MOOD DETECTION ACCORDING TO FACIAL EXPRESSIONS

YASIR MUSTAFA WAIS

JANUARY 2015

MOOD DETECTION ACCORDING TO FACIAL EXPRESSIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED
SCIENCES OF
ÇANKAYA UNIVERSITY

BY
YASIR MUSTAFA WAIS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF
DEPARTMENT OF COMPUTER ENGINERING

JANUARY 2015

Title of the Thesis: **Mood Detection According To Facial Expressions.**

Submitted by **Yasir Mustafa WAIS**

Approval of the Graduate School of Natural and Applied Sciences, Çankaya University.

Prof. Dr. Taner ALTUNOK

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Müslim BOZYİĞİT

Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Sadık EŞMELİOĞLU

Supervisor

Examination Date: 16.01.2015

Examining Committee Members

Assist. Prof. Dr. Sadık EŞMELİOĞLU	(Çankaya Univ.)	
Assist. Prof. Dr. Fahd JARAD	(THKU.)	
Assist. Prof. Dr. Reza ZARE HASSANPOUR	(Çankaya Univ.)	

STATEMENT OF NON-PLAGIARISM PAGE

I hereby declare that all information in this document has 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 material and results that are not original to this work.

Name, Last Name : **Yasir Mustafa, WAIS**

Signature :

Date : 16.01.2015

ABSTRACT

MOOD DETECTION ACCORDING TO FACIAL EXPRESSIONS

WAIS, Yasir Mustafa

M.Sc., Department of Computer Engineering

Supervisor: Prof. Dr. Sadık EŞMELİOĞLU

January 2015, 63 Pages

Facial expression is considered to be one of the universal languages of the human body and an active channel of non-verbal communication that is used to convey large amounts of information. By using this information, we are able to read human internal emotional states, intentions or social communications.

Therefore, this study is conducted, to analyze facial expression and to extract information that is used to detect the emotional state of any person in static images. The study contains two new proposed methods: the first method is a type of template matching that finds a good match between an input face image and four pre-defined templates (Happy, Fear, Surprise and Neutral). The second method is formulated to detect a mood for three emotional expressions (Happy, Neutral and Surprise) by locating four feature points on two regions (mouth and nose) of a detected face. The distances between these feature points are measured and then compared with pre-defined thresholds, after which a decision is made.

The detection efficiency of the proposed method is tested on the JAFFE (Japanese Female Facial Expression) database and the IMM (Informatics and Mathematical Modeling) database. The average percentage of detection rated is between 80 and 85 percent.

Keywords: Skin Color Detection, Face Detection, Facial Expressions Classification.

ÖZ

YÜZ İFADELERİNE GÖRE RUH HALİNİN TESPİTİ

WAIS, Yasir Mustafa
Yüksek Lisans, Bilgisayar Mühendisliği Anabilim Dalı
Tez Yöneticisi: Prof. Dr. Sadık EŞMELİOĞLU
Ocak 2015, 63 sayfa

Yüz ifadesi insan vücudunun evrensel dillerinden biridir ve bir çok bilgi iletmi için kullanılan sessiz iletişimin aktif bir yolu olarak görülür. Bu bilgi sayesinde insanın duygusal durumunu, niyetini ve sosyal iletişimini anlamak mümkündür.

Bu nedenle, bu çalışmada yüz ifadesini analiz edip, statik görüntülerden herhangi bir kişinin duygusal durumunu tespit etmek için kullanılabilir bilgileri çıkarmak için yapılmıştır. Çalışmamız yeni önerilen iki yöntemi içerir : İlk yöntem bir giriş yüzü görüntüsünün varolan ve tanımlanmış dört şablonun (Mutlu, Korkmuş, Şaşırmış ve Normal) arasında kendine en uygun olanı bulur. İkinci yöntem tespit edilen yüzün iki bölgesinde (ağız ve burun) dört özellik noktasını bularak üç duygusal ifadeyi (Mutlu, Normal ve Şaşırmış) bir ruh halini tespit etmek için formüle edilmiştir. Bu özellik noktalarının arasındaki mesafeleri ölçüp daha önceden hesap edip tanımlamış olduğumuz ifadenin ölçüleri ile karşılaştırıp öyle kararımız verilir.

Önerilen yöntemin tespit verimliliği JAFFE (Japon Kadın Yüz İfadesi) veritabanı ve İBB (Bilişim ve Matematik Modelleme) veritabanı üzerinde test edilmiştir. Puan algılama ortalaması yüzde 80 ila 85 arasında olduğu tesbit edilmiştir.

Anahtar Kelimeler: Cilt Rengi Algılama, Yüz Algılama, Yüz İfadeleri Sınıflandırma.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Assist. Prof. Dr. Sadık EŞMELİOĞLU for his supervision, special guidance, suggestions, and encouragement through the development of this thesis.

It is a pleasure to express my special thanks to my family and my friends for their valuable support.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

FIGURES

LIST OF TABLES

TABLES

LIST OF ABBREVIATIONS

HCI	Human Computer Interface
LBP	Local Binary Pattern
ELBP	Extended Local Binary Pattern
AAM	Active Appearance Model
ASM	Active Shape Model
PCA	Principal Component Analysis
NN	Neural Network
MLP	Multi Layer Perception
SOFM	Self-Organized Feature Map
SCG	Scale Gradient Conjugate
ANN	Artificial Neural Network
SVM	Support Vector Machine
k-NN	k-Nearest Neighbor
RGB	Red Green Blue
HSV	Hue Saturation Value
HSI	Hue Saturation Intensity
HSL	Hue Saturation Lightness
JAFFE	Japanese Female Facial Expression
IMM	Informatics and Mathematical Modeling
FJ	Fixed Jacobian
POIC	Project out Inverse Compositional
SIC	Simultaneous Inverse Compositional
RIC	Robust Inverse Compositional
IEBM	Iterative Error Bound Minimization
HFBID	Haar like Feature Based Iterative Discriminative

CHAPTER 1

INTRODUCTION

As human beings, we have an ability to communicate with each other by using two different modalities: verbal communication, which refers to using words or the voice to share thoughts and ideas; and non-verbal communication, which is defined as wordless (visual) communication consisting of different types of apparent behaviors, such as facial expressions, body language, sign language, gestures, touch and eye contact.

Many scientific studies have been proposed with the aim of learning about the extent to which non-verbal communication takes place in our daily communication.

A study conducted by Birdwhistell [1] concluded that 65% of the meaning of a message is transmitted through non-verbal cues. Another study by Albert Mehrabian [2] in two papers concluded that 93% of the meaning of a message is conveyed through non-verbal cues. His statistics showed that the verbal component contributes only 7% to the effect of the message as a whole; the vocal component contributes 38%, while facial expressions of the speaker contribute 55% to the effect of the spoken message.

Moreover, Fromkin and Rodman [3] also found that up to 90% of the meaning of a message is transmitted non-verbally. Even Deborah Tannen [4] had found the same result. Thus, no matter the non-verbal communication percentage, it implies and indicates the extent of the significance of non-verbal communication in our daily communications.

Generally, facial expressions are an important channel of non-verbal communication which we use extensively in our daily activities. This is the case across different cultures. Moreover, facial expressions communicate an extremely large amount of information which may be useful when we wish to reveal attention, personality, intention and psychological state [5] without uttering words. The obvious fact is that facial expressions are universal in comparison to other non-verbal communication types.

Thus, by extracting and analyzing facial expressions, we are able to know how a person feels about a particular situation; for instance, a smile may indicate approval or the happiness of a person, while an expression of surprise could indicate something unexpected happening. Other expressions will presumably indicate other specific emotions. This helps us to interpret facial expressions and correlate them to our conscious and unconscious behaviors. Therefore, the question is: *How can we build a machine (computer) that is able to think like a human?*

Facial Expression analysis started in the 19th century when Darwin proposed the concept of universal facial expressions in humans and animals. In his book, *The Expression of the Emotions in Man and Animals* [6], he noted:

"*...the young and the old of widely different races, both with man and animals, express the same state of mind by the same movements.*"

Since the mid 1970s, automatic facial expression analysis has attracted the interest of many computer vision research groups with the aim of providing a computer with the capability of recognizing facial expressions and classifying them into the six basic expressions (fear, anger, disgust, happiness, surprise and sadness) regardless of context, culture, gender, and so on.

Automatic facial expression analysis may be applicable in many applications, such as security control systems (to predict possible crime), emotion and paralinguistic communication, clinical psychology, psychiatry, neurology, pain assessment, lie detection, intelligent environments and multimodal human computer interaction (HCI).

1.1 What are Facial Expressions?

Facial expressions are an important channel of non-verbal communication which we can use extensively when we communicate with others. They add extra meaning to words in many different ways so that we may know for example, that a person could be joking or perhaps being very serious. For this reason, in email and messaging, we use emoticons, small pictures which express our feelings as reactions toward specific actions taken from the peer side.

Even medically, it has been proved that facial expressions are due to the movements of muscles underneath the skin of the face which are controlled by the brain through facial nerves [7], as shown in Figure 1. The movements of muscles can convey and indicate state of mind such as the emotional state of an individual in a specific situation.

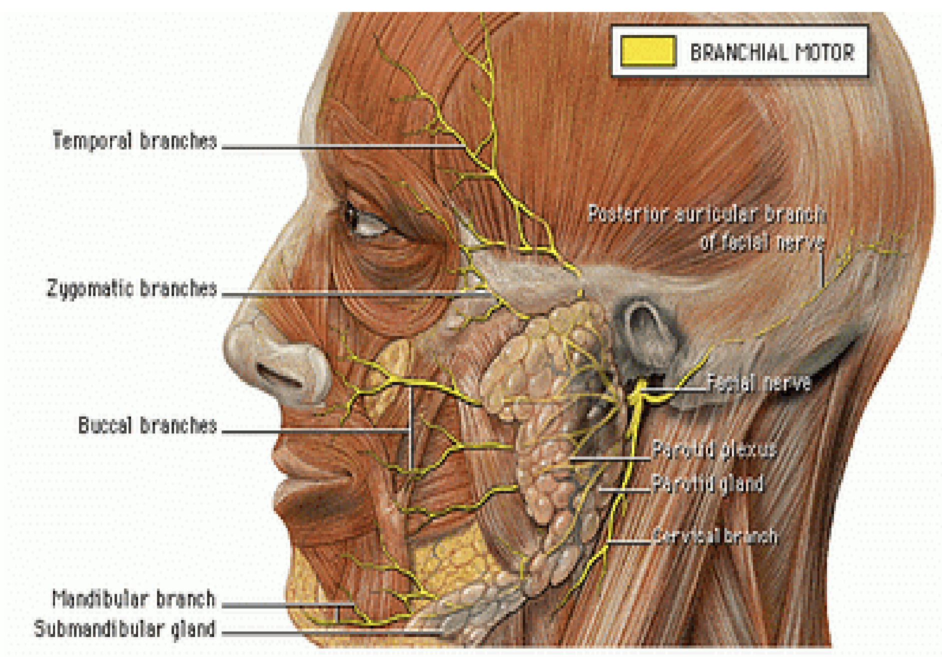


Figure 1: Facial nerves [7]

1.2 Related Works

The origins of facial expression analysis date back to the 19th century. From that time, various techniques have been developed by computer vision research groups with the aim of providing a computer with the capability of automatic facial expression classification or recognition.

Generally, the basic structure of any proposed facial expression technique consists of three main stages: face detection, facial feature extraction and classification, as shown in Figure 2.

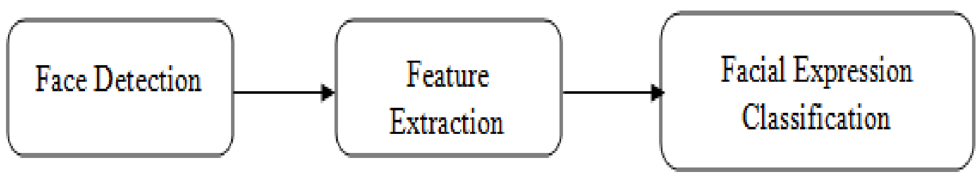


Figure 2: Basic structure of facial expressions classification system

A brief explanation for each stage is listed below:

1. **Face detection stage:**
   This is the pre-processing stage that locates a face region for the input image or sequences of images by using a face detection method. The detected face passes through to the second stage.
2. **Feature extraction stage:**
   This stage is considered to be an essential component of facial expressions classification systems. Its purpose is to extract a unique characteristic of each facial expression class and form it as a feature vector from the detected face. The feature vector then passes down to the last stage.
3. **Facial expressions classification stage:**
   The final stage is used to classify each facial expression class (Neutral, Happy, Surprise, Sad, etc.) according to an extracted feature by using one of a number of image processing classification methods.

As a result of our literature survey, Table 1 lists the common algorithms that have been used in each stage of the facial expression classification system.

Face Detection	Feature extraction	Classifier
Template matching	LBP	SVM
PCA	Gabor Filter	NN
NN	AAM	KNN
Viola-Jones	PCA	

Table 1: Algorithms Used in Facial Expression Classification Stages

1.3 Aim of the Research

Obviously, the communication between human and computer is more active in one direction than in the other, while there is a gap in communication from computer to human. Filling this gap will make computers more active, smarter, and friendlier.

Thus, many academic studies have been conducted in HCI (Human Computer Interaction) with the aim of providing a computer with the ability to emulate human communication with each other by using two different kinds of communication channels: the first is the auditory channel, which allows the computer to be able to analysis human speech, recognize it and take action accordingly; the second is the visual channel, which makes a computer able to analyze and understand human operators and movements throughout the vision; for instance, human body detection, human face detection, facial expression recognition, etc.

Since the human face is the richest source of non-verbal communication as well as the most accessible interface to reveal human emotional states and intentions, the objective of this thesis is to improve upon facial recognition methodologies that automatically detect the emotional mood of any person in static images based on facial expressions.

1.4 System Overview

1.4.1 Method one

The first proposed method that is used to detect the four different kinds of emotion (Happy, Neutral, Surprise, and Fear) can be listed as a template-matching process that is used to find a good match between the input face image and the pre-defined the four face templates.

The idea is basically structured into two main steps without any pre-processing steps that are used to detect a face:

In the first step, the Y directional gradient is found for the input face image by using the MATLAB function (**imgradientxy**) in order to overcome any lighting change.

The second step is a matching process carried out by finding the correlation coefficient between the input face images and the pre-defined four face templates. According to this value, a decision is made to classify the input face image into one of four classes.

1.4.2 Method two

The basic idea behind the proposed second method is automatic detection of the mood of any person based on the facial expression from any input static face image after two major pre-processing steps being used to detect a face.

The first step is skin color segmentation, which is used to detect a skin color from any input color images by using one of the skin color detection techniques. The benefits of this step are the speeding up of the face detection process and overcoming any background complexity. This step is important in the case of the input image being color. However, if the input image is grayscale, as shown in Figure 3, the input image will directly jump to second step.

The second step is face detection, which is a process to detect and locate a face in an input image by using a highly-rated approach known as the Viola-Jones algorithm, which is considered to be one of feature-based methods employed to detect a face according to face features, such as the nose, mouth, eyes etc.

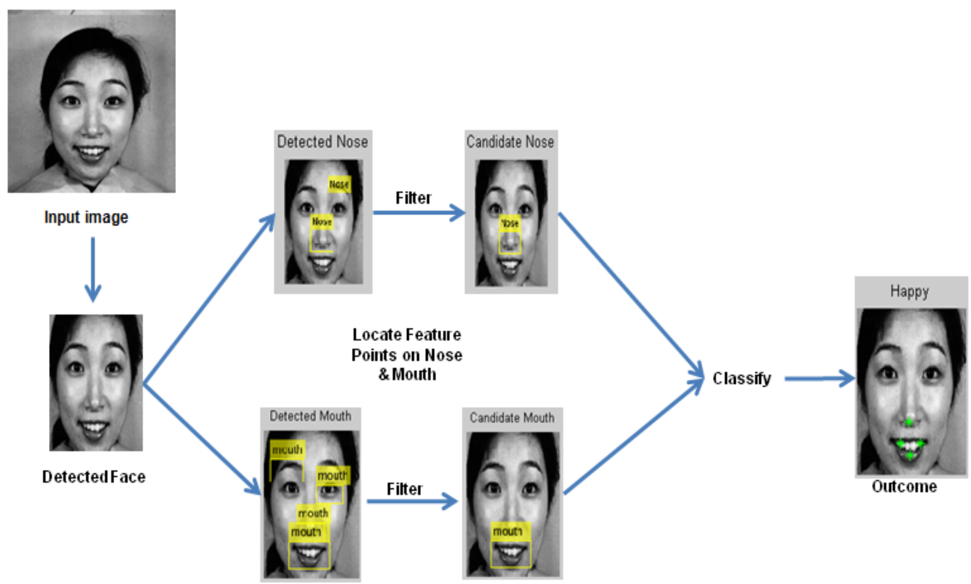


Figure 3: Proposed system steps in cases of the input image being grayscale

Finally, auto detecting the mouth and nose regions from the candidate face is also carried out by using the Viola-Jones algorithm. This is followed by auto locating some of the feature points in these two regions. A decision is made after computing the distance between these feature points so as to classify each class according to the distance values. From the outcome of this classification, we can detect the mood of a person.

1.5 Outline of the Thesis

The remainder of the dissertation is organized as follows:

Chapter Two: A literature survey of the common approaches and studies that are or have been proposed on skin color segmentation, face detection and facial expression classification with brief explanations of some of them in order to gain a better understanding of the existing approaches.

Chapter Three: The details of the first proposed method are presented which aim to classify the four basic facial expressions.

Chapter Four: A full description of the second proposed method together with the results of experiments.

Chapter Five: Includes the conclusion and future plans for the two proposed methods.

CHAPTER 2

LITERATURE SURVEY

In this chapter, we explore some of the published papers on skin color detection, face detection and facial expression classification and recognition in order to gain a better understanding of existing methods and techniques already being used. This will help us to propose new and useful ideas to detect the emotional mood of any person.

2.1 Skin Color Detection

Skin color detection is an image processing approach that is used to locate a skin color pixel in any input color image or sequence of images (video).

Typically, skin color detection requires great care in many vision-based methodologies. They are widely used as a fundamental part in a large number of applications, such as face detection [8], face tracking, human motion analysis [9], and naked image filters [10] in order to find the regions that can potentially be human faces or any other part of the human body.

Human skin color is represented as the most powerful and stable feature of the human body. Moreover, it has been proved that skin color does not fall randomly in any given color space, such as RGB, YCbCr, HSV etc.

Therefore, a majority of the proposed face detection methods use skin color detection as a pre-process step in order to speed up the detection process and to overcome non-skin color pixels, such as clothes and background, which significantly influences performance.

Generally, many skin color detection methods have been suggested which can be classified in two categories:

1. **Pixel-based methods:**

   The ability to detect skin color according to pixel (color) intensity individually in color digital images.

2. **Region-based methods:**

   This is the ability to detect skin color by using a pixel and its neighbors (texture) [11].

Basically, many skin color detection methods are concentrated on pixel intensity for skin color classification rather than attempting to use texture information to classify skin color.

However, experimental results show that the pixel-based methods are a greater influence on the factors listed below [12]:

- **Illumination conditions**: A change in the spectral distribution and illumination level of a light source (indoor, outdoor, highlights, shadows, color temperature of lights).

- **Camera characteristics**: The color reproduced by a CCD camera is dependent on not only the illumination conditions but also the spectral sensitivities of a camera sensor.

- **Ethnicity:** Skin color varies according to ethnic groups.

- **Individual characteristics:** Individual characteristics such as age, sex and body parts affect the skin color.

- **Other factors:** Different factors such as make up, spectacles, background colors, shadows and motion which affect skin color.

Of the above factors, illumination condition is the most important and difficult problem in skin color segmentation which may seriously lower the performance characteristics of the proposed method.

Referring to the literature survey done on skin color detection methods, a proposed method would consist of two main phases:

First phase:

The first phase is to convert the input color image to a suitable color space such as RGB, YCbCr, HSV, etc. This might be a skin color occupying a compact part of the selected color space. It would be easy to define the boundaries of skin color on that space.
A brief explanation of some color spaces that has been attempted on skin color detection is described in the following:

1. Basic color spaces such as

a. RGB: one of the most commonly used color spaces used for processing and storing digital images. Each pixel is a combination of three colored rays (red, green and blue) but mixing from the chrominance and luminance components makes this color space problematic [12] as any change in lighting (luminance) will totally effect the color distribution in the space.

b. Normalized RGB : proposed to reduce the dependency on lighting with the three components *red, blue,* and *green* being reduced to two pure colors: *red* and *green* [13], thus making a normalized RGB better than RGB for colored skin segmentation.

2. Orthogonal Color spaces: such as YCbCr, YIQ and YUV, which are widely used in television transmission and JEPG compression due to the ease of separating the luminance component from the chrominance component; for example, YCbCr, in which Y represents luminance (lighting) and Cb, Cr the chrominance [14].

3. Perceptual color space: such as HSI, HSV and HSL. Hue represents a dominant color, such as (red, green, purple and yellow) of an area. Saturation measures the colorfulness of an area, while (Intensity, Value, Lightness) represent the brightness [14]. The HSV color space tends to be more realistic such that its representation strongly relates to the human perception of color [15].

Basically, a good color space that can increase the performance of any proposed method should meet the following two criteria [13]:

(i) It has to provide a comprehensive coverage of the objects of interest, namely *skin color*.

(ii) It must be compact so that the probability of pixels in the background being classified as having object-like color is minimized.

Second phase:

The skin classifier (model) is used to formulate the decision rules to distinguish a skin pixel from a non-skin pixel. This may be done in different approaches, but all endeavor to define the boundaries of the skin cluster in the color space. These approaches may be classified into three categories:

1. **Non-parametric:**
   The basic idea is to estimate skin color distribution from training data without driving an explicitly model [13], such as a normalized lookup table “histogram based “and Bayes classifier.

2. **Parametric:**
   This model has a specific functional form with adjustable parameters selected to fit the model [16]; for example, a single Gaussian and mixed Gaussian model.

3. **Semi parametric:**
   This is a combination of the parametric and non-parametric approaches, an example of which is a neural network [17].

2.2 Face Detection

Face detection is an image processing technique that is used to determine the location and size of human faces in digital images or in sequences of images (videos).

Face detection is a preliminary step in many applications; for instance, face recognition, facial expression analysis, content based image retrieval, surveillance system and intelligent human computer interaction.

The face detection process is becoming a significant challenge due to various factors associated with the detection process, including the influence on the performance of any proposed algorithm.

These factors include [18]:

1. **Occlusion:** this occurs in cases where the face on an image is partially hidden (occluded) by other objects, which leads to the occlusion of some facial features.

2. **Image orientation:** Face images directly vary relative to different rotations about the camera's optical axis.

3. **Imaging conditions:** When an image is formed, factors such as lighting (spectra, source distribution and intensity) and camera characteristics (sensor response, lenses) affect the appearance of a face.

4. **Pose:** The images of a face vary due to the relative camera-face pose (such as frontal, 45 degree, profile, inverted), and some facial features such as an eye or the nose becoming partially or wholly occluded.

5. **Presence or absence of structural components:** Facial features, such as beards, mustaches and spectacles, may or may not be present and there is a great deal of variability of these components, including shape, color and size.

Therefore, since the 1970s, many research papers have been proposed in face detection with the aim of overcoming these factors and to provide a higher detection rate However, until now no one has been able to devise a method that works at a 100-percent successful detection rate.
Moreover, the literature survey shows that whenever the computation time is increased, the face detection rate will increase and vice versa.
In spite of these difficulties, many approaches have been proposed in this area, which can be classified in four categories.

2.2.1 Knowledge-based methods

The rule-based methods being used to detect a face by encoding human knowledge about a typical face include criteria such as the face shape as an ellipse, two eyes being symmetric, a nose in the middle and a mouth underneath the nose.
"One problem with this approach is the difficulty in translating human knowledge into well-defined rules. If the rules are detailed (i.e., strict), they may fail to detect faces that do not pass all the rules. If the rules are too general, they may give many false positives. Moreover, it is difficult to extend this approach to detect faces in different poses since it is quite a challenge to enumerate all possible cases. On the other hand, heuristics about faces works well when detecting frontal faces in uncluttered scenes" [19].

2.2.2 Template matching methods

These methods are used to find a good match between a pre-defined face (template) and any input sub-regions of the image. Since the pre-defined face is usually a frontal face, such methods do not stand well in cases of occlusion wherein certain facial features, such as the eyes, nose etc. are hidden. Moreover, it cannot effectively deal with variation in scale, pose and shape.
Nevertheless, it has been proved that these methods are easy to implement and work well in the case of frontal faces.
The face template is created by applying different ideas, one of which is to find eigenfaces from training face images, which Smita Tripathi, Varsha Sharma and

Sanjeev Sharma have proposed [19]. Another idea is to find an average face image from multiple grayscale face images as proposed and attempted by S. Kherchaoui and A. Houacine [20], by finding an average for face samples after converting it to binary format [21].

After the face template is prepared, the matching processing is carried out by using different techniques. For instance, cross-correlation is used in order to find the degree of similarity between the face template and the candidate regions. It is then compared with the fixed threshold to make a decision whether or not the candidate regions contain a face [20]. Another method for matching is to find the Hausdorff distance between the face template and each sub-window in order to determine whether or not an image contains a face [19].

2.2.3 Machine learning-based methods

These methods are learned to detect a face by using training sets containing positive (face) and negative (non-face) samples. The performance of these methods has been evaluated by using many metrics, including the number of training samples being used in the training stage (which totally has an impact in the accuracy of detection), learning time, execution time and detection rates. Moreover, it cannot deal effectively with any variation of scale, pose or shape.
In the literature survey, common examples of learning-based methods are listed below:

1. **Neural Network:**
   This method is commonly applied in many pattern recognition problems, such as object recognition and face detection.
   There are different kinds of neural network structures that are used in face detection, such as MLP (Multi-Layer Perception) [22], SOFM (Self-Organized Feature Map) [23]. There are also many algorithms used to train neural networks, such as back-propagation [24] and Scale Gradient Conjugate (SCG) [25].

2. **PCA (Principal Component Analysis):**
   Principal Component Analysis is one statistical method that is widely used to reduce a data space [26] and to find a smaller representation for a large data space without losing a general characteristic. Therefore, this method is used in image compression.
   Furthermore, PCA has become a face detection method aimed at finding a face space that is a better description for a face. This consists of three main stages: training the image, reconstructing the image, and comparing the image.

2.2.4 Feature-based methods

These methods are used to detect a face according to exiting facial features on a digital image for instances of skin color, face shape, eyes, mouth, eyebrows, lips etc. individually or by finding a relationship between these facial features.
The Viola-Jones algorithm is considered to be one of the common feature based methods that are used to detect a face by scanning a sub-window over the image at different scales with the aim of searching for specific Haar features of the human face. Moreover, it has been proved that this is a highly rated method in comparison to other proposed methods [27].

However, based on literature surveys, it has been shown that using one category is not sufficient to achieve a higher detection rate due to complexity and the factors already mentioned influencing face detection. Therefore, only some methods are applicable to every situation. Thus, several features and methods should be considered simultaneously for any suggested face detection method.

2.3 Facial Expression Classification

Facial expression plays an important role in our daily activities. The human face is a rich and powerful source full of communicative information about human behavior and emotion [28].

Recently, various techniques have been developed with the aim of providing to a computer the capability of automatic facial expressions classification or recognition. A majority of these techniques are carried out in two stages: the *face detection stage*, which used to locate a face region for the input digital image or sequences of images (video), after which the detected face is pass down to second stage which is the *facial expressions classification stage* in order to classify each class (neutral, happy, surprise, etc.) according to the extracted facial features.

A brief survey of some of the existing proposed techniques in facial expressions classification is presented below:

1. Ioanna-Ourania Stathopoulou and George A. Tsihrintzis [29] proposed a system that is used to classify three types of facial expression (neutral, smile, surprise). The system consists of two modules: firstly, face detection based on features extracted from Sinha's template as a pre-process. The candidate image is then fed into the Artificial Neural Network (ANN) to determine whether or not there is a face in the image. Secondly, the facial expression classification is carried out by extracting a feature vector from the detected face after converting it into a binary format.

   This feature vector consists of sixteen feature points (four corner points of the mouth, eight corner points for the two eyes and four corner points for the eyebrows). The Euclidean distances between these points are calculated and the results are fed as an input vector into a trained Neural Network (NN) to make a decision.

2. Abu Sayeed Md. Sohail, and Prabir Bhattacharya [5] proposed a method that is used to classify six basic facial expressions (anger, disgust, fear, happiness, sadness and surprise) from static images by automatically detecting eleven feature points on the eyebrows, eyelids, mouth and nose, after which a set of

eleven measurement distances is performed between these points which are used as input vectors for three classification methods: Support Vector Machines (SVM), k-NN, Neural Network (NN), Naïve Bayes Classifier, which achieved an average successful recognition rate of 89% and 84%.

3. Renuka R. Londhe, Vrushsen P. Pawar [30] proposed a method which is able to classify six universal facial expressions by using the Local Binary Pattern (LBP) technique to find the binary representation for each pixel for the input face image after which a feature vector is computed by finding the histogram to feed it into a trained artificial neural network in order to classify it into different classes, such as anger, disgust, fear, happy, neutral, sad and surprise. Their approach achieved 93% on average.

4. S L Happy, Anjith George, Aurobinda Routray [31] proposed an approach which is applicable in real time facial expression classification and is able to classify the six basic emotions in frontal grayscale face images.
   Initially, their algorithm attempts to localize the face in an image by using the Haar classifier based method and then the blocks of the Local Binary Pattern histogram features are extracted from the detected face by using the Local Binary Pattern method. Finally, the trained PCA is used to classify each emotion according to these Local Binary Pattern blocks. The performance of their approach exceeds 97%.

5. Ying Zilu1, Fang Xieyan [32] proposed an idea for facial expression classification which is tested in the JAFFE database to yield a recognition rate of 65.71%.
   Their approach consists of a feature extraction which is carried out by using a local binary pattern after which the feature selection is obtained by using an advanced Adaboost algorithm in order to discard any unusable features and increase classification accuracy and performance. Finally, the support vector machine is used for classification.

6. Turan Güneş and Ediz Polat [33] proposed a method that is used for facial expression classification consisting of three main steps: first, feature extractions are carried out by using Gabor filters with different orientations; second, a feature selection process to overcome unusable features; and finally, classification is carried out by using a support vector machine.

7. Seyed Mehdi Lajevardi and Margaret Lech [34] used a set of characteristic features obtained by averaging the outputs from the Gabor Filter Bank with 5 frequencies and 8 different orientations rather than using the old proposed method of the entire Gabor Filter Bank outputs, which improves the efficiency of the method and decreases the complexity of computation.
   Furthermore, PCA is used for further reduction of the feature dimensionality after which classification is carried out by using the K-Nearest Neighbors method. The recognition rate is 94.52% for the full Gabor Filter Bank outputs, whereas it is 93.15% for the averaging Gabor Filter banks.

8. Luning. Li, Jaehyun So, Hyun-Chool Shin, Youngjoon Han [35] proposed an approach which is able to classify six basic facial expressions according to face shape. They began with the assumption that people have three different face shapes; melon seed, round and square.
   Thus, they used two facial expressions databases to find the AAM model and train the SVM, which has a total of 1485 images and to check the performance of the system. They used 180 images which were collected randomly. The recognition rate is 90% for the melon seed, 93.3% for the round and 85% for the square face shapes.

9. Odoyo O. Wilfred, Geum-Boon Lee, Jung-Jin Park, Beom-Joon Cho [36] used AAM to obtain the shape and texture model for six basic facial expressions by manually landmarking points for each image on the training data in each class after which PCA was used to define the mean shape of each class.

Subsequently, the Mahalanobis distance algorithm was used to find the similarity between the test image and the mean shapes, according to which classification was carried out.

From the literature survey, it is observed that facial expression classification approaches consists of two major steps: the first being the facial feature extraction step, which is a process to extract facial changes caused by a person reacting to a particular event. This can be reflected on facial expressions by using one of the feature extraction methods. The second major step is classification, in which each class is classified according to the extracted features in the previous step by using one of the classification methods.

2.3.1 Facial feature extraction

Feature extraction is considered to be an essential and crucial step in many applications such as face recognition, facial expression classification and access control systems. Moreover, it has been proved that humans are able to show emotions through facial expressions; therefore, extracting such information would be useful to read human internal emotional states, intentions or social communications.

In the past two decades, many methods have been proposed for facial feature extraction with the aim of extracting the most powerful features that do not contain irrelevant data and are less noisy. This has a major influence on the performance of any classification method which uses these features as an input vector.

In facial feature extraction for expression analysis, there are two categories of these methods: geometric feature-based methods and appearance-based methods. The geometric facial features present the shape and locations of facial components (including the mouth, eyes, brows and nose). The facial components or facial feature points are extracted to form a feature vector that represents the geometry of the face. With appearance-based methods, image filters, such as the Gabor filter or Local Binary Pattern (LBP), are applied to either the whole face or to specific regions in a face image in order to extract a feature vector.

Subsequently, this feature vector is used to feed into one of the classification methods, such as the Support Vector Machine (SVM), Neural Network (NN), Decision Tree and Bayesian.

2.3.1.1 Local binary pattern

The Local Binary Pattern operator represents one of the most popular texture descriptors. It was first introduced by T. Ojala, M. Pietikäinen, and D. Harwood [37] in 1994, and was used to extract local texture characteristics. It has subsequently become widely used in various applications for instances face image analysis, image and video retrieval, environment modeling, visual inspection, motion analysis, biomedical and aerial image analysis, and remote sensing [38].

The Local Binary Pattern operator has the advantage of robustness to monotonic illumination change and computational simplicity, thereby making it attractive in comparison to other feature extraction methods.

Generally, the original LBP operator is used to produce binary representations for any pixel in an image by shifting 3×3-sized window over the image. The window center pixel represents a threshold value, while the adjacent eight pixels around the center pixel and their gray scale values are used for comparison. If the value of the pixels around the center pixel is greater than the center pixel value; the pixel location is marked as 1, otherwise it is marked 0. Figure 4 shows and explains how to compute the LBP value for one pixel.

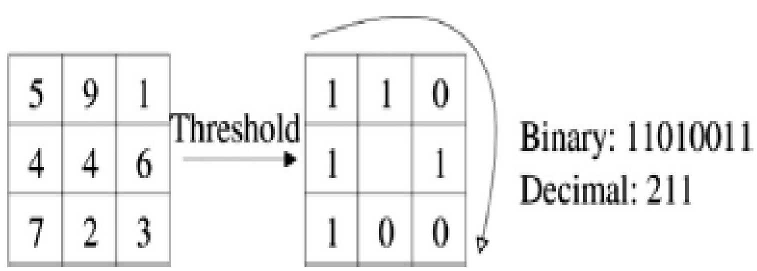


Figure 4: Example of the basic LBP operator [38]

Thus, the 3×3 neighborhood of 8 points in comparison produces an 8-bit binary number (usually converted to a decimal number that LBP codes, a total of 256 texture patterns can be produce).

Another approach to compute the LBP code of each pixel in the image is to subtract the center pixel from the neighborhood pixels by using the following equation:

$$LBP_{N,R} = \sum_{i=0}^{N-1} S(N_i - N_c)2^i, S(x) = \begin{cases} 1, x \leq 0 \\ 0, x > 0 \end{cases}$$

where N_c is the gray value of the central pixel, N_i the gray value of neighboring pixel, i=0,…, N–1, N the total number of involved neighboring pixels and R being the radius of the neighborhood which determines how far the neighboring pixels are located away from the center pixel. S(x) = 1 if x ≥ 0 else S(x) = 0 [30].

However, in the past few years, researchers have been trying to improve and optimize the basic local binary pattern operator to overcome the limitation of the 3×3 window size that cannot capture dominant features with large-scale structures.

Therefore, researchers proposed the idea known as Extended Local Binary Pattern (ELBP), which is generalized to use neighborhoods of various sizes as shown in Figure 5.

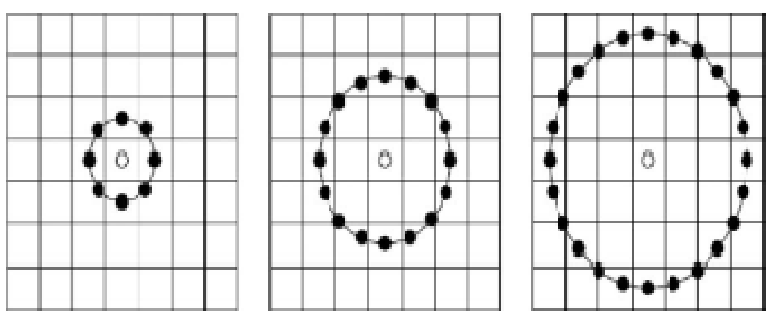


Figure 5: Examples of the ELBP operator, the circular (8, 1), (16, 2), and (24, 3) neighborhoods [38]

2.3.1.2 Gabor filter

The Gabor filter is a common feature extraction method originally introduced by Dennis Gabor (Gabor, 1946). It is an image processing and computer vision method that has been successfully used in many applications, such as facial expression classification [39], texture segmentation/classification, character recognition, fingerprint recognition, face recognition, edge detection [40], image analysis and compression.

Gabor filters are basically band pass filters which operate in a frequency domain to extract a feature from any input face by defining a filter bank containing filters with different orientations to extract a feature with different characteristics [33].

In spite of the popularity of Gabor filters for feature extraction, it may suffer from the loss of low and high frequency information. Furthermore, computational complexity high resource requirements make this inapplicable in real-time applications [31].

2.3.1.3 Active appearance model

The Active Appearance Model algorithm is an image processing algorithm and an extension of Active Shape Models (ASM) [41] [42], which are widely used in the field of pattern recognition and facial expressions classification as a feature extraction method.

The Active Appearance Model algorithm is able to extract a single model of appearance by combining the shape and texture information of images and this model is built during the training phase by manually landmarking the points for each image in the training set. The final model (template) is produced and is used for the matching process by using one of a number of statistical methods, such as Principal Component Analysis (PCA).

Subsequently, this extracted model or template is used to fit with any test image to find a similarity by using one of the fitting algorithms, such as the Fixed Jacobian Method (FJ), Project Out Inverse Compositional Method (POIC), Simultaneous Inverse Compositional Method (SIC), Robust Inverse Compositional Method (RIC), Iterative Error Bound Minimization Methods (IEBM) and Haar-like Feature Based Iterative Discriminative Method (HFBID) [43].

2.3.2 Facial feature classification

The last step of any facial expression classification system is to classify the extracted feature vectors from the input face image to one of the six universal emotional expressions (fear, anger, disgust, happiness, surprise and sadness) by using one of the classification methods, for instance Neural Network (NN) [29], Support Vector Machine (SVM) [33], Principal Component Analysis (PCA) [31], Naïve Bayes Classifier [44], *k*-Nearest Neighbor [34], Rule Based method.

Referring to the paper published by Abu Sayeed Md. Sohail, and Prabir Bhattacharya [5], we see that they attempted to use different classification methods to classify extracted facial features, and according to the experimented result, indicate that the performance of Support Vector Machine (SVM) is better than other classification methods.

CHAPTER 3

METHOD ONE

Referring back to the literature survey, the main concept of the first proposed method is derived from template matching methods that are used for face detection. Therefore, in our first attempt, we applied the same concept to facial expressions classification.

The basic idea behind the method is to define a face template for each basic facial expression class (Happy, Fear, Surprise and Neutral) and then use it to detect the emotional mood of any input face image by finding a good match between them.

The method does not contain any pre-processing steps, such as skin color detection or face detection. Even the classification is carried out by using a rule-based classification method rather than using any learning-based classification methods, such as Support Vector Machine, Neural Network etc.

For this reason, the method is somewhat faster than other proposed facial expressions classification methods, thereby making it applicable to any real-time application.

The method is implemented by using MATLAB version R2013a with the detection efficiency being evaluated by using the JAFFE database.

3.1 Image Gradient

Image Gradient is one of the image processing techniques that is widely used in image analysis, including object detection, edge detection, robust feature extraction and texture matching [45].

The Image Gradient technique is able to extract two pieces of information from any digital image: the magnitude of the gradient, which tells us how quickly the intensity of the image is changing; and the direction of the gradient, which (self-explanatorily) most rapidly tells us in which direction that the intensity of the image is changing.

In our proposed method, we use this algorithm to apply an image gradient to the input face image before the matching process in order to overcome the lighting change and camera properties which have a significant influence on the matching algorithms.

MATLAB provides us with two functions (**imgradient, imgradientxy**) with which we can easily compute the gradient magnitude and the (X, Y) directional gradient, as shown in Figure 6, for a test face image.

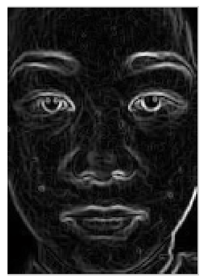
Neutral	Gradient magnitude	X-Directional Gradients	Y-Directional Gradients
			

Figure 6: Gradient and directional gradient

According to the conclusion of our data analysis, which was reached by using different face images and then using the outcome of the gradient and directional gradient for the matching process, we found that applying the Y-directional gradient to the face image would increase the percentage of matching because the extracted information is less noisy than the gradient magnitude and X-directional.

Therefore, with this method, the Y-directional gradient is used as a filter before the matching process.

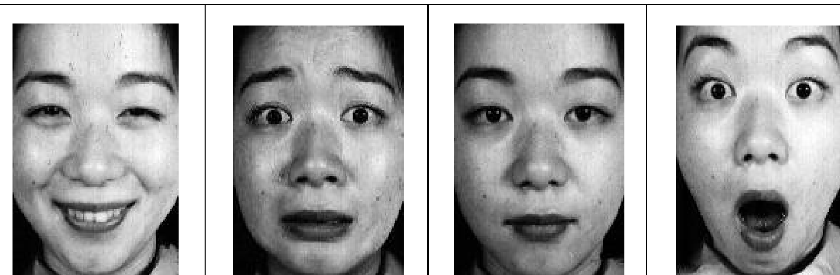
3.2 Method Description

The basic structure of the method consists of two main stages:

1. Template definition stage

The aim of this stage is to create a face template from one face image for each facial expressions class, these being (Happy, Fear, Surprise and Neutral) by using the following three steps:

a) In the first step, four faces from the **JAFFE** database are randomly selected as a template for the required facial expression class, as shown in Figure 7.



Happy	Fear	Neutral	Surprise

Figure 7: Selected face template for the four classes

b) The second step is used to find the Y-directional image gradient using the MATLAB function **imgradientxy** for each selected face template. The results for the four selected face templates are shown in Figure 8.

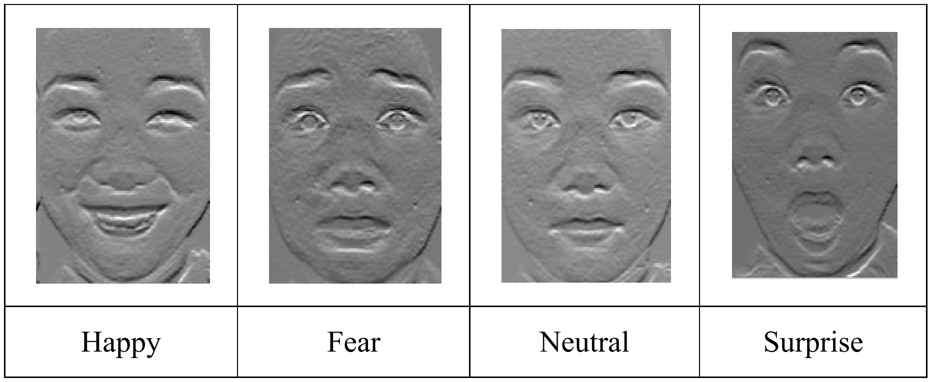


Figure 8: Applying the Y-directional gradient to each face template

c) Finally, the outcomes of the second step are converted to vector form and saved in four different variables for one class, as shown in Figure 9.

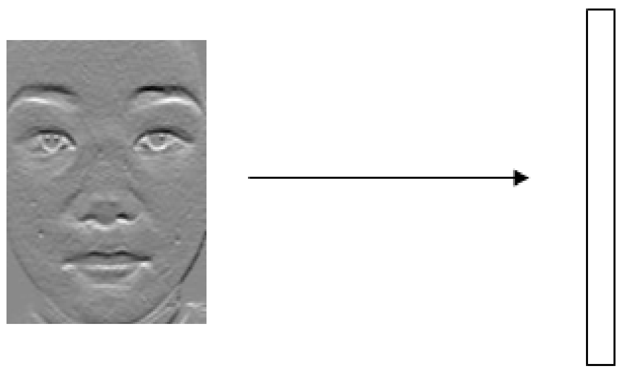


Figure 9: Converting the face template to the vector form

2. Matching stage

After the four face templates are prepared and defined as a vector form on the previous stage, for any input test face, we apply the same steps to the first stage in order to find a vector form. Then by comparing the input vector with pre-defined vectors, the matching process is carried out.

Many techniques are used to find a good match between the two vectors. A common technique is to find a correlation coefficient which measures the degree of similarity between the two vectors.

In our method, the correlation coefficient is found by using the MATLAB function (**corrcoef**) between the input test face vector and the four pre-defined face templates. According to this coefficient, the decision is made to detect the emotional mood of the input face.

3.3 Experimental Results

The proposed method is implemented using the MATLAB environment. The performance of the method is evaluated using the JAFFE database (APPENDIX A), which is considered to be one of the common facial expression databases widely used for this purpose.

Initially, we endeavored to match the input face image and the pre-defined four templates as a whole. The outcome for the Happy database is twelve true detections from twenty face images, as shown in Table 2, in which the first row of the table represents a number of the face, while the second row represents one of these symbols: (H=Happy, F=Fear, S=Surprise, N=Neutral).

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N	H	H	N	H	H	H	H	H	F	F	H	H	H	S	H	N	S	N	H

Table 2: Outcomes for the Happy Database

The outcomes for the Neutral database are nine true detections from twenty face images, as shown in Table 3.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N	N	S	H	N	F	F	N	N	F	N	N	S	S	N	F	N	F	H	F

Table 3: Outcomes for the Neutral Database

The outcomes for the Surprise database are nine true detections from fifteen face images, as shown in Table 4.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
H	S	N	S	S	S	S	S	F	F	S	S	N	N	S

Table 4: Outcomes for the Surprise Database

Finally, the outcomes for the Fear database are five true detections from sixteen face images, as shown in Table 5.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
F	F	S	N	N	F	N	H	N	H	F	N	N	F	N	S

Table 5: Outcomes for the Fear Database

According to the outcomes, the detection rate is 60% for Happy, 45% for Neutral, 60% for Surprise and 31% for Fear, which did not meet with our expectations. Therefore, we improved the idea by dividing the input face and four pre-defined templates into two parts (upper and lower) rather than using the face as a whole. We then found a correlation coefficient for the corresponding parts, as shown in Figure 10.

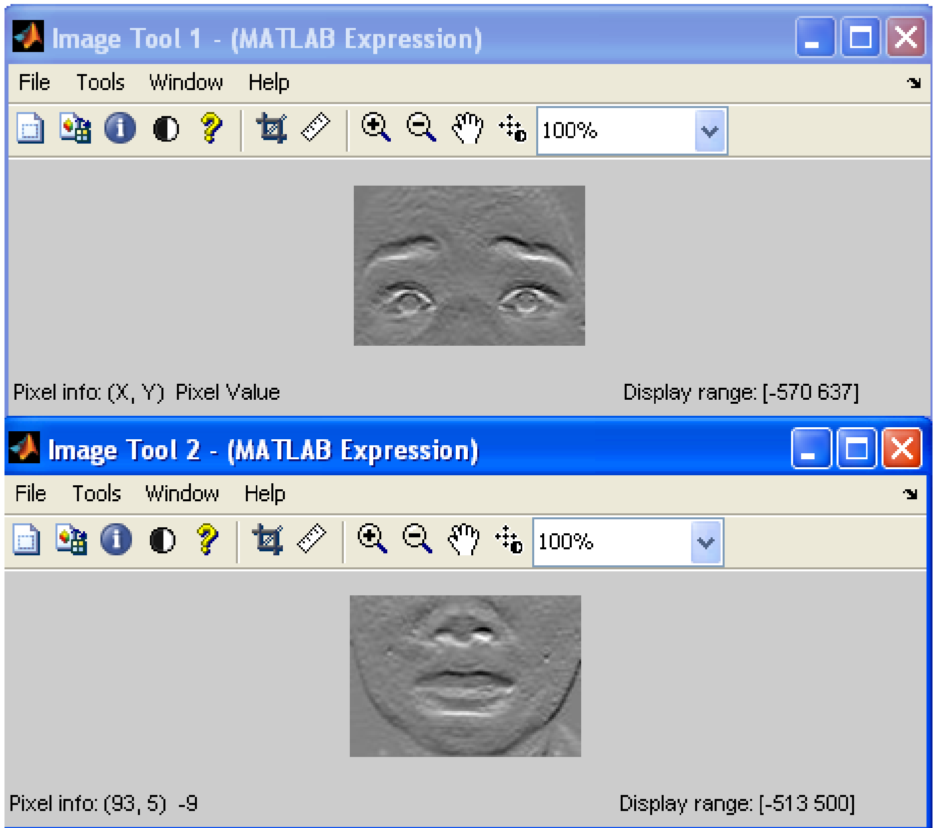


Figure 10: Dividing the face template into two parts

The outcomes for the new improvement in the Happy database, as shown in Table 6, is twelve true detections from twenty face images, which shows no change.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Upper	S	H	H	S	H	F	H	S	F	S	S	H	H	N	S	H	N	S	H	H
Lower	N	H	H	N	H	N	H	H	H	F	F	H	H	H	H	H	F	H	N	N

Table 6: Outcomes for the First Improvement in the Happy Database

The outcomes for the Neutral database are improved, as shown in Table 7, with thirteen true detections from twenty face images.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Upper	N	H	H	H	S	S	F	H	H	H	H	H	S	S	N	S	S	S	H	F
Lower	N	N	S	S	N	N	F	N	N	N	N	F	N	N	F	F	N	F	F	N

Table 7: Outcomes for the First Improvement in the Neutral Database

The outcomes for the Surprise database remains as it did previously; that is, nine true detections from fifteen face images, as shown in Table 8.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Upper	H	S	F	F	S	S	S	S	N	F	S	S	N	F	F
Lower	H	S	N	S	S	S	S	S	S	S	H	H	H	S	S

Table 8: Outcomes for the First Improvement in the Surprise Database

Finally, the outcome is six true detections from sixteen face images for the Fear database, which indicates a slight change, as shown in Table 9.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Upper	N	F	N	N	S	S	F	H	N	H	F	F	S	S	N	S
Lower	F	F	F	F	N	N	N	H	F	F	F	N	N	N	F	S

Table 9: Outcomes for the First Improvement in the Fear Database

The first improvement increases in the detection rate for Neutral from 45% to 65% and for Fear from 31% to 37%, while no change occurs in the Happy and Surprise databases.

The final improvement is to divide the input face and the four pre-defined templates into four parts after which a correlation value is calculated for each part, as shown in Figure 11.

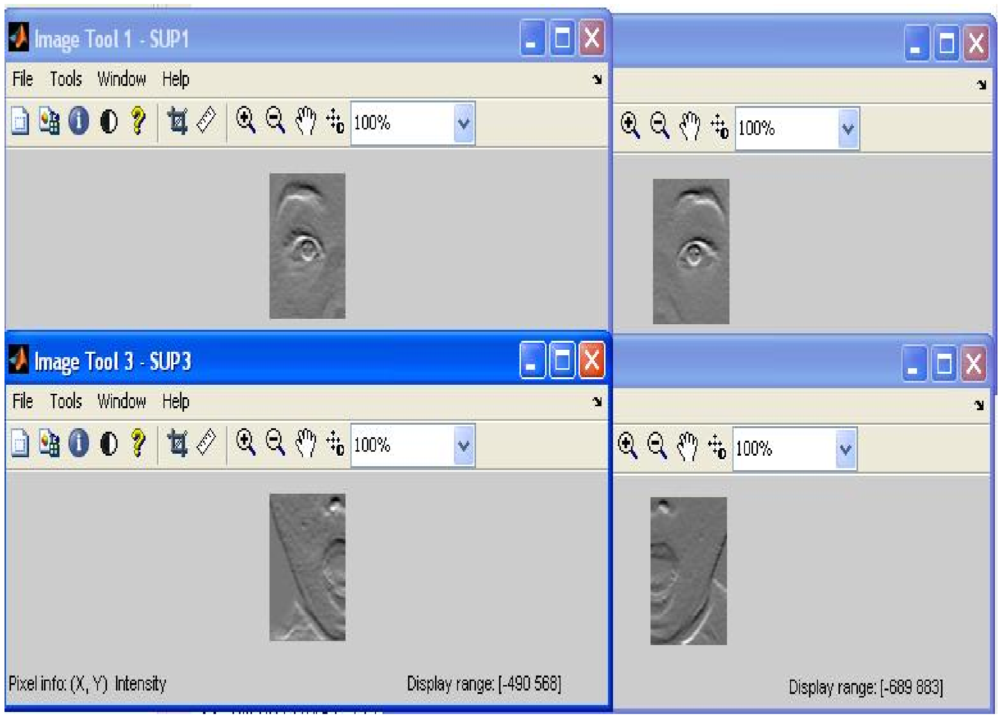


Figure 11: Dividing the face template into four parts

The outcome is twelve true detections from twenty face images for the Happy database, as shown in Table 10.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N	H	H	N	H	F	H	H	H	S	N	H	H	H	H	H	F	H	N	N

Table 10: Outcomes for the Second Improvement in the Happy Database

The outcome is ten true detections from twenty face images for the Neutral database, as shown in Table 11.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N	N	S	S	N	F	F	N	N	N	N	N	N	S	F	F	S	F	N	F

Table 11: Outcomes for the Second Improvement in the Neutral Database

The outcome is ten true detections from fifteen for the Surprise database, as shown in Table 12.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
H	S	N	S	S	S	S	S	S	H	S	S	H	U	S

Table 12: Outcomes for the Second Improvement in the Surprise Database

Finally, the outcome for the Fear database is seven true detections from sixteen, as shown in Table 13.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
F	F	H	F	F	F	F	H	H	F	N	N	N	N	H	S

Table 13: Outcomes for the Second Improvement in the Fear Database

The last improvement is increasing the detection rate for Surprise from 60% to 66% and Fear from 31% to 43%, while no changes occur in the Happy and Neutral databases.

CHAPTER 4

METHOD TWO

This chapter explains the second proposed method on mood detection according to facial expression for three basic expressions (neutral, happy and surprise).

The second idea differs totally from the first proposed method and consists of three major steps, first of which is skin color detection in cases where the input image is a color image, while this step is neglected if the input image is grayscale. The second step is face detection, which is used to locate the face region. Finally, the mood detection step is used to detect the emotional mood of a person according to the extracted feature information from the facial expression.

Basically, the extracted feature information is formed as a vector consisting of four feature points: three of the feature points are located in the mouth region while the fourth point is located in the mid-region of the nose.

Generally, the method is implemented using MATLAB version R2013a and the performance of the method has been evaluated by using JAFFE and IMM databases.

4.1 Skin Color Detection

Skin color detection is commonly used as a pre-process step in many computer vision applications, such as face detection, face tracking, human motion tracking, hand segmentation for gesture analysis, filtering of objectionable Web images and other human-related image processing applications.

Since skin color is one of the stable features in the human face, we use a skin color detection step with the aim of accelerating the face detection process by removing non-skin pixels from any input color image. However, this step is rendered no longer useful when the input image is grayscale or black-white color.

Referring back to the literature survey, any skin color detection method comprises two main steps:

1. Extracting a feature vector by finding a suitable color space such as RGB (Red Green Blue), NTSC, YCbCr, HSV, CMY, HIS, YUV, CIE-Lab etc.

2. A skin color classifier (Model) being used to define the skin color boundaries in a specified color space in order to distinguish skin color from non-skin color.

In our thesis, the YCbCr color space is used in which Y represents the luminance component while the Cb and Cr together represent the chrominance component (blueness and redness). With this useful property of this color space, the luminance and chrominance components can be easily separated. In this case, being able to overcome the lighting variation will have a significant influence on skin color detection. Another useful property is that the composition principle of this color space is identical to human visual perception processes [46].

Moreover, skin color in the YCbCr color space is a more compact cluster relative to other color spaces. It has been proved [1] that the chrominance component of skin color for the different human races are ranged thus: $77 \leq Cb \leq 127$ and $133 \leq Cr \leq 173$. The luminance component is scattered according to the lighting conditions on the scene, which can therefore be ignored since it lacks useful information.

Thus, to check the proposed range having been imposed by Chai, and Ngan [47] for (Cb and Cr), we download a picture from the National Geographic web page, as shown in Figure 12, [48] which contains the skin colors of different human races and is used as a skin color database.

Figure 12: Human rainbow of skin colors

We manually cut off a skin region for each human race in the downloaded picture in order to obtain an image that unifies all these different skin tones, which we can use as a skin color database, as shown in Figure 13.

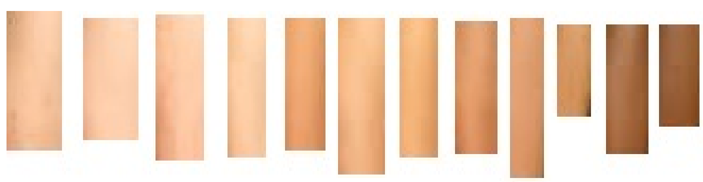


Figure 13: Skin color database

Then we find a histogram for the extracted skin color database (Figure 13). The outcome, as shown in Figures 14, 15 and 16 respectively, show that the chrominance component (Cb and Cr fall on the range as mentioned by Chai, and Ngan [49].

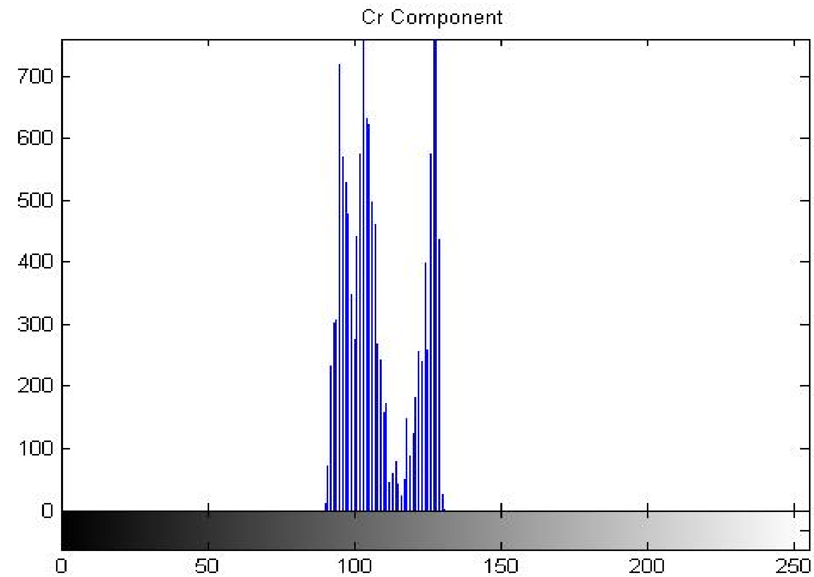


Figure 14: Histogram of Cr chrominance values of the skin color database

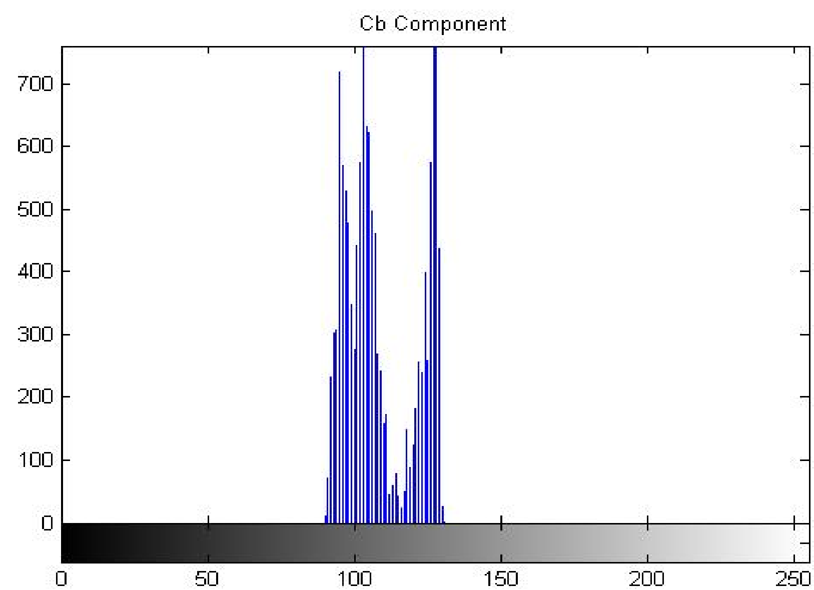


Figure 15: Histogram of Cb chrominance values of the skin color database

While the luminance component (Y) is scattered over a wide range, as shown in Figure 4.4, it can be ignored due to its containing unusable information.

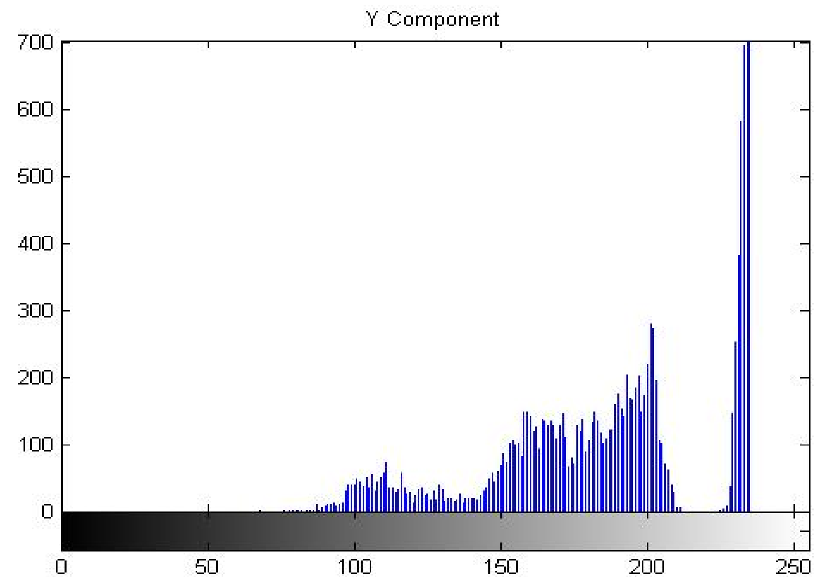


Figure 16: Histogram of luminance (Y) of the skin color database

This implies that skin color on the YCbCr color space is clustered in a specific range. In order to distinguish a skin color from a non-skin color, we used one of the simplest skin color classification approaches (explicitly model), whereby the boundaries of the skin color cluster are defined on the YCbCr color space. The outcome for the skin color detection in Figure 12 is shown in Figure 17.

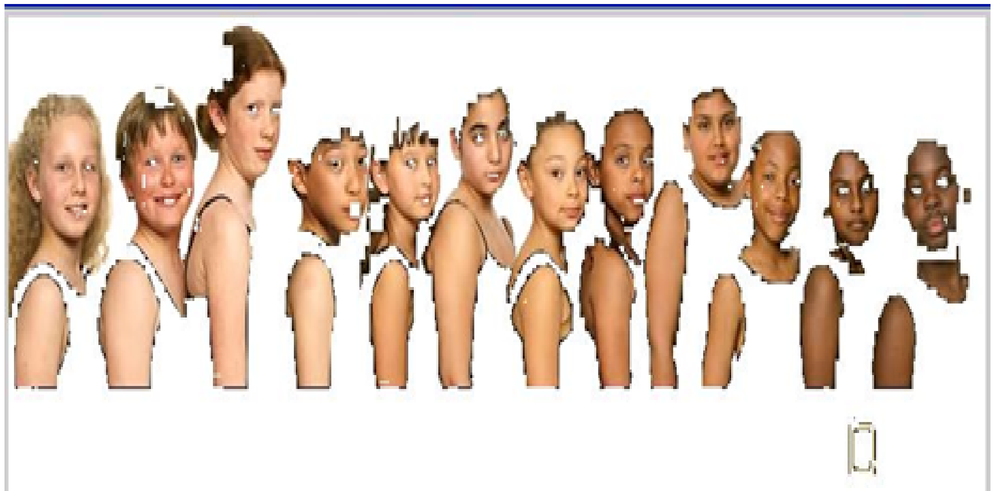


Figure 17: Skin color detection for the first test image first approach

Moreover, the same code is applied to the second complex test image (Figure 18) in order to evaluate the method.

Figure 18: Second test image

The outcome of skin color detection by using same the approach is shown in Figure 19.

Figure 19: Skin color detection for the second test image first approach

The outcome shows that there is some false detection which might be due to the skin color modeling approach.

Therefore, we used the same color space YCbCr with another skin color modeling approach, namely the neural network. The structure of the neural network consists of two input layers, twenty hidden layers and two output layers.

Initially, we arranged the input and output vectors to train a neural network. The input vectors consist of two columns (Cr and Cb) for skin and non-skin values, which are extracted from a skin color database (Figure 4.2) after converting it to a YCbCr color space, while the non-skin values are extracted from different images not containing skin color values. The output vectors consist of two columns: 1 for skin, and 0 for non-skin color values.

For implementation, the MATLAB graphical user interface (**nftool**) tool is used to design the structure of the neural network and to select the Levenberg-Marquardt (**trainlm**) algorithm to train it after feeding it with the extracted input and output vectors.

Subsequently, the trained neural network is used to detect skin color from the first test image and the outcome is shown in Figure 20.

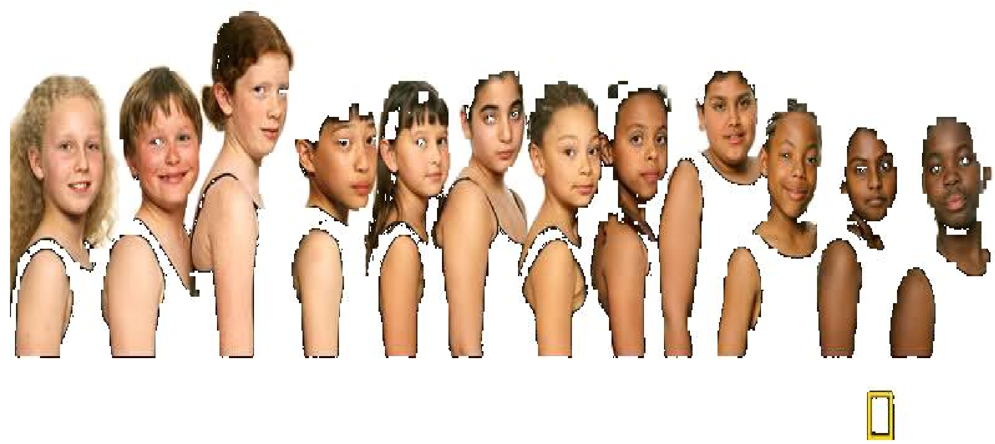


Figure 20: Skin color detection for the first test image second approach

The outcome for the second test image is shown in Figure 21.

Figure 21: Skin color detection for the first test image second approach

By comparing the outcomes of the two approaches, we can conclude that the neural network skin modeling approach is better than the explicit modeling in the same color space. Therefore, it was used in our project as a pre-processing step for any input color static image before passing it down to the next step of face detection in order to accelerate the face detection process and overcome background complexity.

4.2 Face Detection

The Second permanent step prior to detecting the mood of any person according to facial expression is to detect a face from any input static image, whether color or grayscale image.

In the past few years, there have been several approaches published on face detection, one of which is the Viola-Jones algorithm, which is listed as a feature-based approach that is used to detect a face by locating facial features, such as the eyes, nose, mouth, etc.

Additionally, the Viola-Jones algorithm is considered to be one of the most powerful algorithms with a high rate of detection when compared with other existing face detection methods, it is clear to see why it is used in our proposed system to detect faces.

The Viola-Jones algorithm [49][50] was published in 2001 and it was the first object detection classifier that has been used widely to detect different objects, such as faces, stop signs or cars in digital images.

The inner component of this algorithm consists of features known as Haar-like features; Figure 22 shows some basic Haar-like features. However, the algorithm needs to be trained by using a set of positive (containing interesting objects) and negative (not containing interesting objects) samples before using it to detect a face.

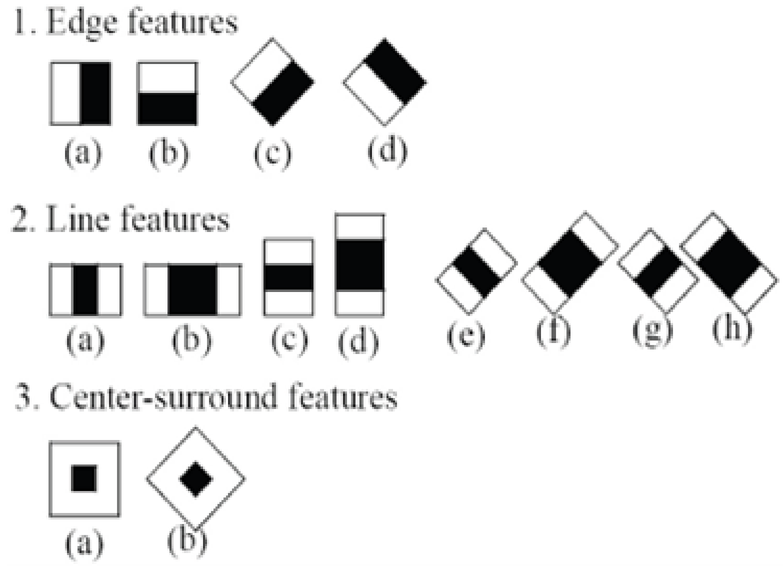


Figure 22: Prototypes of Haar-like features

Generally, the algorithm is operated by sliding a sub-window over a digital image with different scales with the aim of finding Haar features. The human face contains some of these, as shown in Figure 23. We use different sub-window scales in order to find the faces of different sizes that the image might contain.

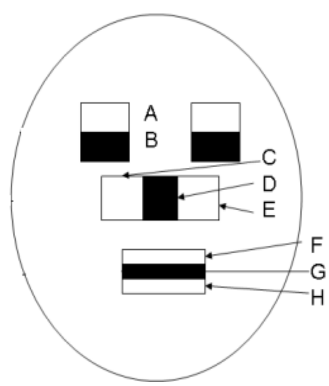


Figure 23: Example of face features

Furthermore, there are two important usages of this algorithm which lead to high detection rates: the first being the integral images technique, which is used to convert the input image into an integral form. This has the advantage and benefit of speeding up the calculation of the Haar features values on each sub-window. We are able to overcome any computational complexity with this feature.

The other important use of this algorithm is the cascading technique (as shown in Figure 24), which is the internal component of this algorithm and is divided into many stages where each stage consists of a number of weak features. The main advantage of designing the stages is to reject the negative samples as quickly as possible in the early stages.

If the sub-window that slides over the input image contains the object of interest, such as a human face, the sub-window should pass through all these stages in order to evaluate and label it as a positive sample. Otherwise, it is rejected at any stage. In this case, the sub-window does not contain the required object of interest and is thus labeled as a negative sample.

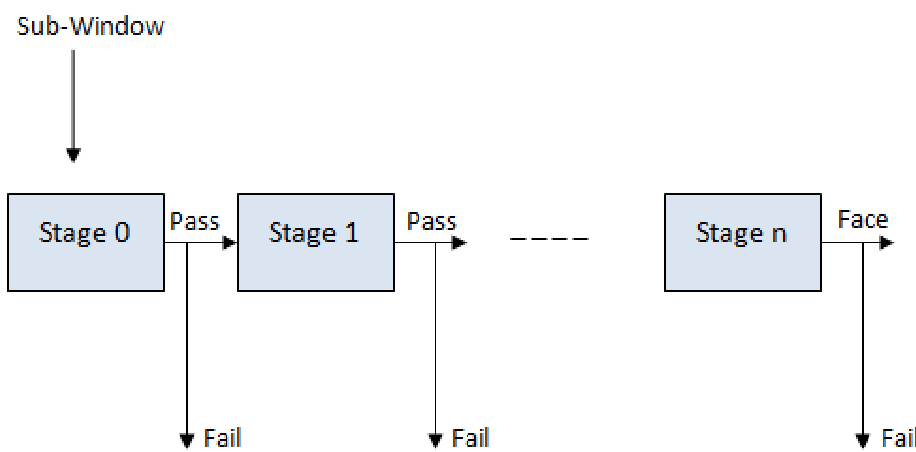


Figure 24: Structure of cascading

Regarding the implementing of the Viola-Jones algorithm, the MATLAB package (vision.CascadeObjectDetector) (which was developed by the MATLAB computer vision team) is used to detect people's faces, noses, eyes, mouth or upper body. Therefore, we used this package to detect a face in the computer vision team image. The result of the face detection is shown in Figure 25.

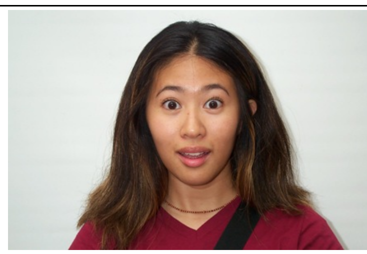
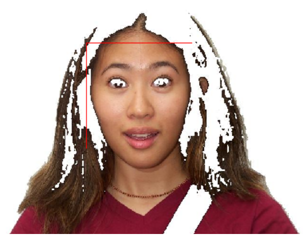
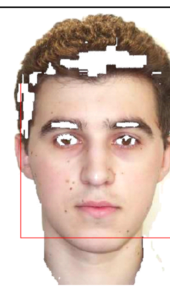
False detection

Figure 25: Detected faces using the Viola-Jones algorithm

According to the outcome, one false detection occurred, which is due to the algorithm mistakenly classifying that region as a face. In order to overcome this false detection and increase the face detection rate, we apply the skin color detection first as a pre-processing step, after which we pass down the outcome of the skin color detection to the face detection process by using the same algorithm. The outcome for the same vision team image is shown in Figure 26, which indicates no more false detections. Additionally, the face detection process is now faster.

Figure 26: Combination of skin color detection and face detection

Table 14 shows other samples that were collected from the Internet in order to check face detection for different races, ages and genders.

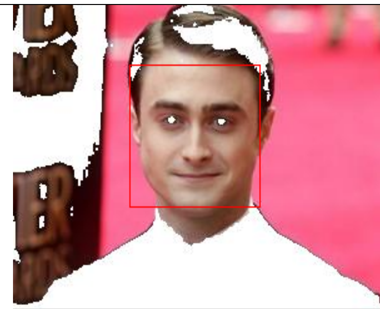
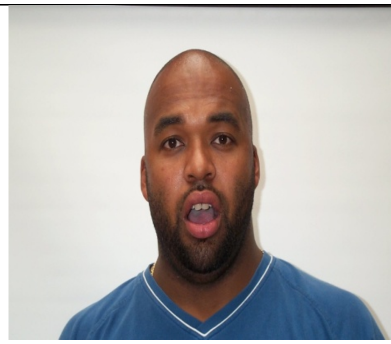
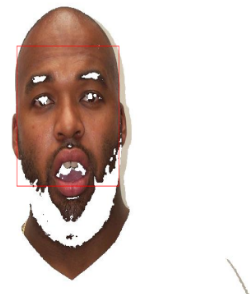
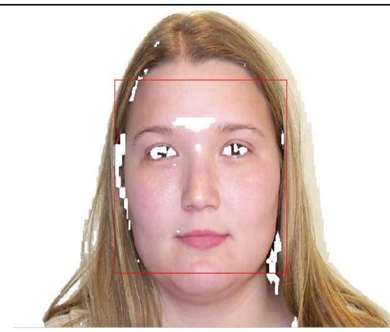
	
	
	

Table 14: Face Detection for the Different Samples

4.3 Mood Detection

After any face has been detected in the previous step, the final step of the proposed method is to detect a mood according to the extracted feature from the facial expression. In order to accomplish this on static images, an idea consisting of two main steps was implemented.

4.3.1 Feature point locating

The first step concentrates on locating four feature points in two different regions of the detected face. Three variable feature points are located on the mouth region (one point on the middle of the lower lip and two points on the right and left mouth corners) while one fixed point is located on the midpoint of the nose, as shown in Figure 27.

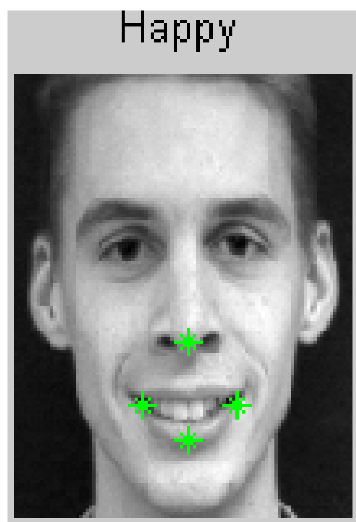


Figure 27: Four feature points

4.3.1.1 Located feature points on the lip (mouth region)

An appearance analysis on the three basic facial expressions (happy, neutral and surprise) proves that there is an obvious change in the mouth region when somebody changes his/her emotion from neutral to happy or from neutral to surprise and vice versa, as shown in Figure 28. Thus, by using this apparent property, we can easily discriminate between these expressions.

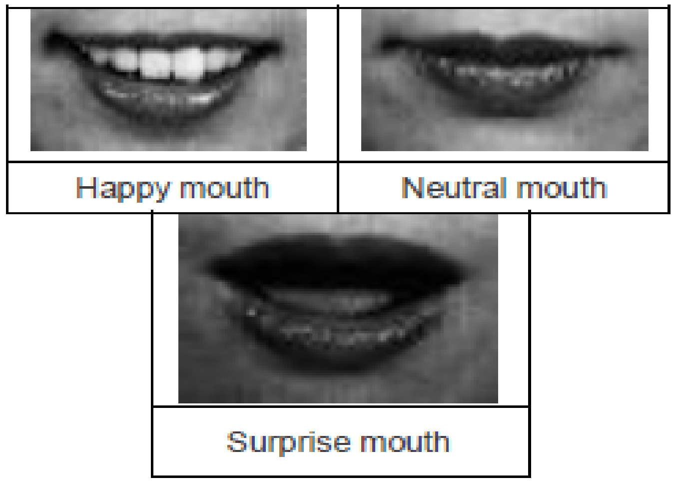


Figure 28: Mouth region changes on the three basic expressions

Therefore, our goal is to find the mouth region on the detected face for these three basic expressions and then locate three feature points on it, as shown in Figure 29.

Figure 29: Feature points on the lip

To implement this, the Viola-Jones algorithm is used to detect the mouth region on the detected face, as shown in Figure 30. This algorithm has one weakness point when we apply it to the face; it is found that more than one mouth has been detected on a single face.

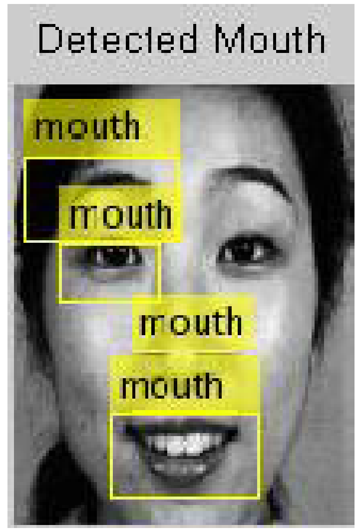


Figure 30: Multi-mouth detection

To overcome the multi-mouth detection problem, we make an analysis by applying the algorithm to different face images of different age and race. The correct location of the mouth was found to be on the lower part of a typical face. Therefore, we formulate a rule to pick up the mouth that has greatest row value. The result is shown in Figure 31.

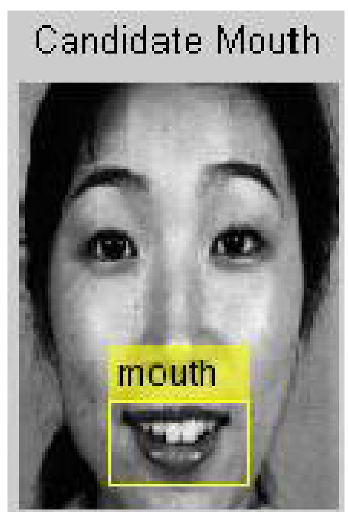


Figure 31: Candidate mouth

Subsequently, in order to locate the three feature points on the candidate mouth, two important steps are used:

1. Apply the Y directional gradient to the mouth region by using the MATLAB function (**imgradientxy**) in order to overcome the brightness. The outcome for the candidate mouth is shown in Figure 32.

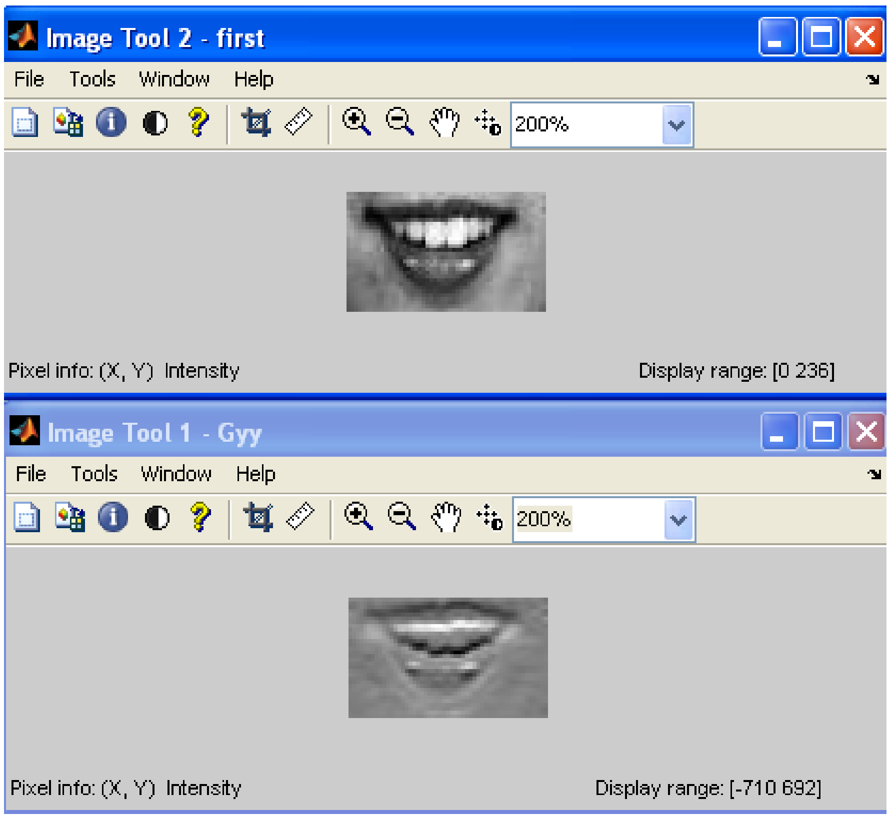


Figure 32: Applying the Y-directional gradient in the mouth region

2. Prior to locating the three feature points, we do a data analysis for the outcomes of the previous step by making a 3D plot of the mouth region for different faces using the MATLAB function (**mesh**) as shown in Figure 33.

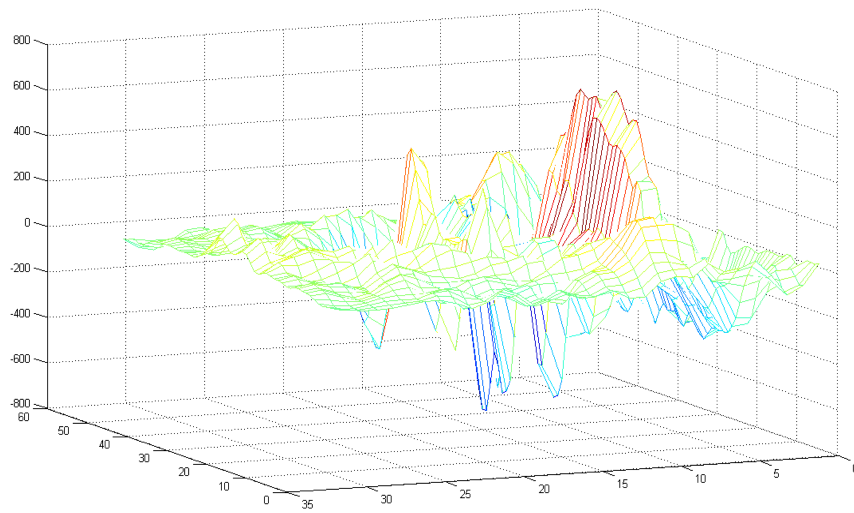


Figure 33: 3D Plot of mouth region data

This data analysis helps us to locate the three feature points on the candidate mouth region by using statistical methods (**mean, max**). The results are shown in Figure 34.

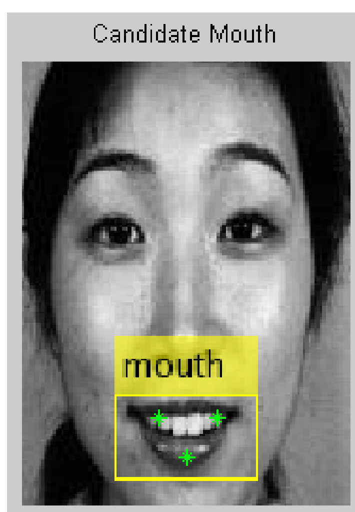


Figure 34: Located feature points on the candidate mouth

4.3.1.2 Located feature point on the nose

Since locating the three feature points in the mouth region is not sufficient to make a decision on mood detection, it is necessary to obtain another fixed feature point to become involved.

According to appearance analysis for the mentioned three basic expressions, we conclude that this fixed feature point will be on the midpoint of the nose because there no obvious changes in this region. Therefore, this necessitates finding the nose region first and then locating the fixed feature point.

Additionally, the Viola-Jones algorithm is used to find a nose region on the detected face. However, the same multi mouth problem occurs. This algorithm has the same issue on the nose, namely multi nose detection, as shown in Figure 35.

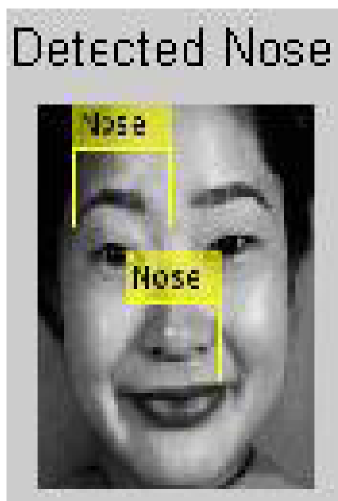


Figure 35: Multi-nose detection

To overcome the multi-nose problem, we apply a rule to pick up the nose that is coordinated in the middle of the face from the detected noses because the correct location of the nose is in the middle of a typical face. The outcome of the rule is shown in Figure 36.

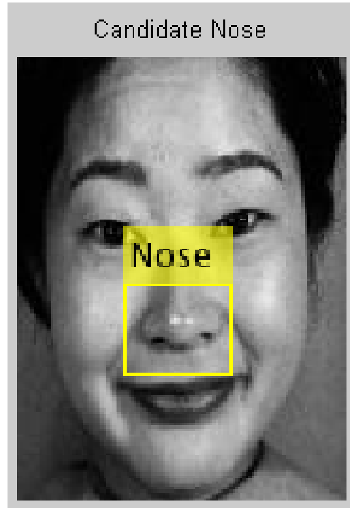


Figure 36: Candidate nose

The feature point is located in the candidate nose by using the same method already being used to locate the feature points in the mouth region. The outcome is shown in Figure 37.

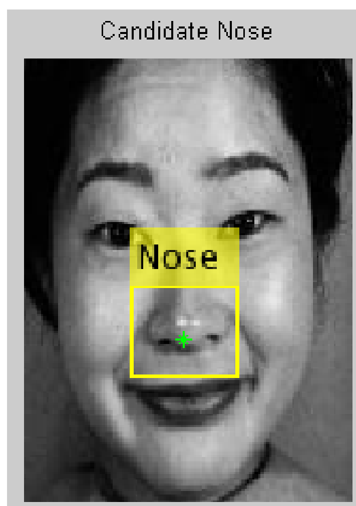


Figure 37: Located feature point on the candidate nose

4.3.2 Classification

Referring back to the literature survey, many classification methods have been tested on facial expressions classification or recognition. A majority of these methods are learned-based such as SVM, NN, PCA, etc.

However, these methods have two major limitations: training complexity and the need for more resources, such as CPU, RAM, etc, in order to handle the huge amount of data not applicable to real-time applications.

Since the extracted feature vector contains only four feature points in our proposed method, we decided to use the rule-based classification method rather than the learned-based method.

Therefore, for classification, it is necessary to find two distances: the first value V1 is used to measure a distance between the midpoint on the nose and a point on the lower lip of the mouth, while the second value V2 is used to measure the distance from the midpoint of the nose to the point on the right corner of the mouth, as shown in Figure 38.

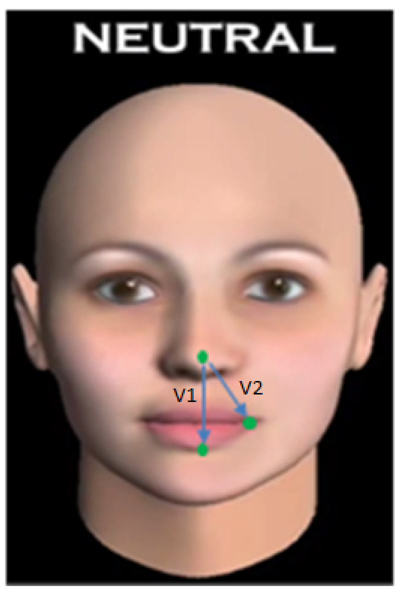


Figure 38: V1 and V2 distance vectors

We compare these two distance values with the pre-defined two threshold values.

Classification
If V1 > Threshold1 → Surprise
Else
  If V2 > Threshold2 → Neutral
  Else → Happy

We can easily detect the emotional mood of the detected face. The threshold values are found according to our data analysis on different facial expressions for different faces.

4.4 Experimental Result

The proposed second method has been implemented using MATLAB R2013a. The performance of the system is evaluated by using the first common facial expression database, namely the Japanese Female Facial Expression database (JAFFE). The detection rate is shown in Table 15.

Class	Number of faces	False detection	Detection rate
happy	29	4	86%
neutral	29	5	82%
surprise	9	1	88%

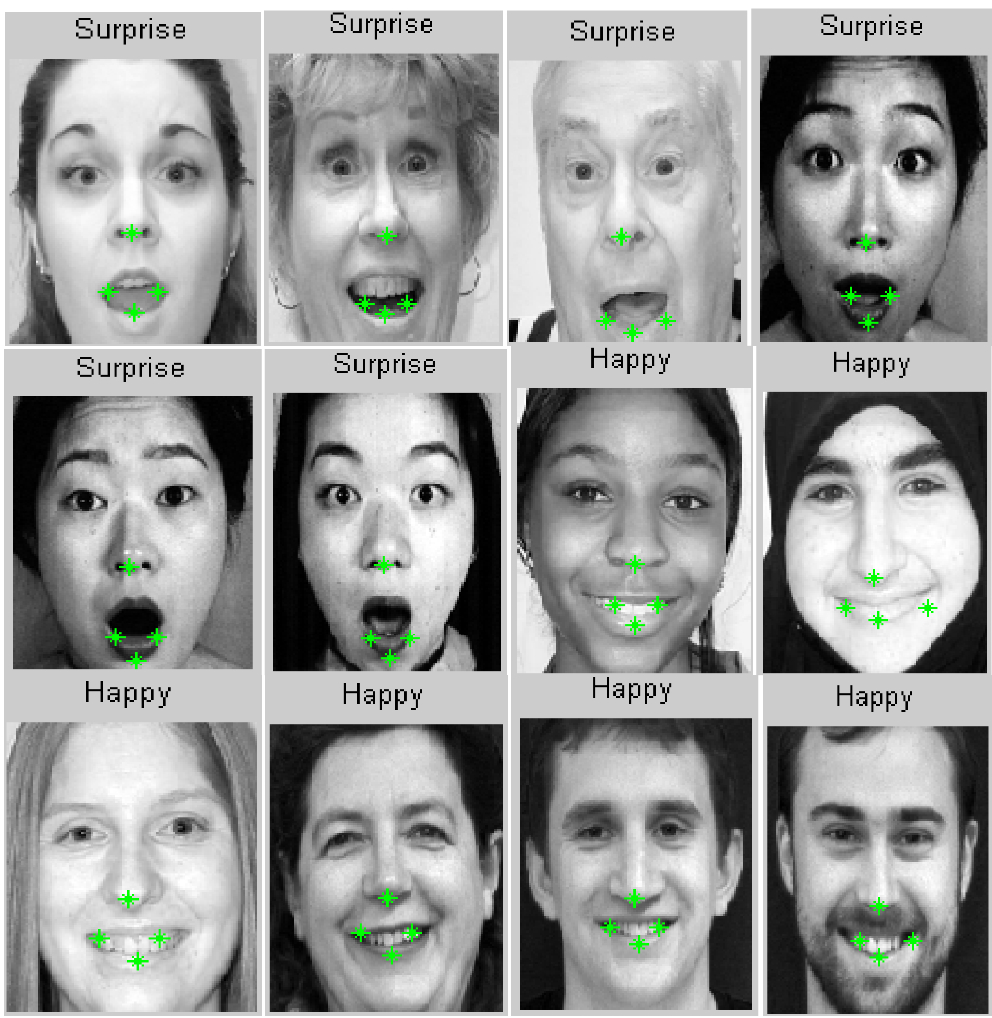
Table 15: Detection Rate on the JAFFE Database

Moreover, the method has been tested using another database, namely the Informatics and Mathematical Modeling (IMM) face database, which contains two types of facial expressions (see APPENDIX B). The detection rate is shown in Table 16.

Class	Number of faces	False detection	Detection rate
happy	34	5	85%
neutral	34	7	79%

Table 16: Detection Rate on the IMM Database

In spite of the idea being tested by using two famous facial expression databases, the idea also has been tested by using some static color images which were collected from the Internet randomly for different race, age and gender. Figure 39 shows some correct detection samples.



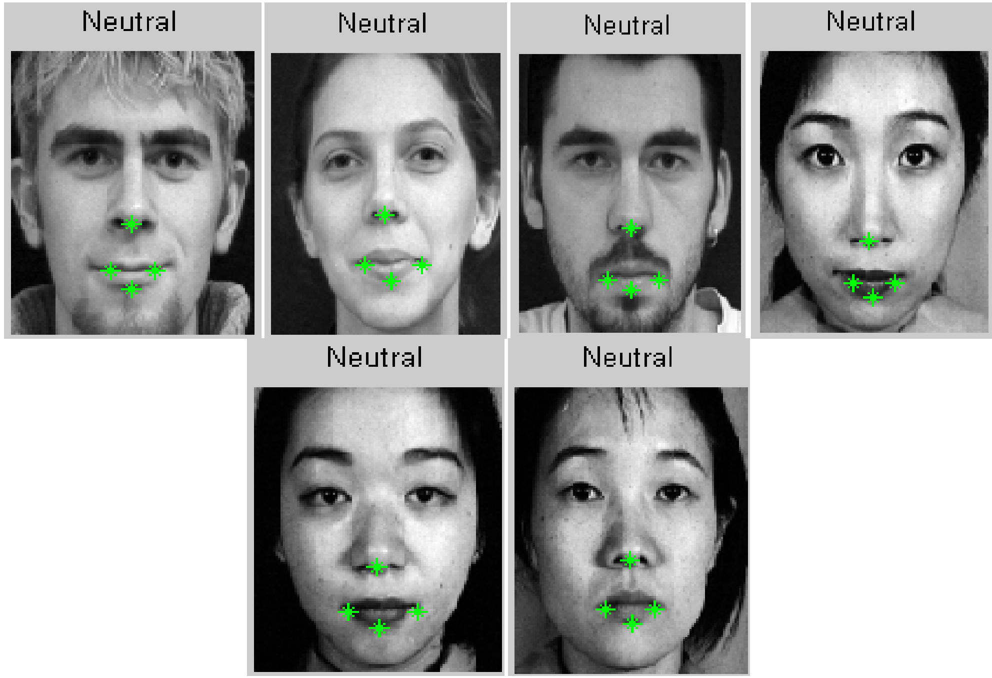


Figure 39: Samples for correct detection

A point of weakness in this method occurs when there is no obvious change on the mouth region apparently in the case of two classes, happy and surprise, which leads the method to classify it as a neutral class since the distance values between the feature points does not exceed pre-defined thresholds (see Figure 40). On the left, the first and second person was happy, but the method mistakenly labeled it as neutral, while in contrast (on the right) for the same two persons, the method correctly detected it because of the obvious change. The same applies for the case of surprise in the third person, when the change occurs in the mouth region. The method classifies it correctly.

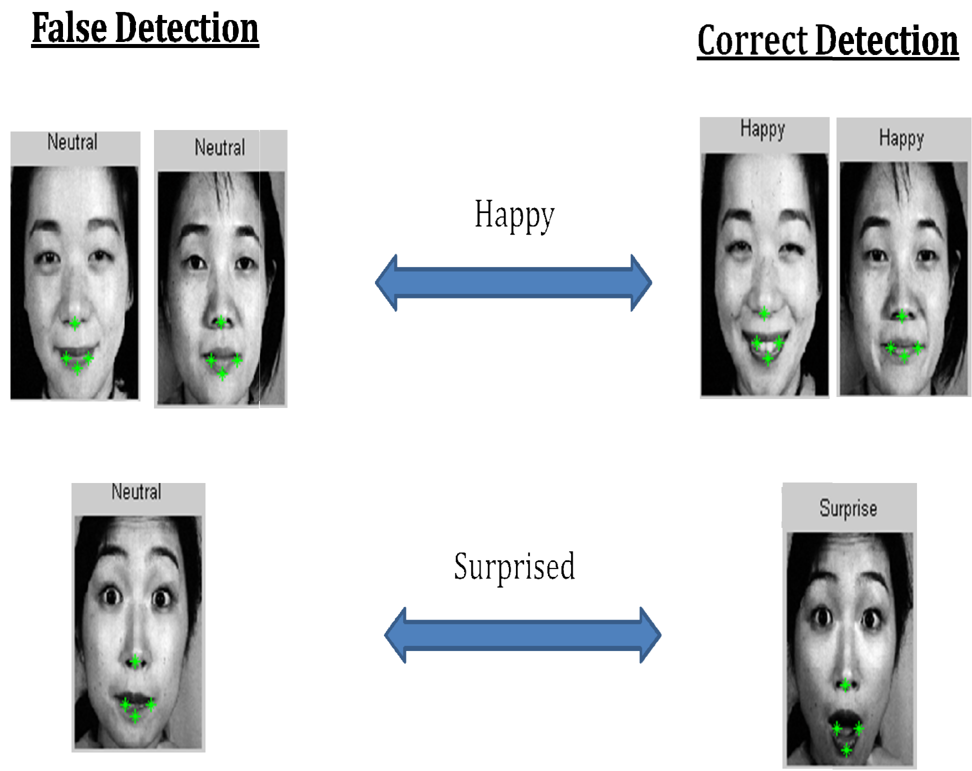


Figure 40: Samples for false detection

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Method One

5.1.1 Conclusion

The new simple idea of a facial expressions classification method has been presented. The proposed method is an example of a template matching method which is used to define a face template for four classes (Happy, Neutral, Surprise and Fear), after which the input face image is compared with these four templates by finding correlation values between them and according to this value, a decision is made to classify the face.

The performance of this idea is evaluated by using the JAFFE database, which yields 49% on average when comparing the whole face. Then we make an improvement by dividing the input face and the four templates into two parts, which yields 55% on average and a second improvement when we divide the input face and four templates into four parts, which yields 60% on average.

5.1.2 Future work

The detection rate for the first proposed method seems to increase when we subdivide the input face and the four pre-defined face templates further. Therefore, as a future plan, we believe that by dividing the input face and templates into small pieces and then finding the correlation coefficient for these pieces and then by using any learning-based method as a classification method, we would increase the detection rate.

Another idea for improvement is to extract a template by using PCA from the database containing samples for different races and ages, thereby making a template more universal than using one face (template) as carried out in our proposed method.

5.2 Method Two

5.2.1 Conclusion

In this chapter, the second idea of automatic mood detection in static images according to facial expression has been studied for three basic universal emotions (neutral, happy and surprise).

The study used the auto locating of one feature point on the midpoint of the nose as a fixed point and three feature points on the lower part of the lip (mouth region). It then finds the distance value between these feature points to classify it by using a rule-based method.

According to the experimental results, when the idea is tested by using two common databases (JAFFE and IMM), it is proved that the performance of the second method is between 80 and 85% on average.

5.2.2 Future work

Even though this idea can be improved by using one or all of the below suggested points as future work, we believe this would increase the detection rate of the proposed method.

1. By using one classification method, for instance SVM, NN, *k*-NN rather than using rule based method initially, one would need to build a dataset containing distance values for the different races, ages and genders and then use this dataset to train the selected classification method.

2. By finding more feature points on the lower and upper part of the lip to overcome the dislocation of some feature points which have an impact on the detection rate.

3. By involving and locating more feature points on another part of the face, such as the eyebrow or eye, and by using it in case there is no change in the mouth region.

REFERENCES

1. **Birdwhistell R. L., (1970),** "*Kinesics and Context, Essays on Body Motion Communication",* University of Philadelphia, Philadelphia Press, Philadelphia, pp.40-67.

2. **Mehrabian A., (1972),** "*Nonverbal Communication"*, Aldine-Atherton, Chicago, pp. 12-53.

3. **Fromkin V., Rodman J., (1983),** "*An Introduction to Language"*, CBS College Publishing, New York, pp. 30-66.

4. **Tannen D., (1986),** "*That's not What I Meant: How Communication Style Makes or Breaks Relationship"*, Ballanitne, pp. 70-97.

5. **Sohail A. S. M., Bhattacharya P., (2007),** "*Classifying Facial Expressions Using Point-Based Analytic Face Model and Support Vector Machines",* Systems Man and Cybernetics, IEEE International Conference, pp. 1008-1013.

6. **Darwin C., (1872),** "*The Expression of the Emotions in Man and Animals",* London, John Murray, Albemarle Street, ISBN 0-19-515806-7, pp. 200-240.

7. http://www.medicalook.com/human_anatomy/organs/Facial_nerve.html. (Data Download Date: 21-05-2014).

8. **Hiremath P., Danti A., (2006),** *"Detection of Multiple Faces in an Image Using Skin Color Information and Lines-of-Separability Face Model"*, International Journal of Pattern Recognition and Artificial Intelligence, pp. 39-62.

9. **Wang L., Hu W., Tan T., (2003),** *"Recent Developments in Human Motion Analysis",* Pattern Recognition, vol. 36, pp. 585-601.

10. **Lee J. S., Kuo Y. M., Chung P. C., Chen E., (2007),** *"Naked Image Detection Based on Adaptive and Extensible Skin Color Model"*, Pattern Recognition, vol. 40, pp. 2261-2270.

11. **Ng P., Pun C. M., (2011),** *"Skin Color Segmentation by Texture Feature Extraction and K-mean Clustering"*, Computational Intelligence, Communication Systems and Networks, Third International Conference, pp. 213-218.

12. **Yang U., Kim B., Sohn K., (2009),** *"Illumination Invariant Skin Color Segmentation",* Industrial Electronics and Applications, 4th IEEE Conference, pp. 636-641.

13. **Vezhnevets V., Sazonov V., Andreeva A., (2003),** *"A Survey on Pixel-Based Skin Color Detection Techniques",* Proceeding of International Conference on Computer Graphics, vol. 3, pp. 85-92.

14. **Powar V., Jahagirdar A., Sirsikar S., (2011),** *"Skin Detection in YcbCr Color Space",* International Journal of Computer Applications, pp. 0975-8887.

15. **Naji S. A., Zainuddin R., Jalab H. A., (2012),** *"Skin Segmentation Based on Multi Pixel Color Clustering Models",* Digital Signal Processing, pp. 933-940.

16. **Xu Z., Zhu M., (2006),** *"Color-Based Skin Detection: Survey and Evaluation",* Multi-Media Modeling Conference Proceedings, 12th International, pp. 10-15.

17. **Phung S. L., Chai D., Bouzerdoum A., (2001),** *"A Universal and Robust Human Skin Color Model Using Neural Networks",* Neural Networks, Proceedings International Joint Conference, vol. 4, pp. 2844-2849.

18. **Yang M. H., Kriegman D., Ahuja N., (2002),** *"Detecting Faces in Images: a Survey",* Pattern Analysis and Machine Intelligence, IEEE Transactions, pp. 34-58.

19. **Aiping C., Lian P., Yaobin T., Ning N., (2010),** *"Face Detection Technology Based on Skin Color Segmentation and Template Matching",* Education Technology and Computer Science, Second International Workshop, pp. 708-711.

20. **Kherchaoui S., Houacine A., (2010),** *"Face Detection Based on a Model of the Skin Color with Constraints and Template Matching",* Machine and Web Intelligence, International Conference, pp. 469-472.

21. **Li X., Li Y., Yao Y., Lv X., Chen J., Zhang L., (2010),** *"Study and Realization of Face Detection Based On Skin Segmentation and Template Matching",* New Trends in Information Science and Service Science, 4th International Conference, pp. 375-378.

22. **Mohamed A. S., Weng Y., Ipson S., Jiang J., (2007),** *"Face Detection Based on Skin Color in Image by Neural Networks",* Intelligent and Advanced Systems, International Conference, pp. 779-783.

23. **Kun Y., Hong Z., Ying P., (2006),** *"Human Face Detection Based on SOFM Neural Network",* Information Acquisition, IEEE International Conference, pp. 1253-1257.

24. **Mohamed A., Weng Y., Jiang J., Ipson S., (2008),** *"Face Detection Based Neural Networks Using Robust Skin Color Segmentation",* Systems Signals and Devices, IEEE 5th International Multi-Conference, pp. 1-5.

25. **Sharma S., Sharma N., Malhotra A., (2011),** *"Analysis of A Hybrid Face Detection System Using Feature Ex-traction and Neural Network Methods",* International Journal Computer Technology Application, vol. 2, pp. 300-303.

26. **Han J., Kamber M., (2006),** *"Data Mining: Concepts and Techniques",* 2nd Edition, Morgan Kaufmann, San Francisco, ISBN 1-55860-901-6, pp. 285-347.

27. **Jiang N., Lu Y., Tang S., Goto S., (2010),** *"Rapid Face Detection Using A Multi-mode Cascade and Separate Haar Feature",* Intelligent Signal Processing and Communication Systems, International Symposium, pp. 1-4.

28. **YONG Y., (2007),** *"Facial Expression Recognition and Tracking Based on Distributed Locally Linear Embedding and Expression Motion Energy",* Master Thesis, Department of Electrical and Computer Engineering, National University of Singapore, Singapore, pp. 1-40.

29. **Stathopoulou I. O., Tsihrintzis G. A., (2004),** *"An Improved Neural Network Based Face Detection and Facial Expression Classification System",* Systems Man and Cybernetics, IEEE International Conference, vol. 1, pp. 666-671.

30. **Londhe R. R., Pawar V. P., (2012),** "*Analysis of Facial Expression Using LBP and Artificial Neural Network",* International Journal of Computer Applications, vol. 44, no. 21, pp. 0975-8887.

31. **Happy S. L., George A., Routray A., (2012),** *"A Real Time Facial Expression Classification System Using Local Binary Patterns",* Intelligent Human Computer Interaction, 4th International Conference, pp. 1-5.

32. **Zilu Y., Xieyan F., (2008),** *"Combining LBP and Adaboost for Facial Expression Recognition",* Signal Processing, 9th International Conference, pp. 1461-1464.

33. **Gunes T., Polat E., (2008),** *"Feature Selection for Multi-SVM Classifiers in Facial Expression Classification",* Computer and Information Sciences, 23rd International Symposium, pp. 1-5.

34. **Lajevardi S. M., Lech M., (2008),** *"Averaged Gabor Filter Features for Facial Expression Recognition"*, Digital Image Computing: Techniques and Applications, pp. 71-76.

35. **Li L., So J., Shin H. C., Han Y., (2013),** *"An AAM-based Face Shape Classification Method Used for Facial Expression Recognition",* International Journal of Research in Engineering and Technology vol. 2, no. 4, pp. 2277-4378.

36. **Wilfred O. O., Lee G. B., Park J. J., Cho B. J., (2009),** *"Facial Component Features for Facial Expression Identification Using Active Appearance Models",* AFRICON, pp. 1-5.

37. **Ojala T., Pietikainen M., Harwood D., (1996),** *"A Comparative Study of Texture Measures With Classification Based on Featured Distributions",* Pattern Recognition, pp. 51-59.

38. **Huang D., Shan C., Ardabilian M., Wang Y., Chen L., (2011),** *"Local Binary Patterns and Its Application to Facial Image Analysis: A Survey",* Systems Man and Cybernetics, Part C: Applications and Reviews, IEEE Transactions, pp. 765-781.

39. **Rose N., (2006),** *"Facial Expression Classification Using Gabor and Log-Gabor Filters",* Automatic Face and Gesture Recognition, 7th International Conference, pp. 346-350.

40. **Namuduri K. R., Mehrotra R., Ranganathan N., (1992), "***Edge Detection Models Based on Gabor Filters"*, IEEE 11th IAPR International Conference on Speech and Signal Analysis Proceedings, pp. 729-732.

41. **Cootes T. F., Edwards G. J., Taylor C. J., (2001),** *"Active Appearance Models",* IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681-685.

42. **Cootes T. F., Taylor C. J., Cooper D. H., Graham J., (1995),** *"Active Shape Models - Their Training and Application",* Computer Vision and Image Understanding, vol. 61, no. 1, pp. 38-59.

43. **Asthana A., Saragih J., Wagner M., Goecke R., (2009),** *"Evaluating AAM Fitting Methods for Facial Expression Recognition",* Affective Computing and Intelligent Interaction and Workshops, 3rd International Conference, pp. 1-8.

44. **Azcarate A., Hageloh F., Sande K., Valenti R., (2005),** *"Automatic Facial Emotion Recognition",* Universiteit van Amsterdam, pp. 1-5.

45. **Acharya T., Ray A. K., (2005),** *"Image Processing: Principles and Applications",* pp. 133-138.

46. **Gasparini F., Schettini R., (2006),** *"Skin Segmentation Using Multiple Thresholding",* Proceedings Internet Imaging VII, vol. 6061, pp. 1-8.

47. **Chai D., Ngan K. N., (1999),** *"Face Segmentation Using Skin Color Map in Videophone Applications",* Circuits and Systems for Video Technology, IEEE Transactions, pp. 551-564.

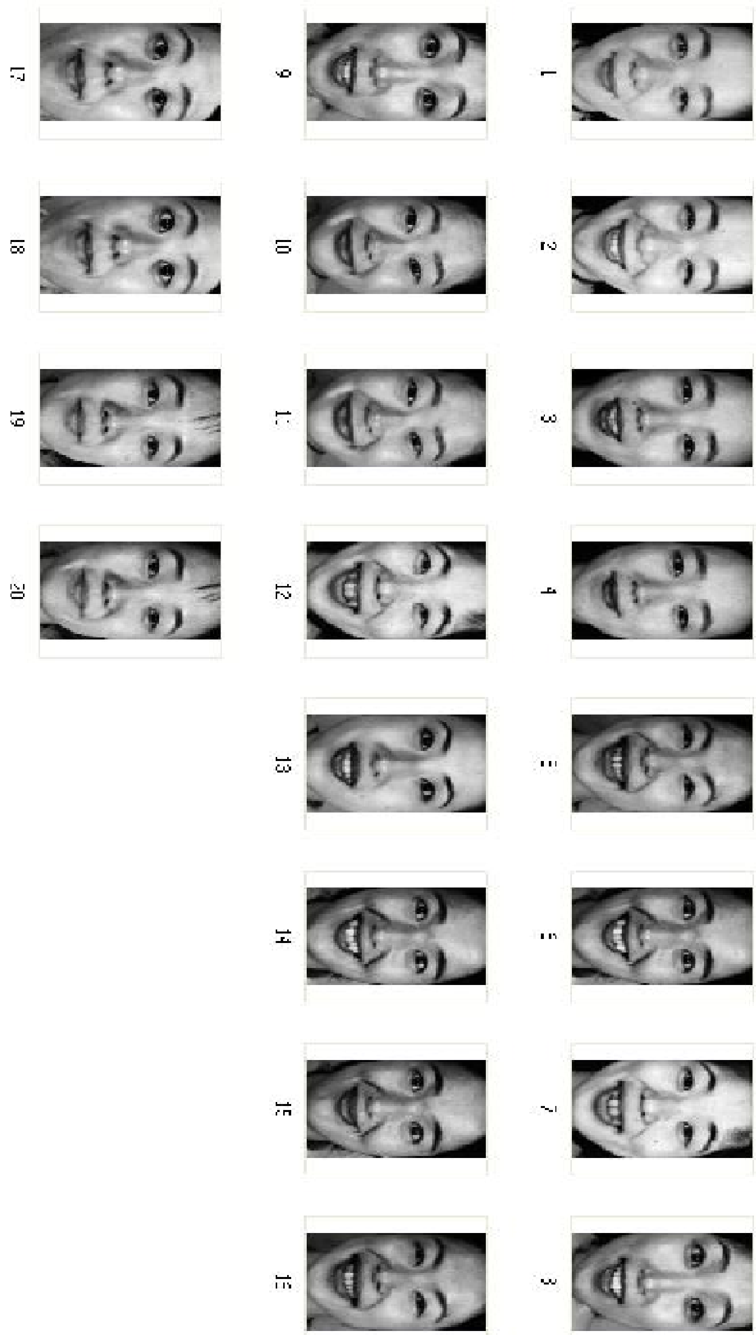
48. http://science.nationalgeographic.com/science/enlarge/skin-colors.html. (Data Download Date: 11-07-2014).

49. **Viola P., Jones M. J., (2004),** *"Robust Real-Time Face Detection",* International Journal of Computer Vision, pp. 137-154.

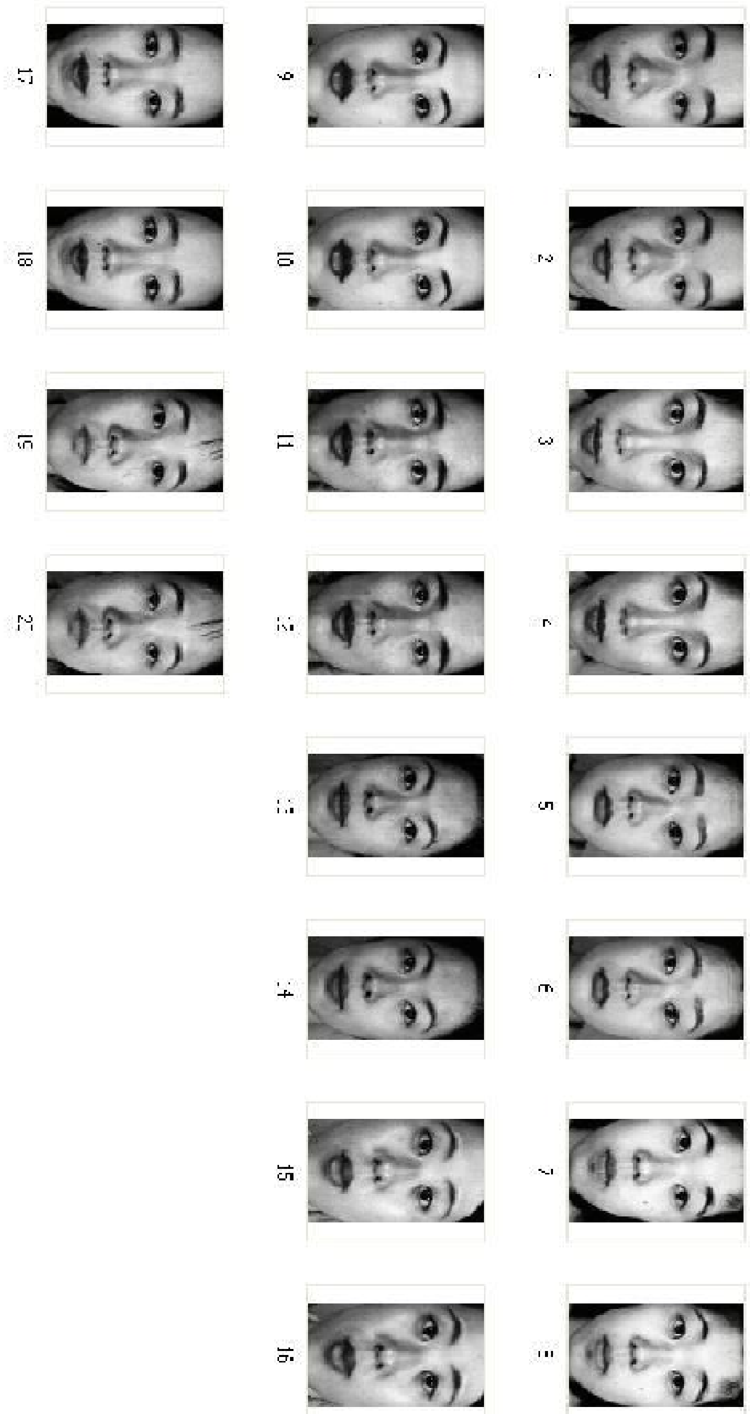
50. **Viola P., Jones M., (2001),** *"Rapid Object Detection Using a Boosted Cascade of Simple Features",* in Proceedings IEEE Conference, Computer Vision and Pattern Recognition, vol. 1, pp. 511-518.

APPENDIX A – LIST OF JAFFE database

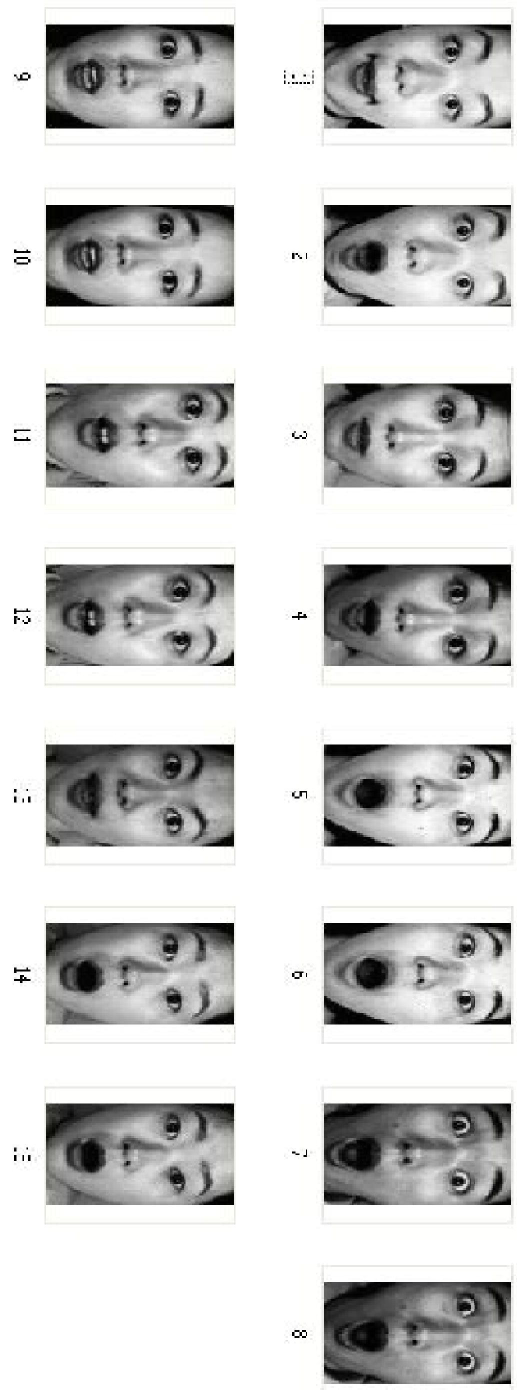
1. Happy Database.



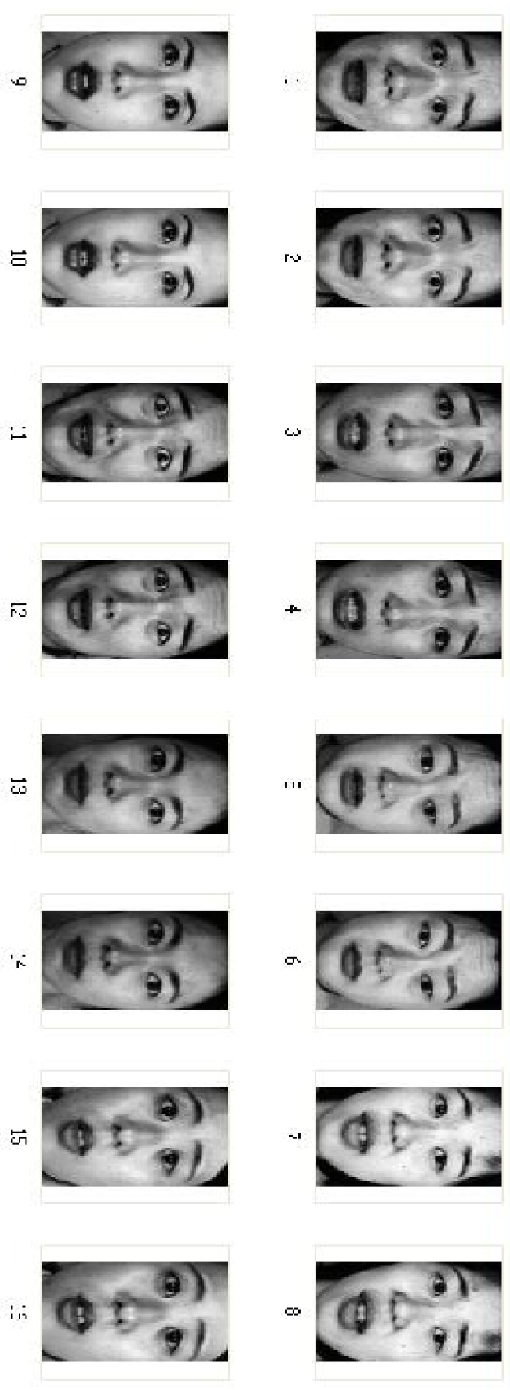
2. Neutral database.



3. Surprise database.

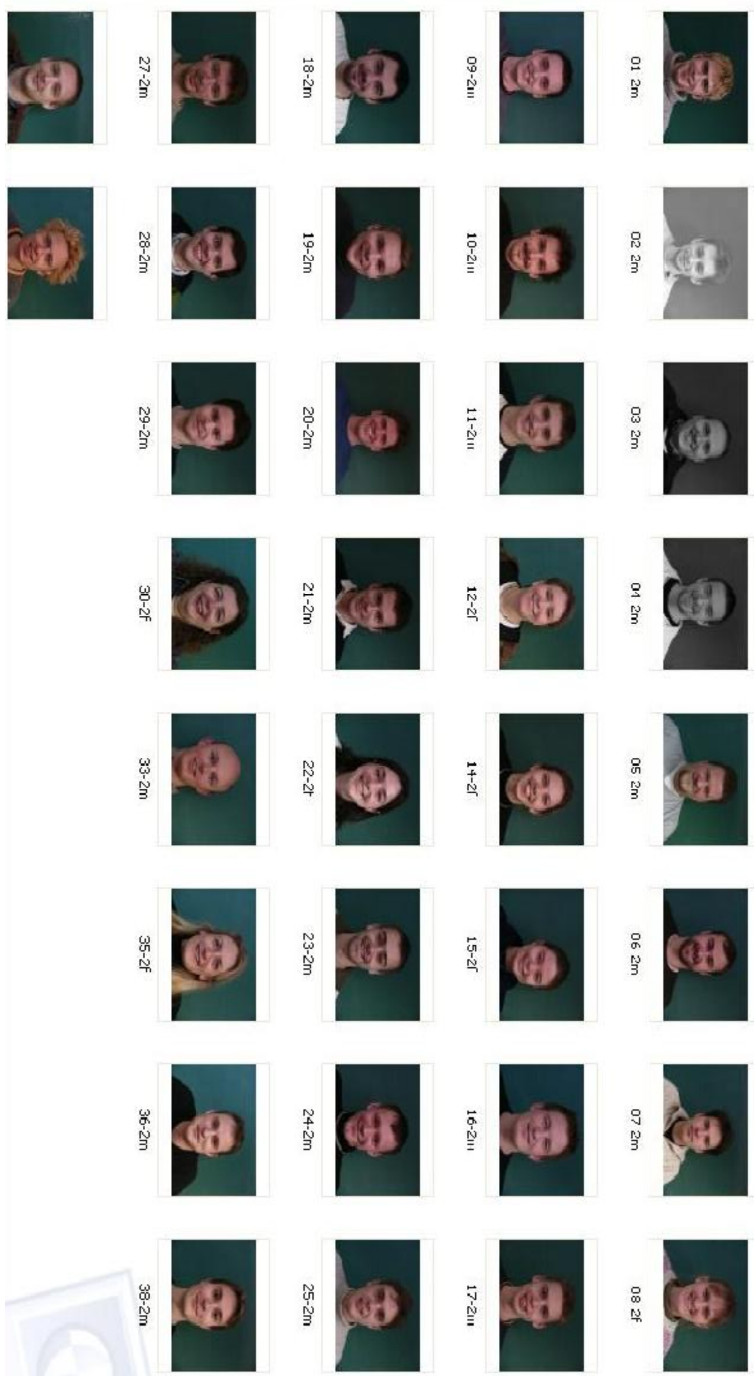


4. Fear database.

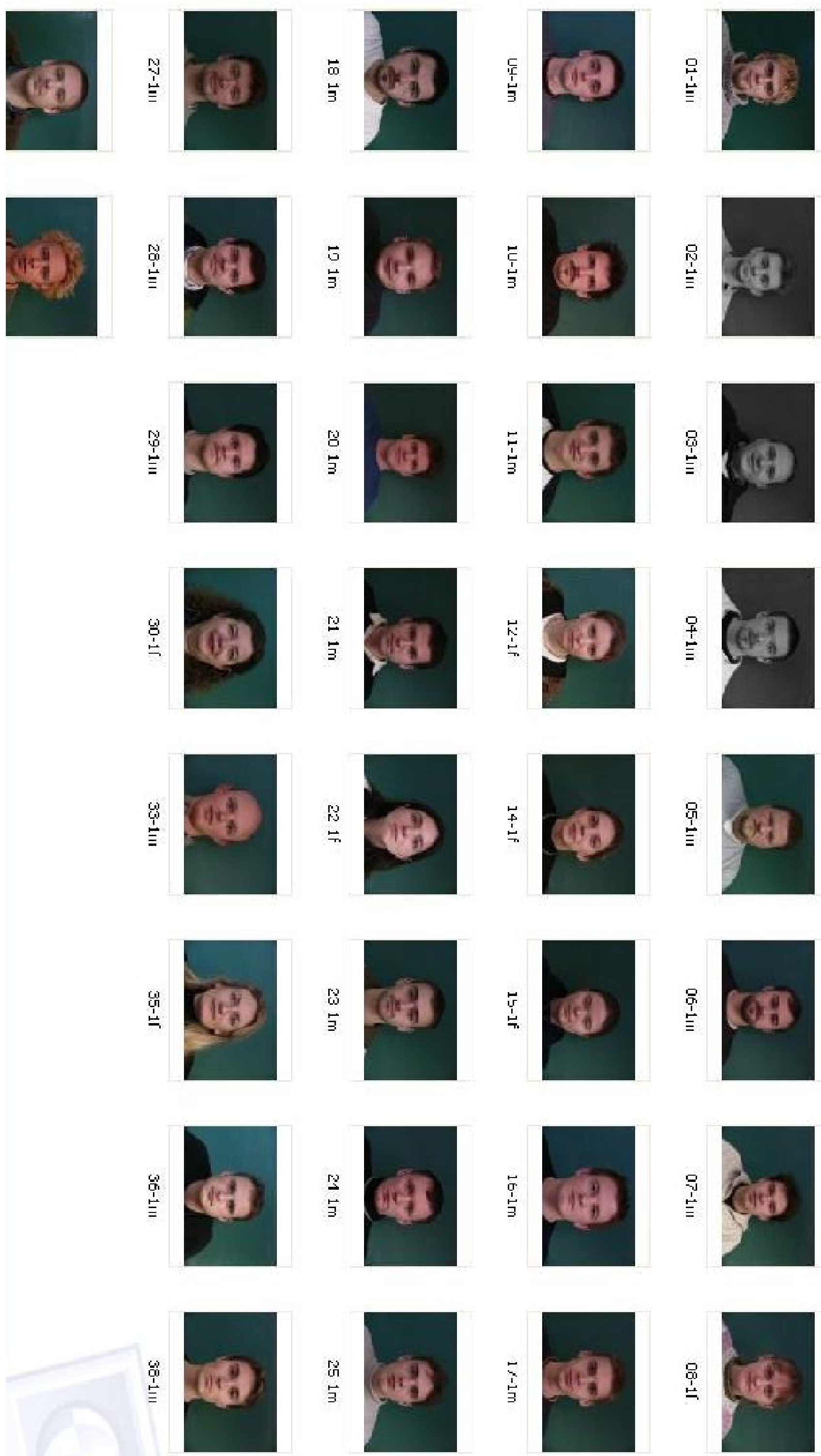


APPENDIX B – LIST OF IMM database

1. Happy database.



2. Neutral database.



APPENDIX C

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: WAIS, Yasir Mustafa

Date and Place of Birth: 30 Jan 1986, Iraq-Kirkuk

Marital Status: Married

Phone: +964 7702382419

Email: yaser_nss@yahoo.com

EDUCATION

Degree	**Institution**	**Year of Graduation**
M.Sc.	Çankaya University, Computer Engineering	2015
B.Sc.	College of Technology Kirkuk, Software Engineering	2007
High School	Al-Waled High School	2003

WORK EXPERIENCE

Year	**Place**	**Enrollment**
2008- Present	Kalimat Telecom Company-Iraq	NSS Engineer
2007-2008	North Oil Company - Iraq	Trainer

FOREIN LANGUAGES

Advanced Arabic, Advanced English.

PUBLICATIONS

1. **WAIS Y. M., EŞMELİOĞLU S., (2014),** *"Mood Detection According to Facial Expression"*, Seventh Engineering and Technology Symposium, Çankaya University, Turkey.
2. **WAIS Y. M., MOHAMMED ALI A. H., (2014),** "*A New Hybrid Method in Watermarking Using DCT and AES"*, International Journal of Engineering Research and Development, vol. 10, issue 11, pp. 64-69.

HOBBIES

Football, Travel, Books, Swimming, Fitness.