UNIVERSITY OF THE WITWATERSRAND

WITS
UNIVERSITY

A study of the temporal relationship between eye actions and facial expressions

by

Moses Rupenga

(1287805)

Supervisor: Dr Hima B. Vadapalli

A dissertation submitted in fulfillment of the requirements for the degree of Master of Science

in the

School of Computer Science and Applied Mathematics
Faculty of Science

August 15, 2017

Declaration of Authorship

I, Moses Rupenga, declare that this dissertation titled, 'A study of the temporal relationship between eye actions and facial expressions' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this dissertation is entirely my own work.
- I have acknowledged all main sources of help.
- Where the dissertation is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ____

Date: ____

"He who asks a question remains a fool for five minutes. He who does not ask, remains a fool forever."

Chinese proverb

UNIVERSITY OF THE WITWATERSRAND

Abstract

Faculty of Science
School of Computer Science and Applied Mathematics

Master of Science

by Moses Rupenga

Facial expression recognition is one of the most common means of communication used for complementing spoken word. However, people have grown to master ways of exhibiting deceptive expressions. Hence, it is imperative to understand differences in expressions mostly for security purposes among others. Traditional methods employ machine learning techniques in differentiating real and fake expressions. However, this approach does not always work as human subjects can easily mimic real expressions with a bit of practice. This study presents an approach that evaluates the time related distance that exists between eye actions and an exhibited expression. The approach gives insights on some of the most fundamental characteristics of expressions. The study focuses on finding and understanding the temporal relationship that exists between eye blinks and smiles. It further looks at the relationship that exits between eye closure and pain expressions. The study incorporates active appearance models (AAM) for feature extraction and support vector machines (SVM) for classification. It tests extreme learning machines (ELM) in both smile and pain studies, which in turn, attains excellent results than predominant algorithms like the SVM. The study shows that eye blinks are highly correlated with the beginning of a smile in posed smiles while eye blinks are highly correlated with the end of a smile in spontaneous smiles. A high correlation is observed between eye closure and pain in spontaneous pain expressions. Furthermore, this study brings about ideas that lead to potential applications such as lie detection systems, robust health care monitoring systems and enhanced animation design systems among others.

Acknowledgements

I would like to take this opportunity to thank my supervisor Dr Hima B. Vadapalli for her unwavering support throughout this study. I also take a moment to appreciate the knowledge she imparted to me and her consistent availability whenever I needed help.
I also want to thank my father Mr F.P Rupenga and the rest of my family for their continued support in all my academic life, I really appreciate it.
Special thanks is also given to Wits university for granting me the Postgraduate Merit Award, it went a long way and I am truly grateful.
I also recognise that this research would not have been possible without the financial assistance of the National Research Foundation (NRF). All opinions expressed and conclusions arrived at are our own and should not necessarily be attributed to NRF.

Publications

This study led to the publication of the following articles listed below and these articles are provided at the end of the document.

1. Title: Investigating the temporal association between eye actions and smiles
   Authors : Moses Rupenga and Hima B. Vadapalli
   Conference: 2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)[1].

2. Title: Automatic spontaneous pain recognition using supervised classification learning algorithms
   Authors : Moses Rupenga and Hima B. Vadapalli
   Conference: 2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)[2].

3. Title: Investigating the temporal association between eye actions and smiles
   Authors : Moses Rupenga and Hima B. Vadapalli
   Journal article: *Extended paper, manuscript in preparation, 2017.*

Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	**Meaning**
AAM	**A**ctive **A**ppearance **M**odels
ANN	**A**rtificial **N**eural **N**etworks
ASM	**A**ctive **S**hape **M**odels
AU	**A**ction **U**nits
C-APP	**C**anonical-Normalized **APP**earance
CCRF	**C**onditional **R**andom **F**ields
CK	**C**ohn **K**anade Facial Expression Database
CK+	**E**xtended **C**ohn **K**anade Facial Expression Database
CSM	**C**osine **S**imilarity **M**easure
DISFA	**D**enver **I**ntensity of **S**pontaneous **F**acial **A**ction Database
EDA	**E**lectro **D**ermal **A**ctivity
EEG	**E**lectro**E**ncephalography
ELM	**E**xtreme **L**earning **M**achines
EOG	**E**lectro**O**culography
FACS	**F**acial **A**ction **C**oding **S**ystem
FER	**F**acial **E**xpression **R**ecognition
FFNN	**F**eed **F**oward **N**eural **N**etworks
HOG	**H**istogram **O**f **G**radient
JAFFE	**J**apanese **F**emale **F**acial **Expression** Database
KNN	**K**-**N**earest **N**eighbour
LDA	**L**inear **D**iscriminant **A**nalysis
LSTM	**L**ong **S**hort-**T**erm **M**emory **R**ecurrent **N**eural **N**etworks
MCC	**M**aximum **C**orrelation **C**lassifier
MLP	**M**ulti**L**ayer **P**erceptron

MLR	**M**ulti-**L**inear **R**egression
NN	**N**eural **N**etworks
PCA	**P**rincipal **C**omponent **A**nalysis
PHOG	**P**yramid **H**istogram **O**f **G**radient
PLBP	**P**yramid **L**ocal **B**inary **P**attern
PSPI	**P**rkachin **S**olomon **P**ain **I**ntensity
QDC	**Q**uadratic **D**iscriminant **C**lassifier
S-APP	**S**imilarity-Normalized **APP**earance
S-PTS	**S**imilarity-Normalized **P**oin**TS**
SLFN	**S**ingle **L**ayer **F**eed **F**oward **N**eural **N**etworks
SVM	**S**upport **V**ector **M**achine
SVR	**S**upport **V**ector **R**egression
sw-SVM	**s**partially **w**eighted **S**upport **V**ector **M**achine

In loving memory of my mother

Rhoda Charuka

1979 – 2012

Chapter 1

Introduction

1.1 Introduction

A facial expression can be said to be an observable exhibition of the cognitive activity, personality and psychopathology of a person [16]. It also describes the emotional state and the intention of a person. Facial expressions transmit communicative cues, which are pivotal in creating good inter-personal relations. Facial expressions and gestures enhance verbal communication by emphasising what a person is trying to say. Facial expression recognition (FER) has been around for a while, but has been exclusive to human-to-human interaction. The concept was studied by researchers like Darwin[17] and Ekman[18] well before it was implemented on machines [19][20]. The concept of FER gave machines the ability to "see" and deduce facial expressions, and this falls under the field of Human computer interaction (HCI). This idea allows machines to view and interact with humans by simply picking facial cues. The primary objective of an FER system is to impersonate the human visual framework in the most similar way [21].

Facial expression recognition focuses on facial behaviour which includes muscle movement, wrinkles and temporary deformations, among others. On the other hand, facial recognition encompasses the process of identifying and recognising a face without processing the underlying facial emotion[22]. An FER system is expected to overcome all problems faced in general facial recognition systems. It is supposed to be completely automatic, person independent and perform robustly in any environmental condition. Newer facial expression recognition systems work better in identifying a person as compared to traditional face recognition systems. Some facial recognition systems cannot distinguish between real and counterfeited faces [23]. A facial expression can be used as a password to a computer application as a security enhancement measure. This can also be implemented for physically challenged people who might not be able to speak or may be having difficulties with typing. Facial expressions may forecast a person's future action(s) and people use them to influence others, that is, make themselves believable, likeable etc. This shows that there is need to deduce the authenticity of these expressions since they may affect people. Moreover, previous studies show that it is possible to deduce a person's emotional state from their facial behaviour at any given time. However, the major problem is the determination and understanding of subtle characteristics of an expression.

It is difficult to infer a person's internal emotion from their facial expression. People tend to use facial expressions to hide their true emotions. By emotion the study refers to the underlying psychological state. It is commonly believed that, one cannot be smiling while they are experiencing body pain, they have to exhibit a facial expression that shows pain. The same applies to happiness, one cannot show a sad face while they are happy. While a person may generally conceal their underlying emotion, the emotion involuntarily comes out through subtle cues, perhaps such as eye blinks etc. This comes back to the facial muscle movement and also certain brain signals which are triggered when one is exhibiting different expressions. The aim of this study is to map the relationship of eye action units to different expressions. This is whereby time related distances between eye actions and facial expressions are evaluated. These eye actions can be either total eye closure or blinks. This temporal relationship will help in deducing if an expression is posed (fake or animated) or non-posed (spontaneous or natural), as it is very difficult for a person to fake an entire expression, subtle clues are always present. The same relationship can also be used with animation designers in the creation of animations that exhibit natural and realistic expressions. Ultimately, once a temporal relationship is deduced it can be used as a framework that contributes towards the determination of an expression if it is deceitful or not especially in security systems.

This research aims to study the idea of using eye action units to deduce and understand expressions as either real or fake. Eye blink or eye closure is usually involuntary it happens while one is not aware. It usually happens as a way of lubricating the eye or as a mechanism to protect the eye from external objects. A fast action of opening and closing of eyelids is regarded as a blink. This is caused by a prompt shrinking and contraction of the orbicularis oculi [24]. From other findings, blinking is mainly influenced by the brain [25], bad feeling and efforts to cover up trickery [26], [27] [14]. With the use of eye blinks, one can note their time-related distance in relation to an expression being exhibited and be able to evaluate that same expression as either posed (fake) or non-posed (spontaneous).

1.1.1 Motivation and applicability

This study seeks to help alleviate human deception by noting the relationship between eye actions and an expression. This will ultimately improve security systems since facial cues will hint if the person is being deceitful or not. A system of this nature can be used in human resources evaluations and high stake situations such as police interrogation rooms and the courts of law. A typical example of a situation where there is a need to authenticate facial expressions is when a patient is admitted in a hospital. Some patients tend to fake to be in pain (malingering) so as to get undeserved attention or as a way of protesting against a set of medication they are receiving. Hence, there is great need for general practitioners to authenticate these gestures and prioritise patients accordingly. With this capability to authenticate expressions, it will be easy to minimise chances of deception. Another area of applicability is in the creation of computer generated characters, whereby facial expressions are created by virtue of placing eye actions and smiles on correlated times. The system can help in the production of much realistic and normal facial expressions that appear natural to viewers.

This system can be of great help to the visually impaired populace. It can help them in denoting the authenticity of facial expressions of the people whom they are communicating with, thereby allowing them to experience a feel of total communication. In general, this study presents an intelligent and real time framework for authenticating facial expressions. The framework will map the temporal relationship between eye actions and facial expressions such as smiles and pain. From related studies, it was noted that there might be a temporal relationship between eye blinks and smiles[14] and thus gives a starting point.

1.1.2 Hypothesis

It is hypothesized that it is possible to deduce the temporal relationship that exists between eye actions and an expression.

1.1.3 Main Research question

Can the temporal relationship between eye actions and an expression be deduced?

.

1.1.4 Conclusion

In this chapter the the hypothesis and the main research question to prove it was introduced. A detailed motivation has been presented. In the next chapter, a literature survey is conducted. Chapter 3 gives a detailed description of algorithms and databases that were considered in this study. In Chapter 4, a detailed methodology of the study is presented. Chapter 5 gives experiments, results as well as a detailed discussion regarding the findings. Then lastly, Chapter 6 concludes the thesis and gives a brief summary of the study.

Chapter 2

Literature survey

2.1 Introduction

In exceedingly interactive species, such as humans, facial expressions have greatly advanced to convey rich information [28]. Due to the way human brains are assembled, individuals can reenact emotions that they are not encountering, so effectively that they deceive other individuals. Indeed, deceptions are a part of everyday life [29], and there are considerable adaptive advantages, including social acceptance, to deliberately manipulating, suppressing, and dissembling emotional expressions [30]. It is a well-known practice to hide actual feelings by expressing a different facial expression than the actual one. It is also of major concern, security-wise to know whether the expressions shown are real or fake. Most people tend to suppress their actual emotions especially when they want to say something they do not mean or which is not true. The study of facial expressions and emotion recognition can be of great value in lie detection systems.

Most FER studies have focused on evaluating the existence and non-existence of expressions. For instance, to check if a smile exists or not in an image or video file. Facial expression recognition (either intrusive or not-intrusive) is a way of checking the emotional state of a person based on the different appearances on the face at any given time. To date, people have managed to get away by faking expressions while they conceal their actual emotions. Most facial recognition systems in literature would only recognize the existence of an expression but were unable to deduce whether the expression is real (non-posed) or fake (posed). In a 2014 research article, some researchers tried to distinguish real and fake pain expressions using the shape of mouth muscles [28]. This shows that studies are now shifting towards evaluating expressions as either posed or non-posed. Furthermore, there are quite a number of papers that have explored deceptive smiles as well. Another article that hinted on the difference between real and fake smiles was published by Trutoiu et al. [14]. In their studies they mainly brought up the correlation between blinks and smiles. In light of this, this study was motivated to shed more light in the area of establishing the differences in expressions exhibited by humans in different circumstances as either posed or non-posed. This study seeks to check the correlation between eye action units and displayed expressions. The study will focus mainly on smiles and pain expressions. These expressions carry a lot of information and it is important to distinguish real and fake ones. A typical example is when a person smiles before he or she lies or when a person starts to scream or show signs of pain before pain is actually induced to him or her. Therefore, it is hypothesized that it is worthwhile to pursue this study to provide a model for security systems based on FER. This study is going to focus on eye blinks and eye closure for effectively understanding the intrinsic characteristics of an expression. By doing so, the relationship of facial expressions relative to eye blinks and eye closure can be analyzed.

The following section takes a look at some of the past and existing approaches used for facial expression and emotion recognition. It also gives a brief background of existing facial expression recognition systems.

2.2 History of Facial Expression Recognition

The foundation studies that structure the basis of today's research on facial expressions can be followed back to the seventeenth century. In 1649 John Bulwer wrote a comprehensive article concerning different expressions and head muscle movements [31]. Some of the most vital findings on facial expression analysis that has an immediate relationship to the modern day advanced investigations of automated FER were done by Charles Darwin[31]. He also wrote an article that built up the general and fundamental principles of expressions and their meanings in both humans and animals [17]. He went on to create a systematic list of facial deformations that had different classes of expressions and their meanings.

An experimental solution to their study was carried out by Paul Ekman in the 1960s, he began to study outward facial expressions and made this inquiry, "Whether people from different cultures agreed on the meaning of different facial expressions?" [19]. Ekman, later on, carried out an experiment on the meaning of expressions to different parts of the world and he came up to a certainty that facial expressions were universal across different cultures and countries. However, the most initial move towards automatically recognizing facial expressions was done by Suwa et al.[32]. They came up with a model that recognized different expressions from video sequences by tracking the face. Their idea was further developed by other researchers later in the 1990s to aid in the building

of FER systems.
With recent advancements in robotics and automated software, the need for accurate and efficient FER systems may be high. People impart adequately and are receptive to one another's emotional states, hence, the need for computers to attain this capability. Today, human computer interaction is at an advanced stage because machines can now potentially read and understand human facial expressions. Artificial Intelligence as a field has managed to outperform humans in some contexts e.g Chess, Go etc and they can potentially outperform humans at FER.
The next section brings about different data collection methods as either intrusive or not-intrusive. A review on static and temporal classifiers is also presented. A literature survey on posed and non posed is also presented. Lastly, section 2.6 gives a motivation of the study.

2.3 Data Collection Methods

This subsection takes a look at data collection methods. Data collection is widely subdivided into intrusive and non-intrusive methods. A review of articles that fall under each technique is presented to give a detailed picture of how techniques have evolved over the years.

2.3.1 Data collection using intrusive methods

Intrusive data collection methods are methods of data collection that somehow invades personal space or that becomes too involved with the object or person being examined and at times they cause annoyance. These include Electroencephalography (EEG), Electrooculography (EOG) and Electro dermal activity (EDA) among others. Electroencephalography is the measurement of electrical signals generated in the human brain by connecting electrodes to the scalp[33].

EEG has greatly evolved over the years, for instance, clinical EEG uses a head gear with a lot of nodes and takes approximately 40 minutes to prepare and use. A conductive gel is applied to each node to enhance skin conductance. To date, preparation and collection of EEG signals has improved and it no longer takes considerable time. In 2011, Brown et al.[34] designed an experiment to monitor the emotion of a person using an EEG sensor. The study was to check the existence of emotions as either being positive, negative or neutral. In their study, they incorporated a wireless EEG transmitter that monitored levels of emotion in different situations. To conduct the study, 11 subjects watched slide shows of images collected from the IAPS database [35]. The wireless EEG sensor measured brain signals while another wireless sensor measured the skin conductance. Extracted features were transformed to avoid the challenges and complexities associated with high dimensionality. A statistical analysis method was then incorporated for feature selection. For classification, they used support vector machines (SVM), the quadratic discriminant classifier (QDC) and K- nearest neighbour (KNN). They obtained an 82% accuracy rate using film clips. They concluded that their results showed high potential applicability of EEG to real life systems. The head gear described in this experiment shows that the collection of EEG signals has improved from the primitive days of Dr. Hans Berger.

In 2012 Harischandra et al. [33] took up an experiment as a way of enabling people who were paralyzed and those who had severe movement disorders to help them communicate. They wanted them to use expressions to communicate amongst themselves and the outside community. In their work, they investigated the possibility of recognizing emotions using EEG signals. This was to be achieved using primarily an inference system. The main signal was decomposed into different frequency bands. The decomposed signal in the frequency range of 8 to 16 Hz was used for measuring different emotions which were happy, fear and disgust. EEG recordings for this study and setup were collected from the database for emotion analysis using physiological signals (DEAP) [36]. Harischandra et al. [33] designed a tool that helped in interpreting eye movements in real time. The system was tested on the distance between the camera and the subject, lighting, camera placement and time of the day. The best distance between the subject and the camera was found to be about 1 to 3 feet, while the optimum lighting distance was about 2 to 6 feet. The best camera placement was said to be proportionate to the vertical position of the user's face while the best time of the day was between 8 am and 6 pm. These findings were said to vary depending on the type of the camera used. The one used in this experiment was a 16 mega pixel Altech USB camera. Obtained results were satisfactory but there were a number of improvements that could be implemented to enhance its effectiveness and usefulness. One area that was mentioned for improvement was the eye tracking which will be able to detect all emotions and provide the user with entertainment while they perform other activities with their eyes. Improvements to EEG technology got so advanced that researchers wanted to try and control objects using their brains. This is whereby brain signals would be transmitted via EEG, interpreted and sent as instructions to devices.

In 2013, Al-Zubi [37] wanted to create a natural way of controlling computers besides using voice recognition or the traditional way of a mouse and hands. In their work, they suggested a replacement of the traditional way with a different concept that incorporated eye blinks and head movements. To capture the head movements, they used a gyroscope and for emotion and eye movement they used an EEG head gear. They used a low-cost EEG device for measuring both EEG and EOG (electrooculography) signals. This headset acquired raw EEG signals and sent them over Bluetooth. They used a gyroscope to collect data about head movements in real time. This data was then transmitted to a computer where an algorithm would control the mouse movement. Results of this system were generally based on the system usability and accuracy. The system accuracy was acceptable though there was a need for improvement. Its usability was satisfactory since most subjects managed to use it.

In 2014, Soleymani and his colleagues designed an experiment to continuously monitor emotion induced by watching video clips. In their study, they used the MAHNOB-HCI[38] database. Soleymani et al.[39] collected EEG signals and extracted facial landmark features as the subjects watched the videos. This was done so as to have both facial and brain signals collected simultaneously. Extracted features were fed into regression models which are long short-term memory recurrent neural networks (LSTM-RNN), support vector regression (SVR), multi-linear regression (MLR) and conditional random fields (CCRF). On average, these models achieved comparable emotion recognition accuracies. However, the CCRF produced a much smoother regression line than the other models. Based on the results obtained by the LSTM, Soleymani et al.[39] observed that EEG based recognition was as good as facial expression recognition, contrary to their initial expectation. From the studies carried out over the years it is certain that EEG technology has been used for different problems and providing remarkable solutions.

2.3.2 Data collection using non-intrusive FER

Non-intrusive methods are ways that do not involve human contact or human awareness. They retrieve information in a non-invasive manner. The following sections take a look at FER as a non-intrusive method of data collection. In this subsection, an examination of the past studies using non-intrusive based FER is presented. These experiments were performed with data collected using cameras. It comprises of the aims, algorithms and results obtained.

Jeni et al.[40] designed a FER system that is robust to pose and illumination. The study was conducted using near infrared cameras and 3D shape. In their study, they created their dataset using the infrared-visible light camera array. In the process, they created short videos of 5 subjects. The subjects were told to enact different head movements as recording progressed. For comparison purposes, they used the Karolinska directed emotional faces (KDEF) database and the Oulu-Casia near-infrared database. In 2013, Trutoiu et al.[14] sought to prove the existence of a time-based relationship that exists between smiles and blinks. In their study, they used videos that contained spontaneous smiles. These videos were taken from an existing dataset called the Cohn-Kanade facial expression[41] database. That very same year, Sadeghi et al.[42] presented an approach that eliminated geometric variability in emotion based expressions. They introduced this approach so that appearance features can be accurately and effectively used for FER. In their study, they used images collected from the extended Cohn-Kanade (CK+)[8] database.

Malawaski et al.[43] designed a system that uses a low-resolution 3D Kinect sensor to recognize facial expressions. To evaluate their algorithm, they recorded 2520 images of 10 subjects using a Kinect for Windows and a Kinect for Xbox 360 sensor. Generally, images were recorded in normal and poor lighting and at frontal and different poses. Patil and Bailke[44] designed a system that evaluates emotions using facial expressions. They manually captured images of facial expressions from their participants. The process was done using an Intel RealSense SR300 camera. Their dataset contained 90 images of neutral, happy and surprise facial expressions. These were recorded from 30 subjects at 30 frames per second.

2.4 Classification

Classification can be defined as the process of categorizing feature vectors into appropriate categories[45]. Each category is called a class. This section focuses on different classification methods used in FER. It contains temporal and static classification techniques used in either of the two data collection methods mentioned above. Algorithms involved are fully explored to have a full understanding of how they can be used as well as their associated pros and cons.

2.4.1 Static classifiers using intrusive methods

Static classifiers, for example, Multilayer Perceptrons (MLP), do not consider temporal information during classification as they classify a single feature vector. Static classifiers work well with fixed length input vectors while Temporal classifiers can handle changing input length vectors. Haselsteiner and Pfurtscheller[46] designed a model that used a standard MLP to classify EEG data. In their experiment, they did a comparison between static and temporal classification. To do static classification, they converted temporal data to static input data. They observed that while doing static classification, there were no correlations amongst patterns. Hence, each individual pattern could be evaluated on its own. Nevertheless, some of the major advantages observed during static classification experiment were that the creation of a neural network was quite easy and straightforward. The learning process of MLP was quite easy as well. However, the performance of static classifiers was not as good as that of temporal classifiers where finite impulse response filters were used.

Alzoubi et al.[47] tested different classifiers on how best they can classify EEG data. One of the aims of this study was to do a comparative analysis between static and a combination of static and adaptive classification. Adaptive classification is a process whereby a classification model built earlier on is constantly updated with new data to reflect any pattern changes. In their study, they used a wireless EEG head gear to measure brain signals. This was done on 3 subjects who were asked to self-induce a series of emotions based on recollections of real life events. Algorithms that were used in this experiment were Naive Bayes, K-nearest neighbour (KNN) and SVM. In the setup for static classification, the best classification rate was achieved by KNN followed by a linear SVM and lastly Naive Bayes. Ultimately, the performance of the classifiers in combination with an adaptive classifier was better than static classifiers. Hema et al.[48] did a study whereby they tried to classify different tasks using signals recorded from the sensorimotor cortex. To do the experiment, subjects were told to imagine their arms or hands moving in the forward, left or right direction. They would perform the "relax" task where they would not think of anything in particular. The main objective of their study was to compare the performance of dynamic neural networks to that of static feed forward neural network. Using band power features, static neural networks achieved a recognition rates of 89.4% while dynamic neural nets achieved 97.65% using the same features. In the case of Parseval features, static neural nets achieved a recognition rate of 92.2% while dynamic neural nets scored a recognition rate of 97.7%.

2.4.2 Static classifiers using non-intrusive methods

These type of classifiers basically work on static images. There are a lot of static classifiers that can be used in classifying expressions which include the tree augmented naive Bayes classifier, cosine similarity measure (CSM), maximum correlation classifier (MCC), MLP and the SVMs among others. SVMs are based on statistical theory and as a popular machine learning method, they can be effectively used for static classification[49]. Yang et al.[50] brought up the notion that a facial event naturally evolves in three phases that are the onset, then the apex and finally the offset. These phases generally affect the facial expression as well as the action units involved. Based on the results, they noticed that temporal features were more effective than static features in both training and testing of their system.

However, Gold et al.[51] designed an experiment to check the importance of temporal information in understanding one's emotional state. In their experiment, they measured the amount of information available to an observer as they recognize facial expressions on static and dynamic images. They went on to see how exactly that information would aid them when performing similar expressions. From their study, they noticed that temporal features had a little effect in helping an observer to understand another person's emotional state based on facial expressions. Furthermore, they noticed that static images that contained vivid facial expressions had more distinguishing details than an evolving dynamic video. Hsu et al.[52] presented a method of recognizing discriminative facial features from static images. They used the Extended Cohn-Kanade (CK+)[8] database to conduct their experiments. Their study aimed at distinguishing all the 7 basic expressions. In their study, they used distance based features. To improve and have more illustrative features, they combined holistic and local features. They passed these features to the K-means algorithm for vector quantization. The features were finally passed on to an SVM for classification and a hit rate of 87.7% was achieved. For faces that contained the happy expression they scored an average of 95.6%.

Wang et al.[53] designed an AAM and SVM based automatic FER system. In their study, they used the Japanese female facial expression (JAFFE)[54] database to train and test their model. They divided face regions into 3 categories that looked at the full face as a whole, the upper face (eyes) separated from the lower face (mouth) and lastly the segmented face consisting of the two separated eyes (left and right) as well as the lower region containing the mouth. AAMs were used to segment the facial regions while feature extraction was done using Gabor wavelet transform. Classification was then done using SVMs. Their system had a general recognition rate of 91.4%. For the happy expression, the system achieved an average accuracy rate of 66.66%.

Suk and Prabhakaran[55] designed a FER mobile application system to notice 6 basic expressions in real-time. In their work, they did a ten-fold cross validation test using samples collected from the CK+ [8]. For face detection and feature extraction, they used the Haar cascade and Active shape models (ASM), respectively. The ASM created landmark points that demarcated the faces. Landmark points for neutral and non-neutral expressions were classified using a linear SVM. The non-neutral expressions were further classified to pick if they belonged to any one of the six basic expressions. This process was done using an SVM with an RBF kernel. From the 10 fold cross validation, the system had an accuracy rate of 86% and for the happy expression, it had an accuracy rate of 98.6%.

Blinski et al.[56] designed a system to determine the gender of a person based on their smile. In their study, they used the UvA-NEMO smile dataset. Their system was to capture all the information from a video sequence. Input features were spatio-temporal ones based on the dense trajectories, represented and encoded by Fisher vectors. These features were then passed on to a linear SVM for classification. For adolescents, it gave a recognition rate of 86.3% while a recognition rate of 91% was obtained for adults above 20 years. Recently, Iglesias et al.[57] did an experiment to compare the performance of a static to a temporal approach in recognizing 7 basic facial expressions. In the temporal approach, a conditional random field (CRF) was used to recognize the dynamic properties of image sequences. The CRF was also used for classification whereby gestures were classified based on a voting system. In the static approach, they used a binary descriptor to extract features on each and every frame. These features were then classified using a linear SVM. The temporal approach reached an average recognition rate of 78% while the static approach reached an average of 86.86%. Ultimately, the static approach outperformed the temporal one.

Sanjula and Gowrishankar[58] decided to design a facial recognition system that identifies patients before and after a surgery. The idea was to use ELM to recognize altered face images due to plastic surgery procedures. In this process, extended uniform circular local binary pattern (EUCLBP) and Scale invariant feature transform (SIFT) were implemented to identify all relevant features. Based on results obtained, it could be noted that ELMs were very efficient at both face detection and recognition. The ELM performed extremely well with an accuracy rate of 95% against Linear discriminant Analysis (LDA), Principal component analysis (PCA), K-nearest neighbour (KNN), LTP and Local binary pattern (LBP) which scored 66.7%, 88%, 76.4%, 77% and 47.8%, respectively.

Ghimire and Lee[59] designed an experiment that used a collection of ELMs by implementing a bagging algorithm for facial expression recognition. This experiment was performed on two data sets which are the JAFFE[54] and the CK+[8] databases. The Histogram of gradient orientation (HOG) was used for feature extraction. These features were fragmented into many smaller groups and used to create unique training data files. These were passed to individual ELMs to create training models. This setup yielded a performance accuracy of 94.37% on the JAFFE and 97.39% on the CK+ [8] databases. Liu et al.[60] designed a facial recognition system based on ELMs. In their experiment, they used the JAFFE[54] and the CK+[8] database. Gabor filters were used for feature extraction while two dimension principal component analysis (2DPCA) was used for dimensionality reduction. In their study, they compared the performance of the ELM to that of the SVM and the KNN algorithm. The ELM outperformed both classifiers achieving a recognition rate of 94% and 95% on the JAFFE and CK+ databases, respectively.

Cui et al.[61] designed a system to recognize smiles in real life applications. In their study, they used the GENKI-4K [62] database to test their system. Pair-wise distance vectors were used as features. Their idea was to pick landmark points surrounding the mouth region. In light of that, they decided to create a pair-wise group of each smile image and its corresponding neutral face. This was done to help in calculating and observing landmark displacements as the person smiles. These calculated distance features were then fed into an ELM model which in turn achieved an average recognition rate of $93.42 \pm 1.46\%$. Recently, Afshar and Salah[63] proposed a system that sought to resolve the problem of recognizing emotion on real world videos. Their system used improved dense trajectory, local Gabor binary patterns- three orthogonal planes (LGBP-TOP) and geometric features for training and testing. Their study was done on the CK+[8] database and the third emotion recognition in the wild (EmotiW) 2015 challenge database. Procrustes analysis was implemented to register facial images from each frame of the video. For classification, they used the ELM instead of the SVM due to its faster training times and its capability to achieve much better results. The ELM scored an accuracy rate of 93.58% while the SVM achieved 80.73%. In this experiment, they noted that their implementation was able to recognize small facial changes that were tough to detect even for human annotators.

Static classifiers do not use any temporal information that is available in image sequences. They use information of a single image. They can easily denote the apex of an expression, which makes it easy for them to pick up expressions when people are normally expressing themselves or making an emphasis using facial expressions. Static classifiers are easy to implement and can effectively work on a smaller number of parameters per image. Generally, they are deemed to be not ideal for noting the progression of a facial expression since they work best on static images. However, static images are able to capture eye opening, eye closure and other facial expressions such as smiles and pain as they progress at different instances.

2.4.3 Temporal classifiers using intrusive methods

Most intrusive FER systems such as brain computer interface use temporal classifiers to denote trends of signals. Temporal classifiers, for example, Hidden Markov Model (HMM), can classify an ordered chain of feature vectors and ultimately catch temporal dynamics [64]. Jrad and Congedo[65] designed an experiment to distinguish brain activities. In their work, they wanted to test the performance of their proposed method, that is, a combination of temporal features and ensembled classifiers against SVM and spatially weighted SVM (sw-SVM). To do this, they selected highly distinctive spatio-temporal features. After getting spatio-temporal features, they went on to resample the EEG signal so that they can extract temporal features. A statistical test was then incorporated to evaluate the performance of their suggested method against the use of SVM and sw-SVM. They found that the use of temporal features in collaboration with ensemble of classifiers yielded much better performance.

Zhao et al.[66] did an experiment that sought to classify electrocorticorgraphic (ECoG) signals using Coupled Hidden Markov Models (CHMM). These signals were captured in such a way that best describe any dynamic linkages that might occur between brain signals during different hand movements could be noticed. In their study, they compared the performance of the CHMM against a standard HMM and an autoregressive model. They discovered that CHMMs were much better than the other two algorithms in denoting temporal dynamics of ECoG.

Based on the brief review between static and temporal classifiers, it can be noted that temporal classifiers outperformed static classifiers in classifying time related EEG signals. Temporal classifiers probably were successful in brain computer interfaces because they could easily recognize relevant time-related variations present in the extracted features[67].

2.4.4 Temporal classifiers using non-intrusive methods

There is much attention on the use of dynamic classifiers to work on video frames because they recognize the transition of the facial expression as it evolves. Ambadar et al.[68] found out that faint expressions which were not easily recognized in singular images, abruptly got distinctly obvious when seen in a video sequence. Yang et al.[50] designed an approach that used dynamic haar-like features to represent images. Their experiment focused on video based facial action units and expression recognition. Haar-like features are simple to use as compared to Gabor features. They are less computationally expensive since they are built on additive operators. In a video sequence, dynamic classifiers generally seek to recognize the temporal pattern that exists between features describing each and every frame of the sequence. Examples of dynamic classifiers include Hidden Markov models (HMMs) and dynamic Bayesian networks (DBN) and Recurrent neural networks (RNN) among others.

Vadapalli et al.[69] designed a system to recognize facial action units using temporal features. In their study, they extracted Gabor features from image sequences and passed them onto an Elman RNN for classification. Recurrent neural nets have an advantage of being able to memorise as it traverses over different states. This is mainly due to the internal feedback mechanism it possesses. Meng et al.[70] proposed the use of the SVM on two Kernels (Two view SVM_2K). This classification algorithm is good when dealing with many features. Its performance is better than individual SVMs especially in human action and object recognition tasks. Hedaoo et al.[71] designed a system that recognizes facial expressions on images. In their study, they tried to capture facial interactions that existed at feature tracking, extraction and expression recognition. To achieve this they used the DBN for feature tracking while the Adaboost did AU extraction. The extracted features were then finally passed to a DBN for classification. Temporal classifiers are more suitable for person-dependent tasks due to their higher degree of variability in expression in humans as well as the variation in the dynamics of each expression. However, temporal classifiers are quite difficult to train than static ones as they require a larger training set and more parameters in order to train them adequately.

2.5 Differentiation of posed and non-posed expressions

In this subsection, a review of papers that have done studies concerning posed and non-posed expressions is presented. Smile and pain expressions are the main expressions under consideration since they are the main expressions to be evaluated in the current study.

2.5.1 Smile expressions

Liu and Wu[72] evaluated different machine learning algorithms to distinguish between true (spontaneous) and fake (posed) smiles. Feature extraction was performed using Gabor filters and feature selection was performed using Adaboost algorithm. Three different machine learning algorithms were used for the classification task that included SVMs, Back propagation neural networks and Linear discriminant analysis (LDA). The highest recognition rate of 85% was obtained using SVMs. The obtained classification output was then further sent for evaluation check for the existence of AU6 (cheek raiser) or AU12 (lip corner puller). The best recognition rate for these individual action units was 71.8%. In their study, they suggested that true smiles were generally made up of a combination of AU6 and AU12.

Hoque et al.[73] classified smiles by exploring temporal patterns in video sequences. In their first set of experiments, participants were involved in imitating delighted and frustrated expressions. However, in the second set of experiments, participants were exposed to videos containing frustrating and delightful events to induce natural frustration as well as delight. An average accuracy rate of 82.3% was achieved in recognizing acted expressions. However, experiments performed on elicited data obtained a lower accuracy rate of 41.8%. It was further noted that 90% of the subjects tested upon smiled under natural frustration which is contrary to the absence of smiles in acted expressions of frustration.

The relationship between smiles and blinks was investigated in the work done by Trutoiu et al.[14]. The main aim was to use this relationship in animation designs for the generation of natural expressions. Sample video sequences collected from the Cohn-Kanade [74] database were used. These samples contained natural and spontaneous smile expressions (AU12) obtained from female subjects. Blinks were determined by the AAM feature points surrounding the eye region. Time-related distances between smiles (activation and deactivation) and blinks were then calculated. Their results show a strong correlation between eye blinks and the offset of a smile. From their findings they hinted that this temporal relationship could be handy in noting the difference between real and fake expressions.

One of the disadvantages of their study was that they manually recognized smiles and this process was time consuming. Gunadi et al.[75] designed an experiment to detect fake smiles. In their study, they looked at the state of the mouth and that of the eyes. They monitored the muscle (orbicularis oculi) surrounding the eye region as well as the muscle with greatest effect to mouth movement (zygomatic major) when a person is smiling. The Laplacian of Gaussian (LOG) was used for edge detection of the mouth while eyes were evaluated by calculating the eye elongation. Unfortunately the study does not specify the actual dataset that was used to conduct experiments. Nevertheless, a 4-fold cross validation test was conducted and in each fold 25 images were used for training and 75 images were used for testing. Classification was done using a linear SVM and a recognition rate of 86% was achieved in distinguishing both real and fake smiles. This result was an average of all the results obtained in 4 rounds of tests.

2.5.2 Pain expressions

Hill and Craig[76] designed an experiment to detect deceptive pain expressions. Their study was to distinguish genuine and disguised pain expressions by looking at temporal patterns of AUs and occurrences of deceptive cues among others. 40 patients (23 Male, 17 Female) who suffered from lower back pain were videotaped as they exhibited pain expressions. These expressions were categorized as either genuine, faked or masked. Their study proved that some AUs were directly related to pain expressions. From their findings, they noted that faked pain had longer peak intensities and generally longer duration than genuine ones. They further proved that there is a biological difference between genuine and faked expressions due to consistent organized pattern of motoric display observed in genuine pain that did not occur in faked pain. However, there were no major differences between masked pain and neutral expressions besides the more pronounced frequency of mouth opening and eyebrow movement noted in masked pain.

Larochette et al.[77] studied genuine, masked and faked pain expressions in children. In their experiment, a cold pressor was used to stimulate facial expressions on 25 boys and 25 girls whose age ranged from 8 to 12 years. Cold water at 10 °C was used for genuine and masked pain expressions, while warm water at 30 °C was used for fake ones. In this study, FACS AUs were also incorporated in analyzing facial expressions. It was noted that children exhibit more pain related facial actions and they are very good with hiding pain expressions than faking them. Ashraf et al.[78]'s work involves deducing pain in terminally ill people and young children where self reporting is not viable. Facial indicators were identified to automatically detect severe pain. In this study, FACS coded video sequences extracted from the UNBC-McMaster shoulder pain expression database[12] were used. Feature extraction was done using AAM and classification was done using SVM. Their studies suggested that non-rigid features contained more information as compared to rigid ones.

Littlewort et al.[79] studied the use of machine learning techniques in differentiating real and fake pain. They collected data from 26 subjects as they went through real pain, fake pain as well as baseline conditions. To induce pain on their subjects, they immersed the arms of their subjects in cold water at a temperature of 3 °C. For baseline and fake pain, they used water at 20 °C. One of their main goals was to check if automated testing was inline with human experts evaluation. Results based on human analysis were compared to computer generated results. In the second part of the experiment which included the training of an automated system, they used 3 posed data sets which are the Cohn-Kanade DFAT-504 dataset[74], a portion of directed facial actions collected by Ekman and Hager and lastly the MMI database[80]. Automated systems achieved a recognition rate of 88%, contrary to human observers who achieved 49% in differentiating real and fake pain. Lucey et al.[81] used the UNBC-McMaster shoulder pain expression database[12] to model an automatic pain assessment system. Different feature descriptors derived from AMM were used in this study. These include S-PTS: similarity normalized shape, S-APP: similarity normalized appearance and C-APP: canonical appearance features. These features were then passed on to a linear SVM for classification. A combination of all the three types of features performed well compared to using an individual type of feature. However, their system worked well with stationery patients who were showing pain intensity levels greater than 10.

Boerner et al.[82] designed an experiment to compare and see how well can different caregivers detect deceptive pain expressions in children. The different caregivers who were used for the study were pediatricians, pediatric nurses, and parents. They were made to watch 48 video sequences showing children going through genuine, suppressed, fake and no pain. The caregivers would tell how much pain the child is succumbing to. However, a rating was also done to check the amount of confidence they have as they declare the amount of pain they assume the child was going through. Generally, all caregivers did not perform well in recognizing genuine pain. However, they were more accurate in identifying faked and suppressed pain. Amongst the three types of caregivers, pediatric nurses showed good judgement than the rest.

Khan et al.[83] designed and implemented a computer vision system that detected pain by analyzing facial attributes. Sample video sequences for their study were collected from the UNBC-McMaster shoulder pain expression database[12]. Both geometric and appearance features were extracted using pyramid histogram of gradient (PHOG) and pyramid local binary pattern (PLBP) and were further classified using different machine learning techniques such as SVM. The highest recognition rate of 96% was achieved using the simple 2 nearest neighbour (2NN) algorithm. Bartlett et al.[28] studied the effectiveness of computer vision based systems to human observers. Trained human observers achieved an average recognition rate of 55% in distinguishing real and fake pain. On the other hand, computer based systems achieved an accuracy of 85% which indicates the effectiveness of automated systems compared to humans.

2.6 Motivation for the current study

Based on the two classification models presented (static and temporal), it can be noted that there were relative pros and cons of the two classifiers. Based on how the present study is going to be performed it can be noted that static classifiers are much more appropriate in describing the state of the face at each and every instant. This implies that, static classifier will evaluate and classify each and every frame till we get a full sequence that depicts the onset and offset of an expression. The term onset describes the start of an expression while the term offset describes the end of an expression. This study proposes the use of AAMs for feature extraction and the use of SVMs for classification. Extreme learning machines will be used in the classification of spontaneous smiles and pain expressions while SVMs will be used as the baseline model.

From previous studies, it is been noted that little work has been done on the issue of authenticating expressions as either posed or non-posed. Based on pain deception studies presented in past studies, it can be noted that a knowledge gap exits due to the fact that they only focused on comparing human judgement to computer judgement in recognizing pain expressions. However, this study attempts to find the temporal relationship between eye actions and pain expressions so as to have a better understanding of spontaneous pain expressions. Another knowledge gap was noted in a different paper by Trutoiu et al.[14]

where they sought to find the temporal relationship between blinks and smiles. They used these temporal relationships to model accurate expressions for implementation on computer generated characters. Since they managed to find a relationship between blinks and smiles, it may be possible to map eye action units to different expressions such as pain. This is whereby time distances between eye actions and facial expressions are noted and studied to see if there exists a relationship. Moreover, this study aims to automate the process of recognizing blinks as well as smiles using machine learning techniques. This study will try to compare temporal relationships of these eye action units to real (non-posed) and fake (posed) expressions. Based on previous studies, many researchers are in agreement that facial expressions convey rich information. Therefore, it can be concluded that if a facial expression is posed it may be conveying untrue information, hence the motive to understand differences in facial expressions.

2.7 Conclusion

This chapter gave a brief history of FER and two main data collection methods which are intrusive and not-intrusive techniques. A detailed discussion of various papers that used different classifiers and classification techniques was also presented. Previous studies on fake and real expressions were given, as well as the main motivation towards this study. Chapter 3 presents algorithms and databases that were considered for this study.

Chapter 3

Background

3.1 Introduction

This chapter presents different image processing and machine learning techniques that were implemented in this study. These were selected based on their good performance and reliability as reported in the literature survey. Active Appearance Models (AAM) and SVMs were selected as predominant algorithms in smile and pain detection, while ELMs are introduced as a relatively new algorithm. Section 3.2 describes AAM, which is a feature extraction technique. Section 3.3 follows and discusses SVMs which is a classification algorithm. Section 3.4 discusses ELM which is also a classification technique that was implemented in the study. Section 3.5 describes the FACS AU system, which is a method that was used in evaluating facial expressions in the study. Section 3.6 follows, and describes the PSPI scale which was used in calibrating pain expressions. Lastly, Section 3.7 presents different databases that were used in performing this study as well as those that were previously used by other researchers in FER studies.

3.2 Active Appearance Models (AAM)

Active appearance models come in two forms which are independent and combined active appearance models. Independent AAMs are standalone AAMs which can either be described according to shape or according to appearance, while combined AAMs uses both shape and appearance models simultaneously. Independent AAMs are explained in subsection 3.2.1 while combined AAMs are explained in subsection 3.2.2.

Active Appearance Models are nonlinear, generative and parametric models of a certain visual phenomenon [4]. They generally fit a model of an image created from a training data set. Active appearance models are an extension of Active Shape Models (ASM) which focus mainly on the shape or geometry of an object in an image. Cootes et al.[84] used ASM to search for elastic and stretchy objects in images. Lanitis et al.[85] extended the same concept by warping and normalizing the face image. A model based on absolute correct points of the face, was used to make sense of new images. This method was further improved by Edwards et al.[86] to incorporate not just models of shape but also grey level appearance. However, the new method was highly dependant on the ASM to pick up faces in new images. The AAM is now an extension of the previous work that has been done, whereby all shape and appearance information is combined to create a model that fits to a new image [87]. To build an AAM model according to Cootes et al.[88] a training set of images with correctly marked points is needed (Figure 3.1). Procrustes analysis is then applied to align all feature points and build a shape model. Each training image is then scaled and rotated so as to match all points to those of the mean normalized shape thereby creating a "shape-free patch". The shape is scanned into a texture vector(g), which is then normalized using a linear transformation:

$$g \longrightarrow \frac{(g - \mu_g 1)}{\sigma_g} \tag{3.1}$$

where 1 is a vector of ones and μ_g and σ_g are the mean and variance of elements of g, respectively. After normalization, $g^T 1 = 0$ and $\mid g \mid= 1$. Eigen analysis is applied to build a texture model. Ultimately, correlations between shape and texture are learned to produce a combined appearance model.

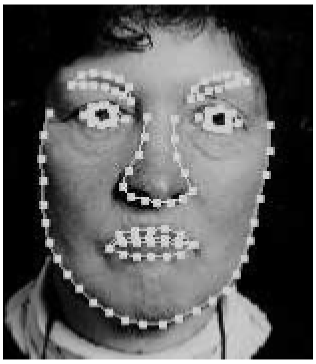


FIGURE 3.1: An example of an annotated face image [3].

3.2.1 Independent AAMs

Independent AAMs can be explained in terms of shape and appearance. Subsection 3.2.1.1 explains the shape of an independent AAM while subsection 3.2.1.2 explains the appearance of an independent AAM, respectively.

3.2.1.1 Shape

The shape of an independent AAM can be described by a model which has a set of vertex locations interconnected to create a mesh (Figure 3.2). The shape s of a mesh is defined as the interconnected points of v vertices, which is denoted as follows[4]:

$$s = (x_1, y_1, x_2, y_2, ..., x_v, y_v)^T \tag{3.2}$$

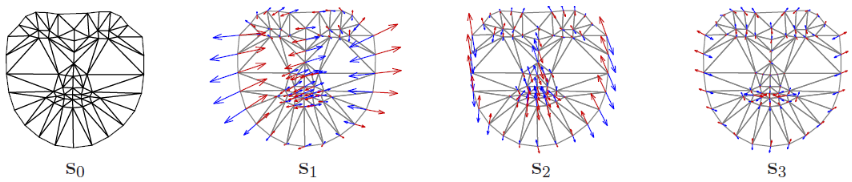


FIGURE 3.2: Shows a shape model. Where s_0 is the base mesh while s_1, s_2 and s_3 are the first three shape vectors showing different deformations of s_0 [4].

Active appearance models permits for linear shape changes, that is, the shape s can be expressed as a base shape s_0 plus a linear set of n shape vectors s_i (Figure 3.2):

$$s = s_0 + \sum_{i=1}^{n} p_i s_i \tag{3.3}$$

where p_i is the set of shape parameters

At this point, it can be assumed that vectors s_i are both orthogonal and normalized, since linear reparameterization can be applied. AAMs are a product of manually labeled training images, it is best to normalize them before PCA is applied to reduce the dimensionality. This removes variations resulting in a PCA output that perfectly focuses on the local and non-rigid shape deformations.

3.2.1.2 Appearance

The appearance model is defined with the base mesh s_0. Let s_0 also denote the set of pixels $x = (u, v)^T$ that lie inside the base mesh s_0. The appearance of an AAM is then an image $A(x)$ defined over the pixels $x \in s_0$ (Figure 3.3). The appearance $A(x)$ can be defined as the base appearance $A_0(x)$ plus a linear series of m appearance images $A_i(x)$:

$$A(x) = A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x) \forall x \in s_0 \tag{3.4}$$

where coefficients of λ_i are the appearance parameters. A_i is assumed to be orthonormal since linear reparameterization can be applied whenever necessary. Similar with the shape component, the base appearance A_0 and the appearance images A_i are computed by applying PCA to a set of shape normalized training images. Each training image is then shape normalized by warping training mesh onto the base mesh s_0. The base appearance is set to be the mean image and the images A_i to be the m eigenimages corresponding to the m largest eigenvalues.

FIGURE 3.3: The linear appearance variation of an independent AAM. Where A_0 is the base appearance image on the pixels inside the base mesh s_0 plus a linear combination of m appearance images A_i also defined on the same set of pixels[4].

3.2.2 Combined Active Appearance Models

These are formulated by merging a shape variation model and an appearance variation model in a shape normalized frame [3]. They combine shape and grey-level in one model. While independent AAMs have separate parameters for shape and appearance, combined AMMs use a single set of parameters $c = (c_1, c_2, ..., c_l)^T$ to parameterize shape:

$$s = s_0 + \sum_{i=1}^{l} c_i s_i \tag{3.5}$$

and appearance:

$$A(x) = A_0(x) + \sum_{i=1}^{l} c_i A_i(x) \tag{3.6}$$

Therefore, the shape and appearance parts of the model are coupled. This formulation is more general and is a strict super set of the independent formulation. Whereby, set $c = (p_1, p_3, ..., p_n, \lambda_1, \lambda_2, ..., \lambda_m)^T$ and s_i and A_i are chosen appropriately. One of the greatest advantages of this implementation is that the number of parameters needed to represent a similar visual phenomenon is reduced, that is $l \leq m + n$. However, at this stage, it cannot be assumed that the vectors s_i and $A_i(x)$ are orthonormal. Ultimately, combined active appearance models are useful for tracking and identification. They are also able to represent any face within the bounds of the training set and are very thorough when it comes to image interpretation.

3.3 Support Vector Machines (SVM)

Support vector machines are derived from ground work done in statistical learning theory that emerged in the 1960's. In the 1990's, this theory gave birth to many other learning algorithms which include SVMs[89]. Support vector machines are good at extreme classification and are quite prominent in visual pattern recognition. They are highly effective learning machines that offer robust and accurate performance in binary classification and regression estimation tasks. They attempt to deduce the global minimum and have a good generalizing capability[90]. In a binary classification task, the SVM seeks to find the best hyperplane that recognizes and separates individual members of the two classes in the training data. The metric for the concept of the best classification function can be found geometrically. For a linearly separable data set, a linear classification function corresponds to a separating hyper plane $f(x)$ that passes through the middle of the two classes while separating the two. The SVM does vector subtraction between the two closest points from either of the two classes (support vectors) and draws a crossing line connecting them. The best hyperplane is the one that maximizes the margin between two classes and is perpendicular to the crossing line between two closest support vectors.

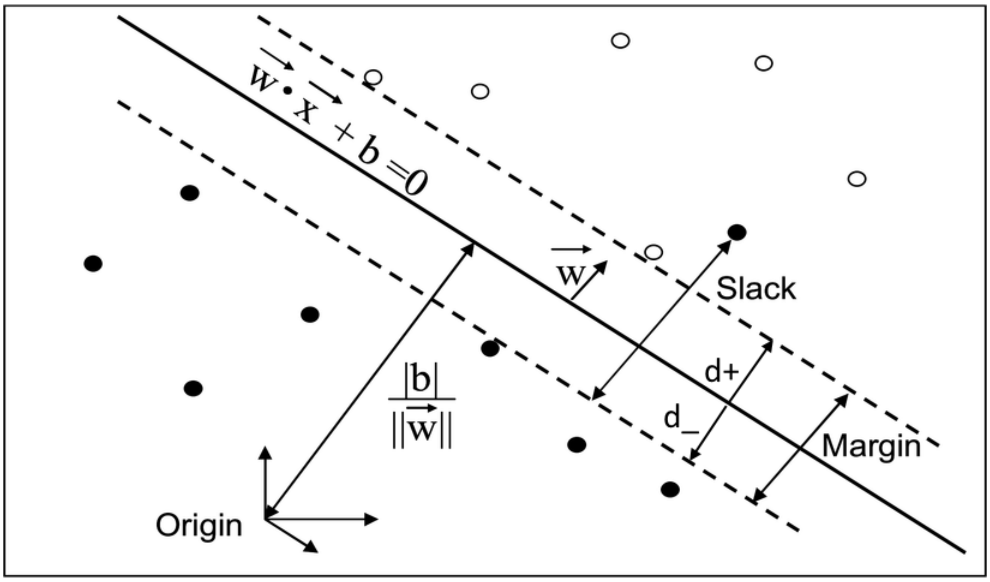


FIGURE 3.4: An SVM separating two classes using a hyperplane[5]

For instance, there is a training data set expressed as x_i, y_i where $i = 1 \ldots N, y_i \in -1, 1, x \in \Re^D$

Assuming there are N training points where x_i has D attributes and falls in either of the two classes y_i = -1 or +1.

Assuming that the data is linearly separable, a line can be drawn between $x_i^{(1)}$ and $x_i^{(2)}$

when $D = 2$.
A hyperplane can be introduced on graphs of $x_i^{(1)}, x_i^{(2)} \ldots x_i^{(D)}$ when $D > 2$.
The hyper plane can be described as

$$w \cdot x_i + b = 0 \tag{3.7}$$

where w is the weight vector perpendicular to the hyper plane, x_i is the input sample and b is the threshold. From Figure 3.4 an SVM selection of w and b can be described by:

$$w \cdot x_i + b \geq +1 \quad \text{for } y_i = +1 \tag{3.8}$$

$$w \cdot x_i + b \leq -1 \quad \text{for } y_i = -1 \tag{3.9}$$

Combining the two equations:

$$y_i(w \cdot x_i + b) - 1 \geq 0 \quad \forall i \tag{3.10}$$

3.3.1 Determining the optimal hyper plane

For instance, there are two classes of data where the margin distance is made up of $(d_+ + d_-)$ Figure 3.4. Where d_+ is the distance from the closest positive support vector to the center of the margin and d_- is also the same distance but from the negative class. The maximum margin distance is when $(d_+ = d_-)$. Denoting H_- as the hyper plane that satisfies $x_i \cdot w + b = -1$ and H_+ as the hyper plane that satisfies $x_i \cdot w + b = 1$, then the margin becomes $\frac{2}{\|\vec{w}\|}$. Therefore, the best hyperplane is the one that minimises $\| w \|^2$.

Introducing Lagrangian multipliers α, $i =, \ldots, l$ and achieving the following Lagrangian,

$$L_P \equiv \frac{1}{2}\| w \|^2 - \sum_{i=1}^{l} \alpha_i y_i (x_i \cdot w + b) + \sum_{i=1}^{l} \alpha_i \tag{3.11}$$

Minimizing L_P w.r.t b and w

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \tag{3.12}$$

$$w = \sum_{i=1}^{l} \alpha_i y_i x_i \tag{3.13}$$

and re-substituting them into Equation 3.11:

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \tag{3.14}$$

Training is accomplished by maximizing L_D with respect to Equations 3.12 and 3.13. All training points with $\alpha_i > 0$ are called support vectors and they lie closest to the separating hyper plane. All other points with $\alpha_i = 0$ are not critical and do not influence the shape of the hyper plane boundary.

3.3.2 The idea of Kernels

In cases of data being nonlinear, SVMs do a mapping of the data into a higher dimensional feature space where linear algebra and geometry may be used to separate the data using nonlinear rules in the input space. To cater for this, the SVM is formulated in such a way to use kernels allowing for efficient computations of dot products in the feature space without having to directly work in the higher dimensional space. A kernel function (K) is a function comparable to a dot product in a feature space. It is given below as

$$K(x_i, x_j) = (\phi(x_i)^T \cdot (\phi x_j)) \tag{3.15}$$

where ϕ is a nonlinear mapping function which maps input features into a higher dimensional feature space. Kernel functions that are mostly used are as follows[91]:

- Linear: $K(x_i, x_j) = x_i^T x_j$
- Polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$
- Radial basis function (RBF): $K(x_i, x_j) = \exp(-\gamma \parallel x_i - x_j \parallel^2), \gamma > 0$
- Sigmoid: $K(x_i, x_j) = tanh(\gamma x_i^T x_j + r)$

where γ, r and d are kernel parameters.

3.4 Extreme Learning Machines (ELM)

Extreme learning machines are single hidden layer feed forward neural networks (SLFN). The concept behind them back dates to Frank Rosenblatt's multilayer perceptron. He believed that perceptrons can enable computers to "walk, talk, see, write, reproduce itself and be aware of its existence"[6]. Extreme learning machines aims to remove the hurdle between artificial and biological learning capability. From Huang et al.[92], extreme is defined as a move from artificial learning to brain-like learning. The notion behind the ELM is that, input weights and hidden layer biases of single hidden layer feed forward neural networks can be randomly assigned if the activation functions in the hidden layer are infinitely differentiable. The output weights of a single hidden layer feed forward neural networks can be analytically determined through simple generalized inverse operation of the hidden layer output matrices[93]. Extreme learning machines overcome limitations such as slow learning, human intervention and expensive computations faced by other algorithms like the artificial neural network with back propagation among others[6]. The training speed of the ELM is extremely fast relative to the conventional artificial neural networks and also obtains better generalization performance. A diagram of an ELM implementation is presented in Figure 3.5.

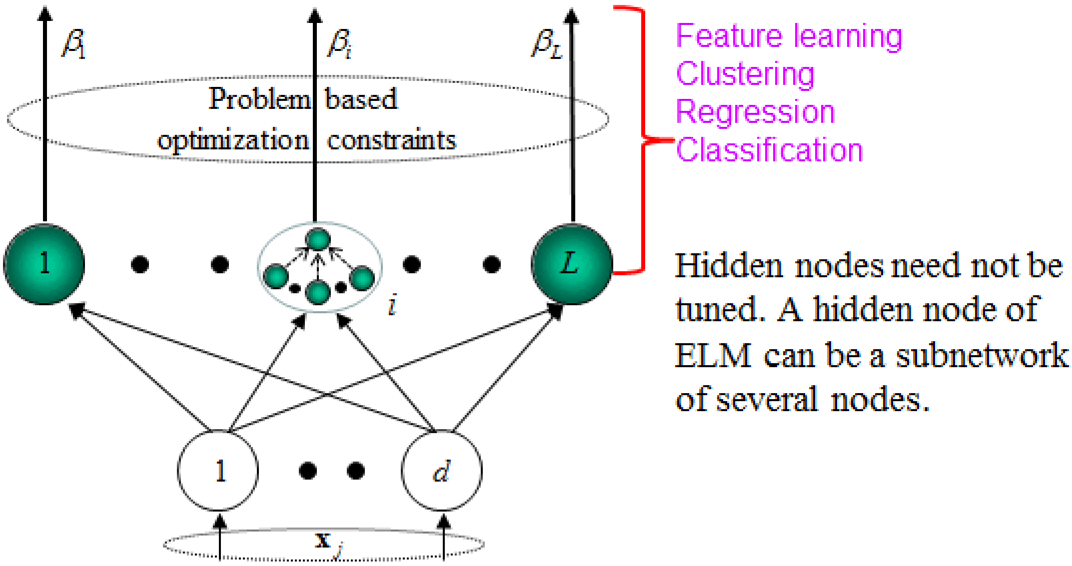


FIGURE 3.5: A general ELM diagram[6]

The ELM algorithm can be described as a SLFN with random hidden nodes as follows: For N distinct samples (x_i, t_i) where $[x_i^1, x_i^2, \ldots, x_i^n]^T \in \Re^n$ and $t_i = [t_i^1, t_i^2, \ldots, t_i^m]^T \in \Re^m$, general SLFNs with $\hat{N}$ hidden nodes and activation function $g(x)$ are mathematically modeled as

$$\sum_{i=1}^{\hat{N}} \beta_i g_i(x_j) = \sum_{i=1}^{\hat{N}} \beta_i g(w_i \cdot x_j + b_i) = o_j, \quad j = 1, \ldots, N \tag{3.16}$$

where $w_i = [w_i^1, w_i^2, \ldots, w_i^n]^T$ is the weight vector connecting input nodes to the i^{th} hidden node. $\beta_i = [\beta_i^1, \beta_i^2, \ldots, \beta_i^m]^T$ is the weight vector connecting the i^{th} node and the output nodes and b_i is the threshold of the i^{th} hidden node. $w_i \cdot x_j$ denotes the inner product of w_i and x_j.

The standard SLFNs with $\hat{N}$ hidden nodes and activation function $g(x)$ can approximate these N samples with zero error means that $\sum_{i=1}^{\hat{N}} \|o_j - t_j\| = 0$ that is, there exist β_i, w_i and b_i such that:

$$\sum_{i=1}^{\hat{N}} \beta_i g(w_i \cdot x_j + b_i) = t_j, \quad j = 1, \ldots, N \tag{3.17}$$

The above equations can be written as :

$$H\beta = T \tag{3.18}$$

where

$$H(w_1, \ldots, w_{\hat{N}}, b_1, \ldots, b_{\hat{N}}, x_1, \ldots, x_N) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_{\hat{N}} \cdot x_1 + b_{\hat{N}}) \\ \vdots & \cdots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_{\hat{N}} \cdot x_N + b_{\hat{N}}) \end{bmatrix}_{N \times \hat{N}} \tag{3.19}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\hat{N}}^T \end{bmatrix}_{\hat{N} \times m} \quad and \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \tag{3.20}$$

where H is called the hidden layer output matrix of the neural network. The i^{th} column of H is the i^{th} hidden node output with respect to $x_1, x_1, \ldots, x_N$

3.4.1 Minimum norm least squares solution to SLFNs

The common understanding of conventional neural networks stipulates that all parameters of SLFNs need to be tuned or adjusted. However, in this case, the input weights w_i and the hidden layer biases b_i are not necessarily tuned and the Hidden layer matrix H can actually remain the same once random values have been assigned to the parameters in the beginning of learning. For fixed weights and biases, to train an SLFN one has to find a least square solution $\hat{\beta}$ of the linear system $H\beta = T$:

$$\| H(w_1, \ldots, w_{\hat{N}}, b_1, \ldots, b_{\hat{N}})\hat{\beta} - T \| = \min_{\beta} \| H(w_1, \ldots, w_{\hat{N}}, b_1, \ldots, b_{\hat{N}})\beta - T \| \tag{3.21}$$

If $\hat{N}$ number of hidden nodes is equal to the number of N distinct training samples, then matrix H is square and invertible when the input weight vectors and biases are randomly chosen. This implies that SLFNs can approximate these training samples with zero error. In most cases the number of hidden nodes is usually less than the number of distinct training samples, $\hat{N} < N$, H is a non square matrix and there may not exist $w_i, b_i, \beta_i(i = 1, \ldots, \hat{N}$ such that $H\beta = T$. Therefore, the smallest norm least squares solution of the above linear system is

$$\hat{\beta} = H^{-1}T \tag{3.22}$$

where H^{-1} is the Moore-Penrose generalized inverse of matrix H

In a nutshell, the presented learning method for SLFNs called extreme learning machines is given as follows: Consider a training set

$$(x_i, t_i)|x_i \in R^n, t_i \in R^m, i = 1, \ldots, N$$

, with an activation function $g(x)$ and $\hat{N}$ hidden nodes,

- Randomly assign input weights w_i and bias b_i $i = 1, \ldots, \hat{N}$
- Calculate the hidden layer output matrix H
- Calculate output weights β
$$\beta = H^{-1}T \tag{3.23}$$
where $T = [t_1, \ldots, t_N]^T$

In theory, this algorithm works for any infinitely differentiable function $g(x)$ such as sigmoidal function, radial basis, sine, cosine and many other non-regular functions.

3.5 Facial Action Coding System (FACS) and Action Units (AU)

The facial action coding system (FACS) [94] is undeniably the most used method for coding facial expressions. It was created by Ekman and Friesen in 1978 to taxonomies facial movements by their appearance on the face. The system explains facial expressions in terms of 46 component movements, which correspond to the individual facial muscle movements. It came about as a standard measure of quantifying facial expressions and facial behavior as a whole. The system has proven to be useful to psychiatrists, animators and also psychologists among others. The coding system divides the face into 11 regions or parts as shown in Table 3.1 below. It includes recognizing individual or sets of facial muscles that have an effect on facial behaviors. Different muscles have different effects on the face and these lead to different facial changes. As a result, different muscles involved in this process are called Action Units (AU). Action units are numerical codes that calibrates the movement of facial muscles. Action units can occur singly or in combination. As they occur in combination they can be said to be additive or non-additive. Action units are said to be additive if the appearance of each AU has no effect on the other and they are said to be non-additive if they have an effect on each other's appearance [95]. For instance, AU1 + AU5 are additive since the combination does not alter the appearance of the constituents. On the other hand, AU1 +AU4 is a non-additive combination since it changes the appearance of the constituents. AU4 appears differently when alone and when in combination with AU1. The FACS system divides AUs into the upper face, lower face and miscellaneous. The miscellaneous set involves head-eye movement AUs. Action units from 1 to 7 are said to represent the upper face while 8 to 46 represent the lower face and the remaining 12 are the head-eye movement AUs. AUs come in different intensities and they are given below in Figure 3.6. In this study, FACS AU was used to decode all observed facial expressions. For instance, AU6 (Cheek riser) and AU12 (Lip corner puller) denoted a smile.

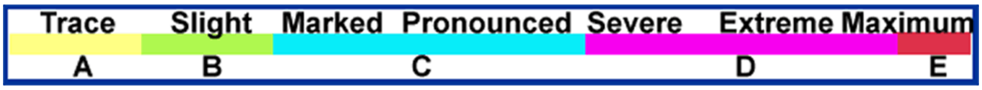


FIGURE 3.6: Shows FACS AU intensity levels

TABLE 3.1: Facial areas and their respective definitions[15].

Facial Area	Definition
Glabella	Area of forehead between the eye brows
Root of nose	The beginning of the nose between the eyes; also called the nasal root.
Eye Aperture	The degree to which the eye is open; the eye opening.
Eye cover fold	The skin between the eyebrows and the palpebral part of the upper eyelid(the part that contacts the eyeball), which folds into the eye socket.
Lower eyelid furrow	A place below the lower eyelid where a line or wrinkle may appear. A line or wrinkle may be permanently etched into the face; if so, it will deepen with certain AUs. If not, it should appear when these AUs are contracted.
Infraorbital furrow	A place where a line or wrinkle may appear parallel to and below the lower eyelid running from near the inner corner of the eye and following the cheek bone laterally.
Nostril wings	The fleshy skin of the side of the nose that forms the outside of each nostril.
Nasolabial furrow	A place where a line or wrinkle may appear which begins adjacent to the nostril wings and runs down and outwards beyond the lip corners. In some people, it is permanently etched in the face; if so,it will deepen with certain AUs. If not, it will appear on most people's faces with certain AUs.
Philtrum	The vertical depression in the center of the upper lip directly under the tip of the nose.
Chin boss	The skin covering the bone of the chin
Sclera	The white part of the eyeball.

3.6 Prkachin and Solomon pain intensity (PSPI) scale

In 1992, Prkachin[96] discovered that there were consistent facial AUs that existed as people went through pain. These AUs were brow lowering, tightening and closing of the eye lids and nose wrinkling or upper lip raising. In 2008, Prkachin and Solomon[97] did a further analysis on patients that suffered from shoulder pain. They confirmed that indeed there are consistent facial actions exhibited as humans go through pain. In that study, they brought about a ground breaking model called the Prkachin and Solomon pain intensity PSPI scale. The PSPI scale is one of the greatest contributions towards pain detection to date that complements self-reporting. Its calibration is solely based on the FACS AU intensities. The model sums up certain active facial AUs along with their intensities and gives an ultimate pain score. Pain intensity is derived from a set of facial action units that co-exist as a person is going through pain (Figure 3.7). The

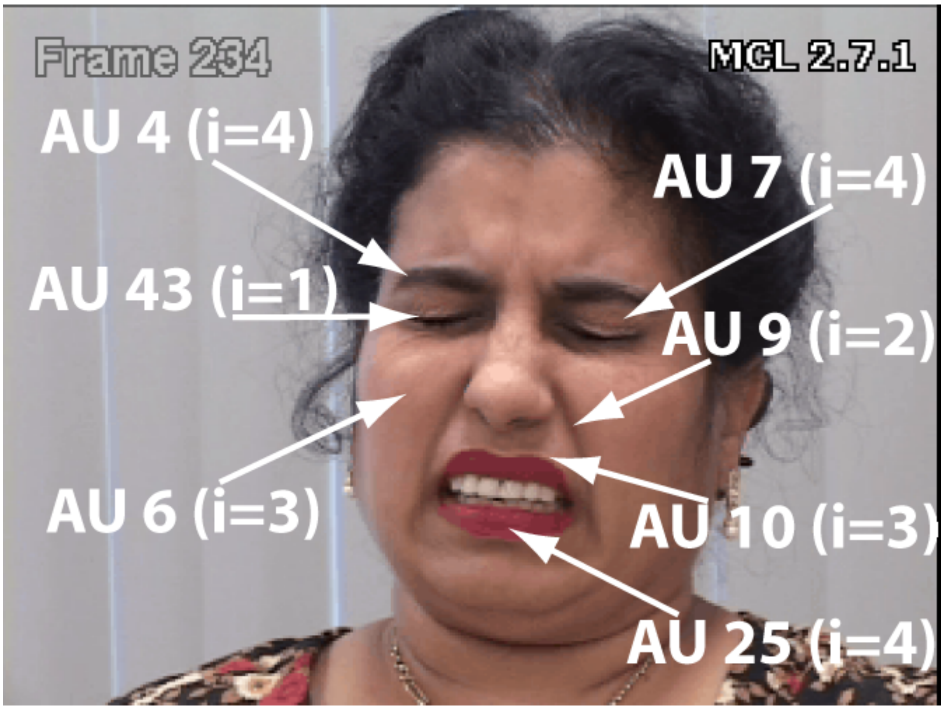


FIGURE 3.7: Shows a face showing all relevant Action Units (along with their intensities) that describe a pain expression[7]

pain intensity equation is given as:

$$Pain \quad Intensity = I(AU4)+max(I(AU6)\|I(AU7))+max(I(AU9)\|I(AU10))+I(AU43) \tag{3.24}$$

where I represents the intensity, AU4 is brow lowerer, AU6 is cheek raiser, AU7 is lid tightener, AU9 is nose wrinkler, AU10 is upper lip raiser and AU43 is eyes closed. The sum of all the AUs depending on its intensity from A to E, where $A = 1, B = 2, ..., E = 5$ respectively, are used to get a single score. For AU6 and AU7 as well as AU9 and AU10, only one with the higher or greater intensity is considered per pair. AU43 will not be calibrated according to intensity, its value is either a 1 or a 0 depending on its existence or absence, respectively.

3.7 Databases

Insights on the existing data sets and some that were used in this study are presented. Cohn-Kanade AU-Coded Facial Expression Database, the Japanese Female Facial Expression (JAFFE) database, Yale Database, the Denver Intensity of Spontaneous Facial Actions (DISFA) database and the UNBC-McMaster shoulder pain archive database are briefly explained in the following section.

3.7.1 Cohn-Kanade AU-Coded Facial Expression Database

The Cohn-Kanade AU-Coded facial expression [74] database was created using subjects aged between 18 and 30 years. It contains image sequences of subjects from different backgrounds, race, and ethnicity. Two Panasonic WV3230 cameras, each connected to a Panasonic S-VHS AG-7500 video recorder with a Harita synchronized time-code generator were used for sample collection. Recordings were done with a camera positioned directly in front of the subject, while another one was placed at a 30 degree angle to the subject. However, only images showing the frontal view are provided for research purposes. Subjects were told and shown by an instructor, on how to perfectly reenact a series of 23 facial expressions that varied from single to combinations of action units such as AU6(Cheek raiser) + AU12(Lip corner puller). Most facial expressions were based on the prototypic basic emotions such as joy, anger, fear, disgust, sadness, and surprise. The Cohn-Kanade was later enhanced into the extended Cohn-Kanade (CK+) database[8]. This database contains 593 sequences from 123 subjects. Each video sequence starts from a neutral (no expression) to a peak intensity of an expression. Only the last frame is FACS coded for presence of AUs and most videos contain posed facial expressions. Grayscale sequences were recorded at a resolution of 640×490 pixels while 28-bit color sequences were recorded at 640×480 pixels. Figure 3.8 shows a subject in a neutral state and at the peak of an expression.

3.7.2 Japanese Female Facial Expression (JAFFE) database

This database contains 213 images collected from 10 Japanese female models [54]. Each subject was recorded showing 7 facial expressions composed of 6 basic facial expressions

FIGURE 3.8: Extract from the CK+ database showing a neutral face (Frame 1) and peak of an expression (Frame 19). Each sequence begins with a neutral expression and proceeds to a target expression. In the example shown, the peak expression is happiness, AU (6+12+25) [8].

and a single neutral expression. Each image was rated by 60 Japanese subjects for each of the six basic emotions. Images are labeled based on the most dominant expression in that image. The images were taken in a controlled environment at the Psychology Department at Kyushu University. However, considerably low-resolution images were captured (256×256) relative to other facial expression data sets. Figure 3.9 shows examples of images along with their labels in the JAFFE database.

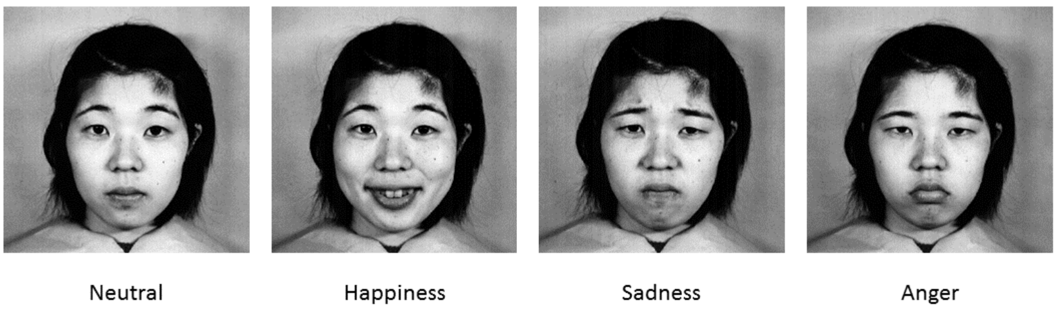


FIGURE 3.9: showing examples of Images in the JAFFE database and relevant labels for each image [9].

3.7.3 The Yale database

The Yale Face Database [98] contains 165 gray scale images in GIF format of 15 individuals with 60 facial expression images. There are 11 images per subject, one per different facial expression or configuration: center-light, with glasses, happy, left-light, with no glasses, normal, right-light, sad, sleepy, surprised, and wink. The extended Yale face Database B contains 16128 images from 28 human subjects under 9 poses and 64 illumination conditions. The images were recorded using a purpose-built illumination rig. This apparatus is fitted with 64 computer controlled strobes. 64 images per subject in a specific pose were recorded at 30 frames per second in a period of about 2 seconds. This was done so as to cater for any changes in head pose and facial expression. Figure 3.10 shows sample images extracted from the YALE database.

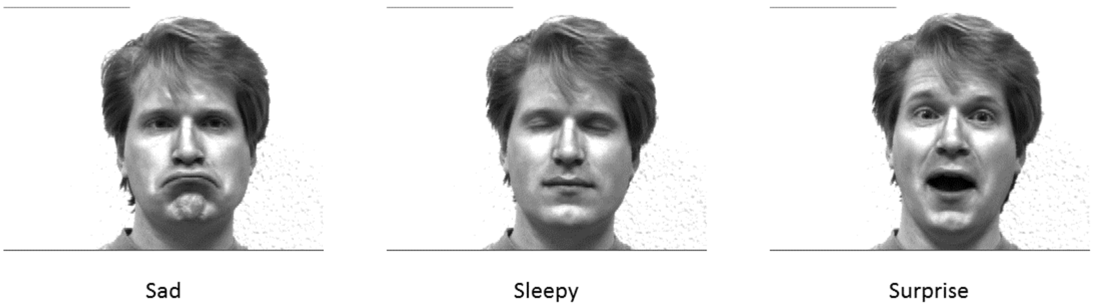


FIGURE 3.10: Sample images from the Yale database [10].

3.7.4 Denver Intensity of Spontaneous Facial Actions (DISFA) Database

The Denver Intensity of Spontaneous Facial Actions (DISFA) [11] database contains approximately 130,000 annotated frames from 27 adult participants (15 men and 12 women) aged between 18 and 50 years. Out of these subjects, 3 are Asian, 21 are Euro-American, 2 are Hispanic, and one is an African-American [11]. Participants watched a 4-minute video clip so as to elicit spontaneous AUs in response to videos. They then captured a range of facial expressions of emotion. The clip consisted of nine segments taken mostly from YouTube. All videos were recorded using a high-resolution camera (1024×768 pixels) at 20 fps and each frame is manually FACS coded for the presence of AUs. For each video sequence, a total of 4845 frames were recorded. Figure 3.11 below shows images of the 27 participants. Two refused to have their images used in publications.

3.7.5 The UNBC-McMaster Shoulder Pain Expression Archive Database

The UNBC-McMaster shoulder pain expression archive database[12] contains 200 sequences of spontaneous pain expressions recorded from 25 subjects and these have been given out for research purposes. This database was recorded at a resolution of 320×240 pixels. Every frame is FACS coded and pain score calibrated. 66 AAM landmark points for each and every frame are readily available. Video sequences in the database were recorded as participants were activating injured shoulders to make sure that genuine

FIGURE 3.11: Facial images of 25 of the 27 subjects from the DISFA database [11].

pain expressions were exhibited. Associated pain self-report and observer ratings are provided in the database as well. Figure 3.12 shows samples extracted from the UNBC-McMaster database.

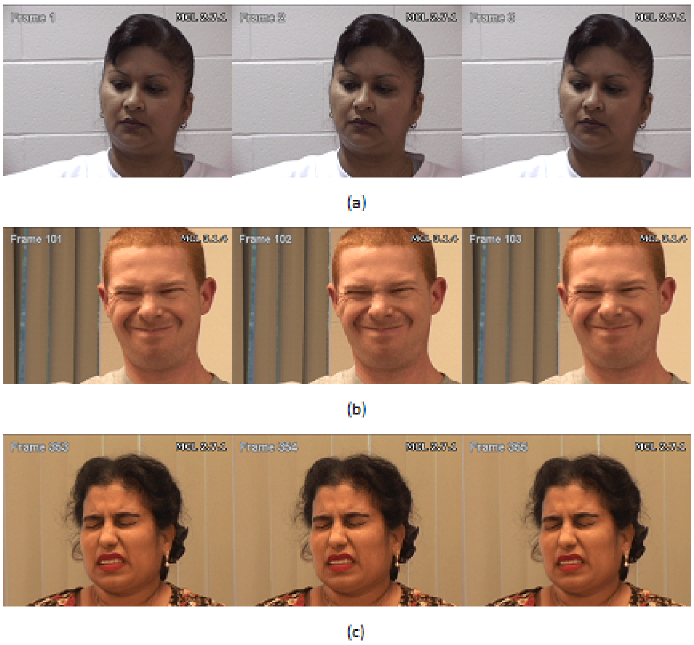


FIGURE 3.12: Samples from the UNBC-McMaster shoulder pain expression archive[12]. Figure 3.12(a) showing subject 042 with PSPI score = 0, Figure 3.12(b) showing subject 52 with a PSPI score = 8 and lastly, Figure 3.12(c) showing subject 107 with a PSPI score = 15.

3.8 Conclusion

This chapter focused on algorithms that are commonly used in FER and some that were mostly used in literature survey. It also presented some databases that are commonly available for FER. In the following chapter, a detailed sequence of steps that were followed during experimentation is presented.

Chapter 4

Methodology

4.1 Introduction

The main aim of this research is to be able to find the temporal relationship between eye actions and facial expressions. The findings of the study may contribute towards the authentication of expressions based on relevant subtle changes in features as opposed to stand-alone authentication. This chapter presents the hypothesis, main research question, sub questions as well as respective objectives. The sequence of steps that were followed in carrying out the study are also presented.

4.2 Hypothesis

It is hypothesized that it is possible to deduce the temporal relationship that exists between eye actions and facial expressions.

4.2.1 Main Research question

Can the temporal relationship between eye actions and an expression be deduced?
In order to study the answer the above question, two different sets of experiments were conducted. The study investigated the time-related distances that exits between smiles and blinks as well as between pain expressions and eye closure. In subsection 4.3.1, steps taken to study the smile-blink relationship are presented, while in subsection 4.3.2, steps taken to study pain-eye closure relationship are presented.

4.3 Experimental methodology

At this stage, the sequence of steps that were taken for smile authentication are separated from the ones that were taken for pain detection studies. However, both studies follow the same steps such as data collection, feature selection and classification. Both studies where done at frame and sequence level. Frame level studies involved, individual frames that were evaluated on how they were classified and in sequence level studies, a series of frames where evaluated to see if they produce the overall desired expression.

4.3.1 Study 1: Smiles and eye blink detection

Here a detailed procedure on how smile studies were conducted is presented. As previously mentioned in the hypothesis, this study aims to find the temporal relationship between eye actions and an expression. In smile-blink study, eye actions relate to eye blinks and expressions relate to smiles. The greatest motivator to use eye blinks in this study is based on previous studies done by Trutoiu et al.[14], it was noted that there was a temporal connection between blinks and smiles. So this study sought to use eye blinks and their mapping with smiles as a way to enhance the understanding of smile expressions and their authenticity.

Subsection 4.3.1.1 provides all the research sub-questions to be answered under smile-blink studies. The respective objectives are given in Subsection 5.2.1.1. Subsections 4.3.1.3 to 4.3.1.6 provide finer details on the steps followed in data collection, feature extraction, eye action labeling and classification of blinks and smiles, respectively

4.3.1.1 Research sub-questions

1. Can an active appearance model (AAM) extract all the relevant features for smile expression and eye blink recognition?
2. Can support vector machines effectively classify eye blinks and smile expressions?
3. Can a temporal mapping between eye blinks and smiles be found?
4. Can the extracted temporal relationship between eye blinks and smiles help in understanding the difference between posed and spontaneous smiles?

4.3.1.2 Objectives

1. To build an AAM-SVM model that uses AAM points to recognize eye blinks and smile expressions.

2. To build a framework, based on the temporal relationship between eye blinks and smiles, for understanding smiles.

4.3.1.3 Data collection

The CK+ database[8] and DISFA database[11] were used in this experiment and they both contain AAM landmark points. Landmark points that were considered for eye blink and smile expression are depicted in Figure 4.1. Each landmark point is made up of x and y coordinates which represents a pixel's position on a normalized image. These coordinates change per frame as the facial expression evolves. These displacements are descriptive of the changes that occur on the face as an expression is being activated.
From the CK+ database, which comprises of posed expressions, a total of 467 samples (containing open eyes and closed eyes) were collected from 30 subject and used for the blink experiment while 545 samples (positive i.e containing smiles and negative i.e containing no smile) extracted from 35 subjects were used in the smile experiment.
From the DISFA database, which consists of spontaneous expressions, a total of 121114 samples (positive and negative) extracted from 25 subjects were used for the smile experiment. These samples included 13 925 positive and 107 189 negative samples. For the blink experiment a total of 1000 samples were used which were made up of 500 positive and negative samples each.

Figure 4.1: Feature points extracted for both smile and blink expressions. Figure 4.1a: 66 smile feature points. Figure 4.1b: 12 blink feature points.

4.3.1.4 Feature selection

Active appearance model landmark points given along with the databases were used for all experiments. These were extracted using active appearance models and packaged with the database images. The DISFA database had 66 landmark points while the CK+ database had 68 landmark points. Two extra AAM points from CK+ database were not considered so as to maintain a uniform number of landmark points from both databases. For evaluation of blinks, a total of 12 landmark points surrounding the eye region were selected (Figure 4.1b). For smiles, all 66 points were considered so that all facial muscles constituting a smile (AU6 + AU12 and/or AU25) could be examined (Figure 4.1a). Based on previous studies AU6 and AU12 are shown to be inter-related since they activate the same muscle[24]. The orbicularis oculi is responsible for the squinting of eyes as a person smiles. The same muscle is also responsible for the eye closure during the blinking process of an eye[24]. Some parts of the orbicularis oculi are also involved in demarcating Duchene smiles. The Zygomaticus major muscle is responsible for the U shape of the mouth as a person smiles (Figure 4.2). Looking at these major muscle effects that are associated with smiling, the study chose to consider all 66 AAM landmark points that describe the face so that no information about any muscle contraction is lost.

4.3.1.5 Eye action labeling

Since FACS AU coding is available only for first and last frames of samples available in CK+ database, a method was incorporated that would help in labeling eye blinks (eye open and eye close). After feature selection, inter-eyelid distances were calculated to determine AU labels for each frame as suggested by Trutoiu et al.[14]. This is whereby the absolute distance between the horizontally center upper eyelid marker and the horizontally center lower eyelid marker was calculated. The local minima in the inter-eye blink distance series implied that the eyelids were close together and this event was regarded as a blink.

The same process was applied to samples from DISFA database since no eye action unit labeling was provided. This technique gave a guide on where to expect blinks in a video sequence before classification using algorithms is done. Euclidean distances between the upper and lower eyelids were calculated. The max inter-eyelid distance between the two eyes was considered so as to avoid picking one eye closure (a wink). Geometric points that were considered for blink calculations are given in Figure 4.3.

In the CK+ database a blink was evaluated using a 80% eye closure threshold relative to the initial neutral frame per subject. Anything beyond the mark was recorded as an eye closure as shown by the red points in Figure 4.4. However for samples from DISFA database, a threshold value of 10 pixels and below was used to indicate an eye closure action as some videos did not start with a neutral position and the recordings had a high resolution. This pixel metric is specific to DISFA database samples, any new samples

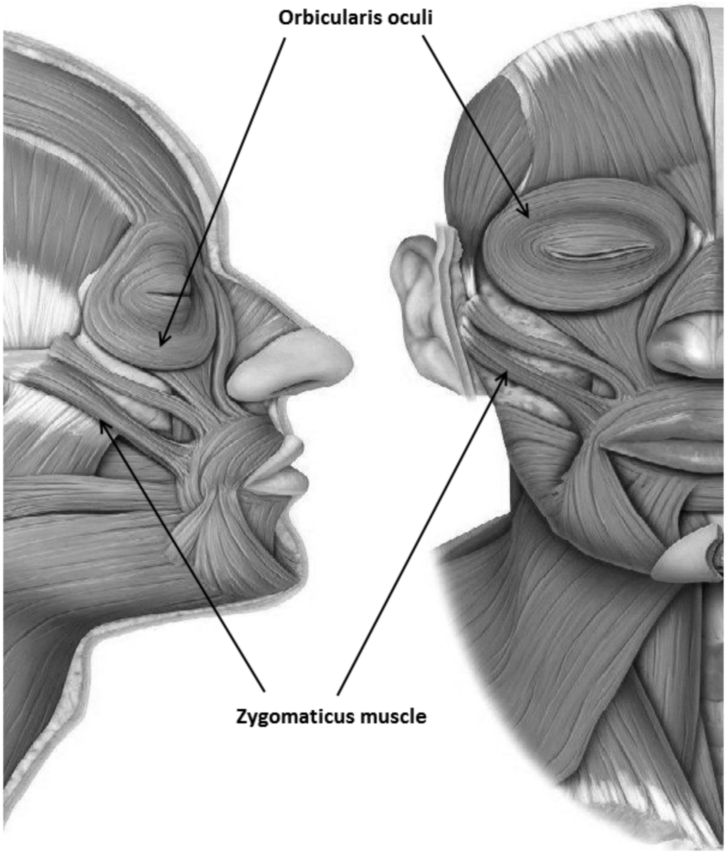


FIGURE 4.2: Shows the muscles involved in blinking and smiling which are the orbicularis oculi and the zygomaticus major muscle[13].

may need to be normalized to 1024×768 pixels per frame for this system to work. To affirm the correctness of this system, individual frames were manually checked to see if they truly showed eye closure or eye opening in both databases. Problems like variations in test subjects did not affect the results since the images were all normalized to a fixed size per frame. In addition, the documentation of both databases mentions that AAM landmark points were extracted on uniformly normalized images.

4.3.1.6 Classification

Classification was performed using SVMs and LibSVM[99] was used for this purpose. Support vector machines are powerful classifiers and are robust to the curse of dimensionality. The curse of dimensionality is a problem that occurs when feature size is greater

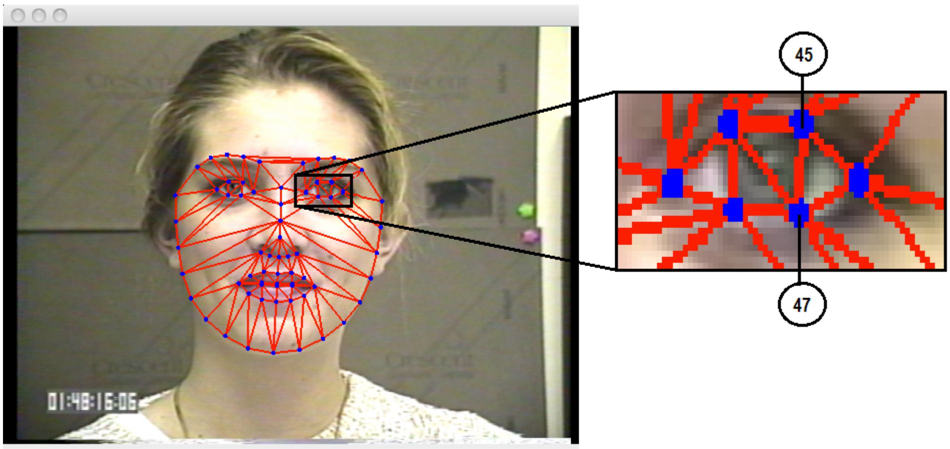


FIGURE 4.3: Image showing relevant points for inter-eyelid distance calculations. Points 45 and 47 on the left eye and points 38 and 42 on the right eye of the subject were used[14].

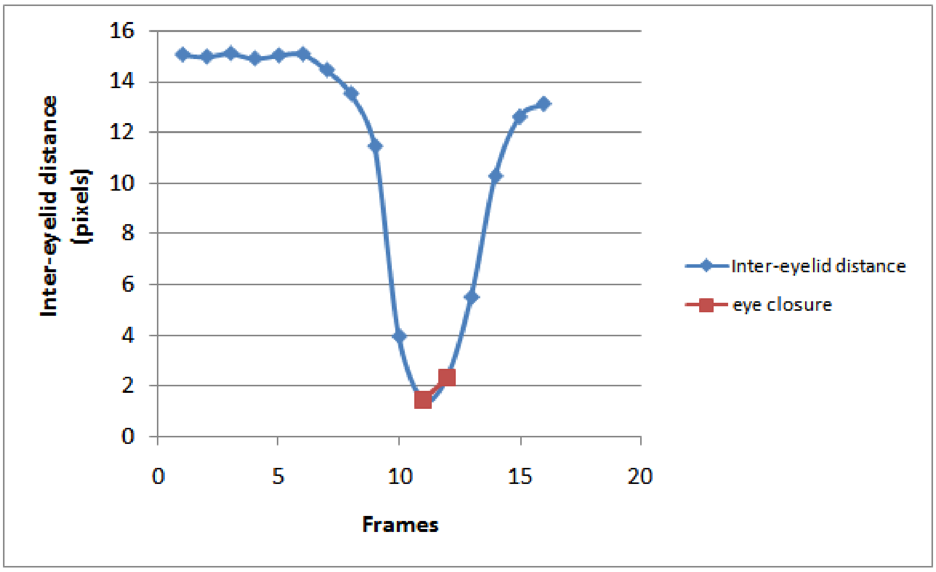


FIGURE 4.4: Example of the inter-eyelid distance calculation for a single subject.

than the number of training samples. In most cases this leads to over-fitting and poor generalization of new samples[100]. Additionally, it also increases the computational complexity and sensitivity to noise because of a high dimensional feature space[100]. Since SVM uses kernels it has a capability of overcoming problems associated with the curse of dimensionality. They seek to maximize the hyper plane margin between two classes. Support vector machines are superior to neural nets as their formation allows them to have one global minimum and also possess a high generalization capability [90]. SVMs with an RBF Gaussian kernel were implemented in all the experiments. Selection of optimum values for C and γ was done using the grid search. This selects the optimum parameters for the kernel through cross validation and these parameters are only based on the training set. For each fold, 132 iterations were conducted in search for best C

and γ values. Different C and γ values obtained for each fold yielded an average training accuracy rate of 98.21% when tested on the CK+ database. Samples from DISFA database were all classified correctly, leading to a 100% accuracy rate. Results from test sets are provided in the following Chapter.

Individual classifiers were used for classification of either eye open or eye close and smile detection. This means that a single SVM was used to recognize eye close and eye open. Another single SVM was used to recognize the existence of smiles as well as neutral faces. This allowed to separately observe the onsets and offsets of both expressions in a single video sequence. It also clearly revealed the dynamic classification sequence per frame for each expression.
Extreme learning machines were also incorporated as classification techniques for the classification of DISFA database samples. An ELM model made up of 132 input neurons (66×2 features), 300 hidden neurons (optimized from experiments), a Sigmoid activation function and 2 output neurons (smile or no smile classification) was used. For the determination of optimum number of hidden neurons, a range of 1 to 600 neurons was considered. Training and testing was done at an interval of 50 hidden neurons with the best performance being achieved at 300 hidden neurons. An average training accuracy of 99% was achieved. Results from the test sets are provided in the following Chapter.

4.3.1.7 Classification of Blinks

A five-fold subject independent test was carried out where each training set contained samples from 24 subjects. In the creation of training data, neutral frames (open eyes) and frames where both eyes were completely closed (closed eyes) were selected. 12 landmark points that surrounded the eye region (points 37 to 48, Figure 4.1) were selected for evaluation of eye status (either open or closed) per frame. The 12 AAM landmark coordinates were converted into a single 1×24 feature vector. Training data file in each fold was made up of a $m \times n$ feature vector, where m was the number of samples and n was the number of features. These values were constant for each training data file with $m = 48$ and $n = 24$. +1 and -1 labels were used to represent open eyes and closed eyes, respectively. 24 samples for each eye status (open and closed) were used. Test data had different values of m because total number of frames per video sequence varied. Nevertheless, n was constant for each and every video sequence at 24. A prediction sequence consisting of ($+1, -1, +1$) in less than $\frac{1}{2}$ of a second was used to show a correct blink detection while the model on test sets. This is based on the description of a blink event given in the FACS manual[15]. From the DISFA database, 25 subjects were considered. A subject independent test was done for the tracking of blinks. Training data in each fold was made up of 1000×24 feature vector. The size of the test data varied per subject but ranging between 4843 and 4845 frames. Similar labeling as that of CK+ based experiments was used.

4.3.1.8 Classification of Smiles

For experiments using samples from CK+ database, a five-fold subject independent cross validation test was carried out. For instance, a single video frame was represented as an $m \times n$ feature vector where $m = 1$ was the number of samples in the set and $n = 132$ was the number of features. Therefore, a single training data file was made up of a 56×132 feature vector, where $m = 56$ and $n = 132$. Out of the 56 samples, 28 were positive and 28 were negative. The labels used on the training data were -1 for no smile and +1 for existence of a smile (AU6 + AU12) per frame. For samples taken from the CK+ database, a prediction sequence from -1 to +1 was regarded as a correct recognition of the existence of a full smile expression. This is due to the fact that expressions in the CK+ run from neutral to peak intensity.
For experiments conducted using samples from DISFA database, a five-fold cross validation test was carried out with the size of the training data being 96891×132, where $m = 96891$ and $n = 132$. Similar labeling as that of CK+ was used. Negative and positive training samples per fold were approximately 85751 ± 2 and 11140 ± 1, respectively.

4.3.2 Study 2: Pain and eye closure experiment

In this section the sequence of steps on how pain studies were conducted is presented. In this study, eye closure is used as the respective eye action. The study of the temporal relationship between eye closure and pain is highly motivated by the findings of Prkachin and Solomon [97], where they found a number of action units that were consistent with pain expressions. One of these action units was AU43 which depicts eye closure. Based on their findings, this current study sought to understand the temporal relationship between eye closure and spontaneous pain. Thereby, contributing towards evaluation of posed and non-posed pain based on eye closure.
Subsection 4.3.2.1 details the research sub-questions to be answered under pain and eye closure studies as well as their respective objectives in Subsection 4.3.2.2. Subsections 4.3.2.3 to 4.3.2.6 provide finer details on the steps followed in data collection, feature extraction, eye action labeling and classification of eye closure and pain expressions, respectively.

4.3.2.1 Research sub-questions

1. Can an active appearance model extract all the relevant features for pain expression and eye closure recognition?
2. Can extreme learning machines help in improving classification of pain expressions as compared to standard predominant classifiers such as support vector machines?
3. Can the temporal relationship between eye closure and spontaneous pain be determined?

4.3.2.2 Objectives

1. To build models that use AAM+SVM and AAM+ELM to recognize pain expressions.
2. To compare the performance of AAM+SVM and AAM+ELM models, and determine a better model for pain detection.
3. To find correlations between eye closure and spontaneous pain.

4.3.2.3 Data collection

The pain-eye closure experiment was done using SVMs and ELMs on the UNBC-McMaster database[12]. This database comprises of 200 spontaneous pain videos recorded from 25 subjects. Criteria for selection of frames was based on the PSPI scale. All frames that contained PSPI scores ≥ 8 and PSPI score $= 0$ were used for training and testing. The reason behind using higher pain intensity scores is attributed to the possibility of effectively depicting pain expressions when all action units associated with pain are active. In instances of pain with lower intensity, a single AU may contribute an intensity of 5 while others are non existent. This will make it complex in training a classifier for a single AU does not necessarily describe a face going through pain. For instance, AU7 may exist with an intensity of 5 in combination with AU43 with an intensity of 1 while all other AUs are absent. In such a case, the face might not best describe a pain expression hence the motive to use higher pain scores. As previously mentioned, the study was conducted at frame and sequence level. At frame level the idea was to assess the classification of expressions per frame and to assess if a series of frames produce the overall desired expression at sequence level. For the frame based test, a total of 461 frames were used and for the sequence based test 8 video sequences (2809 frames) were evaluated. In the frame based test, each frame is checked for the existence of either open eyes or closed eyes. In the sequence based test the whole sequence is evaluated to see if the classifier recognises the full expression being exhibited by the person in the video. This process was done using proposed classifiers and results are tabulated and graphically presented in the next chapter.

4.3.2.4 Feature selection

In this study, AAM feature points were used for representing relevant features. Active appearance model landmark points used in this study are pre-extracted and were given with the database. 66 landmark points are given for every frame. All landmark points were considered since they best describe the face as a person exhibits an expression of pain. A single video frame was converted into a $1 \times n$ feature vector, where n (66×2) is the number of feature points retrieved using AAM. A single training data set was made up of $m \times n$ feature vectors, where m is the number of samples which varied depending on the number of frames with the desired PSPI score. However, (n) the number of features remained constant at 132.

4.3.2.5 Eye action labeling

The same process of calculating inter-eyelid distances to determine eye AU label for each frame as suggested by Trutoiu et al.[14] was incorporated. This helped in understanding the actual state of the eye at any given instance in the video. Though in the pain study, much focus was on eye closure instead of eye blinks.

4.3.2.6 Classification

Both SVMs and ELMs were used for the detection of pain expressions. Both algorithms are quite fast and robust in terms of classification. A brief description on how they were implemented is given below. Subsections 4.3.2.7 to 4.3.2.9 provide finer details on the SVM and ELM algorithm, as well as the classification procedure that was followed, respectively.

4.3.2.7 Support vector machines

In this study, an SVM package called LibSVM[99] was used. Support vector machines are very good with extreme classification, hence the motivation to use them as a baseline architecture. They seek to find the global minimum and have a good generalizing capability[90]. An SVM with a nonlinear kernel called the radial basis function (RBF) was used for classification. Selection of optimum values for C and γ was done using the grid search. This selects the optimum parameters for the kernel through cross validation and these parameters are only based on the training set. 110 iterations were performed per fold while seeking for the best C and γ. The average training accuracy rate obtained on training dataset was 100%.

4.3.2.8 Extreme learning machines

In this experiment, an ELM made up of 132 input neurons (66×2 features), 20 hidden neurons (optimized from experiments), a Sigmoid activation function and 2 output neurons (pain or no pain classification) was used. For the determination of optimum number of hidden neurons a range of 1 to 50 neurons was considered. Training and testing was done at an interval of 5 neurons with the best performance being achieved at 20 neurons.

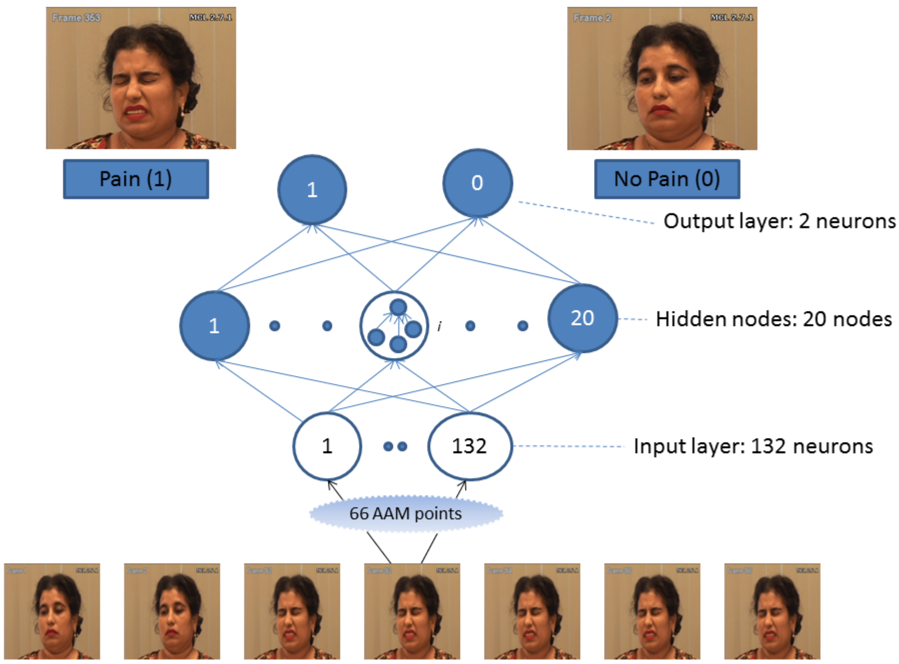


FIGURE 4.5: Representation of how the ELM model was implemented.

4.3.2.9 Classification procedure

A leave-one-out subject independent test was carried out and an average recognition accuracy rate was calculated. This study was done at both frame level and sequence level. A total of 461 frames (both positive and negative) were used in the frame based experiment, and in the sequence based experiment, 8 video sequences were used that contained a total of 2809 frames. Video sequences contained both pain and no pain data with varying PSPI scores ranging from 0 to 15. The labels that were used in both SVM and ELM experiments were 1 and 0 for pain and no pain, respectively. Same number of frames and subjects were used to test both models. The PSPI scores given in the database were used as the ground truth. Results are tabulated and graphically presented in the next chapter.

4.4 Evaluation parameters

This section gives different evaluation parameters that were used in both studies. These came in handy in assessing the performance of classifiers.
Confusion matrix, Accuracy of a classifier, Precision, Recall, $F1score$ and Sample variance are described in Subsections 4.4.1 to 4.4.6.

4.4.1 Confusion matrix

A confusion matrix is a tabular layout that contains information about the actual and predicted classifications done by a classification algorithm. It describes the performance of a model on a test data set where actual true values are already known. Figure 4.6 shows a general diagram of a confusion matrix where:

		Predicted	
		Negative	Positive
Actual	Negative	a	b
	Positive	c	d

FIGURE 4.6: Confusion matrix

- a is the number of correct predictions of a negative sample also known as true negative (TN).
- b is the number of incorrect predictions of a positive sample also known as false positive (FP).
- c is the number of incorrect predictions of a negative sample also known as false negative (FN).
- d is the number of correct predictions of a positive sample also known as true positive (TP).

4.4.2 Accuracy

From the confusion matrix, the accuracy of a classifier can be determined. It is a metric that evaluates the overall effectiveness of a classifier[101]. In binary classification, it measures the proportion of true results relative to the total number of samples examined. This measure is usually given as a percentage. The greater the accuracy the better the performance of the algorithm under evaluation. Accuracy (AC) is calculated as follows:

$$AC = \frac{a+d}{a+b+c+d} \tag{4.1}$$

4.4.3 Precision

Precision is the proportion of the predicted positive cases that were correct. Precision (P) is calculated as follows:

$$P = \frac{d}{b+d} \tag{4.2}$$

4.4.4 Recall

Recall can be defined as the proportion of positive cases that were correctly identified. It can be called the true positive rate. Recall (R) is calculated as follows:

$$R = \frac{d}{c+d} \tag{4.3}$$

4.4.5 F_1score

The F_1 score value can be described as the ratio between the geometric mean and the arithmetic mean of precision and recall. It is considered a good measure of how reliable a model performs. The greater the F_1 score the better the performance of the model[73]. $F_1 score$ is given as:

$$F_1 score = 2 * \frac{precision * recall}{precision + recall} \tag{4.4}$$

4.4.6 Sample variance

Sample variance is an unbiased estimator of the distribution variance. It helps in assessing the quality of a classifier. The lower the variance, the better the quality of the classifier. Sample variance (σ^2) is given as:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N} \tag{4.5}$$

where μ is the sample mean and N is the number of samples.

4.5 Conclusion

This chapter discussed the research questions along with the objectives of the study. It further gave a detailed sequence of steps that were followed in the experimentation process. Two main studies were carried out, one for smile detection and temporal mapping and the other one for pain detection. Both experiments were done at frame and sequence level. In the next chapter, results of the study are presented and a detailed discussion of those findings is provided.

Chapter 5

Results, analysis and discussion

5.1 Introduction

Overall results of this study are presented in this chapter along with their analysis. A concise discussion of the findings is also presented. Findings of the smile study are presented in section 5.2 while findings of the pain studies are presented in section 5.3.

5.2 Smile study

The blink and smile study was done at frame level as well as at sequence level. Results are presented separately per database. Section 5.2.1 gives the research questions and associated objectives once again just before they are answered. In subsection 5.2.2, the study seeks to answer all the three research questions and fulfill all the objectives of the experiments that were conducted on the CK+ dataset. In subsection 5.2.3, the same process is done for the DISFA dataset.

5.2.1 Research sub-questions

1. Can an active appearance model (AAM) extract all the relevant features for smile expression and eye blink recognition?
2. Can support vector machines effectively classify eye blinks and smile expressions?
3. Can a temporal mapping between eye blinks and smiles be found?
4. Can the extracted temporal relationship between eye blinks and smiles help in understanding the difference between posed and spontaneous smiles?

FIGURE 5.1: Neutral and peak images of subject 52 in the CK+ [8] database. Label -1 is assigned to the first image and +1 to the last image which depicts the peak of a smile expression.

5.2.1.1 Objectives

1. To build an AAM-SVM model that uses AAM points to recognize eye blinks and smile expressions.
2. To build a framework, based on the temporal relationship between eye blinks and smiles, for understanding smiles.

5.2.2 CK+ database experiment results

The initial experiment's main objective was to build an AAM-SVM model that recognizes smiles and blinks. This experiment was to answer question 1 and 2 in one go. Classification was done separately for eye blinks (open eye and closed eye) and smiles. Results tabulated in Table 5.1 and Table 5.2 gives the recognition rate and F_1 scores achieved in classifying each of the expressions rendered in frames. Results are given in terms of T.P (true positive), F.P (false positive), T.N (true negative) and F.N (false negative).

In Table 5.1, eye action represents the recognition rate of frames where eyes were closed or open. Table 5.2 shows the blink sequence classification score. A label sequence that followed a $(+1, -1, +1)$ pattern was classified as a blink. On the other hand, smile sequence expression recognition rate was based on a label sequence of $(-1, ..., +1)$ owing to the structure of the data in the CK+ database (Figure 5.1) i.e start with a neutral face and end with the peak of a smiling face.

TABLE 5.1: F_1 score values for eye and smile action units per frame

Expression	No. of frames	T.P	F.P	T.N	F.N	Accuracy	F_1 score
Eye action	30	14	1	14	1	93.33%	0.93
Smile	70	34	1	35	0	98.60%	0.99

TABLE 5.2: F_1 score values for blink and smile expressions per video sequence

Expression	No. of seq	T.P	F.P	T.N	F.N	Accuracy	F_1 score
Blink sequence	15	10	0	0	5	66.67%	0.80
Smile Sequence	35	34	0	0	1	97.14%	0.99

5.2.2.1 Discussion of results

From the results, it can be noted that classification accuracy deteriorated when evaluating sequences as compared to frames for eye actions. This was due to the fact that a subject would blink (+1,-1) in a manner in which their eyes remained closed for the remained of the video. Such blink trends, that would not follow the expected (+1,-1,...,-1,+1) sequence, were regarded as wrong classifications. However, the classification rate for both expressions was acceptable. By virtue of the AAM-SVM model being able to recognize both smile and blink expressions, the first two questions of this study were successfully answered and the first objective perfectly met.

Comparing the results of this study to findings in the literature review. It can be noted that the AAM-SVM combination presented in this study performed well. A comparative analysis is presented firstly for experiments that were done using the CK+ database then secondly with other datasets. Hsu et al.[52] used distance based features and SVM to check for the 7 basic expressions. They obtained an average accuracy rate of 87.7% but for the happy expression which is mostly characterized by smiling faces, they obtained 95.6%. The AAM-SVM combination in this experiment performed better than theirs. This might have been because they used a codebook which calculated the similarity of an input image to a training image. This process might have clustered input images into wrong clusters and hindered correct recognition. Hsu et al. does not explain the process of optimizing the SVM performance which might have also contributed to a lower recognition rate. Suk and Prabhakaran[55] designed a mobile application to recognize 6 basic expressions. In their study, they used Haar features and ASM then SVMs for classification. They got an overall accuracy rate of 86% and for the happy expression they achieved an accuracy rate of 98.6% which is similar to our recognition rate. This is due to the fact that they had a much similar approach to the one that was implemented in this study.

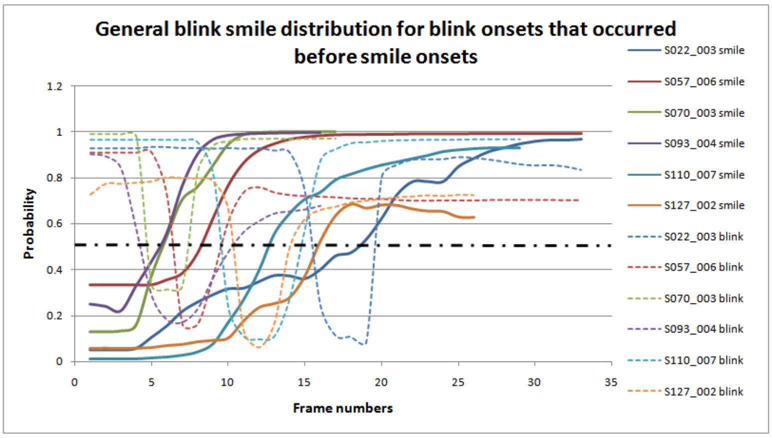
The following references used other data sets though they all focused on smiling expressions. Wang et al.[53] designed an AAM, Gabor wavelet transform and SVM model to try and recognize basic facial expressions from the JAFFE database. Their system achieved an average accuracy rate of 91.4% while the happy expression achieved an average of 66.66%. This method was not as good as the model presented in the current study. Gunadi et al.[75] designed a system to detect fake smiles. They used LOG and thresholding for edge detection. Then used a linear SVM for classification and achieved a recognition rate of 86%. This method was also outperformed by the method presented in this study. Recently, Bilinski et al.[56] tried to determine the gender of a person based on their smile. In their study, they used spatio-temporal features based on dense trajectories encoded by Fisher vectors. For classification, they used a linear SVM. Their system had a highest accuracy rate of 91% which was not as good as the model presented in this study. Overall, the AAM-SVM model presented in this study achieved an average accuracy rate of 98.60% and 97.14% at frame and sequence level, respectively. This was due to the fact that much descriptive feature points that covered the whole face were used and the SVM classifier that was used was highly optimized using the grid search that found the appropriate C and γ values.

5.2.2.2 Relationship between eye blinks and smiles

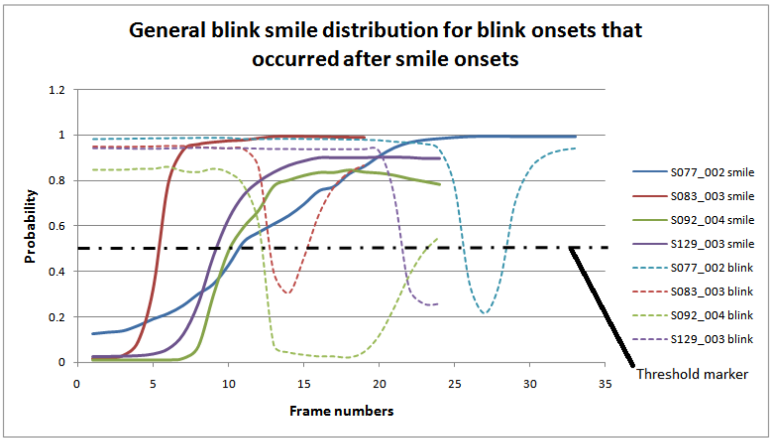
In order to answer question 3 and 4, the created AAM-SVM models were used to recognize smiles and blinks in sequences that contained both expressions. The main objective of this experiment was to find the temporal relationship between eye blinks and smiles. Continuing from the previous initial experiment, the created SVM based model was tested on all the remaining videos in the database and 12 video files were found to contain both blink and smile expressions. Out of the 12, half contained blinks that occurred just before the smile onset. The other 4 videos had blink onsets towards the peak of the smile. The last 2 had blink onsets occurring at the same time with smile onsets. To have a vivid picture of the results, the correlation coefficients between onset times of both expressions were calculated and tabulated in Table 5.3. Data was grouped into 3 sets sequentially as follows:

- blinks whose onset occurred *before* the smile onset.
- blinks whose onset occurred at the *peak intensity of a smile*.
- blinks whose onset occurred *simultaneously* with the smile onset.

Table 5.3 below gives the findings of the correlation coefficients of the tests carried out for the “before” smile onset, “peak intensity of a smile” and “simultaneously” states.



(A)



(B)

FIGURE 5.2: Shows the general relationship between a blink and a smile i.e before and after smile onset. Figure 5.2a is for blinks that occurred before smile onset and Figure 5.2b is for blinks that occurred after smile onset. The dashed and dotted thick black line is the 0.5 probability threshold marker. The lines in the graphs are labeled using a subject number followed by a sequence number as done in the actual database.

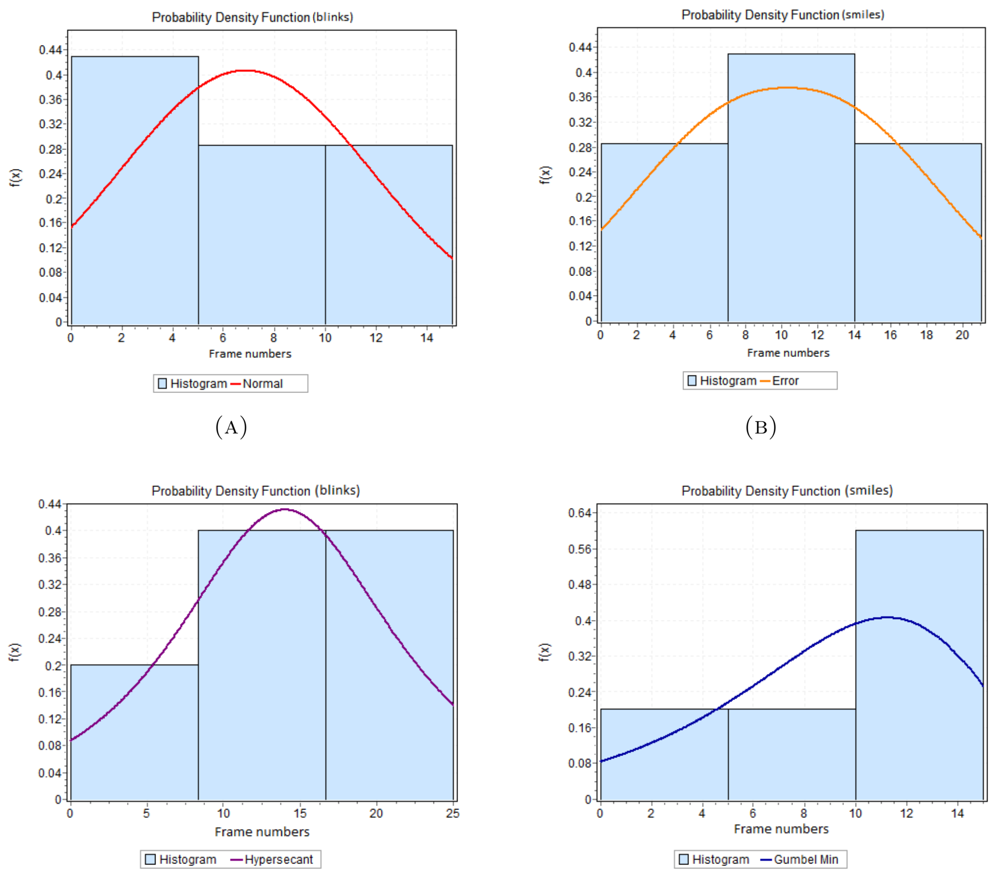
TABLE 5.3: Correlation coefficients at 95% confidence

State	No. of seq	μ_B	μ_S	$p-value$	corr (r)
Before smile onset	6	8	12	9.184×10^{-5}	0.9921651
Peak intensity of a smile	4	17.50	10.75	0.1045	0.8954859
Simultaneously	2	5	5	–	1

Where μ_B and μ_S are the average blink and smile onset frames, respectively.

5.2.2.3 Discussion of results

As given in Table 5.3 it can be noted that there is a high correlation between blinks that occur before a smile onset. From the data tested upon, it took an average of 8 frames (0.27 sec at 30 frames per second) for a blink to activate and about 12 frames (0.4 sec at 30 frames per second) for a smile to activate. There was also a strong correlation for



(A) (B) (C) (D)

FIGURE 5.3: Shows the expected probabilities of occurrence for blink and smile onset respectively, in relation to frame numbers. where $f(x)$ is the probability density function. Figure 5.3a and Figure 5.3b are distributions that occur before smile onset while Figure 5.3c and Figure 5.3d are for after smile onset.

blink onsets that occur at the peak intensity of a smile though it was not as strong as the correlation of blinks that occurred slightly before smile onset. It took an average of 17.50 frames (0.58 sec) for a blink to activate while it took about 10.75 frames (0.35 sec) for a smile to activate. For blink onsets that occurred concurrently with smile onsets, the correlation coefficient $r = 1$ but the $p - value$ was not deduced since only two pairs of samples were under consideration hence no concrete notions were arrived at from that scenario. For blink onsets that occurred before smile onsets, we can conclude that, since the p-value is significant, thus < 0.05 there is a strong relationship between blinks and smile at 95% level of confidence. This is also attributed to by the high correlation coefficient(r) of 0.99.

For blink onsets that occurred at the peak intensity of a smile, there was a positive correlation though the $p - value$ was not significant ($p = 0.1045$). The general blink-smile distribution noted after experiments is given in Figure 5.2 where smile sequences and blink sequences are given for both before and after smile onset. It can be noted from Figure 5.2b that blink onsets occurred at the peak of a smile if they occur after the

smile onset. This gives a clear picture of the general relationship that was found from the study of these two expressions. The study went on to look at the probabilities of having either a blink or a smile onset for different scenarios or groups once mentioned above which are *'before smile onset'* and *'peak intensity of a smile'*. For the *'before smile onset'*, it was noted that there was a high probability of having a blink onset between frame 4 (0.13 sec) and 10 (0.33 sec) and a smile onset between frame 7 (0.23 sec) and 15 (0.50 sec). This notion is attributed to by the distributions given in Figure 5.3a and Figure 5.3b. Taking a look at blinks that occurred at *'peak intensity of a smile'*, it was noted that there was a high probability of having a blink onset between frame 10 (0.30 sec) and 18 (0.6 sec) and also having a smile onset between frames 8 (0.27 sec) and 14 (0.47 sec) as shown in Figure 5.3c and Figure 5.3d.
Based on the findings and analysis given above it can be concluded that questions 3 and 4 were successfully answered with the 2nd and final objective being achieved. Whereby, a temporal relationship has been found between blinks and posed smiles. However, samples that were used to arrive at this notion were a bit limited. Given a much larger data sample, different results may be arrived at.

5.2.3 DISFA database experiment results

For the DISFA dataset experiments, the first objective was to create an AAM-SVM model that recognizes both smiles and blinks. This was done in this first and initial part of the experiment so as to answer the 1st and 2nd questions. However, the study went on to incorporate ELMs for classification only to observe how they would perform on spontaneous smiles. These were not tested on the CK+ database since SVMs were performing very well on posed smiles. Classification results obtained while using SVMs are presented below in Table 5.4. The created SVM model for recognizing eye opener and eye closure achieved a high recognition rate of 99.40% and an F_1 score of 0.994. For smiles, it also achieved a high recognition rate of 92% and an F_1 score of 0.940 at frame level. The recognition rate however deteriorated at recognizing smiles at sequence level to an average of 65.06%. This was mainly due to the fact that subjects exhibited different spontaneous actions which at times were out of plane hence hindering recognition rate. The same data used in the SVM experiment was tested on an ELM and the model achieved an average recognition rate of 99.90% for the eye opener and eye closure recognition. For smiles, it achieved an excellent recognition rate of 100% at frame level and 99.02% at sequence level (Figure 5.5).

TABLE 5.4: F_1 score values for eye actions and smiles obtained using SVM

Expression	No. of frames	T.P	F.P	T.N	F.N	Accuracy	F_1 score
Eye opener and closure	1000	500	6	494	0	99.40%	0.994
Smile at frame level	500	250	30	220	0	92.00%	0.940
Smile at sequence level	121 114	11 558	39 947	67242	2367	65.06%	0.353

TABLE 5.5: F_1 score values for eye actions and smiles obtained using ELM

Expression	No. of frames	T.P	F.P	T.N	F.N	Accuracy	F_1 score
Eye opener and closure	1000	500	1	499	0	99.99%	0.999
Smile at frame level	500	250	0	250	0	100%	1
Smile at sequence level	121 114	13 042	294	106 895	883	99.02%	0.984

5.2.3.1 Discussion of results

Based on the results obtained by the two models, it can be seen that ELMs were better classifiers than SVMs. Taking a look at one of the studies where ELMs were used in classifying smiles, it can be seen that this model performed impressively well. Cui et al.[61] designed a system to recognize smiles in real life applications. However, they used the GENKI-4K [62] database to test their system. They used pair-wise distance vectors as their features and passed them onto an ELM model for classification. The model achieved an average recognition rate of $93.42 \pm 1.46\%$. Their model was outperformed by the ELM model presented in this study. This might have been caused mainly by the input features that were used in this study. AAM features that were used in this study were more descriptive of face changes at every instance.

After recognizing blinks and smiles using trained models, surrogate data was introduced as was done by Trutoiu et al.[14]. Surrogate data is randomly generated data, this data was generated using the actual data. In this instance, blink data was regenerated. This data was based on the inter-eye blink distances, that is, the distance between consecutive eye blinks. This was done so as to decorrelate the blinks from the smiles (Figure 5.4). The figure shows the original smile sequence and associated inter-eye blink intervals (IBI) on top. That same sequence is then randomly reshuffled to produce sequences shown on the lower part of the diagram. In the CK+ database, the regeneration process could not be performed since a single blink event existed per subject in most video sequences. As a result, this process was highly exclusive to the DISFA database. For each subject, 1000 samples of surrogate time series were created and the average distances of blinks immediately before smile onset, immediately after smile onset, immediately before smile offset and immediately after smile offset were noted. This process was done three times per subject and average distances were collected. These calculated distances became the expected values depicting where blinks were to occur naturally. In the next paragraph, results of tests between the expected value based on the surrogate date is compared to the measured values. The measured values are actual blink positions in the database.

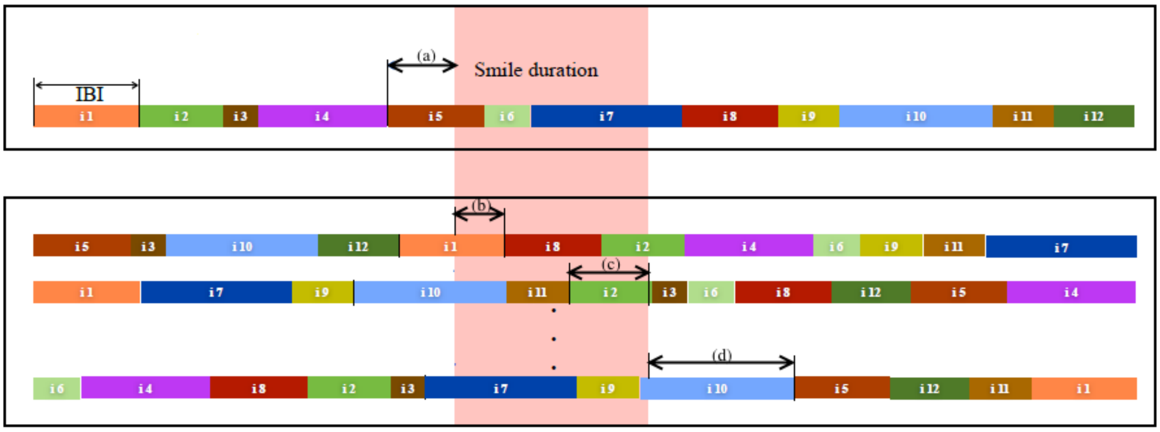


FIGURE 5.4: Shows all considered inter-eye blink intervals (IBI). All values in colored boxes with an i prefix are labels for inter-eye blink intervals between two consecutive blinks. Where (a) is the distance of the blink immediately before smile onset, (b) is the distance of the blink immediately after smile onset, (c) is the distance of the blink immediately before smile offset and finally (d) is the distance of the blink immediately after smile offset. This process was done as to how Trutoiu et al.[14] did it in their study.

Four two-tailed paired t-tests were conducted at 95% level of confidence. Blinks that occurred immediately before smile onset were measured to be 49.14 frames while the expected value was 53.43 frames away from the smile onset. The difference was not significant backed by a high p- value of 0.7230. For blinks that occurred immediately after smile onset, the measured value was 21.80 frames while the expected value was 23.06 (p = 0.9391) frames away from the smile onset. Taking a look at blinks occurring immediately before smile offset, the measured value was 11.29 frames while the expected value was 60.99 frames away from the smile offset. The difference was marginally significant (p = 0.0670 , Cohen's d= 0.00864428). For blinks that occurred after smile offset, the measured value was 39.25 frames while the expected value was 32.00 ($p = 0.5307$) after the smile offset.
Based on these findings, the system could recognize smiles and blinks, thereby successfully answering the 1st and 2nd question.

5.2.3.2 Relationship between blinks and spontaneous smiles

To answer the 3rd and 4th question, an analysis of the findings from the first part of the experiment was conducted. This section describes how the process of checking for the temporal relationship between blinks and smiles in the DISFA database was done.
Since there were no significant differences between the measured and expected values as proven by the t-tests, the Pearson's correlation was calculated on the measured data. For blinks that occurred before smile onset, there was an insignificant correlation ($r = 0.4493, p = 0.312$). For blinks that occurred after smile onset there was a marginal

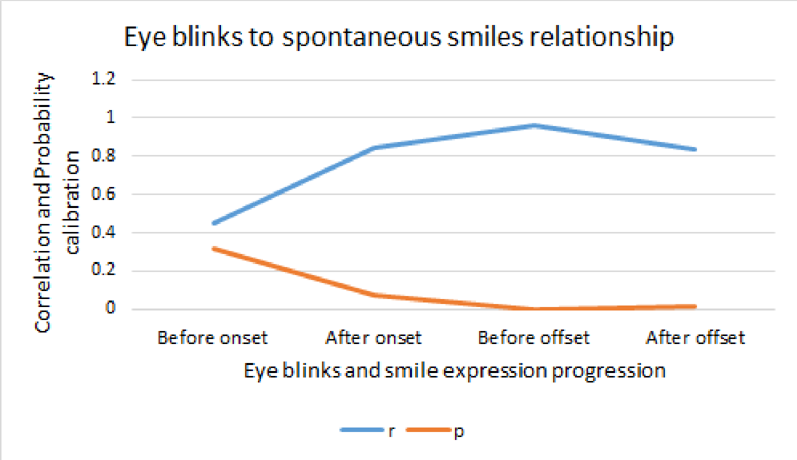


FIGURE 5.5: The temporal relationship between eye blinks and spontaneous smiles. Where "r" is the correlation coefficient and "p" is p value

correlation ($r = 0.8418, p = 0.074$). For blinks that occurred towards the smile offset, there was a strong correlation ($r = 0.9582, p = 0.0006$). Lastly, for blinks that occurred after a smile offset there was a strong correlation as well ($r = 0.8318, p = 0.010$). Based on this analysis it can be concluded that blinks are highly correlated with the end of a smile in spontaneous smiles (Figure 5.5). This is in line with the findings discovered by Trutoiu et al.[14] where they suspected that blinks demarcate the end of a smile. However, they had a fine tuned implementation that manually demarcated smiles (AU12) as compared to the one in this current study that looked at a combination of AU6 and AU12 for smiles.

5.2.4 Analysis of posed and spontaneous smiles

To fully answer question 4, an analysis that clearly shows the difference between posed and spontaneous smiles is presented. Based on the findings presented, it can be noted that there was a high correlation coefficient for blinks that occurred before smile onsets in posed smiles. However, the relationship between blinks and posed smile offsets was not studied due to the orientation of the database. Video sequences in the CK+[8] database run from a neutral, to a peak of an expression. In comparison, the findings reported in spontaneous smiles study showed that there was a high correlation between blinks and smiles for blinks that occurred immediately before and immediately after a smile offset. This clearly shows that there is a difference between in posed and spontaneous smiles based on the temporal relationship that they have with eye blinks. Therefore, the study can safely say question 4 was successfully answered as this finding helps in understanding the difference between posed and spontaneous smiles.

5.3 Pain study

This study was conducted so as to give an insight on the characteristics of spontaneous pain based on the temporal relationship between eye closure and pain expressions. The database did not contain posed pain hence studies focused on spontaneous pain.
Subsection 5.3.0.1 gives all the research questions and associated objectives to be met in the study. Section 5.3.1 gives the experimental results achieved by the study. Section 5.3.2 gives the relationship that exists between eye closure and pain expressions.

5.3.0.1 Research sub-questions

1. Can an active appearance model extract all the relevant features for pain expression and eye closure recognition?
2. Can extreme learning machines help in improving classification of pain expressions as compared to standard predominant classifiers such as support vector machines?
3. Can the temporal relationship between eye closure and spontaneous pain be determined?

5.3.0.2 Objectives

1. To build models that use AAM+SVM and AAM+ELM to recognize pain expressions.
2. To compare the performance of AAM+SVM and AAM+ELM models, and determine a better model for pain detection.
3. To find correlations between eye closure and spontaneous pain.

5.3.1 Experimental results

In this subsection, the main aim was to answer question 1 and 2 and also to fulfill objective 1 and 2. During the classification stage, both SVM and ELM techniques were implemented and their recognition results are given in Table 5.6 and Table 5.7. These results are divided into two categories which are frame level and sequence level.

As clearly depicted in Table 5.6 it can be noted that the ELM outperformed the SVM at frame level classification. The ELM had an average accuracy rate of 95.96% while the SVM had 79.78%. Taking a closer look at individual results, it can be noted that the

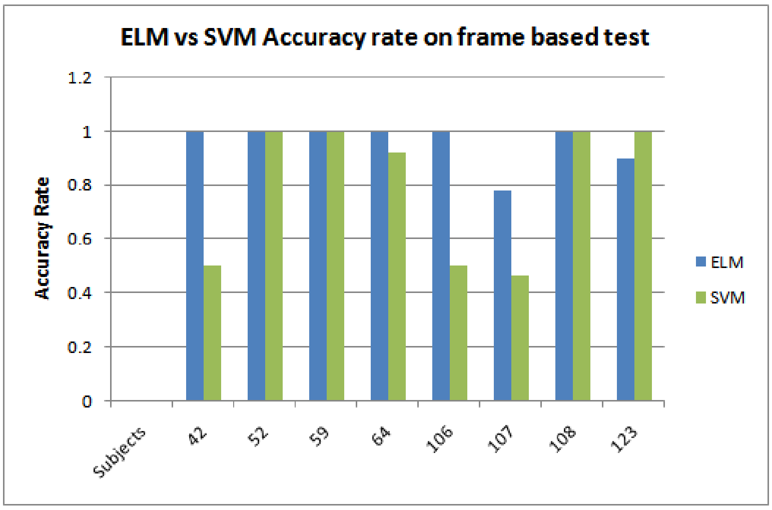
TABLE 5.6: Accuracy and F_1 score values for ELM and SVM per frame

Subject	No of frames per subject	ELM		SVM	
		Accuracy	$F_1 score$	Accuracy	$F_1 score$
42	168	100%	1.00	50%	0.67
52	468	100%	1.00	100%	1.00
59	368	100%	1.00	100%	1.00
64	397	100%	1.00	92%	0.91
106	352	100%	1.00	50%	0.00
107	412	78%	0.74	46%	0.00
108	285	100%	1.00	100%	1.00
123	359	90%	0.89	100%	1.00
Average	-	95.96%	95.44%	79.78%	69.76%
Sample variance	-	0.007	-	0.067	-

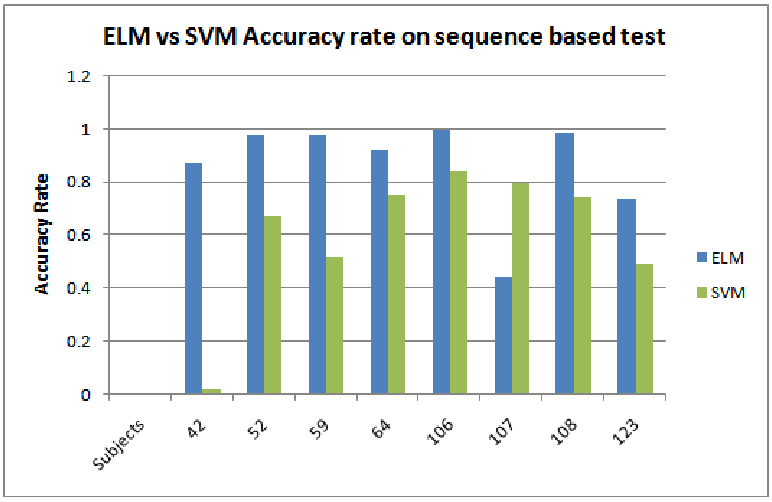
TABLE 5.7: Accuracy and F_1 score values for ELM and SVM per sequence

Subject	No of frames per subject	ELM		SVM	
		Accuracy	$F_1 score$	Accuracy	$F_1 score$
42	168	87.5%	0.222	1.80%	0.035
52	468	97.9%	0.783	67.1%	0.189
59	368	97.8%	0.840	51.6%	0.191
64	397	92.2%	0.710	75.1%	0.393
106	352	99.7%	0.976	83.8%	0.00
107	412	44.4%	0.224	79.4%	0.00
108	285	98.6%	0.964	74.4%	0.597
123	359	73.4%	0.323	49.3%	0.222
Average	-	86.44%	63.02%	60.30%	20.34%
Sample variance	-	0.036	-	0.071	-

ELM had higher accuracy rates in 4 instances over SVMs and had the same accuracy rates in 3 instances and suffered a loss to the SVM in 1 instance only (Figure 5.6a).
Referring to Table 5.7, which depicts sequence based classification results, it can be noted that the average performance of both algorithms deteriorated when compared to frame level classification. This was mainly due to the coming in of new frames that contained lower PSPI scores, that is, between 1 and 7. Since the models were created to observe PSPI scores ≥ 8 as positive, most misclassifications happened on frames whose PSPI scores were between 1 and 7. Head pose was also not taken into consideration in the creation of the models, hence any change in the orientation of the head might have led to an adverse effect on the classification results. The ELM accuracy rate deteriorated by 9.52% to about 86.44% to an F_1 score of 0.63. The SVM average accuracy rate also deteriorated by 19.48% to 60.30%. However, looking at these results, it can be noted that the ELM had higher accuracy rates in 7 instances over the SVM. The SVM had a higher accuracy rate over the ELM in a single instance only (Figure 5.6b). Generally, ELMs outperformed SVMs in pain recognition at both frame and sequence level test. ELMs had a low accuracy rate on subject 107 in sequence based test maybe due to



(A)



(B)

FIGURE 5.6: Shows the comparison between ELM and SVM in pain recognition. Figure 5.6a: shows the rate at frame based level while Figure 5.6b: shows at sequence based level.

the orientation of the data. The subject went through high pain for 3 instances in the same video file. However, the ELM did not manage correctly classify all those instances. In sequence based classification, subject 42 achieved a low accuracy rate with SVMs because they recognized the whole sequence as though the subject was going through pain.

The sample variance which helps in assessing the quality of a classifier was also calculated. The lower the variance the better the quality of the classifier[2]. Based on the calculated accuracy values in Table 5.6, the ELM based model had a sample variance of 0.007 while the SVM based model had a sample variance of 0.067. From the accuracy rates given in Table 5.7, the ELM based model had a sample variance of 0.036 while the SVM based model had a sample variance of 0.071. This clearly shows that ELMs were better classifiers in this study than SVMs. From observation of sequence based

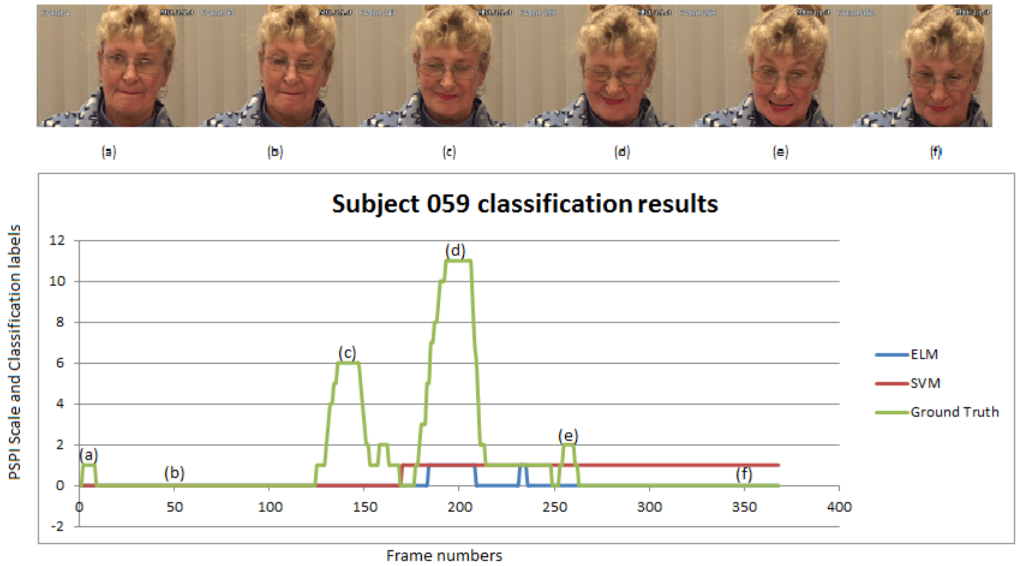


FIGURE 5.7: Sample of how classification occurred for subject 059 using ELM and SVM. NB letters a, b, c, d, e and f corresponds to frames 4, 50, 141, 198, 258 and 350, respectively. The ground truth is given according to the PSPI scale while values for ELM and SVM are classification labels 0 or 1.

classification results, it was noted that ELMs were good at noticing transitions between different PSPI scores as compared to SVMs (Figure 5.7).

Taking a look at results presented in the literature review, it can be seen that ELMs outperformed most algorithms that have been used in the past. A comparative analysis is given firstly for experiments that were done using the UNBC-McMaster database then secondly for the ones using other datasets. Ashraf et al.[78] used AAMs and SVMs to check for extreme pain. They achieved a hit rate of 81% and 77% at frame and sequence level test, respectively. However, the ELM model in the current study had better results than theirs in both tests. Lucey et al.[81] tried using SPTS, SAPP and CAPP features separately for pain detection and their results were not as good as when all features were combined. Results obtained by CAPP were satisfactory, scoring slightly above 80%. A combination of SPTS+SAPP+CAPP features yielded a better rate of 84.7%, which was also obtained through frame by frame evaluation. However, their results did not manage to surpass those attained by the ELM model presented in this study. Khan et al.[83] used a combination of PHOG and PLBP features in their pain detection experiment. In this experiment, they tested performance on four classifiers which are SVM, 2NN, decision tree and random forest. The 2NN had an average accuracy rate of 96.6% followed by the SVM that had an average accuracy rate of 94.4% for all the three levels of testing designed. This might have been due to the feature extraction methods that were implemented.

The following references used different datasets though they were all working on pain recognition. Littlewort et al.[79] tried to differentiate real vs fake pain. Their computer vision system implemented a frame based analysis and scored an average accuracy rate of 88%. The ELM in the current study achieved a higher accuracy rate as compared to theirs. However, they had a real time face and feature detector that performed at 90% accuracy rate. In Bartlett et al.[28] experiment they tried to differentiate fake and real expressions. In their experiment, they tested human capability against computer vision system capability. The computer vision system using an SVM classifier achieved an accuracy rate of 85% while using temporal features. This method was also outperformed by the ELM method presented in the current study. Overall, the ELM model presented in this study achieved an average accuracy rate of 95.96% at frame level and 86.44% at sequence level, which shows that ELMs can yield much better results in this field. In conclusion, questions 1 and 2 were successfully answered in this study. Whereby, AAM feature points gave the descriptive information in terms of pain and eye closure expressions and ELMs and SVMs were compared on how they perform in classifying pain expressions. ELMs ultimately came out victorious than SVMs.

5.3.2 Relationship between eye closure and pain

To answer question 3, the study further analyzed the temporal relationship between eye closure and pain expressions.
After the pain detection experiment. Eye closure periods for the same subjects (2809 samples) used in the pain detection experiment were measured. The average eye closure and pain period was measured to be 63.1 frames and 43.5 frames, respectively. The Pearson's correlation coefficient (r) was calculated to understand the relationship between eye closure onsets and offsets to pain onsets and offsets. The main stages that were assessed were:

- Eye closure onsets that occurred before pain onset.
- Eye closure onsets that occurred immediately after pain onset.
- Eye closure offsets that occurred just before pain offset.
- Eye closure offsets that occurred immediately after pain offset.

For eye closure onsets that occurred immediately before pain onset, a high correlation coefficient was noted ($r = 0.9988, p = 4.7326 \times 10^{-5}$). A high correlation coefficient was also noted for eye closure onsets that occurred immediately after pain expression onset ($r = 0.9973, p = 1.6963 \times 10^{-4}$). A high correlation was noted for eye closure offsets that occurred just before the pain expression offset. For eye closure offsets that occurred before pain expression offset, a high correlation coefficient was also noted($r = 0.9793, p = 6.3855 \times 10^{-4}$). A high correlation was observed for eye closure offsets that occurred after pain expression offset ($r = 0.9637, p = 0.0363$). Figure 5.8 gives a graphical illustration of the relationship as pain and eye closure progresses.

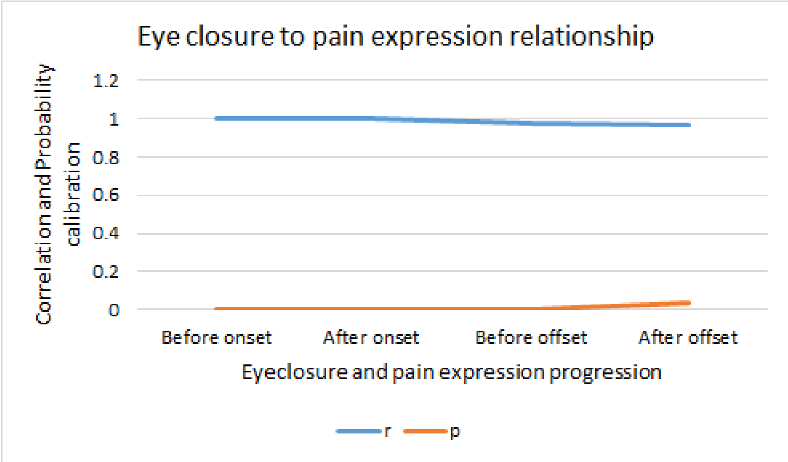


FIGURE 5.8: The general temporal relationship between eye closure and pain. Where "r" is the correlation coefficient and "p" is the p value.

Taking a look at these results, it can be noted that for all stages of assessment the eye closure onsets and offsets are highly correlated to pain expression onset and offsets, which might attribute and compliment to the reason why eye closure is highly associated with pain expressions as from studies previously done by other researchers[96][97]. Eye closure periods are however generally longer than periods of high pain as observed in the study. This analysis successfully answered question 3 by denoting the temporal relationship between eye closure and spontaneous pain.
Finally, it can be concluded that all the research sub questions were successfully answered in these experiments.

5.4 Conclusion

This chapter gave a detailed outline of the results that were obtained in both the studies. All questions and objectives were explained on how they were answered and fulfilled, respectively. In the next chapter, a brief summary of the findings is presented as well as the overall answer to the main research question. Future work that can be pursued, is also presented.

Chapter 6

Conclusion and Future work

6.1 Introduction

In this dissertation, the study of temporal patterns between eye actions and facial expressions is suggested. We have demonstrated the use of AAM feature points to denote eye blinks and smiles, these feature points were then passed on to the SVM which did the classification. AAM feature points were also used to demarcate eye closure and pain expressions. These features were sent to the SVM and ELM for classification. We looked at the temporal relationship that exists between eye blinks and smiles as well as the temporal relationship that exists between eye closure and pain expressions.
In this chapter a detailed answer to the main research question is given. The chapter also concludes the study and gives ideas of intended future work.

6.2 Main Research question

Can the temporal relationship between eye actions and an expression be deduced?

To answer this question a series of studies where conducted to try and reveal the relationship between eye actions and expressions. The overall study managed to clearly depict different characteristics that exists between posed and non posed expressions. One of the major observations was that blinks were highly correlated with the onset of smiles in posed smiles and highly correlated with the end of a smile in spontaneous smiles. The study further looked at pain expressions. In the pain study, much focus was on the temporal relationship that exists between eye closure and pain expressions. The study revealed that there is a high correlation between eye closure and spontaneous pain expressions. Tests were performed on the onsets and offsets of both eye closure and pain expressions and for all instances there was a strong positive correlation detected. This might explain why eye closure is a key expression in pain expressions and also a key AU in the PSPI scale. Furthermore, ELMs were incorporated to both spontaneous smile and pain expressions. They proved to be very good classifiers in both disciplines. A comparative analysis between the ELM model in this study and other models that were previously used was done. Nevertheless, the ELM results of this study remained outstanding.

Looking at the overall findings of this study we can safely say that the main research question was successfully answered.

6.3 Limitations

For both experiments presented, head pose and occlusions were not catered for hence any head movements out of plane hindered the recognition rate. This has an adverse effect if the system is to be implemented in real time. However, this problem can be minimised by extracting 3D landmark points, incorporating pose estimation of the face region as well as deriving an adaptive pose classification method based on the trained images[102].

6.4 Future work

In as much as the study has managed to answer successfully all the research questions, there is still room for improvement. Based on some of the limitations faced in this study, it is necessary to incorporate factors like head pose and occlusions as part of the future work. The same temporal mapping done in this study can be implemented on other facial expressions such as posed pain and it can be further extended to help in authenticating facial expressions as either real or fake. The CK+ database samples were not optimum to properly derive solid trends hence the study suggests the use of a much bigger and appropriate database for future studies.

6.5 Conclusion

The present study managed to find the temporal relationship between eye actions and expressions. Specifically, it deduced the temporal relationship that exists between eye blinks and smiles. It also revealed the temporal relationship that exists between eye closure and pain expressions. Looking at the overall findings, the study managed to answer all the research sub questions thereby fulfilling the ultimate hypothesis that says, “It is possible to deduce the temporal relationship that exists between eye actions and an expression.”

Bibliography

[1] Moses Rupenga and Hima B Vadapalli. Investigating the temporal association between eye actions and smiles. In *Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech), 2016*, pages 1–6. IEEE, 2016.

[2] Moses Rupenga and Hima B Vadapalli. Automatic spontaneous pain recognition using supervised classification learning algorithms. In *Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech), 2016*, pages 1–6. IEEE, 2016.

[3] Tim Cootes, ER Baldock, and J Graham. An introduction to active shape models. *Image Processing and Analysis*, pages 223–248, 2000.

[4] Iain Matthews and Simon Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

[5] Robert Niese, Ayoub Al-Hamadi, Faisal Aziz, and Bernd Michaelis. Robust facial expression recognition based on 3D supported feature extraction and svm classification. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–7. IEEE, 2008.

[6] Guang-Bin Huang. What are extreme learning machines? filling the gap between Frank Rosenblatts dream and John Von Neumanns puzzle. *Cognitive Computation*, 7(3):263–278, 2015.

[7] Zakia Hammal and Jeffrey F Cohn. Automatic detection of pain intensity. In *Proceedings of the 14th ACM international conference on Multimodal interaction*, pages 47–52. ACM, 2012.

[8] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain 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. IEEE, 2010.

[9] Michael Lyons, Shigeru Akamatsu, Miyuki Kamachi, and Jiro Gyoba. Coding facial expressions with gabor wavelets. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 200–205. IEEE, 1998.

[10] Rajkiran Gottumukkal and Vijayan K Asari. An improved face recognition technique based on modular pca approach. *Pattern Recognition Letters*, 25(4):429–436, 2004.

[11] Seyed Mohammad Mavadati, Mohammad H Mahoor, Kevin Bartlett, Philip Trinh, and Jeffrey F Cohn. DISFA: A spontaneous facial action intensity database. *Affective Computing, IEEE Transactions on*, 4(2):151–160, 2013.

[12] Patrick Lucey, Jeffrey F Cohn, Kenneth M Prkachin, Patricia E Solomon, and Iain Matthews. Painful data: The UNBC-Mcmaster shoulder pain expression archive database. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 57–64. IEEE, 2011.

[13] Laurenm11234. Lab 11 and 12 practical, 2015. URL http://www.proprofs.com/flashcards/story.php?title=lab-11-12-practical.

[14] Laura C Trutoiu, Jessica K Hodgins, and Jeffrey F Cohn. The temporal connection between smiles and blinks. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–6. IEEE, 2013.

[15] Paul Ekman and Wallace V Friesen. *Manual for the facial action coding system.* Consulting Psychologists Press, 1978.

[16] Gianluca Donato, Marian Stewart Bartlett, Joseph C. Hager, Paul Ekman, and Terrence J. Sejnowski. Classifying facial actions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(10):974–989, 1999.

[17] Charles Darwin. 77ie expression of the emotions in man and animals, 1872.

[18] Paul Ekman and Wallace V Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124, 1971.

[19] Malcolm Gladwell. The naked face. *The New Yorker*, 5:38–49, 2002.

[20] Karsten Wolf. Measuring facial expression of emotion. *Dialogues in clinical neuroscience*, 17(4):457, 2015.

[21] Virginio Cantoni, Dimo Dimov, and Massimo Tistarelli. *Biometric Authentication.* Springer, 2015.

[22] Mette T Posamentier and Hervé Abdi. Processing faces and facial expressions. *Neuropsychology review*, 13(3):113–143, 2003.

[23] Battista Biggio, Zahid Akhtar, Giorgio Fumera, Gian Luca Marcialis, and Fabio Roli. Security evaluation of biometric authentication systems under real spoofing attacks. *IET biometrics*, 1(1):11–24, 2012.

[24] C Evinger, MD Shaw, CK Peck, KA Manning, and R Baker. Blinking and associated eye movements in humans, guinea pigs, and rabbits. *Journal of Neurophysiology*, 52(2):323–339, 1984.

[25] Yuu Tanaka and Kiyoshi Yamaoka. Blink activity and task difficulty. *Perceptual and motor skills*, 77(1):55–66, 1993.

[26] Stephen Porter and Leanne ten Brinke. Reading between the lies identifying concealed and falsified emotions in universal facial expressions. *Psychological Science*, 19(5):508–514, 2008.

[27] Sharon Leal and Aldert Vrij. Blinking during and after lying. *Journal of Nonverbal Behavior*, 32(4):187–194, 2008.

[28] Marian Stewart Bartlett, Gwen C Littlewort, Mark G Frank, and Kang Lee. Automatic decoding of facial movements reveals deceptive pain expressions. *Current Biology*, 24(7):738–743, 2014.

[29] Paul Ekman and Wallace V Friesen. Felt, false, and miserable smiles. *Journal of nonverbal behavior*, 6(4):238–252, 1982.

[30] Bella M DePaulo, Deborah A Kashy, Susan E Kirkendol, Melissa M Wyer, and Jennifer A Epstein. Lying in everyday life. *Journal of personality and social psychology*, 70(5):979, 1996.

[31] Ali KK Bermani, Atef Z Ghalwash, and Aliaa AA Youssif. Automatic facial expression recognition based on hybrid approach. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 3(11):102–107, 2012.

[32] Motoi Suwa, Noboru Sugie, and Keisuke Fujimora. A preliminary note on pattern recognition of human emotional expression. In *International joint conference on pattern recognition*, pages 408–410, 1978.

[33] Janani Harischandra and MUS Perera. Intelligent emotion recognition system using brain signals (EEG). In *Biomedical Engineering and Sciences (IECBES), 2012 IEEE EMBS Conference on*, pages 454–459. IEEE, 2012.

[34] Lindsay Brown, Bernard Grundlehner, and Julien Penders. Towards wireless emotional valence detection from EEG. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 2188–2191. IEEE, 2011.

[35] Veronica Hollis. Reconviction analysis of interim accredited programmes software (iaps) data. *Research Development Statistics, National Offender Management Service, London*, 2007.

[36] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *Affective Computing, IEEE Transactions on*, 3(1):18–31, 2012.

[37] Nayel Al-Zubi. Detecting facial expressions from EEG signals and head movement for controlling mouse curser. In *Innovative Computing Technology (INTECH), 2013 Third International Conference on*, pages 551–554. IEEE, 2013.

[38] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *Affective Computing, IEEE Transactions on*, 3(1):42–55, 2012.

[39] Mohammad Soleymani, Sadjad Asghari-Esfeden, Maja Pantic, and Yun Fu. Continuous emotion detection using EEG signals and facial expressions. In *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, pages 1–6. IEEE, 2014.

[40] Laszlo A Jeni, Hideki Hashimoto, and Takashi Kubota. Robust facial expression recognition using near infrared cameras. *JACIII*, 16(2):341–348, 2012.

[41] T Kanade. Cohn-Kanade AU-coded facial expression database, 2000.

[42] Hamid Sadeghi, Abolghasem-A Raie, and Mohammad-Reza Mohammadi. Facial expression recognition using geometric normalization and appearance representation. In *Machine Vision and Image Processing (MVIP), 2013 8th Iranian Conference on*, pages 159–163. IEEE, 2013.

[43] Filip Malawski, Bogdan Kwolek, and Shinji Sako. Using kinect for facial expression recognition under varying poses and illumination. In *International Conference on Active Media Technology*, pages 395–406. Springer, 2014.

[44] Jayashree V Patil and Preeti Bailke. Real time facial expression recognition using realsense camera and ANN. In *Inventive Computation Technologies (ICICT), International Conference on*, volume 2, pages 1–6. IEEE, 2016.

[45] Anastasios Koutlas and Dimitrios I Fotiadis. Image processing and machine learning techniques for facial expression recognition. URL http://www.qbase.gr/sites/default/files/FINAL_WITH_NOTES.pdf.

[46] Ernst Haselsteiner and Gert Pfurtscheller. Using time-dependent neural networks for EEG classification. *Rehabilitation Engineering, IEEE Transactions on*, 8(4): 457–463, 2000.

[47] Omar AlZoubi, Rafael A Calvo, and Ronald H Stevens. Classification of EEG for affect recognition: an adaptive approach. In *Australasian Joint Conference on Artificial Intelligence*, pages 52–61. Springer, 2009.

[48] Chengalvarayan Radhakrishnamurthy Hema, Murugesa Pandiyan Paulraj, and AH Adom. Improving classification of EEG signals for a four-state brain machine interface. In *Biomedical Engineering and Sciences (IECBES), 2012 IEEE EMBS Conference on*, pages 615–620. IEEE, 2012.

[49] Li Wang, Ruifeng Li, and Ke Wang. A novel automatic facial expression recognition method based on AAM. *Journal of Computers*, 9(3):608–617, 2014.

[50] Peng Yang, Qingshan Liu, and Dimitris N Metaxas. Boosting coded dynamic features for facial action units and facial expression recognition. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–6. IEEE, 2007.

[51] Jason M Gold, Jarrett D Barker, Shawn Barr, Jennifer L Bittner, W Drew Bromfield, Nicole Chu, Roy A Goode, Doori Lee, Michael Simmons, and Aparna Srinath. The efficiency of dynamic and static facial expression recognition. *Journal of vision*, 13(5):23–23, 2013.

[52] Fu-Song Hsu, Wei-Yang Lin, and Tzu-Wei Tsai. Automatic facial expression recognition for affective computing based on bag of distances. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific*, pages 1–4. IEEE, 2013.

[53] Li Wang, Ruifeng Li, and Ke Wang. Automatic facial expression recognition using SVM based on AAMs. In *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2013 5th International Conference on*, volume 2, pages 330–333. IEEE, 2013.

[54] Michael J Lyons, Shigeru Akamatsu, Miyuki Kamachi, Jiro Gyoba, and Julien Budynek. The Japanese female facial expression (JAFFE) database. 1998.

[55] Myunghoon Suk and Balakrishnan Prabhakaran. Real-time mobile facial expression recognition system-a case study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 132–137, 2014.

[56] Piotr Bilinski, Antitza Dantcheva, and François Brémond. Can a smile reveal your gender? In *Biometrics Special Interest Group (BIOSIG), 2016 International Conference of the*, pages 1–6. IEEE, 2016.

[57] F Iglesias, P Negri, ME Buemi, D Acevedo, and M Mejail. Facial expression recognition: a comparison between static and dynamic approaches. In *Pattern Recognition Systems (ICPRS-16), International Conference on*, pages 1–6. IET, 2016.

[58] S Sanjula and C Gowrishankar. A novel approach for recognizing facial structures using extreme learning machine algorithm. *IJIRSET*, 3:218–225, 2014.

[59] Deepak Ghimire and Joonwhoan Lee. Extreme learning machine ensemble using bagging for facial expression recognition. *JIPS*, 10(3):443–458, 2014.

[60] Zhen-Tao Liu, Gui-Tian Sui, Dan-Yun Li, and Guan-Zheng Tan. A novel facial expression recognition method based on extreme learning machine. In *Control Conference (CCC), 2015 34th Chinese*, pages 3852–3857. IEEE, 2015.

[61] Dongshun Cui, Guang-Bin Huang, and Tianchi Liu. Smile detection using pairwise distance vector and extreme learning machine. In *Neural Networks (IJCNN), 2016 International Joint Conference on*, pages 2298–2305. IEEE, 2016.

[62] MPLAB. The mplab genki database, genki-4k subset. URL http://www.mplab.ucsd.edu.

[63] Sadaf Afshar and Albert Ali Salah. Facial expression recognition in the wild using improved dense trajectories and fisher vector encoding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 66–74, 2016.

[64] Lawrence Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[65] Nisrine Jrad and Marco Congedo. Identification of spatial and temporal features of eeg. *Neurocomputing*, 90:66–71, 2012.

[66] Rui Zhao, Gerwin Schalk, and Qiang Ji. Coupled hidden Markov model for electrocorticographic signal classification. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 1858–1862. IEEE, 2014.

[67] Fabien Lotte, Marco Congedo, Anatole Lécuyer, and Fabrice Lamarche. A review of classification algorithms for EGG-based brain–computer interfaces. *Journal of neural engineering*, 4, 2007.

[68] Zara Ambadar, Jonathan W Schooler, and Jeffrey F Cohn. Deciphering the enigmatic face the importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, 16(5):403–410, 2005.

[69] HB Vadapalli, H Nyongesa, and CWP Omlin. Recurrent neural networks for facial action unit recognition from image sequences. In *Proceedings of the 2009 International Conference on Image Processing, Computer Vision, & Pattern Recognition, IPCV 2009, July 13–16, 2009, Las Vegas, Nevada, USA*, volume 2, pages 351–367, 2010.

[70] Hongying Meng, Bernardino Romera-Paredes, and Nadia Bianchi-Berthouze. Emotion recognition by two view SVM_2k classifier on dynamic facial expression features. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 854–859. IEEE, 2011.

[71] Sonali V Hedaoo, MD Katkar, and SP Khandait. Feature tracking and expression recognition of face using dynamic bayesian network. *Int. J. Eng. Trends Technol.(IJETT)*, 8(10):517–521, 2014.

[72] Hong Liu and Pingping Wu. Comparison of methods for smile deceit detection by training AU6 and AU12 simultaneously. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 1805–1808. IEEE, 2012.

[73] Mohammed Ehsan Hoque, Daniel J McDuff, and Rosalind W Picard. Exploring temporal patterns in classifying frustrated and delighted smiles. *Affective Computing, IEEE Transactions on*, 3(3):323–334, 2012.

[74] Takeo Kanade, Jeffrey 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. IEEE, 2000.

[75] I Gede Aris Gunadi, Agus Harjoko, Retantyo Wardoyo, and Neila Ramdhani. Fake smile detection using linear support vector machine. In *Data and Software Engineering (ICoDSE), 2015 International Conference on*, pages 103–107. IEEE, 2015.

[76] Marilyn L Hill and Kenneth D Craig. Detecting deception in pain expressions: the structure of genuine and deceptive facial displays. *Pain*, 98(1):135–144, 2002.

[77] Anne-Claire Larochette, Christine T Chambers, and Kenneth D Craig. Genuine, suppressed and faked facial expressions of pain in children. *Pain*, 126(1):64–71, 2006.

[78] Ahmed Bilal Ashraf, Simon Lucey, Jeffrey F Cohn, Tsuhan Chen, Zara Ambadar, Kenneth M Prkachin, and Patricia E Solomon. The painful face–pain expression recognition using active appearance models. *Image and vision computing*, 27(12): 1788–1796, 2009.

[79] Gwen C Littlewort, Marian Stewart Bartlett, and Kang Lee. Automatic coding of facial expressions displayed during posed and genuine pain. *Image and Vision Computing*, 27(12):1797–1803, 2009.

[80] Maja Pantic, Michel Valstar, Ron Rademaker, and Ludo Maat. Web-based database for facial expression analysis. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 5–pp. IEEE, 2005.

[81] Patrick Lucey, Jeffrey F Cohn, Iain Matthews, Simon Lucey, Sridha Sridharan, Jessica Howlett, and Kenneth M Prkachin. Automatically detecting pain in video through facial action units. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 41(3):664–674, 2011.

[82] Katelynn E Boerner, Christine T Chambers, Kenneth D Craig, Rebecca R Pillai Riddell, and Jennifer A Parker. Caregiver accuracy in detecting deception in facial expressions of pain in children. *Pain*, 154(4):525–533, 2013.

[83] Riaz A Khan, Alexandre Meyer, Hubert Konik, and Saida Bouakaz. Pain detection through shape and appearance features. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*, pages 1–6. IEEE, 2013.

[84] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.

[85] Andreas Lanitis, Christopher J. Taylor, and Timothy F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and machine intelligence*, 19(7):743–756, 1997.

[86] Gareth J Edwards, Christopher J Taylor, and Timothy F Cootes. Learning to identify and track faces in image sequences. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 260–265. IEEE, 1998.

[87] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor. Active appearance models. In *European conference on computer vision*, pages 484–498. Springer, 1998.

[88] Timothy F Cootes, Gareth J Edwards, Christopher J Taylor, et al. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23 (6):681–685, 2001.

[89] Vladimir N Vapnik. An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5):988–999, 1999.

[90] R Ramanathan, KP Soman, Arun S Nair, V Vidhya Sagar, and N Sriram. A support vector machines approach for efficient facial expression recognition. In *2009 International Conference on Advances in Recent Technologies in Communication and Computing*, pages 850–854. IEEE, 2009.

[91] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University, 2003. URL http://www.csie.ntu.edu.tw/~cjlin/papers.html.

[92] Guang-Bin Huang. An insight into extreme learning machines: random neurons, random features and kernels. *Cognitive Computation*, 6(3):376–390, 2014.

[93] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: theory and applications. *Neurocomputing*, 70(1):489–501, 2006.

[94] E Friesen and P Ekman. Facial action coding system: a technique for the measurement of facial movement. *Palo Alto*, 1978.

[95] Jeffrey F Cohn, Zara Ambadar, and Paul Ekman. Observer-based measurement of facial expression with the facial action coding system. *The handbook of emotion elicitation and assessment*, pages 203–221, 2007.

[96] Kenneth M Prkachin. The consistency of facial expressions of pain: a comparison across modalities. *Pain*, 51(3):297–306, 1992.

[97] Kenneth M Prkachin and Patricia E Solomon. The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. *Pain*, 139 (2):267–274, 2008.

[98] Peter N Belhumeur, Joao P Hespanha, and David J Kriegman. Recognition using class specific linear projection, 1997.

[99] Chih-Chung Chang and Chih-Jen Lin. LibSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3): 27, 2011.

[100] Pierre Dupoint. The blessing of dimensionality for kernel methods. 2006.

[101] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, 2009.

[102] Shireesha Chintalapati and MV Raghunadh. Illumination, expression and occlusion invariant pose-adaptive face recognition system for real-time applications. *arXiv preprint arXiv:1403.1362*, 2014.