Master Thesis
Electrical Engineering

Micro-Expression Extraction For Lie Detection Using Eulerian Video (Motion and Color) Magnification

Submitted By

Gautam Krishna , Chavali

Sai Kumar N V , Bhavaraju

Tushal , Adusumilli

Venu Gopal , Puripanda

This thesis is presented as part of Degree of Master of Sciences in Electrical Engineering

BLEKINGE INSTITUTE OF TECHNOLOGY
AUGUST,2014

Supervisor : **Muhammad Shahid**
Examiner : **Dr. Benny Lövström**
Department of Applied Signal Processing
Blekinge Institute of Technology;
SE-371 79, Karlskrona, Sweden.

This thesis is submitted to the Department of Applied Signal Processing at Blekinge Institute of Technology in partial fulfilment of the requirements for the degree of Master of Sciences in Electrical Engineering with emphasis on Signal Processing.

Contact Information
Authors:

Gautam Krishna.Chavali
E-mail: gautamkrishna.chavali@gmail.com

Sai Kumar N V.Bhavaraju
E-mail: saikumar.bhavaraju@gmail.com

Tushal.Adusumilli
E-mail: tushal.adusumilli@gmail.com

Venu Gopal.Puripanda
E-mail: venugopal1035@gmail.com

University Advisor:
Mr. Muhammad Shahid
Department of Applied Signal Processing
Blekinge Institute of Technology
E-mail: muhammad.shahid@bth.se
Phone:+46(0)455-385746

University Examiner:
Dr. Benny Lövström
Department of Applied Signal Processing
Blekinge Institute of Technology
E-mail: benny.lovstrom@bth.se
Phone: +46(0)455-38704

School of Electrical Engineering
Blekinge Institute of Technology
SE-371 79, Karlskrona
Sweden.

Internet : www.bth.se/ing
Phone : +46 455 38 50 00

Abstract

Lie-detection has been an evergreen and evolving subject. Polygraph techniques have been the most popular and successful technique till date. The main drawback of the polygraph is that good results cannot be attained without maintaining a physical contact, of the subject under test. In general, this physical contact would induce extra consciousness in the subject. Also, any sort of arousal in the subject triggers false positives while performing the traditional polygraph based tests. With all these drawbacks in the polygraph, also, due to rapid developments in the fields of computer vision and artificial intelligence, with newer and faster algorithms, have compelled mankind to search and adapt to contemporary methods in lie-detection.

Observing the facial expressions of emotions in a person without any physical contact and implementing these techniques using artificial intelligence is one such method. The concept of magnifying a micro expression and trying to decipher them is rather premature at this stage but would evolve in future. Magnification using Eulerian Video Magnification(EVM) technique has been proposed recently and it is rather new to extract these micro expressions from magnified EVM based on Histogram of Oriented Gradients (HOG) features. HOG features is the feature extraction algorithm which extracts local gradient information in an image. Till date, HOG features have been used in conjunction with SVM, and generally for person/pedestrian detection. A newer, simpler and modern method of applying EVM with HOG features and Back-propagation Neural Network jointly has been introduced and proposed to extract and decipher the micro-expressions on the face. Micro-expressions go unnoticed due to its involuntary nature, but EVM is used to magnify them and makes them noticeable. Emotions behind the micro-expressions are extracted and recognized using the HOG features & Back-Propagation Neural Network. One of the important aspects that has to be dealt with human beings is a biased mind. Since, an investigator is also a human and, he too, has to deal with his own assumptions and emotions, a Neural Network is used to give the investigator an unbiased start in identifying the true emotions behind every micro-expression. On the whole, this proposed system is not a lie-detector, but helps in detecting the emotions of the subject under test. By further investigation, a lie can be detected.

Keywords: Micro Expressions, Emotions, Eulerian Video Magnification, Histogram of Oriented Gradients, Voila-Jones Algorithm, Artificial Neural Network.

Acknowledgments

We would like to express gratitude to our supervisor Mr.Muhammad Shahid for introducing us to the topic and for the heartfelt support and encouragement on this pursuit of our masters. Furthermore, we would like to thank Dr. Benny Lövström for his useful comments and remarks through the learning process of this master thesis.

The Department of Signal Processing has provided the support and equipment that are mandatory to produce and complete our thesis.

In our daily work we have been blessed with a friendly and cheerful group of fellow students (PVK Chaitanya and V Revanth). We will forever be grateful for all your love and help.

Finally, we thank our family members and relatives for supporting us throughout all our studies at University and helping us to move across the seas. Also, for providing a second home here, wherein we could complete our writing up.

Contents

List of Figures

List of Tables

List of Abbreviations

AdaBoost	Adaptive Boosting
ANN	Artificial Neural Network
EVM	Eulerian Video Magnification
EMFACS	Emotion Facial Action Coding System
FACS	Facial Action Coding System
FPS	Frame rate Per Second
GUI	Graphical User Interface.
GUIDE	Graphical User Interface Development Environment.
HOG	Histogram of Oriented Gradients
LMS	Least Mean-Squared Algorithm
MATLAB	Matrix Laboratory
ROI	Region Of Interest
RAM	Random Access Memory
SVM	Support Vector Machines
VJ	Voila-Jones algorithm
YCbCr	Yellow Chrominance blue Chrominance red

Chapter 1

Introduction

Lie detection, in general, is referred to as a polygraph. A polygraph is a device that measures various parameters such as respiration, blood pressure, pulse and sweat which are used as indices in estimating a lie. The drawback of the polygraph is that it triggers false positives, when the subject under test is anxious or emotionally aroused. A new design is being created where emotions play a crucial role in determining the lies and overcoming the difficulties posed by the traditional polygraph. Also, the traditional lie detection techniques rely on wired system which induces panic in the subject under test. This new study is designed to overcome the drawbacks of the traditional polygraph and to help the investigator in the process of detecting lies by not involving any physical contact with the subject under test.

Emotions play a very prominent and purposeful role in day-to-day life. Emotions directly reveal the exact feelings of a person at any given time. This new study also works as a tool for deciphering a person's present emotional state around with ease. A technique, where emotions play a crucial role in the process of detecting lies, is more reliable as emotions are universal and don't change with caste, culture, creed, religion and region. At any particular instance of time, the emotion felt by a person can only be deciphered through the expression put up by that person.

A person's 80 facial muscular contractions and their combinations, give rise to thousands of expressions. A major class of expressions are categorized into 7 basic emotions such as anger, disgust, fear, happy, surprise, sadness and contempt. Contempt is an emotion which has been recently added to the list of universal emotions. As of now, the given study in hand is confined to six basic emotions (with neutral expression) leaving behind contempt. The reasons for eliminating contempt emotion are elaborated in the chapters that follow. In general, few predominant emotions such as fear, anger and sadness are mostly observed in the process of lie detection [17]. Thus, this new study helps the investigators, in deciphering the true feelings of subject under test, and to common people, as a tool for understanding other people and their feelings easily.

There would be a high level of uncertainty observed, in estimating the hidden emotion, within an expression that are elicited during low or normal stakes. This uncertainty of emotion estimation, in low or normal stake, is due to the fact that the subject under test can control or tamper his expressions and emotions in such situations since the person would be conscious of his actions. But, when a subject under test is at a high stake situation, there would be a leakage of expression, involuntarily. Thus,high stake situations provide more probability in estimating the emotion correctly. Micro expressions occurring at high stake situations are the basis for these kind of involuntary emotions. Micro expressions occur in a fraction of a second and are hard to recognize in real time without good expertise. Emotions in conjunction with micro expressions play a crucial role in the process of detecting lies. Generally, when a person tries to hide the truth, he feels the pressure inside, which indeed increases his heart rate. Thereby, measuring the heart rates while questioning the subject would strengthen the emotion predictions. This study does both of them simultaneously without any physical contact.

1.1 Objectives and Scope of work

The prime objective is to detect the 6 universal and primary emotions (plus neutral expression) based on micro expressions of a subject under test, which is thereby useful in the process of detecting lies. Another salient feature is to detect the pulse rate of the subject under test. This additional feature of detecting pulse of a subject brings authenticity to the study. Wherein emotions, based on micro expressions and pulse rate, can be observed at any given instance. Results from these emotions and pulse rates are to be exhibited in a Graphical User Interface (GUI). GUI accommodates videos and their corresponding graphs. This study cannot be called as a lie-detector, since, it does not explicitly detect any lies, but it extracts an emotion which would be helpful in the process of lie detection.

The system specifications are high, as the processing required is more and lengthy.

1.1.1 Pre-requisites for the methodology

- Windows 8
- i3/i5/i7, or Xenon E3/E5/E7
- 8Gb RAM (Minimum), 16Gb (recommended)
- MATLAB 2013a or higher versions
  - Signal Processing Toolbox
  - Image Processing Toolbox
  - Computer Vision System Toolbox
  - Neural Network Toolbox
  - Parallel Processing Toolbox

1.2 Research Questions

1. How to find the subtle changes in micro expressions of a subject under test from a video using motion magnification of EVM?
2. How to recognize the 6 universal and primary emotions (with neutral expressions) based on micro expressions using Back-propagation Neural Network (NN) from a motion magnified video?
3. How to find the pulse rate of a subject under test from a video using color magnification of EVM?
4. How to create a GUI that accommodates and exhibits all the results?

1.3 The Method

The micro expressions observed from an interview setup/experiment are hard to analyze. The muscle movements occurring for that small fraction of a second makes it impossible to discover those micro expressions. Eulerian Video Magnification magnifies those subtle changes, thus, helping in overcoming those ambiguities [25]. For the given input video, EVM algorithm has two variants, the one which magnifies motion and the other one, which amplifies the color. Motion magnification, magnifies the small and subtle changes of muscle movements that occur in the face. Color magnification, magnifies the facial color changes caused due to the blood flow through the blood vessels, thus, used in finding out the pulse rate of the subject under test. Hence, color magnification adds an authenticity to the given emotion by finding the variations in the pulse rate of a subject under test while experiencing that particular emotion.

Magnified motion and amplified color changes are the outputs from EVM. The motion magnified video is fed into a Voila-Jones algorithm for face recognition [22] [23]. The recognized face is cropped so as to have only the front face. This recognized face is given as input to feature extraction block that extracts the orientation of gradients and creates a histogram, in the facial region of each frame in the motion magnified video; this type of feature extraction is known as Histogram of Oriented Gradients (HOG). HOG extraction are feature descriptors [3]. These features are inputs for Neural Network (NN) using back-propagation algorithm and NN classifies whether the current frame given at hand, corresponds to any particular emotion in it or not [19].

Thus, results are represented as continuous graphs showing emotions and pulse rate of motion magnified video and color magnified video respectively. These four results, i.e. Motion magnified video and color magnified video with their corresponding emotion and pulse rate graphs are presented in a Graphical User Interface (GUI). The GUI design is very simple and has the basic play, pause and stop buttons. These play/pause/stop button perform their operations on a set of two videos comprising of magnified video and the graph video at the same time. Operational flow of the methodology is represented in the figure 1.1 as shown below in block diagram section.

1.4 Block Diagram

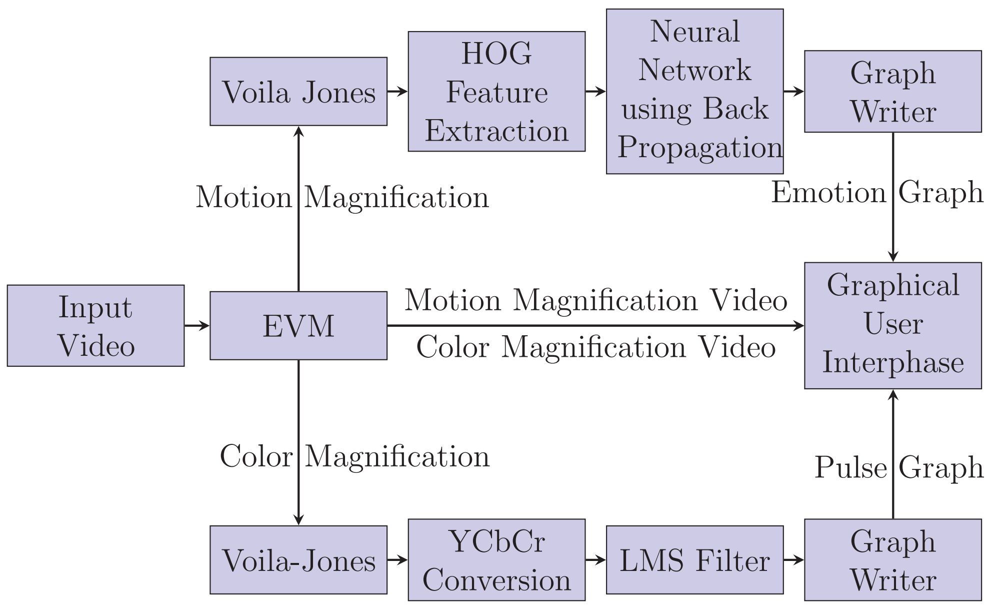


Figure 1.1: Block diagram representing methodology

The recorded video is taken and is used to magnify color and the motion. In the second stage face detection is done using the Viola-Jones Algorithm for both color and motion stages. Then, for motion magnification mode, the resultant video is processed for HOG Feature Extraction, and the features are extracted to be fed into a neural network using Back-Propagation Algorithm.The results of them are graphed and constructed into a video for a moving graph representation.

For color magnification mode, the face-detected video is converted into an YCbCr color space for the reason specified in Section 6.2. Later, an adaptive filter is used to extract the pulses and are graphed and constructed into a video for GUI to access. All the videos consisting of motion and color magnified videos and their corresponding graph videos are accessed through a GUI and displayed in that GUI.

1.5 Overview of Thesis

The following report has been organized into several chapters. Chapter 1 is the introduction where it introduces methodology, Chapter 2 deals with the conceptual learning about lies, expressions and emotions. Eulerian video (color and motion) magnification is discussed in chapter 3. Chapter 4 deals with face recognition and feature extraction using Voila-Jones algorithm and HOG features respectively. Chapter 5 discusses about the database of images that are used to train the Neural Network. Neural Network, pulse extraction and moving graphs are discussed in chapter 6. Chapter 7 deals with graphical user interface. Chapter 8 shows the results. Chapter 9 concludes the work and also throws some light on the future works.

Chapter 2

Lies, Expressions and Emotions

Great poets in their literary works have described romance as a form, where couples developed and maintained myths about each other [9]. Similarly, magicians make the audience believe, accept and get people amused by their acts. Also, fortune tellers make people believe that they are able to predict ones future, just by looking into palms or your face. In our day-to-day life, it is often easy and happy to live in a lie than to comprehend a bitter truth. Human beings always live, believe and accept the lies of oneself and others. Is it possible to have romances without myths? How do people believe the acts performed by a magician? Why do people accept their fates decided by a fortune teller? It is very much inherent for an average person not to identify and perceive a lie. So, is there any way that an average person can train himself to identify a lie?

There is always a necessity to understand the true intentions of others, to make our life simpler, meaningful and truthful, eventually, making the world a better place to live. To represent and quote this truth and lie in a simpler way, if a lie is considered as 0 and truth as 1. Everything other than 1, implying all the quantized values from 0 to 0.9 are considered as lies. People always try to cover their inner insights and demeanor, says Freud. This literally means, our day starts with a lie and also ends with a lie. In life, people should always move and proceed ahead with hope but not in falsification.

2.1 Lie

A lie is concealment or falsification of truth [9]. The terms deceit and lie are interchangeable. A liar is one who betrays or misleads others with his lies [6]. While lying, no prior information is given by a liar. Most people get away with their lies because the victim is unable to differentiate whether the intentions and emotions of a liar are feigned or real [6].

The obvious way of perceiving a person while lying is to look closer for the liar to fail. Evidently, if the reasons for which a lie can fail are apprehended, people can get closer in catching a liar. A liar or lie can fail due to two reasons [6].

1. Thinking: Because the liar has not prepared sufficiently or a liar could not apposite the situation.

2. Feelings: Because the liar could not administer his emotions.

Thus, when people are cautious enough and apprehend the reasons for lier's failure, they can easily catch a liar. Paul Ekman, a famous scientist and a pioneer in the studies of facial expressions and emotions, found three more techniques to detect a lie [9]. In general, behavioral cues are not under a conscious control of any person. Observing and understanding the behavioral cues of a liar, which leak out without his own knowledge determines and tells a lot about that person. This study focuses on a few instances wherein a person's own non-verbal behavior, such as, micro expressions reveals the underlying emotion even though the liar verbally tries to conceal or mask the truth. To reveal the major aspects of non-verbal cues requires a deeper insight of dealing and understanding the concepts of facial expressions and emotions.

2.2 Facial Expressions

Expressions are outward manifestations of the internal changes that are occurring or occurred in the mind [10]. These expressions are signatures of the changes in the mind. Identifying an expression is very easy since it involves a corresponding muscular changes at a particular region in the face. Consider the figure 2.1 shown below and try to decipher the meaning of the expression that the subject has posed.

Figure 2.1: Macro facial expression of a subject [17].

Deciphering the expression posed by the subject might mean either the subject is unaware of something or the subject is trying to refuse something. The meaning of an expression also depends on the context. The human mind has an inherent ability to respond to a stimulus without any explicit training. In other words, human mind intuitively tries to deduce a conclusion to a person's behavior by associating his/her expressions with a particular meaning [10]. The meaning derived from an expression may or may not be correct. For example, in a social gathering or at a party, when a husband does something wrong, his wife gets angry. But being in such a social gathering/party she tries to hide the anger with a smile on her face. This smile can be easily inferred as she being happy, which is a misapprehension. Hence, it concludes that a face is a multi-signal system [10]. A face can blend two emotions at the same time, for example a person can feel both sad and surprised at the same time. It is not mandatory to have only a trace of a particular emotion in a persons face. A human face can elicit two or more emotions at the same time.

Words might sometimes deceit people, but facial expressions are the abundant sources of revealing the truest intentions [12]. Facial expressions are the *sine qua non* sources of emotions. Sometimes, facial expressions are made deliberately to communicate information [10].

Paul Ekman developed a comprehensive facial expression scoring technique called the Facial Action Coding System (FACS) [15]. FACS categorizes each and every expression in the bunch of thousand expressions that are produced from the combination of one or more facial muscle(s). Expressions are generally categorized into three types.

1. Macro Expressions: The general type of expressions that occur in 4 to 5 seconds of time.

2. Micro Expressions: Involuntary expressions that occur in a blink of an eye. The duration of this kind of expressions are from 1/5th of a second to 1/25th of a second.

3. Subtle Expressions: Involuntary expressions that only depends on the intensity of the expression rather than the duration.

2.3 Emotions

Emotions are evolved to adapt and deal with day-to-day activities and situations [8]. Expressions are the richest sources of emotions and emotions are subsets of expressions [10]. This Implies that an expression need not contain an emotion, but the converse is not true. For example, a girl blushing about something or a boy winking at a girl have no emotion in them, they are just expressions.

In linguistic choices, an emotion is always referred as a single word [18], but actually, an emotion is not a sole entity of affective state, but it is the epitome of emotion family of related states [2]. Each emotion or emotion family, has a variety of associated but visually non-identical expressions. For example, anger has 60 visually non-identical expressions with core properties being same [5]. This core property differentiates family of anger with the family of fear. An emotion family is distinguished from another emotion family based on 8 different characteristics [8]. Universal emotions are also called as basic emotions. Every investigator of emotions agreed on universality of six basic emotions: anger, disgust, sadness, happy, fear and surprise. In the recent years, there is one more addition of universal emotion called contempt [7].

2.3.1 Anger

The response when a person feels annoyed or the response when a person is attacked or harmed by someone. Anger can also be response developed from extreme hatred [2]. Anger emotion is represented in the figure 2.2.

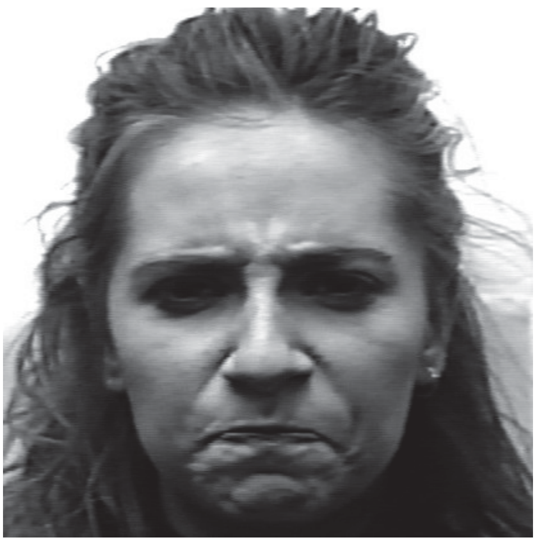


Figure 2.2: Angry- facial emotion [17].

Example: A common man frustrated with the ruling government shows his/her anger in elections.

2.3.2 Disgust

The response when a person feels repulsively provoked by something offensive or revulsion itself [2]. Disgust emotion is represented in the figure 2.3.

Figure 2.3: Disgust - facial emotion
[17].

Example: When a person smells something unpleasant, he/she feels disgusted

2.3.3 Fear

The response when a person feels a threat, harm or pain [2]. Fear emotion is represented in the figure 2.4.

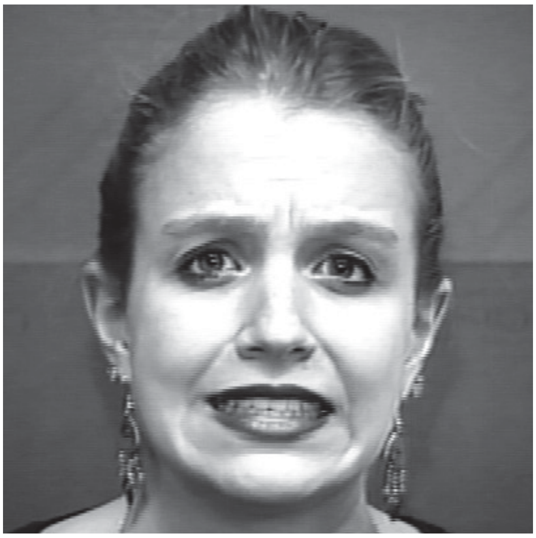


Figure 2.4: Fear- facial emotion
[17].

Example: Most people while watching a horror movie might feel this emotion of fear.

2.3.4 Happy

The response when a person feels contended or feels happy or feels pleasure [2]. Happy emotion is represented in the figure 2.5.

Figure 2.5: Happy- facial emotion [17].

Example: When a person gets recognized for his hard work and is rewarded with a promotion, he/she feels happy.

2.3.5 Sadness

The response when a person feels unhappy or loss of someone or something [2]. Sadness emotion is represented in the figure 2.6.

Figure 2.6: Sadness- facial emotion [17].

Example: when a person loses his/her parents or loved ones, he/she feels sad.

2.3.6 Surprise

The response when a person feels something sudden and unexpected [2]. Surprise emotion is represented in the figure 2.7.

Figure 2.7: Surprise- facial emotion
[17].

Example: When a birthday party is planned without the prior knowledge of a person, it makes him/her feel surprised.

2.3.7 Contempt

The response when a person feels he/she is superior to another person [2]. Contempt emotion is represented in the figure 2.8.

Figure 2.8: Contempt- facial emotion
[17].

Example: A person may feel contempt for his boss.

Every family of emotion contains a large number of expressions representing them, but contempt is the only emotion which has just two expressions. Thus, family of contempt is very limited [18]. The data required for working on contempt is also quite inadequate, since, the evolution of this emotion is very recent. As of now contempt emotion is not considered for some time.

2.4 Micro Expressions

According to Darwin, there are few muscles which are impossible to activate voluntarily and these muscles reveal the true intentions of others [12]. There are 7 characteristics such as duration, symmetry, speed of onset and etc, which distinguish voluntary from involuntary facial expressions [11]. A liar may elicit an emotional cue which betrays the plausibility of the lie itself [6]. Micro expressions of emotions are typically represented in all these involuntary classes.

Micro expressions occur due to involuntary facial muscles which are impossible to interfere with and hard to feign them deliberately [11]. The person who elicits micro expressions never intends in fabricating/making them. Micro expressions or facial expressions of emotions are thus highly informative.

Micro expressions are universal and flash on and off for less than a second in the face [14]. Micro expressions in general occur as peak experience or heated exchange [14].

Two major reasons for micro expressions occurrences are:

1. When a person tries to conceal or mask an emotion, then leakage occurs in the form of micro expressions.

2. As Darwin suggested, micro expressions are also formed due to the involuntary muscle actions.

Thus, micro expressions are the relevant sources where the emotions can truly be revealed. Micro expressions help people in detecting lies and also helps in interpreting the persons intentions and the world around. The actual problem starts with people finding it difficult to recognize these micro expressions as they occur for a brief duration, with quick onsets and, more dominantly at random and unexpected bursts. There are so many minute things that go unnoticed, which requires keen observation skills and proficiency in deciphering them.

1. How can normal people with nominal observational skills can understand and analyze these micro expressions of emotions?

   Micro-Expressions are in fact hard to analyze in real time and with nominal observational skills. People have practiced observing and learning them for years but with limited success rates. This is due to the split-second occurrence of it by nature. Slowing this reaction timing to an observable time-period shall enhance the success rates drastically. This is done by recording it with high frame rates and slowing them down, discussed in detail in Section 3.3.

2. How to design a tool which not only helps investigators working on detection of lies, but also helps people in understanding others around to make life simpler?

   A tool is required to make people understand the emotion behind every micro-expression. Since every micro-expression is associated with its corresponding Facial Action Units, continuously examining them and the micro-expression to derive the emotion behind it is tedious. Therefore a system has to be designed which learns them and estimates the emotion behind a micro-expression automatically. This would help people in practicing the observing and estimating the exact emotion corresponding to a micro-expression. The design of the tool and its working is discussed in detail in further chapters.

3. Can there be a technique which works without any physical contact with the subject under test?

Yes. Facial cues offer a lot of information of the present state of the mind of a subject under test. It requires a trained observation and experienced person to do this. We attempt to create a neural network to do this estimation of the emotion behind the micro-expression.

Chapter 3

Eulerian Video Magnification

In a report it has been said that ”an investigator has more accuracy in detecting lies while observing a video rather than confronting a subject in live action” [13]. In other words, a video can provide a better visualization of the minute and rapid changes that are occurring in the face. These rapid changes occurring in the face are rather hard to observe and analyze in real time. For an expert, it can take days of time for determining the right emotion behind a particular micro expression for a given video of a very short duration. These videos containing micro expressions which are of short duration have to be played several times at a slower rate before concluding any emotion. It not only takes a long time, but also takes lots of energy and an unbiased mind to conclude a certain emotion, generally, which is difficult to achieve. Traditional lie-detection schemes demand a physical contact with the subject under test, this triggers extra consciousness in that subject. Also, sometimes traditional lie-detection schemes generate false-positives if there is any kind of arousal in the subject being tested. Therefore, video based techniques create no physical contact with the subject under test which is an advantage over the traditional systems.

Micro expressions are hard to identify and so it is almost impossible for an average person to detect and decipher micro expressions because of its impulsive nature. But more recently, a design called Eulerian Video Magnification (EVM) has been proposed, which can be used for observing these micro expressions. EVM magnifies small and subtle motion that is captured in a video, which are generally impossible to identify with naked eyes [25]. EVM uses the method of spatial decomposition on input video and then temporal filtering is applied to each and every decomposed frame of the video. EVM technique not only magnifies small and subtle motions, but when the decomposition method and the filters are changed, it also magnifies the color intensity variations. Motion magnification is used to magnify the subtle changes occurring in the face. Micro expressions are the subtle changes which occur for a small amount of time, these changes can be magnified with motion magnification. Color magnification is used to magnify the color changes in the face, which helps in finding the pulse rate of the subject under test. Pulse rate of the subject under test acts as add-on for this micro expression detection. Thus, without any physical contact, EVM helps in detection of micro expressions. Before investigating the deeper insights of EVM, spatial decomposition of an image using pyramids has to be understood.

3.1 Conceptual learning about Pyramids

A Pyramid is a structured and successively condensed information of images [24]. A pyramid structure represents an image at more than one resolution. Pyramids are generally used in motion estimation. A pyramid structure contains an original image and consecutive images with lower resolution of the original image. This consecutive images are formed by passing the original base image into a low pass filter and sub-sampling the result. The new image, thus formed is called the first level image [24].

This first level image has half the resolution of the original image. First level image, thus obtained is again passed into a low pass filter and then it is sub sampled. This process of forming image levels continues. Top image of the pyramid structure has an image with smallest size and lowest resolution [24]. There are two kinds of pyramids, Gaussian pyramid and Laplacian pyramid.

3.1.1 Gaussian Pyramid

In the process of pyramid structure, a low pass filter with separable function of 5X5 point impulse response and down-sampling are used to form a Gaussian pyramid. In other words, low pass filtered image is down-sampled to get the next layers in Gaussian pyramids. The low pass filter has to be separable with 5X5 impulse function as shown below:

$$h(n_1, n_2) = h(n_1)h(n_2) \tag{3.1}$$

$$h(n) = \begin{cases} a & n = 0 \\ \frac{1}{4} & n = \pm 1 \quad [19] \\ \frac{1}{4} - \frac{a}{2} & n = \pm 2 \end{cases}$$

where, a=0.3 to 0.6 and at 0.4 it has Gaussian shape.

The Gaussian pyramid structure representationof the original figure 3.1 is shown below in the figure 3.2

Figure 3.1: Gaussian pyramid structure representation [17].

Figure 3.2: Gaussian pyramid structure representation [17].

3.1.2 Laplacian pyramids

Major drawback with Gaussian pyramid structure is that it has high redundancy. This high redundancy is because of the low pass filtering. In Laplacian pyramids, Gaussian pyramid structures are directly used. The first level decomposed image of the Gaussian pyramid structure is up-sampled to the size of the original base image. This up-sampled image is subtracted from the original image, which results in an image with sharp edges [24]. The resultant image has characteristics of a high pass filtered image. In other words, the difference between 'i+1'th level image of the Gaussian pyramid structure and the 'i' th level up-sampled image, is the output. The Laplacian pyramid structure is thus formed by all these levels of images with sharp edges. The Laplacian pyramid structure representation is shown below in the figure 3.3

Figure 3.3: Laplacian pyramid structure representation [17].

3.2 Video specifications and prerequisites of EVM

The video specifications considered in the EVM are

1. The size of the video frame is 640x480.
2. The frame rate for videos are 30 FPS.
3. The videos are considered either in '.avi' or '.mp4' formats.
4. The videos to be analyzed are recorded in standard one-to-one interview setup format.

3.2.1 Standard one-to-one interview setup format

Videos are recorded in standard one-to-one interview to overcome few artifacts.

1. Irregular ambient lighting conditions have a significant effect on the detection of micro expressions. Lighting should be relatively uniform.
2. The tracking of a person's face is affected by background disturbances, which might cause misclassifications to occur.
3. Accuracy of face detection and emotion recognition gets influenced, when people are around. So, videos recorded at one-to-one interview setup are considered.
4. When there is no proper standardization and stability in videos, which means videos that are recorded without tripod has many artifacts. EVM magnifies even the slightest jerks that occur while recording without a tripod.

Thus, to overcome all the above mentioned problems, videos recorded at standard one-to-one interview setup are considered.

3.3 Eulerian Video Magnification

In EVM, certain spatial locations are selected to amplify the variations in temporal frequency bands. Temporal filtering amplifies both color and motion. Video frame decomposition is done using Laplacian pyramid structure, due to two different reasons as specified [25]. After decomposition of these spatial bands, temporal filtering is applied to each band. In temporal filtering, a pixel value in the frequency band with the corresponding time series is being considered and later a band-pass filter is applied [25]. Extracted band-pass signal is then multiplied with an amplification factor of α, where, α is user and application specific. This amplified signal is added to the original. All these spatially decomposed frames are again reconstructed to form the final output, where the motion or color is magnified. The methodology of EVM is shown below in the figure 3.4:

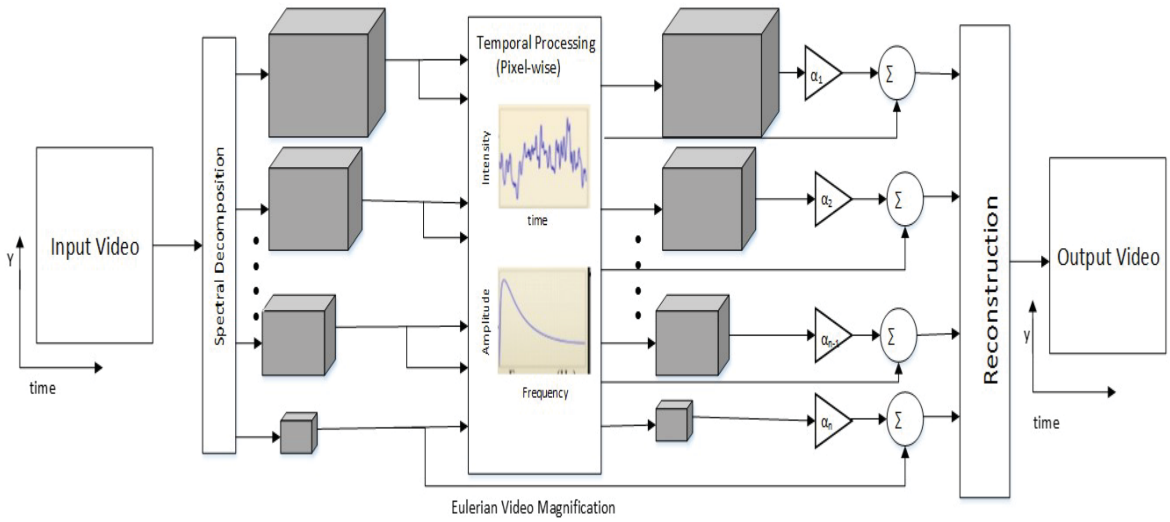


Figure 3.4: Methodology of EVM [25].

There are four steps in practical application of EVM (motion and color)

1. Selecting the temporal band-pass filter [25].
2. Selecting the amplification factor α [25].
3. Selecting spatial frequency cut-off (using cut-off wavelength, λ_c) [25].
4. Beyond this cut-off frequency, the amplification factor α is attenuated for $\lambda < \lambda_c$, i.e., amplification factor is forced to zero or amplification factor is linearly scaled down to zero [25].

This amplification factor and cut-off frequencies are application and user specific.

Micro expressions of emotions are non-periodic changes in the face. EVM magnifies non-periodic motions of the face, when these changes are within the pass-band of the temporal pass-band filter [25]. Thus, EVM magnifies non-periodic movements with smaller magnitudes that are exhibited by the face. EVM not only works for long duration videos, but it also works for videos with very short durations.

3.3.1 Motion Magnification

Subtle motions that are invisible to the naked eye are amplified using motion magnification of EVM. This motion magnification helps in observing those subtle and spontaneous movements in the face (micro expressions) that can easily go unnoticed. In motion magnification, exaggerating the motion by amplifying the changes in temporal color at fixed pixel values is done, rather than using the traditional motion estimation algorithms. Laplacian pyramid structures are used for spatial decomposition of motion.

Temporal changes occurring in motion magnification are analyzed using first order Taylor series expansion [25]. Motion magnification is demonstrated for both small and large motions. For large motions, higher frequencies and large amplification factor is used [25].

The amplification factor α is represented as a function of spatial wavelength λ and motion magnification of video motion $\delta(t)$.

$$(1+\alpha)\delta(t) < \tfrac{\lambda}{8} \tag{3.2}$$

In general, motion magnification uses temporal filtering with the broad pass-band. Also, sometimes a low order IIR filter (of order of 1, 2) are used [25].

The parameters concidered for motion magnification are given below:

α	15
λ level	4
Lower Cut-off Frequency	1 Hz
Upper Cut-off Frequency	2 Hz
Sampling Rate	30
Chrome Attenuation	2
Temporal Filter Used	Ideal Filter

Table 3.1: Parameters concidered for motion magnification

The motion magnified video frame is shown below in the figure 3.5:

Figure 3.5: Motion Magnified Video frame.

3.3.2 Color Magnification

Color magnification is used to find out the blood flow in the face, which is invisible to the naked eye. Thus, without any physical contact, pulse rate is calculated. The process for color magnification is same as that of motion magnification. Color magnification varies with motion magnification with the choice of temporal filter and the pyramid

α	30
λ level	4
Lower Cut-off Frequency	0.833 Hz
Upper Cut-off Frequency	1.2 Hz
Sampling Rate	30
Chrome Attenuation	2
Temporal Filter Used	Ideal Filter

Table 3.2: Parameters concidered for colour magnification

structure that it uses for spatial decomposition. Color magnification uses Gaussian pyramid structures for spatial decomposition. The use of Gaussian pyramids in color magnification is done since they reduce the quantization noise and boosts the color changes in the pulse [25]. In general, a narrow pass-band filter is used. Sometimes, ideal band-pass filters are used, since they have sharp pass-band cut-off frequencies [25]. IIR filters having cut-off frequency W_l and W_h with orders 1 or 2 can also be preferred.

The parameters concidered for color magnification are given below:

Videos are always rendered in RGB color space. In color magnification, color spaces are moved from RGB color space to YCbCr color space. YCbCr color space is used in color magnification so as to reduce the artifacts of the pixels and it also it also allows to observe the clearer color changes of blood flow in the face. After working on YCbCr for intensity variations, the color space is again converted back to RGB color space to see the red and green color changes in the face. The color magnified video frame is shown below in the figure 3.6:

Figure 3.6: Color Magnified Video frame.

Thus, motion and color magnifications of the videos provide a platform to observe the changes that are occurring in the face without any physical contact. Motion magnification helps in observing the changes occurring in micro expressions and color magnification helps in observing the pulse rate of the subject under test. The next challenge is to analyze motion and color magnified videos.

Hence, the subtle changes occurring in the face can be observed, but how these observed changes are correlated with an emotion?

Also, the color changes occurring in the face are observed, but how to determine the pulse rate of a subject?

How to analyze this motion and color magnified videos?

The analysis can be done by first extracting a suitable feature descriptors for every frame. These feature descriptors describe the image as a whole by extracting local patterns. The features used are HOG features, which extract the local gradient information, discussed in detail in Section 4.4. This descriptors are given to a trained Artificial Neural Network system for further analysis of the emotion, detailed in Section 6.1. From the colour magnified video the pulse is determined by extracting the colour changes at a specific ROI, which is further discussed in Section 6.2. The results have been graphically presented for easier interpretation in Section 8.4 and 8.5.

Chapter 4

Face Recognition and Feature Extraction

For an analysis of the face that is obtained from the motion magnified video, recognition and extraction of facial features are to be done. Analysis of the face is mandatory for correlating the micro expression with its corresponding emotion. In the process of finding the emotion that a face corresponds to, firstly, the face is been recognized and then features from the cropped face image are extracted. The prerequisites that are necessary for the Artificial Neutral Network (ANN) analysis and classification of the emotions are discussed in this chapter. This chapter deals only with the branch of motion magnified videos. The pulse detection from color magnified videos are discussed later.

Facial feature extraction is a vital aspect of recognizing and classifying the emotions. The feature based system operates much faster than pixel based system [23]. The face is recognized in the given frame so that features can be tracked later, and other unimportant part of an image is excluded. This recognized face is given as input to Histogram of Oriented Gradient (HOG) feature extraction, so as to get a single vector matrix of each image. These matrix vectors are given as inputs to the ANN for analysis, detection and classification of emotions. The first step of recognizing faces, is done by using the popular Viola-Jones (VJ) algorithm.

4.1 Conceptual learning about Voila-Jones algorithm

VJ uses four different algorithms to recognize the face, such as, Haar-Like features, Integral Images, Adaptive Boost (AdaBoost) and Cascade of classifiers. Haar-Like Features and Integral Images are used for getting all the data from each pixel features. AdaBoost and Cascade of classifiers are used for sorting out the exhaustive data that is obtained from Haar-Like Features and Integral Images. Outputs that are required for recognizing the face are obtained after AdaBoosting and Cascade of classifiers.

4.1.1 Haar-Like features

VJ algorithm first calculates the Haar-Like features of an image with a base resolution of 24x24 pixels. VJ algorithm uses four kinds of masks as shown below, on the sub image of 24x24 pixel size to extract features [23]. The difference between the sums of pixels within two rectangular features are calculated by features are calculated using the masks shown in fig4.1 i.e., the sum of the pixels in white region is subtracted from the sum of the pixels in the dark region. An exhaustive set of data ranging from 160k to 180k is obtained from four different kinds of feature sets which are used for calculating the difference [23]. The amount of data, thus obtained from Haar-Like features are very large to deal with.

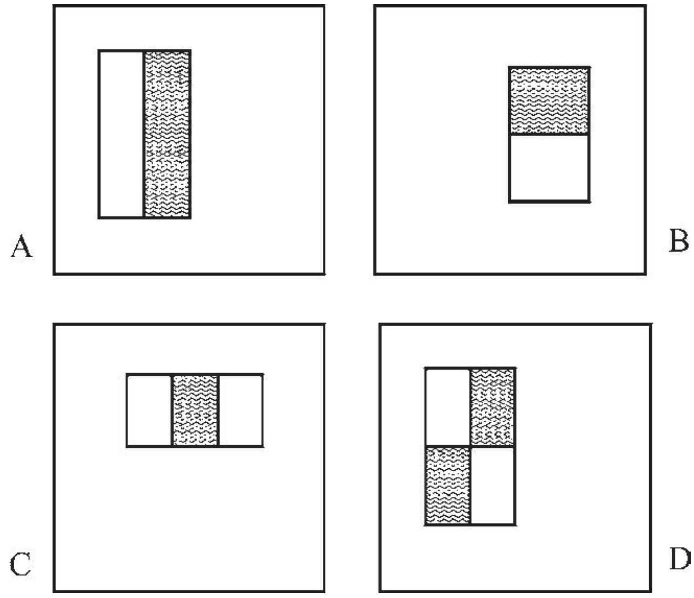


Figure 4.1: Four kinds of rectangular features used in VJ algorithm [23].

4.1.2 Integral Image

The exhaustive data sets that are obtained from Haar-Like features are very difficult to handle. So, the concept of integral images is being introduced by Viola-Jones et al. which directly works on summed image intensities of a given image to reduce the complexity [23].

$$I(x,y) = \sum_{\substack{x' < x \\ y' < y}} i(x', y') \tag{4.1}$$

With the help of summed intensities of integral image, only four corner points for a given sub-image are used for calculating this Haar-Like features, which in fact decreases the processing time to a great extent. In integral images, only four array references are required for computation of any rectangular sum [23]. The figure shown below calculates the sum of image intensities in the region D using only four image references from the sum of intensities 1, 2, 3, and 4. The sum intensities at A is 1; sum of intensities at B is 1+2; sum of intensities at C is 1+3; sum of intensities calculated at D is (4+1)-(2+3) [23].

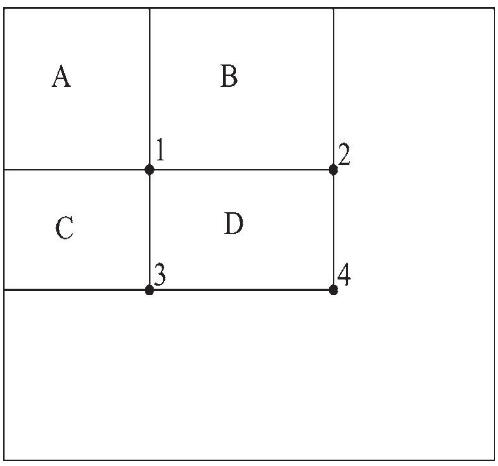


Figure 4.2: The sum of intensities 1, 2, 3 and 4 and regions A, B, C and D [23].

4.1.3 AdaBoost

The data thus obtained from integral images have to be classified. For classification or prediction, ensemble techniques use a group of models, rather than just relying on a single model. Therefore, the predictive performance of the system is increased by this combination of multiple models [23]. AdaBoost is an ensemble technique used in machine learning. In other words, boosting involves the use of weighted sum of weak classifiers. While boosting, misclassified data in the training is considered and is given more priority for next classification. The process of working on misclassified data using different models is called Boosting. The weights at each classification are adjusted based on weighted error.

The models at each classification are called weak classifiers. Allotting weights for each weak classifier and summing all the weak classifiers/models gives rise to a strong classifier. The disadvantage with boosting algorithm is that more training tends to over-fit the data.

4.1.4 Cascade of Classifiers

A cascade of classifiers is used in recognizing the face. This cascade of classifiers achieves increased detection rate and higher performance with reduced computation time. At first, classifiers with lower thresholds are used and the ones below them are rejected. Later, classifiers with higher thresholds are used to achieve low false positives [22]. In other words, the positive response from first classifier triggers the second classifier, a positive response from second classifier triggers the third classifier and so on. A negative result is rejected by classifier at any stage.

These stage classifiers are built using the AdaBoost algorithm. Using Haar-Like features, face is recognized by the trained system in a cascade of classifiers. In the process of recognizing, if Haar-Like features fail, this failure determines that the face is very likely not present at that location. This means, the current location is eliminated for all other Haar-Like features without any further introspection and processing is done to another location for recognition of face.

4.2 Recognition of Face using Voila-Jones Algorithm

Using these four different algorithms, Voila-Jones detects and recognizes the face. The practical application of the VJ algorithm in MATLAB is to use the built-in object 'vision.CascadeObjectDetector' for face detection. This built-in object detection framework is found in the 'Computer Vision Toolbox'. When the framework is applied on an image, the output is given as a 1x4 matrix i.e. [a b c d]

(a, b) are the coordinates of pixel value where the face region starts; c is the height of the face; d is the width of the face.

The values, thus obtained from the output of VJ algorithm are then used in cropping the recognized face only. This cropping is done by using 'imcrop()' function of 'Image Processing Toolbox' in MATLAB.

4.3 Conceptual learning about HOG features

Facial Features such as forehead, eyes, nose, and mouth are important parts of facial data, which describe the face completely. Feature extraction is an ad-hoc process of extracting desired key points from an image which gives a detail analysis of the picture as a whole. In general, features in an image can range from simple pixel intensities at a particular section, to a more complex local gradient magnitudes and so on [3]. Feature extraction is a process of extracting the desired feature using algorithms. Features can be two types.

1. Geometric Features: Features that provide information about the shape and location of facial components. These features are extracted through the geometry of the face [3].

2. Appearance Features: Features that provide information about appearance changes such as wrinkles, bulges, and furrows. These features are extracted by tracking minute facial intensity changes in a particular area using various filters [3].

HOG features are a set of appearance features. HOG features give normalized gradient information that is extracted locally in an image [3]. HOG feature extraction involves, dividing the image matrix into cells and then applying gradient kernels to extract the gradients. Numerous gradients are taken into account to find the overall appearance of the micro expression. HOG feature computation is done in four stages.

4.3.1 Gradient computation

The 1-D gradients of the image are calculated using a 1-D centered point discrete derivative masks. Gradients are calculated either vertically/horizontally or in both the directions. Generally, [-1 0 1] and $[-1\ 0\ 1]^T$ point discrete derivative masks are used [3].

4.3.2 Orientation BiANNing

In this step image is divided into cells of sizes 4x4 or 8x8 and the gradient histogram for each cell is calculated. The histogram chaANNels are either spread into unsigned 0-180 degrees or signed 0-360 degrees bins [3]. These bins are then quantized. The histogram of these quantized bins is the output.

4.3.3 Descriptor Blocks

Blocks are a matrix of cells. The block size is generally 3X3 or 6X6. Normalization is performed on these blocks, so as to account for changes in the illumination and contrast [3]. In other words, the gradient must be normalized locally, which requires grouping of cells into blocks. Normalization is done effectively by using block overlapping. Generally, Rectangular blocks (R-HOG) are preferred over Circular blocks (C-HOG).

4.3.4 Block Normalization

Block normalization is done either by considering L1-norm or L2-norm. The possible ways of normalizing blocks are given below:

$$L1-norm: \quad f = \frac{v}{||v||_1 + e} \tag{4.2}$$

$$L1-sqrt: \quad f = \sqrt{\frac{v}{||v||_1 + e}} \tag{4.3}$$

$$L2-norm: \quad f = v/\sqrt{||v||_2^2 + e^2} \tag{4.4}$$

where ‘e’ is an infinitesimal arbitrary value and ‘v’ is the non-normalized histograms of all the blocks.

4.4 Feature Extraction using HOG features

Images of the database or the video frames obtained after recognizing the face are resized to 64x64 pixel size using nearest neighbor interpolator. This resizing of an image/frame is done so as to have a uniform size for all the images or frames. The resized image / frame is given to HOG features as input to extract the features of the face as a single vector. The parameters considered in HOG are as follows:
Image/Frame size: 64x64;
Cell-size: 4x4;
Block size: 2x2;
Block overlaps: one block;
Block Normalization: L2- Normalization;
Number of bins: 9, unsigned orientation.

The output matrix is reshaped into an 18x450 matrix so that each column consists of orientations of histogram bins. Mean values of these histograms i.e., 1x450 matrix is given as inputs to the AANN. This 1x450 matrix is the input vector for single image of the database that is given to AANN for its training. The mean values of micro expression patterns of test videos are also considered. These extracted features of the test video are given to the ANN to find the matching emotion in the training set.

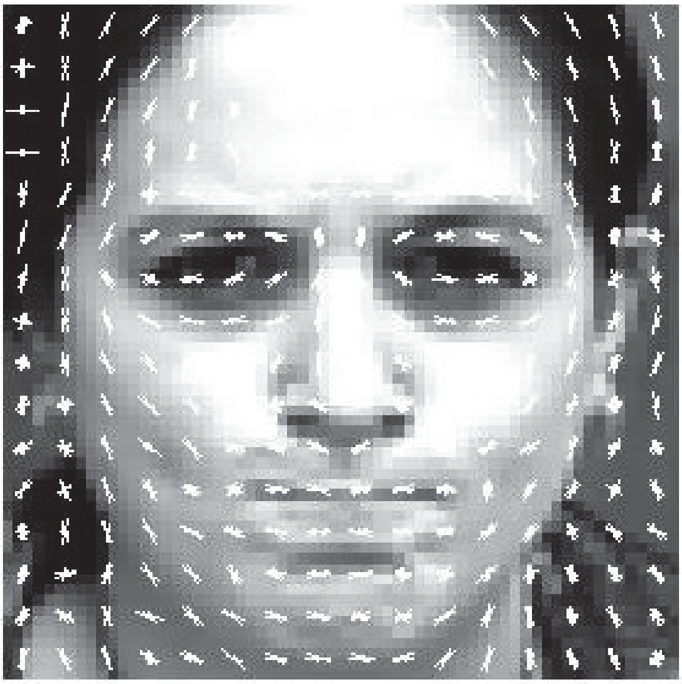


Figure 4.3: HOG feature orientation
[17].

Two feature extractions are done here. The first one is VJ algorithm that deals with extracting the image intensities which corresponds to a face. VJ algorithm is a geometric feature extraction. Secondly, the HOG features extract the intensity variations from the extracted geometric features. The HOG feature extraction algorithm is an appearance feature extraction.

The geometrical feature extraction based VJ algorithm extracts the image intensity matrices i.e., it can extract nose, eyes and mouth regions. After extraction of each region, these features are given to ANN for identification. The outputs, thus obtained from ANN are identified as FACS numbers. Now another ANN is to be trained in identifying the emotions based on these FACS numbers. Also, the features extracted from VJ algorithm are not enough to cover a few more important corners of the face such as the cheeks, chin and the upper forehead. The input set would be inadequate without them and thus, leads to misclassification of emotions.

To avoid these problems, the whole face is used as an input to ANN. But, providing the ANN with mere image intensities of the whole face might lead to lots of errors while predicting the emotions. Other features such as edges and corners haven't given the desired results, leaving HOG features to be more accurate and appropriate. Appearance feature extraction based HOG features with a 4x4 pixels/cell-size improved the results drastically. Thus, a single ANN is sufficient when HOG features are used as inputs.

Therefore, for analysis and detection of micro expressions, prerequisites of face detection and feature extraction is done using VJ algorithm and HOG feature extraction algorithm respectively. Now, to correlate the micro expression with its emotion, the ANN needs to be perfectly trained. Hence, a comprehensive image database with classified emotions is required for ANN training.

Chapter 5

Database of Images

Micro expressions always contain emotions in them. Micro expressions are coded with emotions using FACS [15]. If a database of images has micro expressions and emotions labelled according to FACS, then by using a Artificial Neural Network (ANN), emotions can be trained to classify them easily in the video frame. That means, a ANN has to be trained perfectly before testing it. Therefore, a ANN requires a right kind of database not only for training, but also for validating and testing the outputs. This database design should also consider various factors such as transition among expressions, reliability, eliciting conditions, lighting conditions etcetera [16]. The generalized image database is used for various applications in the field of facial expression recognition and also, comparative tests on a common image database helps in determining the strengths and weakness of various methodologies [16].

5.1 Cohn-Kanade image database

In the year 2000 an image database, with 486 sequences of facial expression images from various subjects have been collected to create this database and is popularly known as Cohn-Kanade database. Initially, Cohn-Kanade image database is designed only for facial expression analysis and these expressions are classified using a method called FACS [16]. FACS coding is verified manually, which makes the database quite reliable. This image database includes samples from various backgrounds, sex, varying skin color and people with eyeglasses [16]. The disadvantage with this database is that emotion labels are not specified, which means, FACS are to be used manually to code these emotion labels.

5.2 Extended Cohn-Kanade image database

Later, in the year 2010, an extended image database has been introduced to overcome few drawbacks of the original. In extended database, another 107 sequences across 26 subjects have been added to the original image database [17]. Thus, the extended database contains a total of 593 sequences across 123 subjects [17]. The extended image database also has emotion labels. These emotions labels are used in classification of basic universal emotions. Extended database also contains FACS labels and Landmarks. Extended database provides information about how to understand these labels of emotions, FACS and landmarks. All expressions may not have emotions, and FACS labels show the action units involved in the peak of the expression. This means every peak image in the database contains a FACS label, whereas, only few peak images in the database have emotions labelled to them. The database has subjects eliciting expressions which are both spontaneous and forced choice of emotion (or deliberate). These deliberate emotions are elicited because the subjects are prepared to elicit emotions rather than getting involved in the activity while experimenting.

5.3 Discussion about emotion labels in extended Cohn-Kanade database

Only few expressions have emotions in them, it simply means only few image sequences in the database has emotion labelled to them. Extended database also has few sequences of the recently added universal emotion - contempt. All these emotions, including contempt are labeled using the FACS investigators guide as in [15] [17]. There are a few criteria according to which an emotion has to be categorized using a FACS investigators guide as shown below in the table 5.1 [17]. In other words, an expression should have the following criteria to be called as a micro expression, which in turn reflects an emotion.

Emotion	**Criteria of FACS, which are mandatory**
Anger	AU 23 and AU 24
Disgust	Either AU9 or AU10
Fear	AU1+AU2+AU4. If AU4 is absent, then at least AU5 with intensity E must be present
Happy	AU 12
Sadness	Either AU1+AU4+AU15 or AU11
Surprise	Either AU1+AU2 or AU 5. With an intensity less than B for AU5
Contempt	AU14

Table 5.1: FACS criteria for categorizing emotions [17].

Only 327 sequences out of 593 sequences have qualified to meet the criteria posed by discrete emotions. Emotions are classified and tested using Active Appearance Models (AAMs) / multi-class SVM [17]. The results are as shown below in the table 5.2:

Emotion	**Percentage of classification accuracy**
Anger	75.0%
Disgust	94.7%
Fear	65.2%
Happy	100.0%
Sadness	68.0%
Surprise	96.0%
Contempt	84.4%

Table 5.2: Confusion matrix of extended image database [17].

These emotion labels are also verified by coding FACS manually. Classification accuracy specified in the extended database forms the basis for verifying and validating the classification accuracy results that are obtained from the confusion matrix of trained ANN. Validating the classification accuracy values of trained ANN with database classification accuracy is more appropriate because the ANN is to be trained with this extended database of images only. Also, validating the results with some other databases completely falsifies the intention.

5.3.1 Advantages of using ANN over multi-class SVM

An SVM is a two class classifier. A multi-class SVM is built using a number of single SVM's. To classify 6 emotions and neutral, 7 SVM's are to be used independently.

The disadvantages of using such multiple two-class classifiers being multiple firing of the outputs simultaneously for the same input. This multiple firing of the outputs have to be considered and various computational intensive measures have to be taken to clear out the ambiguity. One such being the calculation of highest output function and voting the outputs.

Another disadvantage of using such multiple binary-classifiers for multi-class classification in-case of emotion-extraction is that a micro-expression contains a combination of emotions. In that combination, the expression related to the dominant emotion is elicited by that person, for example: a person might feel anger and fear simultaneously, but when anger is the dominant emotion, the muscles corresponding to anger give a reaction. SVMs try to isolate an emotion purely, and so multiple SVMs fire at the same time since they work independently, e.g.: when the SVM for anger is fired, there is high chance of SVM corresponding to either sad, fear or disgust being fired as they are all related emotions. Neural Networks multi-dimensional outputs gives us an option to view the intensity of each emotion in a specific micro-expression.

5.4 Working with images of extended Cohn- Kanade database

All the available image sequences are classified manually according to the given emotion labels as specified. Micro expressions are symmetric in nature. While observing the sequences, there are few subjects exhibiting emotions deliberately (or forced choice of emotion), which are resulted from asymmetric micro-expressions. The image sequences of these subjects have been ignored. Extended database also has few sequences rendered in color. Those images are converted to grayscale images in MATLAB.

The images have been cropped automatically in MATLAB using Viola-Jones method for face recognition. This means, extracting only the features of the face i.e., cutting down the background elements and focusing on the front face (forehead, eyes, nose, cheek and chin) of the image. In this process few images have been wrongly classified as a face, which therefore have been deleted manually. Again, another round of extraction is done using Voila-Jones method for finer face cropping without any artifacts. Later all the images are resized to 64X64 pixel images using bi-cubic interpolator. The image resizing is done to have a uniform size for all the images in the database. Later, feature extraction for each of these 64X64 sized images is done using HOG features. The output of the HOG features of a single image is a 1X8100 array, which is reshaped to 18X450 matrix and mean value of every column is taken to get a 1X450 matrix. This 1X450 is the input vector of a single image that is given to ANN training.

The number of images that are left after resizing the database are categorized according to the alphabetical order of emotions as specified in the extended database. These categorized images are used in training and testing of the ANN. The number of images for each emotion are shown below in the table 5.3:

Emotion	Number of images that are left after resizing
Anger	714
Contempt	154
Disgust	540
Fear	413
Happy	962
Neutral	1547
Sadness	509
Surprise	847

Table 5.3: Number of images left for each emotion.

It is clear from the table that the images carrying the contempt emotion are very less in number to train when compared to other emotions. Therefore, the contempt emotion has been omitted and the rest six emotions are considered for training the ANN.

Chapter 6

Neural Network and Pulse extraction

Bias is one of the important things to be considered while evaluating a person or a situation. This bias can change the way people perceive a situation or other humans. The human mind is always biased, sometimes very slightly biased, and sometimes a bit more significant. This bias cannot be eliminated from a human mind, and this bias is the basis for a human mind to reach a solution, either by contradicting himself (if he's not in the right direction) or by following the intuition (if he's in the right direction). This existence of this bias need not have a reason, it is just simple intuition in some cases or simply a hypothesis based on the physical appearance.

As discussed earlier, this bias sometimes affects the investigator while perceiving a subject. Since, even the investigator has to deal with emotions for himself when judging a subject and an investigator cannot always judge a subject just by solely depending on his intuition or hypothesis based on physical appearance. Thus, this study helps in taking the right direction from the start of the investigation. Basing on this new study, the investigator can always assess the subject with the right kind of mind. This kind of unbiased start and analysis can be achieved through an artificial system with an ability to learn and judge. So, a Neural Network (ANN) is designed to overcome the problem in hand. This ANN is also useful in helping average people to understand the concepts of micro expressions. ANN helps people in observing and analyzing the micro expressions in day-to-day life. Thus, it acts as a tool and makes people learn and understand these micro expressions.

The motion magnified video obtained from the EVM is ripped into video frames and features are extracted. Video frame ripping is done on a duplicate copy of the original magnified video and the original video is retained for later use. As mentioned earlier, the frames from a duplicate copy of the magnified video is taken and the face area of the frame is cropped using VJ algorithm and HOG features are extracted and these features are fed into the trained ANN, thus, the outputs are plotted and a video of the graphs is created. The original motion magnified video that is retained and the graph video is played in parallel so as to simplify the work of the investigator. This procedure of magnification, graphing and video construction is done for the color magnified video too. But the graph for color magnification is a pulse graph. The four videos, pulse graph, emotions graph, color magnified and motion magnified videos, when played simultaneously would help the investigator get an overall picture of the subject.

ANN design doesn't detect any lies but only helps investigators in deciphering the emotion behind a specific expression thereby fastens the process of detecting lies. ANN output is neither lie nor truth. The outputs for ANN are six emotions and neutral expression. This tool helps the most when the investigator is unable to detect the micro expressions and decipher their meaning, due to various reasons such as emotional turmoil or expectation/biasing factor. The investigator has to further probe into the investigation to get more cues relating to the subject. To detect lies a trained investigator has to examine both the motion magnified and color magnified video to conclude an emotion. For others, the graphs are presented for a pictorial representation of the emotion and the pulse density. These can be used to exactly estimate the emotion of the subject at a particular time.

6.1 Artificial Neural Network using back propagation algorithm

ANN's are computational models. ANN is an artificial mathematical model built to mimic the basic functionality of the central nervous system of a living being [19]. These are networks of a simple computational elements called the perceptron [19]. A perceptron is a binary classifier, in which the weighted sum of the input is compared with a threshold and the corresponding output of either 1 or 0 is given [19]. ANN is widely used for solving machine learning and pattern recognition problems. Since, a rule-based programming isn't suitable for the applications which require to learn and solve problems using the patterns identified in the data specifically in the field of pattern recognition, speech recognition and computer vision as it involves learning each problem differently and demands an ad-hoc solving. Here, ANN is used for analyzing, detecting and observing emotions from micro expressions [21].

The outputs of ANN are six emotions (anger, disgust, fear, happy, sad, surprise) and also neutral (without any emotion). Most of the times, when a subject under test elicits some micro expression, it starts and ends with a neutral expression. ANN has to give output for each and every frame. Falsified outputs are produced when the ANN is not trained with a neutral expression for a given frame with a neutral expression. So, the ANN is also trained for neutral expression. The outputs of ANN are classified in the alphabetical order i.e., anger, disgust, fear, happy, neutral, sadness and surprise.

When the problem in hand is a pattern recognition, Back propagation algorithm is a popular and widely used algorithm specifically for pattern recognition [4]. Back propagation algorithm can solve this emotion recognition problem. Back propagation algorithm, as the names suggest, propagates output error values backwards and adjusts the network weights accordingly [4]. Back-propagation consists of two main sub-networks i.e., feed-forward network and a back-propagation network.

Implementation of ANN using back propagation algorithm is done in MATLAB using Neural Network Toolbox.

There are two steps in emotion recognition of a ANN using back propagation algorithm.

1. **Training**: Training of the ANN is done with an image database consisting of micro expressions classified as emotions.

2. **Testing**: Relying on the trained network, the unknown micro expressions and emotions in a video are identified.

6.1.1 Training phase

A comprehensive image database such as Extended Cohn-Kanade database is used for training ANN. Pre-processing of this image database is done. Viola-Jones (VJ) algorithm is applied twice for recognizing the face and for refining the filtered database. All the images are resized to a standard 64X64 size. Then HOG features are extracted from every single image and given as inputs to ANN for the training of each emotion. The input vectors from HOG features are stacked into a matrix containing 450 rows and each column representing a single sample of the training set. For example, the dataset of micro expressions depicting an angry emotion has 500 samples, then the training set corresponding to that emotion is 450x500 matrix.

Thus, all the matrices of all the emotions are stacked to form a single matrix containing 450 rows and number of columns equal to the total samples of all emotions. While creating this training set, simultaneously, the target values are also created, which is a 7x1 vector containing an output of 1 for the corresponding expression and a 0 for the rest of the elements. In this way it is ensured that the number of targets equals the samples in the training set.

This training set and the targets are given to the neural network and trained using the parameters shown below:
Input Vector Dimension: 450x1 double (450 x No. of Samples)
Hidden Layer Size : 20 Neurons
Network : Pattern Recognition Network
Data Division : 70% Training Set, 15% Validation Set, 15% Testing Set
Target Vector Dimension : 7x1 double

The confusion matrix of the trained ANN is shown below in the figure 6.1.

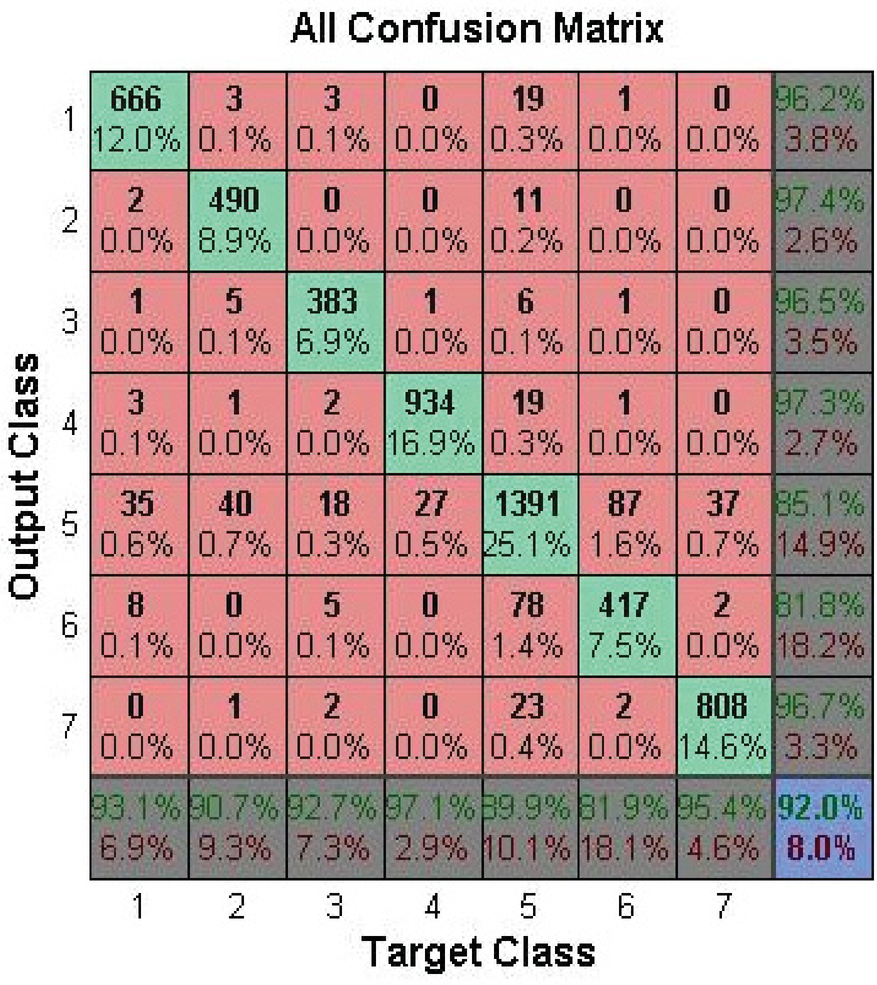


Figure 6.1: Confusion matrix of trained ANN using the Cohn-Kanade image database

The digits shown in the confusion matrix (Fig 6.1) indicate/represent emotions or states in alphabetical order. 1 for Anger, 2 for Disgust, 3 for Fear, 4 for Happy, 5 for Neutral, 6 for Sad and 7 for Surprise.

The average classification percentage of all the emotions with neutral expression is 92.0%

The performance graph of the ANN is shown below in the figure 6.2:

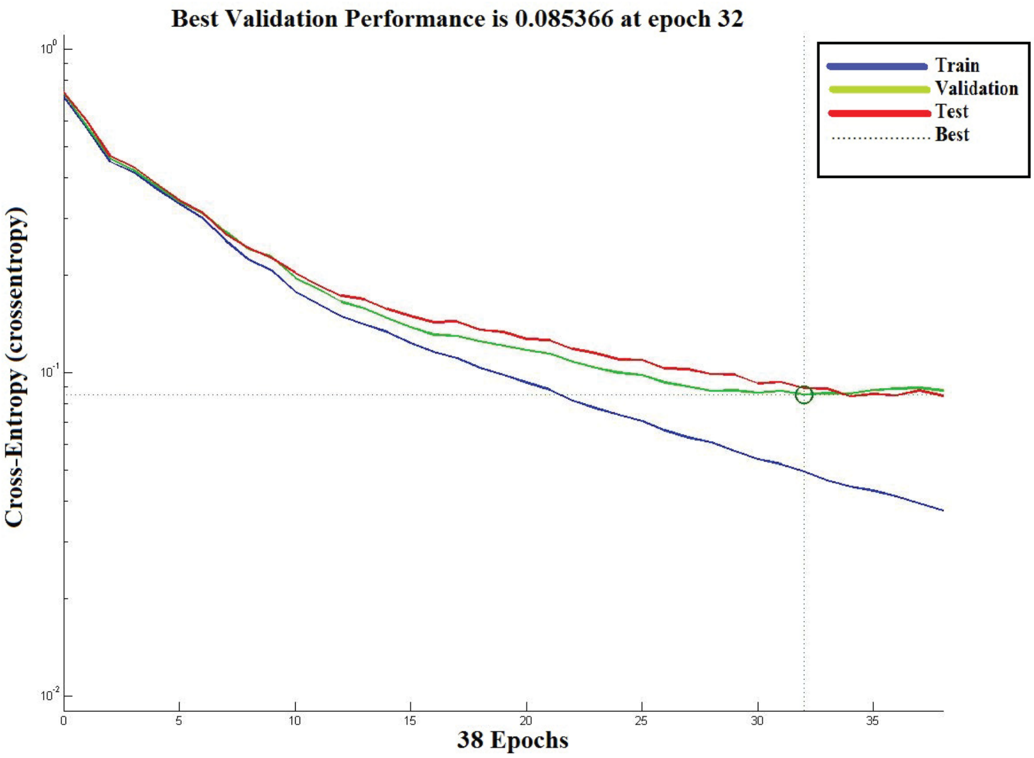


Figure 6.2: Performance plots of ANN.

When the validation curve touches the minimum, the training of the ANN stops. This is done to prevent over-training of the network. "The Neural Network Toolbox" in the MATLAB automatically checks for the minimum and stops.

6.1.2 Testing phase

In the testing phase, ripped frames of a duplicate motion magnified copy video containing micro expressions are considered. Each frame is considered to be an image. Face recognition performed for each and every frame using VJ algorithm. An automated algorithm is written in MATLAB, so that, if no face is detected in the frame using VJ algorithm, that frame is skipped automatically. The inputs of ANN for these skipped frames are zero and so output for those frames are also zeroes. Then all the images that are classified properly are resized to 64x64 pixels. HOG features are extracted and given as input to the trained ANN, this network, which by now is trained, identifies the micro expression in that given frame and fires the corresponding emotion as its output. The image size of 64x64 is from the duplicate copy of magnified video and it is nowhere connected to the output of ANN because the outputs of ANN are only emotions, but not any images. In other words, no frame reconstruction is done on these duplicate images.

When a single frame is processed through the ANN, a ANN may fire all the outputs at the same time but with different intensities. The maximum value out of the seven values in the output vector is considered as the dominant emotion for that particular frame. The output of the ANN is not a vector consisting of only one emotion in an isolated manner, but is a mix of emotions. This shows that a human face doesn't explicitly express only one emotion, but rather a combination of emotions. When all these values are plotted for all the frames, few consecutive frames of a particular emotion would be dominant. Therefore, the inference would obviously be that particular emotion being expressed at that particular time.

6.2 Pulse Extraction

After color magnification on a video through Eulerian Video Magnification, the next important step is the pulse extraction. Pulse extraction is a user-defined MATLAB function created by extracting the temporal intensity variations from the color magnified video.

The function takes a color magnified video as an input. A duplicate copy of the color magnified video is created and this video is ripped into frames. To each of these frames a facial feature extraction algorithm using Viola-Jones Algorithm is applied and the face is cropped. This face feature extraction is done so that no other illumination variations disturb the pulse extraction process. Then the cropped face matrix is converted from the RGB space to YCbCr space as the intensity variations are equally distributed in the Red, Green, and Blue color spaces and pulse variations have to be calculated in each color space to obtain the total intensity variation. To avoid this extra calculations, all the intensity variations are non-linearly mapped into a single color space and the best way is to map them to YCbCr space, since the total intensity variations are concentrated in the luminance component, Y, variations calculated in this single space would provide better results.

After the face feature extraction, the area around the nose, more precisely the area under either eye or the forehead area is considered, since the intensity variations are profound in these areas of a face. The area where the eyes and forehead meet, this area is generally preferred as it is the center of the Region of Interest (ROI, the face). This is because when the video is color magnified the background reflections, due to the light, generally tend to override the blood flow intensity variations. This effect is more in the corners of the ROI, but is less in the center. The intensity variations in the center are only due to the blood flow, so, it's more accurate than the corners. These variations are the mean value of a sub-matrix of pixels under the above specified area. These variations when plotted gives the pulse variation as outputs. The pulse variations doesn't have a perfect pulse peak and consists of abundant noise. This pulse has to be filtered using a Butterworth filter or an adaptive filter. Pulse filtered with an adaptive filter, with a reference to a sine wave of the frequency equivalent to the frame rate has been observed to give a very good smoothening effect and the output looks more like a pulse. The heart beat rate of a normal person is approximately 72 beats/min. The peaks in pulse graph are measured manually and no separate counter is used for measuring the heart beats of the subject under test. The pulse rate counted from the above method is approximately same as that of heart beat that is counted with physical contact (either by stethoscope or just checking the pulse on the wrist). Therefore, the pulse rate of the subject under test is observed and calculated perfectly without any physical contact.

6.3 Outputs rendered as moving graphs

Pulse obtained through the pulse extraction process has to be converted into a moving graph to get an interactive display of the variation of pulse. The pulse data obtained from the previous step of pulse extraction is an array of the intensity variation. This variation is zero-padded to the next multiple of 10.

After the zero-padding, a ten-point rectangular window is made to slide over the array and the ten points are plotted on a figure window with frame numbers on the x-axis and pulse variations on y-axis, and using the ‘getframe()’ function the figure window is converted to a single frame of a video. This is done to the array in total and every interval of 10 points are plotted and framed. All these frames are then combined written into a video file with a frame rate of 30 FPS and then stored. This video when played appears as an ECG being run. When both the color magnified video and pulse video played in parallel, shows the variations in the pulse of a subject precisely.

The emotion result vectors, which are the outputs of the ANN for the motion magnified video is collected. These output vectors are plotted. The plotting is done with frame number on the x-axis and the values of the output emotions on the y-axis. The method for creating the moving graphs of emotions are same as that of moving pulse graph. That is for plotting the emotions, 10 frame’s outputs are considered every time to plot it. This plots on the figure window would be converted into a frame and stacked to generate a video out of them. These graph video, when played, appear like an emotion reading graphs being played. This emotions moving graph when played in parallel with the motion magnified video would make it easier to track the subject’s micro expression and the corresponding emotion simultaneously. The video thus, created in this process has almost the same duration as of the original video. To play these videos simultaneously, a GUI has to be created to make it easier for people to work on micro expressions.

Chapter 7

Graphical User Interface

Graphical User Interface (GUI) has been popular since their introduction, mainly due to the simple presentation and easy execution of a task. GUIs are used in almost every application ranging from a simple installation of a software to the installation of an Operating system and many major applications. The main advantage of a GUI being the ease of using the application without the need for the end-user to know the insights of the commands required to execute a task and GUIs being platform independent, provides an ease for working with just a click of a button.

A GUI is developed for this system so that end-user concentrates more on analyzing the expressions without bothering about the sequence of scripts to be executed. This GUI design provides an extra advantage of playing the magnified video and its corresponding moving graphs simultaneously with one click of a button.

7.1 Working on GUI

Graphical User Interface Development Environment (GUIDE) in MATLAB is used for designing and creating a GUI. The main elements of the GUI design used for this system are the push-buttons and media players. The drag-and-drop feature of GUIDE is used to build the figure window of GUI and place the push-buttons in the right alignment. Using an in-built MATLAB function, ActiveX controls are imported to use the media players in the GUI design. When the GUI is constructed and executed, a script containing the respective call-back functions and a figure window are generated. This generated script is modified, by filling the function with an action to be taken, when the corresponding button is pressed/clicked.

In this GUI, four media players and two sets of 'Play, Pause & Stop' buttons are imported to provide the required playback controls for the given four media players. Each set of the 'Play, Pause & Stop' control two media players simultaneously. Another set of five push-buttons are:

1. 'Open'- Selecting the file to be played.
2. 'Magnify'- Magnifies (both Color and Motion) a video and places the results in a folder.
3. 'Extract'- Extracts the emotions from a Motion magnified video using micro-expressions and pulse from a Color magnified video.
4. 'Import'- Imports all the micro-expression datasets from an image database and creates input and target vectors for the Neural Network (NN).
5. 'Train'- Trains the NN with the imported datasets and stores the network for emotion extraction.

The operations specified above are coded into the call-back functions of the corresponding push-button. For example, the 'Magnify' button contains all the call-back functions for the magnification process to be done, and 'Extract' button contains the call-back functions to perform the extraction process (Extraction of both pulse and emotions). At the end of every call-back function a structure called the 'handles' is updated for storing of data and also for using the data between different callback functions of the same GUI.

Apart from the push-buttons, the four media-players have to be imported into the GUI using their pragmatic identifier, which gives user the controls of that specific application. This is possible only on a Windows platform. These are then controlled using their corresponding methods and interfaces.

The GUI model of the study is shown below in the figure 7.1:

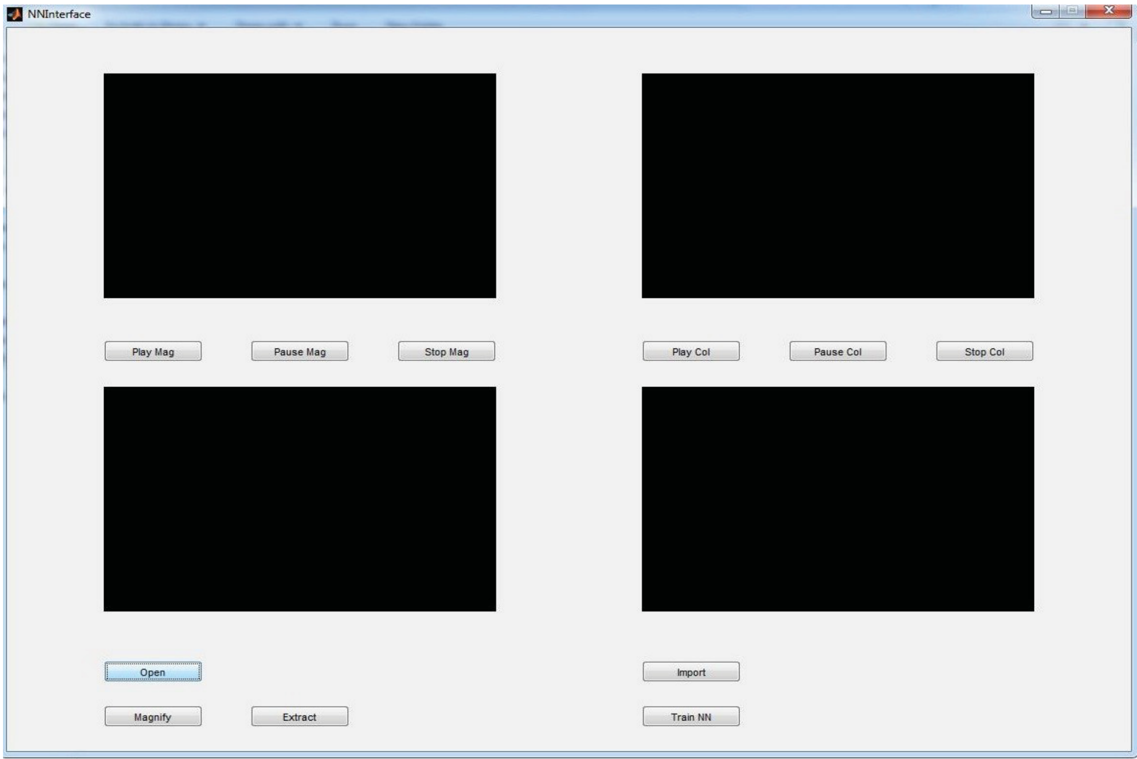


Figure 7.1: GUI model

7.2 Operational flow of GUI

The operational flow for the design using GUI is shown below:

1. Import: The database of images used for NN training are imported.

2. Train: With the imported image database, the NN is trained. Once the NN is trained with a database, there is no need to train a NN every time unless a new database is used.

3. Magnify: A video file with micro expressions which have to be magnified to observe the changes are selected first and later this selected video is magnified. Here the two magnified videos (color magnified video and motion magnified video) are the outputs.

4. Extract: By the use of an automated code, source video, motion magnified video and color magnified video are extracted into another folder with name 'contents

of name (of source file)'. A duplicate copy of motion magnified video is given as inputs to the NN to extract the emotions associated with micro-expressions in the video. From a duplicate copy of the color magnified video, pulse is extracted. Moving graphs of emotions and pulses are the outputs extracted from this motion and color magnified videos respectively. Now these moving graphs are also saved in 'contents of name (of source file)' folder. The names for motion magnified video and moving graph are 'name (of source file) motion mag', 'name (of source file) graph mag' respectively. The names for color magnified video and its moving graph are 'name (of source file) color mag', 'name (of source file) graph col'. Therefore, the 'contents of name (of source file)' folder contains source video, motion magnified video & its corresponding moving emotion graph video and color magnified video & its corresponding moving pulse graph video.

5. Open: Selecting the source video from 'contents of the name (of source file)' folder, automatically imports all the four videos (motion, color and their respective moving graphs) into the GUI. 'Play, Pause & Stop' buttons can either play or pause or stop the magnified videos and its moving graphs simultaneously.

The operational flow chart of the GUI model is shown below in the figure 7.2:

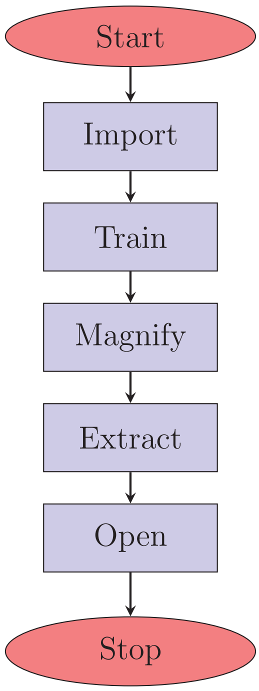


Figure 7.2: Operational Flow Chart of GUI

The GUI model with magnified videos and its corresponding moving graphs are shown below in the figure 7.3:

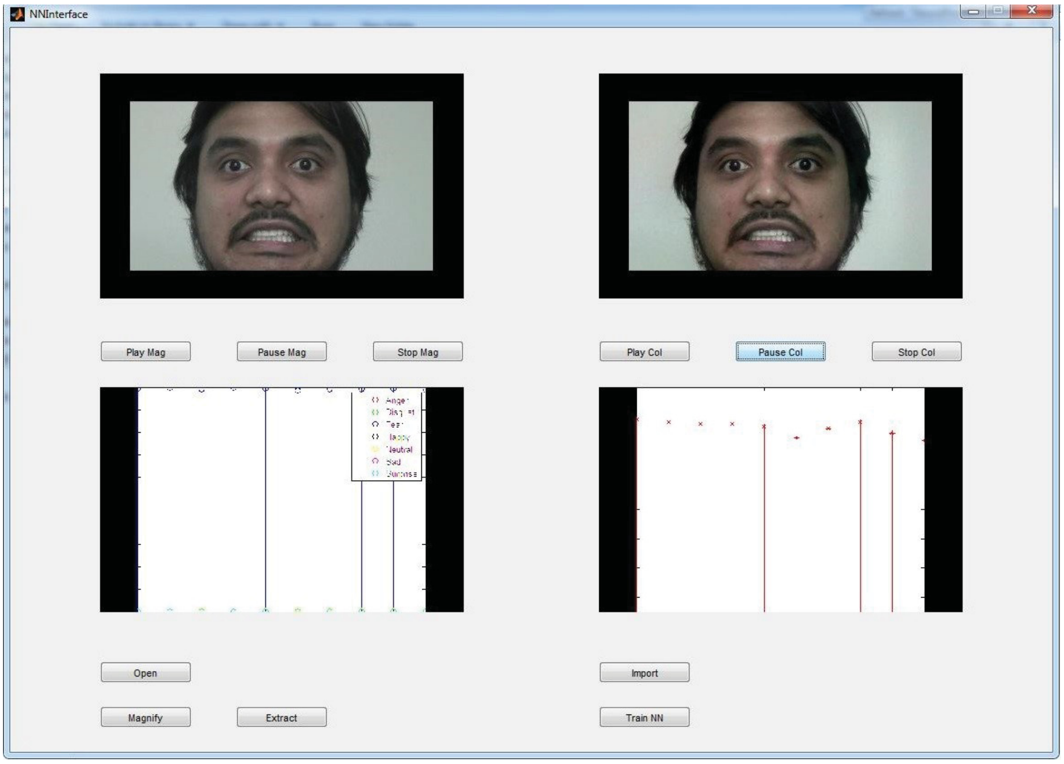


Figure 7.3: Motion and Color magnified videos with their corresponding moving graphs.

Finally, this GUI model presents the total Eulerian Magnification and its Extraction process in a simple and elegant way, so that the end-user need not bother about the background code execution, as the magnification and extraction of videos is done with the simple press/click of a button. Also, the status of each operation is presented to the user using an ample number of wait-bars. Therefore, the GUI model performs all the required operations and acts as a easy-to-use tool for end-user.

Chapter 8

Results

Micro expression extraction using Eulerian (motion and color) magnification is helpful for an investigator in the process of lie detection. If only Motion magnification and its corresponding graphs are considered, they can be used as a tool for learning and understanding micro expressions. Observing, analyzing and learning these micro expressions helps people in designing their life with ease, as it makes people understand what others feel at that given particular instance. When both the color and motion magnification with their corresponding graphs is used jointly, it helps an investigator to detect a deception when posed by the subject under test. More specifically, motion magnification acts as a main tool and color magnification reinforces the results obtained from the motion magnification.

8.1 Performance

Profiling and the time duration for various sub-functions of the system has been evaluated and recorded. This performance estimation has been done using the MATLAB profiler and in-built time estimation functions.

This performance estimation is done on Xenon E5 processor, with 16 GB RAM and MATLAB R2013b version. The time durations of various functions may vary depending upon the processor, RAM specifications and internal processes. The time duration obtained using the below mentioned specifications are as follows:

8.1.1 System Specifications

- Processor: Intel Xenon E5
- RAM: 16 Gb
- Operating System: Windows 7.
- MATLAB R2013b
- Additional Toolboxes: MATLAB Pyramid Tools (MIT), Computer Vision System Toolbox, Image Processing Toolbox, Neural Network Toolbox.

8.1.2 Input Specification

- Total Database samples: 5686 samples (including 1547 Neutral Samples)
- Sample Image Dimensions: 64x64
- HOG Features: 4x4 Cell-Size, 2x2 Block-Size, 9 Bins.
- Training Input Matrix Dimensions to ANN: 450x5686.

- Training Output Matrix Dimensions to ANN: 7x5686.
- Test Video Frame Size: 240x320.
- Test Video Duration: 30 Seconds.
- Test Video Frame-Rate: 30 FPS.

8.1.3 Time Duration

- ANN Input Vector Creation: 44.902s
- ANN Training: 1.841s
- Magnification (Color and Motion): 106.998s.
- Training Input Matrix Dimensions to ANN: 5686x450.
- Expression Extraction (30s x 30fps): 2293.518s (38.2253 minutes)
- Pulse Extraction (30s x 30fps): 64.606s

Major time consumption is found while the expressions are extracted. Since, the 'getframe()' and the video frame stacking function has to work on every frame, the time execution of it increases as the frame number increases. A solution to decrease the time consumption in the extraction process is, to use videos of very short lengths of approximately 10s to 15s.

The time durations shown have been calculated using an un-optimized version of the code. When an overall optimization is done on the code, there could be a considerable decrease in the execution time in total.

8.2 Validating the Neural Network results with the results of the database

Artificial Neural Network (ANN) is used on motion magnified videos to analyze and detect the emotions related to micro expressions. In this process, an ANN is trained using a database called Extended Cohn-Kanade image database. This database has classified the emotions using Support Vector Machines (SVM) and thus, provided the confusion matrix and percentage of classification accuracy. The classification accuracy percentage of ANN is validated with the classification accuracy percentage of databases, because the ANN is trained using the same database. It is appropriate to validate the trained ANN results with the results of the database. Training a ANN with an image database and validating the obtained accuracies with some other image database (with some technique for classification) falsifies the intension. Thus, the validation is always done with the same database which is used for the ANN training. After validation of ANN with the same database, another database can be used for ANN verification/testing. The classification accuracies are shown below in the table 8.1:

In this study, ANN is also trained with Neutral expression, whereas in the database,

Emotion/state	% of classification accuracy for the database	% of classification accuracy for the trained ANN
Anger	75.0	93.1
Disgust	94.7	90.7
Fear	65.2	92.7
Happy	100	97.1
Neutral	—	89.9
Sadness	96.0	81.9
Surprise	84.4	95.4

Table 8.1: Classification accuracy of the SVM method in the database compared to ANN

neutral expression is neither classified nor mentioned. So, the accuracy results from Neutral expression of the database are unavailable. The average classification accuracy percentage of ANN is 92.0%, which means that this system classifies the emotions accurately. Therefore, the accuracy results of this system is comparatively superior to the accuracy results obtained from the database. Therefore, the accuracy results of this system is comparatively higher than the accuracy results given in the database documentation.

8.3 Reasons for not designing and performing an experiment

The reasons for not designing and performing an experiment for the given study are:

1. Micro expressions occur at high stake situations, creating/designing these high stake situations in the laboratory is a difficult task for the given limited time frame. Even if the experiment is designed and performed, the experiment would only have trivial lies with no pressure of punishment, which can lead to degradation of the results [11].

2. When an investigator has no past, present or future relationship or knowledge of the subject under test, then the micro expressions or emotions elicited by the subject cannot be precise [11]. If an investigator creates a relationship with a subject, that makes an investigator to judge the subject to an extent, which in turn tampers the results.

3. When Paul Ekman performed experiments in his laboratory with some subjects, only quarter of them elicited micro expressions. Paul Ekman also found that accuracy in detecting micro expressions is correlated with judging deception [11].

4. Generally, micro expressions are symmetrical and when an experiment is performed deliberately to elicit micro expression, the results obtained are generally asymmetric in nature [11].

So, keeping these factors in mind, an experiment could not be designed for this study. For verification and validation an alternate method is used.

8.4 Verifying the motion magnification of the proposed design

The validation of ANN is done using the Extended Cohn-Kanade image database. Now this study has to be verified/tested for its working.Verification cannot be performed with the same Extended Cohn-Kanade database due to two reasons:

1. When trained and tested with the same database, the accuracy obtained is always 100%. This is because the sequences used for both training and testing are same.

2. The problem with Extended Cohn-Kanade database is that it only provides a sequence of images but not any videos.

So, another image database called STOIC database is used for testing/verification [20]. STOIC database also provides videos directly for verification. Videos of STOIC database have been classified with emotions which are thus useful in verifying the motion magnification of the proposed system. The results of the system working with the STOIC database are shown below in the Table 8.2:

Emotion	**Video Files which have been successfully tested by the method.**
Anger	—
Disgust	DF1di, DF2di, DF3di, DF5di, DM1di, DM2di, DM3di, DM4di, DM5di.
Fear	DF5fe, DM1fe, DM2fe
Happy	DF1ha, DF3ha, DM1ha, DM2ha, DM5ha
Sadness	DF1sa, DM2sa
Surprise	DF1su, DM2su, DM3su, DM4su

Table 8.2: Results for STOIC database [20].

In the video file naming, the last two letters represented in lower case, shows the emotion that the video has in it. For example, consider DF1di, F represents the female actor, 1 represents the actor file number and di represents the disgust emotion. The results for a single sample of each emotion with its input video frame is discussed below:

8.4.1 Anger

The videos of STOIC database representing anger emotion have irregular facial alignment, which causes misclassifications in VJ algorithm and ANN. All these videos are of less than one second duration, which means that the video contains only a few frames of anger emotion, most of the times all this few frames are misclassified as some other emotion due to irregular facial alignment. So, no anger emotion of STOIC database has been classified perfectly.

8.4.2 Disgust - DF1di

The video frame input 'DF1di' and the results after ANN classification are shown below in the figure 8.1 and figure 8.2 respectively.

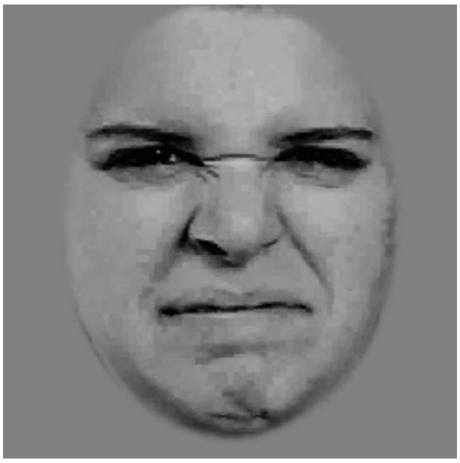


Figure 8.1: Video frame of DF1di [20].

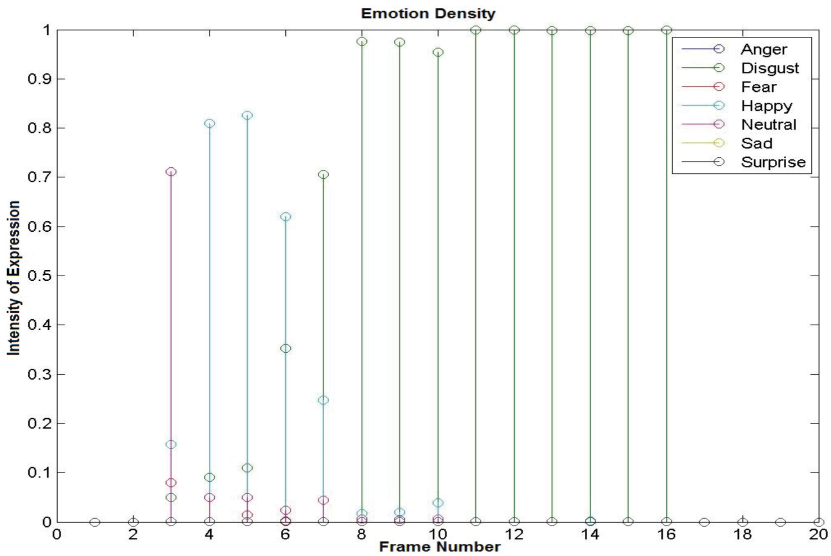


Figure 8.2: The emotion density of DF1di.

The emotion density graph shows that, except for few frames, such as 3rd frame (classified as neutral) and frames from 4th to 6th (misclassified as happy), each other frame has been classified as disgust. Therefore, ANN has perfectly classified disgust emotion for the given input frames of DF1di.

8.4.3 Fear - DM2fe

The video frame input 'DM2fe' and the results after ANN classification are shown below in the figure 8.3 and figure 8.4 respectively.

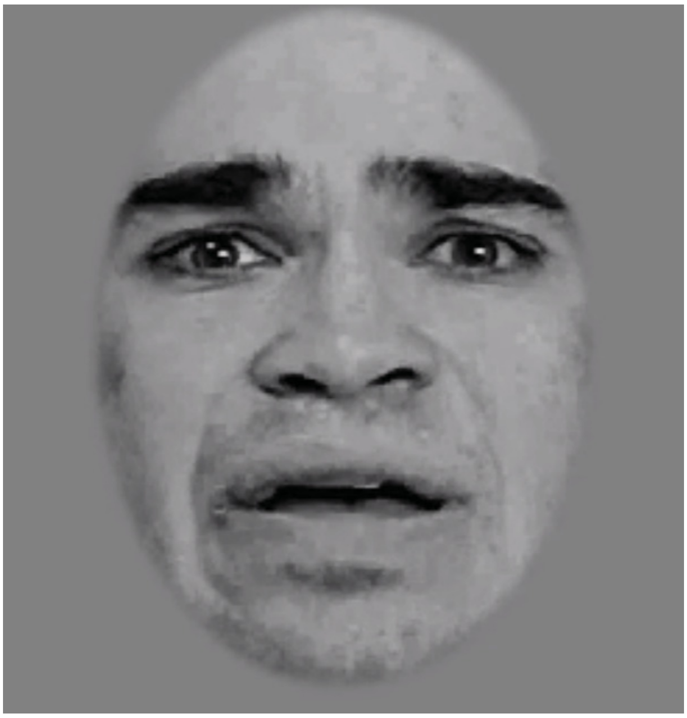


Figure 8.3: Video frame of DM2fe
[20].

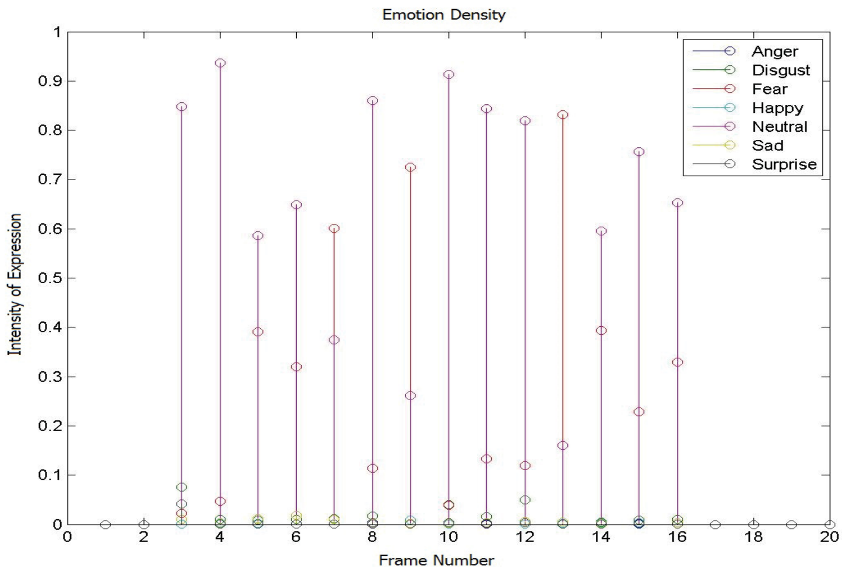


Figure 8.4: The emotion density of DM2fe.

The video is less than one second having only 10 frames in it. Out of the 10 frames, frame number 7, frame number 9 and frame number 13 is the showing the results for fear emotion. The remaining frames are classified as neutral for the video DM2fe. Thus, ANN classified the fear emotion for the given input frame DM2fe.

8.4.4 Happy - DF1ha

The video frame input 'DF1ha' and the results after ANN classification are shown below in the figure 8.5 and figure 8.6 respectively.

Figure 8.5: Video frame of DF1ha
[20].

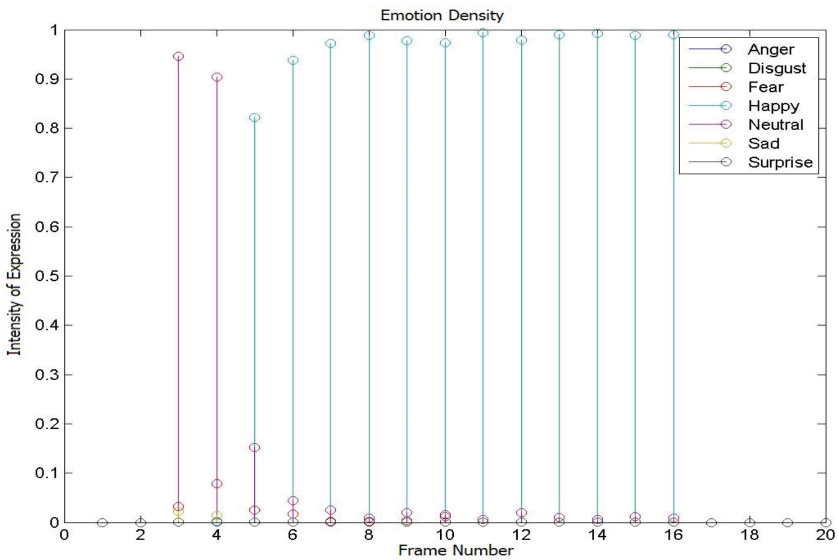


Figure 8.6: The emotion density of DF1ha.

The emotion density graph shows that, except for the first few frames (classified as neutral), every other frame has been classified as happy. Therefore, ANN has perfectly classified happy emotion for the given input frames of 'DF1ha'.

8.4.5 Sad - DF1sa

The video frame input 'DF1sa' and the results after ANN classification are shown below in the figure 8.7 and figure 8.8 respectively.

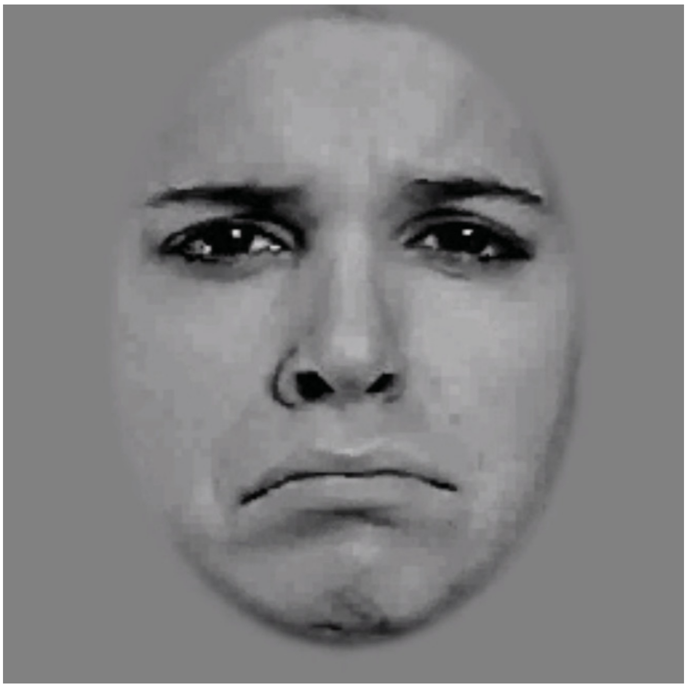


Figure 8.7: Video frame of DF1sa
[20].

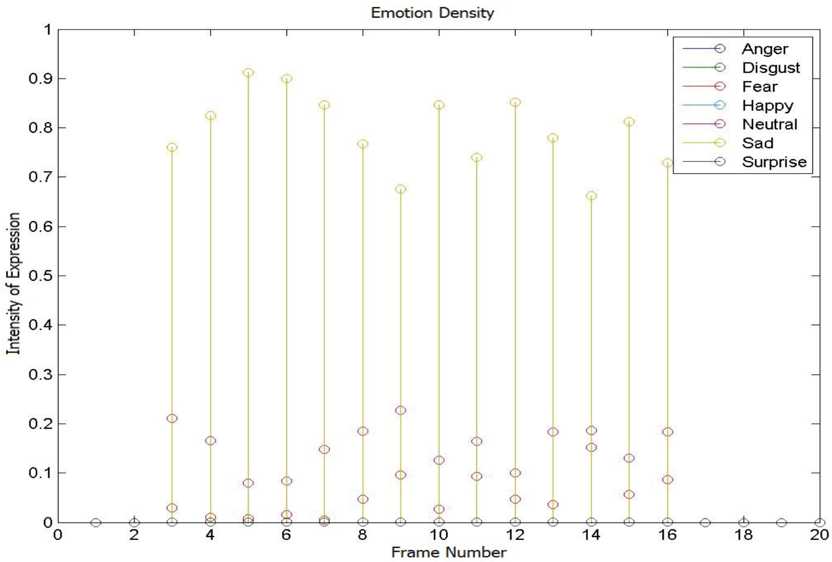


Figure 8.8: The emotion density of DF1sa.

The emotion density graph shows that every frame has been perfectly classified as sad. Therefore, ANN has classified sad emotion perfectly for the given input frames of DF1sa.

8.4.6 Surprise - DF1su

The video frame input 'DF1su' and the results after ANN classification are shown below in the figure 8.9 and figure 8.10 respectively.

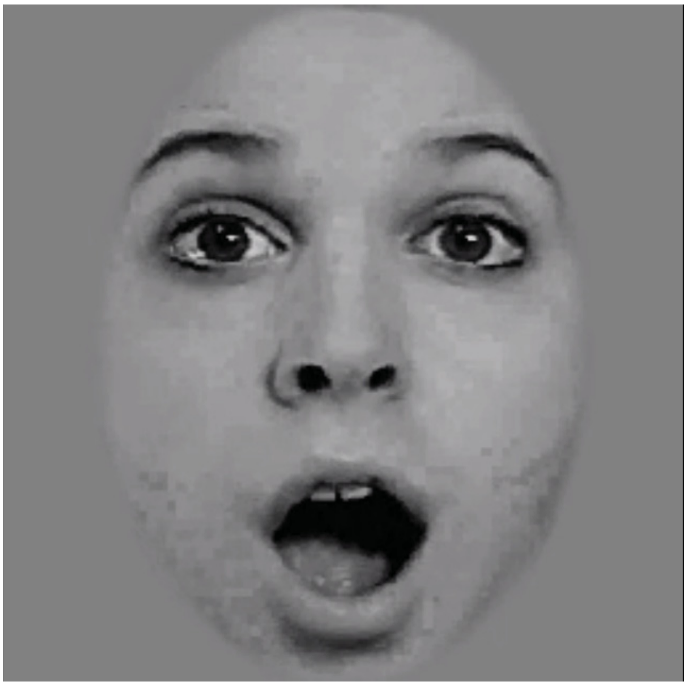


Figure 8.9: Video frame of DF1su
[20].

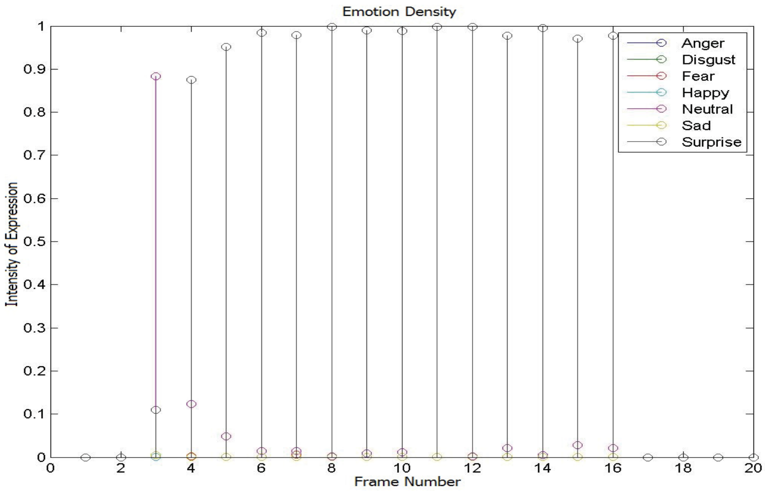


Figure 8.10: The emotion density of DF1su.

The emotion density graph shows that, except for the first frame (classified as neutral), each and every frame has been classified as surprise. Therefore, ANN has classified surprise emotion for the given input frames of 'DF1su'.

Thus, the ANN is trained with Extended Cohn-Kanade database and verified/tested with STOIC database. This study thus shows that it works perfectly and classifies all the emotions except anger. Therefore, this study can be used as a tool for observing, analyzing, learning, and understanding these micro expressions. This analysis can only extract micro expressions and its emotions i.e., videos of STOIC database can only be motion magnified because all the videos are gray scaled. Color magnification is done by observing the color changes of the face. Thus, by using STOIC database videos color magnification cannot be verified/tested.

8.5 Verifying both motion and color magnification of the proposed design

For verifying/testing both motion and color magnification, some videos of subjects eliciting micro expressions are rendered. Two subjects are used for this criteria. These two subjects tried to bring out the micro expressions symmetrically.

8.5.1 Results for angry emotion using motion magnification of subject-1

For subject 1, only anger emotion and its micro expression are rendered. The system study with STOIC database failed to classify anger emotion so, only anger micro expression is evoked so as to test whether it can classify anger emotion or not. Sufficient instructions are given to the subject for eliciting the micro expression according FACS scoring.

Figure 8.11: Motion magnified video frame of subject-1 eliciting anger emotion.

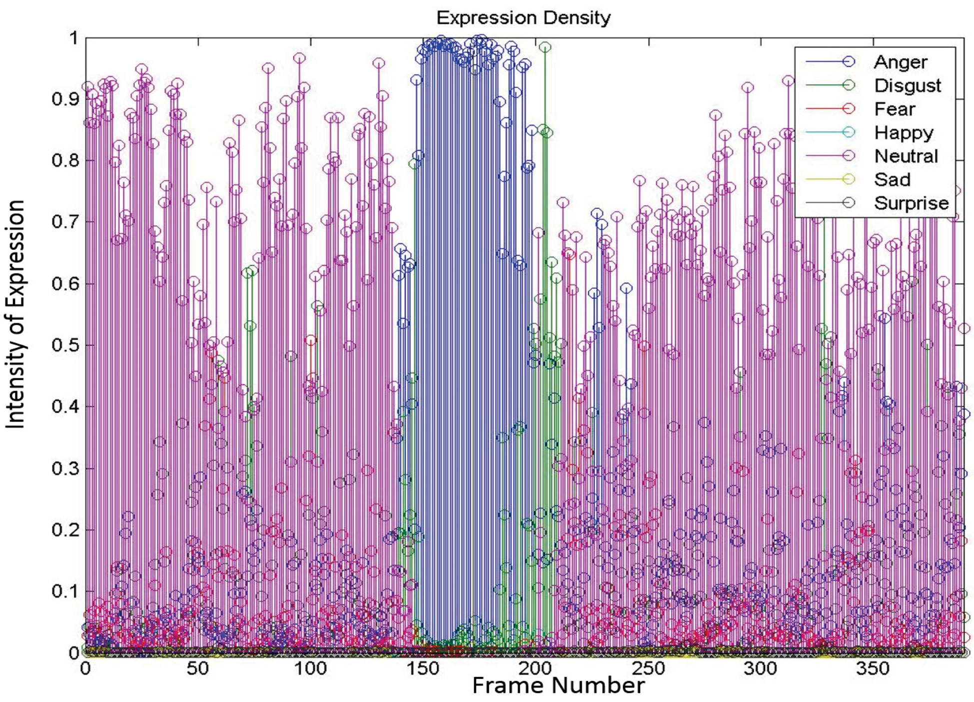


Figure 8.12: Emotion density for the micro expression elicited by subject-1.

Therefore, the anger emotion is working perfectly when subject-1 elicited the corresponding micro expression. The emotion density in the graph shows a predominant emotion of anger and a trace of disgust. The emotion intended is anger and the output of ANN also shows predominant anger. With this verification of anger emotion, it concludes that the study works for all the emotions perfectly.

8.5.2 Results for angry emotion using color magnification of subject-1

Figure 8.13: Color magnified frame of subject-1 eliciting anger emotion.

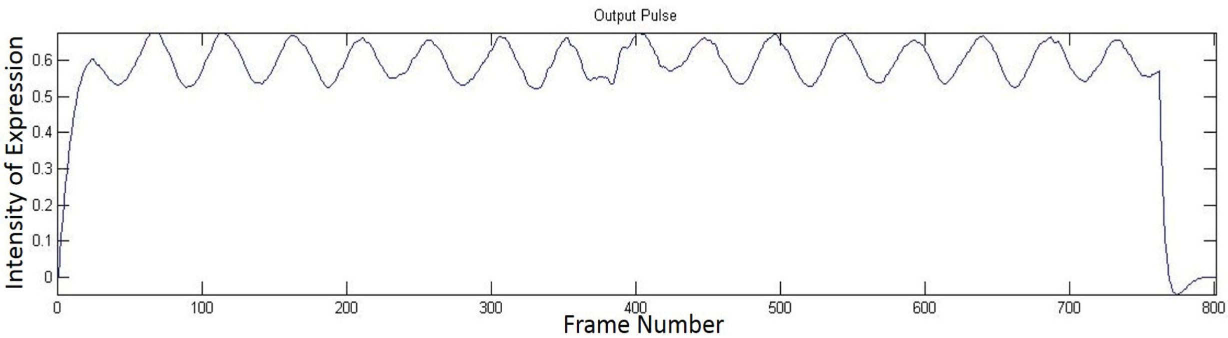


Figure 8.14: Pulse graph of subject-1 eliciting anger.

The video rendered by subject 1 to elicit anger emotion is of duration 12 seconds. The pulse rate observed for 12 second video has 16 peaks, which is approximately correct. This means the color magnification is also working well for subject 1.

8.5.3 Results for various emotions using motion magnification of subject-2

The results of non-magnified emotion extraction wasnt done on subject 1 since the recording of the video had to be done with utmost care and precision. This wasnt possible due to many constraints.

For subject 2, all the micro expressions and emotions are rendered in a single video. Sufficient instructions and good practice is given to this subject for eliciting all the emotions using FACS scores. Thus, the video is rendered with all the required specifications. The time duration of the video is 30 seconds. The only problem with the subject 2 is that the subject is unable to elicit the 'Sad' emotion.

Figure 8.15: Motion magnified frame of subject-2 eliciting various emotions.

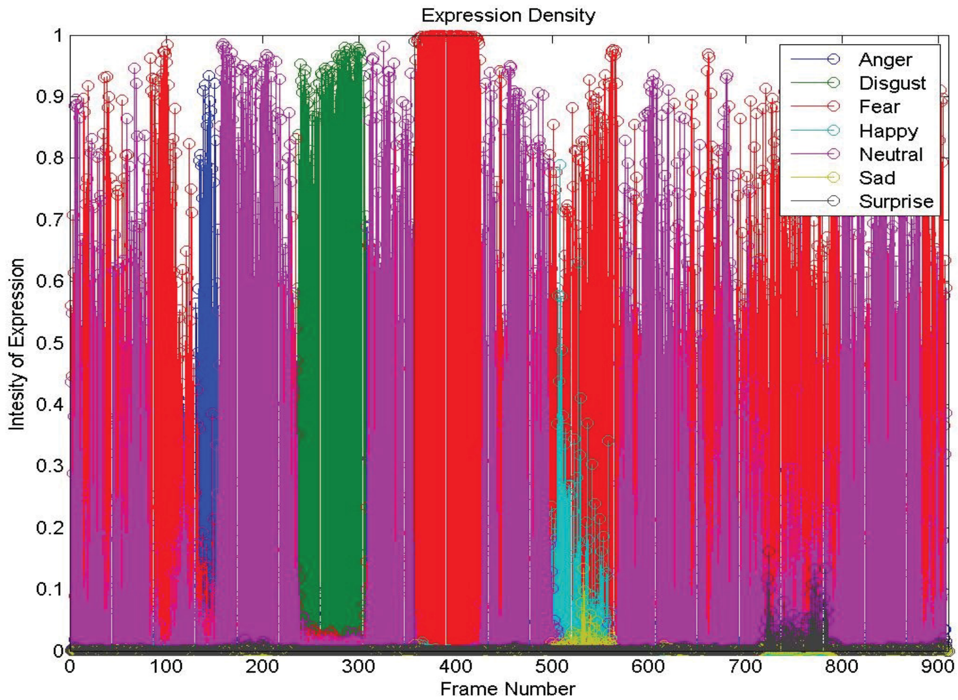


Figure 8.16: Emotion density graph of various emotions elicited by subject-2.

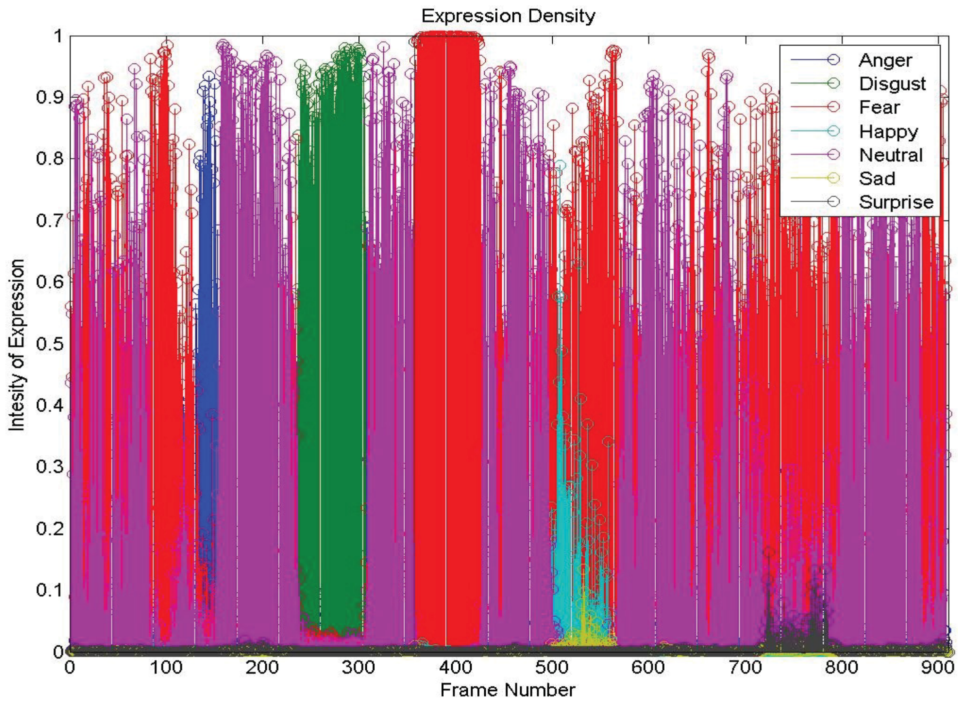
The emotion density graph shows only few traces of anger and surprise because the subject has been able to sustain these emotions for very less duration of time. The subject is able to elicit disgust, fear and happy perfectly, which can be clearly seen in the emotion density graph and finally, the subject is unable to elicit sad emotion which is not at all observed in the graph.

The problem generally arises because the subject is not a method actor. Also, in real life all the emotions are never observed or elicited at the same time. So, this subject is successful to an extent in eliciting all the emotions (except the sad) at the same time. Thus, the study classifies all the emotions when given at once.

8.5.4 Results for various emotions without motion magnification of subject 2

Motion magnification is a vital aspect in detecting micro-expressions. Since the micro-expressions are very imperceptible, these have to be magnified in order to make them perceivable. Magnifying them gives a better chance of identifying them.

To verify the effect of magnification and for identifying the expressions without magnification the video of subject 2 is used. The results shown below clearly state the inability of the neural network to identify the expressions without motion magnification.



The graphs show the reduced classification of the neural network for the non-magnified video. There are also many misclassifications of the emotions in this case. The cases of disgust and fear are exceptional since they are too distinct compared to other expressions in the data set. This experiment concluded the importance of the magnification required to detect the micro-expressions.

8.5.5 Results for various emotions using color magnification of subject-2

The average pulse rate for a normal human being is 72 beats/min. The pulse rate observed for 30 second video is 32 peaks, which is approximately correct. This shows the color magnification working perfectly. The disturbance observed in the pulse graph shown above is due to a sudden jerk while the experiment was being performed.

Thus, the system design works well for different videos. The results obtained from testing STOIC database and also from two other subjects confirm the perfect working of the micro-expressions extraction.

Therefore, the study can be used as a tool for understanding micro expressions and it also helps investigator in the process of detecting lies. With this system, an

Figure 8.17: Color magnified video frame of subject-2 eliciting various emotions.

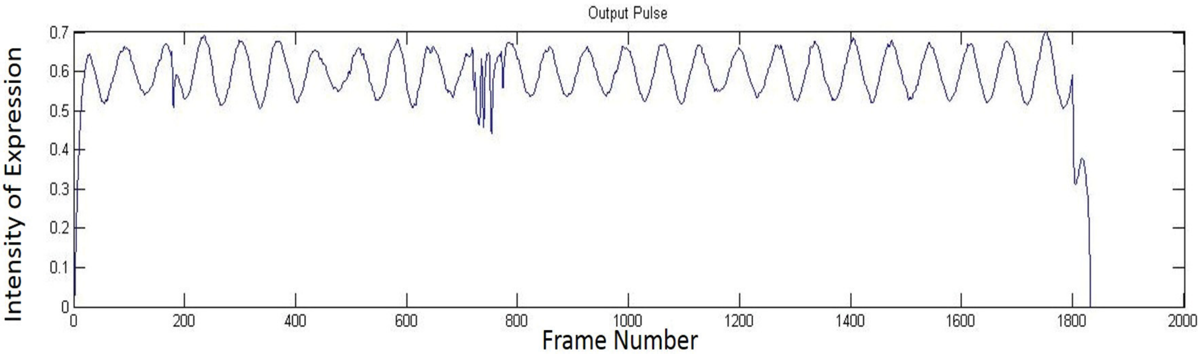


Figure 8.18: Pulse graph of subject-2 eliciting various emotions.

investigator can understand the insights of the subject's mind. When the investigator further prosecutes the subject under test with these obtained lists of emotions from micro expressions and pulse rates, hes more likely to detect the lies. Thus, this study helps only in the process of lie detection rather than itself being a lie detector.

Chapter 9

Conclusion and Future Works

9.1 Conclusion

In a report it has been said that "while detecting deceit or lie, people can reach an accuracy near to 54%. Out of this 54% accuracy, half of them would detect lies just by chance" [1]. This study is not an explicit lie detector, but it extracts micro expressions, which in turn helps both people and investigators in the process of detecting lies. Micro expressions reveal the true intentions of the subject/people. Observing, analyzing and understanding these micro expressions can thus intensify the process of detecting a lie, but comprehending such impulsive expressions is tedious and tough task for an average person. In general, a physical contact with a subject under test can induce a sense of consciousness in that person. Using Eulerian video magnification and Neural Networks, emotion through motion magnification and pulse through color magnification is extracted without any physical contact with the subject under test. The main advantage of no physical contact would result in an unconscious emotion leakage which is captured and deciphered.

Micro expressions are extracted using motion magnification. This motion magnified videos can ideally act as a tool for people to understand these micro expressions which thereby is useful in day-to-day interpretations of people and situations. When both motion and color magnified videos are used, with this extraction tool, an investigator finds it easier to comprehend the inner emotion of the subject more accurately. With further interrogation, the lies can be detected. Therefore, when only motion magnification is considered, it acts as a tool for better interpretations of people, situations, surroundings and circumstances. This makes life simpler and easier. When both motion and color magnification are considered, it helps an investigator of lies in the process of finding the truth from the subject under test.

9.2 Future works

This research work can further be extended in many ways.

1. Contempt can be added into the list of emotions.
2. This study is an off-line design. This requires a pre-recorded video to work. This can be coded into a DSP for on-board processing and to give a result in real-time.
3. This study works only for videos taken in one-to-one interview set-up and has few constraints, like a stable platform with no external jerks or disturbances involved. This study can be made available for all the conditions of videos.
4. This study mainly helps in the process of lie detection. This study can be made handy, when voice analysis is done with video analysis. When the facial cues and vocal cues are combined the accuracy of finding a lie raised to the highest bar [6].

After the addition of voice pitch recognition and analysis, the present study can be called as a lie detector.

Bibliography

[1] Charles F. Bond. Commentary a few can catch a liar, sometimes: Comments on ekman and o'sullivan (1991), as well as ekman, o'sullivan, and frank (1999). *Applied Cognitive Psychology*, 22(9):1298–1300, 2008.

[2] Daniel Cordaro and Paul Ekman. What is meant by calling emotions basic. *Emotion Review*, 3(4):364–370, 2011.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893 vol. 1, June 2005.

[4] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern classification*. Wiley, 2001.

[5] P. Ekman. Facial expression and emotion. *The American Psychologist*, 48(4):384–392, 1993.

[6] Paul Ekman. Why lies fail and what behaviors betray a lie. In JohnC. Yuille, editor, *Credibility Assessment*, volume 47 of *Nato Science*, pages 71–81. Springer Netherlands, 1989.

[7] Paul Ekman. Are there basic emotions? *Psychological review*, 99(3):550–553, 1992.

[8] Paul Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3-4):169–200, 1992.

[9] Paul Ekman. Deception, lying, and demeanor. *States of Mind: American and Post-Soviet Perspectives on Contemporary Issues in Psychology*, page 93, 1997.

[10] Paul Ekman. Should we call it expression or communication? *Innovation*, 10(4):333, 1997.

[11] PAUL EKMAN. Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences*, 1000(1):205–221, 2003.

[12] Paul Ekman. Darwin's contributions to our understanding of emotional expressions. *Philosophical Transactions: Biological Sciences*, 364(1535):3449–3451, 2009.

[13] Paul Ekman and Wallace V. Friesen. A tool for the analysis of motion picture film or video tape. *American Psychologist*, 24(3):240 – 243, 1969.

[14] Paul Ekman and Wallace V Friesen. Nonverbal behavior. *In Communication And Human Interaction*, pages 37–46, 1977.

[15] Paul Ekman, Wallace V Friesen, and Joseph C Hager. Facs investigators guide. *A human face*, 2002.

[16] T. Kanade, J.F. Cohn, and YingLi Tian. Comprehensive database for facial expression analysis. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 46–53, 2000.

[17] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 94–101, June 2010.

[18] David Matsumoto and Paul Ekman. The relationship among expressions, labels, and descriptions of contempt. *Journal of personality and social psychology*, 87(4):529–540, 2004.

[19] Ral Rojas. *Neural networks: a systematic introduction*. Springer-Vlg, 1996.

[20] Sylvain Roy, Cynthia Roy, Catherine Éthier-Majcher, Isabelle Fortin, Pascal Belin, and Frédéric Gosselin. Stoic: A database of dynamic and static faces expressing highly recognizable emotions.

[21] M. Shahid, A Rossholm, and B. Lovstrom. A reduced complexity no-reference artificial neural network based video quality predictor. In *Image and Signal Processing (CISP), 2011 4th International Congress on*, volume 1, pages 517–521, Oct 2011.

[22] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511–I–518 vol.1, 2001.

[23] Paul Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[24] Paul Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[25] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph.*, 31(4):65:1–65:8, July 2012.