Essa and A. P. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. IEEE Transc. On Pattern Analysis and Machine Intelligence, 19(7):757–763, JULY 1997.

[8] Z. Zhang. Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perceptron. International Journal of Pattern Recognition and Artificial Intelligence, 13(6):893–911, 1999.

[9] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying facial actions. International Journal of Pattern Analysis and Machine Intelligence, 21(10):974–989, October 1999.

[10] K. Fukui and O. Yamaguchi. Facial feature point extraction method based on a combination of shape extraction and pattern matching. Systems and Computers in Japan, 29(6):49–58, 1998.

[11] Jyh-Yeong Chang and Jia-Lin Chen, “Automated Facial Expression Recognition System Using Neural Networks” Journal of the Chinese Institute of Engineers, Vol. 24, No. 3, pp. 345-356. 2001.

[12] S.-M. Choi and Y.-G. Kim. An affective user interface based on facial expression recognition and eye-gaze tracking. In Tao et al. [85], pages 907–914. 2014.

[13] M. Suwa, N. Sugie, and K. Fujimora. A preliminary note on pattern recognition of human emotional expression, 1978.

[14] P. Ekman and W. Friesen. The facial action coding system: A technique for the measurement of facial movement, 1978.

[15] P. Ekman, T. Huang, T. Sejnowski, J. Hager, "Final Report To NSF of the Planning Workshop on Facial Expression Understanding", 1992.

[16] A. K. Jain and S. Z. Li. Handbook of Face Recognition. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[17] http://2wanderlust.files.wordpress.com/2009/03/picture-2.png

[18] W. Friesen and P. Ekman. Emfacs-7. Emotional facial action coding system, 1983.

[19] M. Pantic and I. Patras. Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 36(2):433–449, 2006.

[20] A. Ryan, J. Cohn, S. Lucey, J. Saragih, P. Lucey, F. D. la Torre, and A. Rossi. Automated Facial Expression Recognition System. In Proceedings of the International Carnahan Conference on Security Technology, pages 172–177, 2009.

[21] Y. Tian, J. Cohn, and T. Kanade. Facial expression analysis. In S. Li and A. Jain, editors, the handbook of emotion elicitation and assessment, pages 247–276. Springer, New York, NY, USA. 1. 2002.

[22] Y. Tong, W. Liao, and Q. Ji. Facial Action Unit Recognition by Exploiting Their Dynamic and Semantic Relationships. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(10):1683– 1699, 2007.

[23] A. Vinciarelli, M. Pantic, and H. Bourlard. Social Signal Processing: Survey of an Emerging Domain. Image and Vision Computing, 31(1):1743–1759, 2009.

[24] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan. Automated Drowsiness Detection for Improved Driver SafteyComprehensive Databases for Facial Expression Analysis. In Proceedings of the International Conference on Automotive Technologies, 2008.

[25] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan. Towards Practical Smile Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(11):2106–2111, 2009.

[26] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale. A high-resolution 3D dynamic facial expression database. In Proceedings of the International Conference on Automatic Face and Gesture Recognition, 2008.

[27] P. Lucey, J. Cohn, S. Lucey, I. Matthews, S. Sridharan, and K. Prkachin. Automatically Detecting Pain Using Facial Actions. In Proceedings of the International Conference on Affective Computing and Intelligent Interaction, pages 1–8, 2009.

[28] P. Ekman, E. T. Rolls, D. I. Perrett, and H. D. Ellis. Facial expressions of emotion: an old controversy and new findings. Facial Expressions of Emotion: An Old Controversy and New Findings, 335(1273):63–69, Jan. 1992.

[29] I. Eibl-Eibesfeldt. Human ethology. Pages 1–103. Chaps 1 and 2. 1989.

[30] A. K. Jain and S. Z. Li. Handbook of Face Recognition. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[31] M. Bichsel and A. P. Pentland. Human face recognition and the face image sets topology, 1994.

[32] H. A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection, 1998.

[33] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: An application to face detection, 1997.

[34] F. Crow. Summed-area tables for texture mapping, 1984.

[35] H. Schneiderman and T. Kanade. A statistical method for 3d object detection applied to faces and cars, 2000.

[36] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features, 2001.

[37] P. Viola and M. Jones. Robust real time object detection, 2001.

[35] Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar. The Extended Cohn-Kanade Dataset(CK+): A complete dataset for action unit and Emotion-Specified expression. July 2010.

[39] Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting, 1997.

[40] P. Y. Simard, L. Bottou, P. Haffner, and Y. L. Cun. Boxlets: a fast convolution algorithm for signal processing and neural networks, 1998.

[41] T. Cootes, C. Taylor, and A. Lanitis. Multi-resolution searches with active shape models. Pages A: 610–612, 1994.

[42] K.-K. Sung and T. Poggio. Example-based learning for view-based human face detection, 1998.

[43] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. Comput. Vis. Image Underst., 61(1):38–59, 1995.

[44] I. Fasel, M. Bartlett, and J. Movellan. A comparison of Gabor filter methods for automatic detection of facial landmarks, 2002.

[45] A. Lanitis, C. J. Taylor, and T. F. Cootes. Automatic interpretation and coding of face image using flexible models. IEEE Trans. Pattern Anal. Mach. Intell., 19(7):743–756, 1997.

[46] H. Gu, G. Su, and C. Du. Feature points' extraction from faces. Proceedings of Image and Vision Computing New Zealand Conference pages 154–158, November 2003.

[47] M. A. Turk and A. P. Pentland. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1):71–86, 1991.

[48] L. Ding and A. Martinez. Precise detailed detection of faces and facial features. In proceeding of IEEE conference on Computer Vision and Pattern recognition.1-7. In 2007.

[49] S. Strupp, N. Schmitz and K. Berns. Visual-based emotion detection for natural man-machine interaction. In Advance in Artificial Intelligence, ser. Lecture notes in Computer Science. Springe Verlag, Vol. 5243. 2008.

[50] S. Li, Z. Q. Zhang, H.-Y. Shum, and H. Zhang. Floatboost learning for classification, 2002.

[51] S. Z. Li and Z. Zhang. Floatboost learning and statistical face detection, 2004.

[52] S. Z. Li, L. Zhu, Z. Q. Zhang, A. Blake, H. Zhang, and H. Shum. Statistical learning of multi-view face detection, 2002.

[53] S. Xin and H. Ai. Face alignment under various poses and expressions. In Tao Jyh-Yeong Chang and Jia-Lin Chen, "Automated Facial Expression Recognition System Using Neural Networks" Journal of the Chinese Institute of Engineers, Vol. 24, No. 3, pp. 345-356 (2001).

[54] F. Jiao, S. Li, H.-Y. Shum, and D. Schuurmans. Face alignment using statistical models and wavelet features. cvpr, 01:321, 2003.

[55] Y. ShuiCheng and Q. Cheng. Multi-view face alignment using direct appearance models. In FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, page 324, Washington, DC, USA, 2002.

[56] Y. Liu, Y. Li, L. Tao, and G. Xu. Multi-view face alignment guided by several facial feature points. In ICIG '04: Proceedings of the Third International Conference on Image and Graphics (ICIG'04). Washington, DC, USA, 2004. IEEE Computer Society. Pages 238–241. 2004.

[57] P. Aleksic and A. Katsaggelos. Automatic facial expression recognition using facial animation parameters and multistreamhmms. Information Forensics and Security, IEEE Transactions on, 1:3–11, March 2006.

[58] F. B. Claude C. Chibelushi. Facial expression recognition: A brief tutorial overview, 2003. [Online; accessed 20-November-2007].

[59] A. Colmenarez, B. Frey, and T. S. Huang. A probabilistic framework for embedded face and facial expression recognition. Pages 592–597. 1992.

[60] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(6):681–685, 2001.

[61] S. Kimura and M. Yachida. Facial expression recognition and its degree estimation. In CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97), page 295, Washington, DC, USA, 1997. IEEE Computer Society. 1997.

[62] A. Mehrabian. Communication without words. Psychology Today, 2(4):53–56, 1968.

[64] I. Kotsia and I. Pitas. Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Transactions on Image Processing, 16(1):172–187, 2007.

[65] M. A. O. Vasilescu and D. Terzopoulos. Multilinear analysis of image ensembles: Tensorfaces. In ECCV (1), pages 447–460, 2002.

[66] L. Yin and X. Wei. Multi-scale primal feature based facial expression modeling and identification. In FGR ’06: Proceedings of the 7th International Conference On Automatic Face and Gesture Recognition, pages 603–608, Washington, DC, USA, 2006. IEEE Computer Society.2006.

[67] J. Yu and B. Bhanu. Evolutionary feature synthesis for facial expression recognition. Pattern Recogn. Lett., 27(11):1289–1298, 2006.

[68] F. Crow. Summed-area tables for texture mapping, 1984.

[69] J. Jones and L. Palmer. An evaluation of two-dimensional gabor filter model of simple receptive fields in cat striate cortex. Journal of Neurophysiology, 58(6):1233–1258, 1987.

[70] M. A. O. Vasilescu and D. Terzopoulos. Tensortextures: Multilinear image-based rendering, 2004.

[71] M. A. O. Vasilescu and D. Terzopoulos. Multilinear independent components analysis. In CVPR ’05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05). Washington, DC, USA, 2005. IEEE Computer Society. Vol 1, pages 547–553. 2005.

[72] A. Shashua and T. Hazan. Non-negative tensor factorization with applications to statistics and computer vision. In ICML ’05: Proceedings of the 22nd international conference on Machine learning. New York, NY, USA, 2005. Pages 792–799. 2005.