Data-Driven Facial Expression Analysis from Live Video

by

Wee Kiat, Tay

A thesis
submitted to the Victoria University of Wellington
in partial fulfilment of the
requirements for the degree of
Master of Science
In Computer Graphics.
Victoria University of Wellington
2017

Abstract

Emotion analytics is the study of human behavior by analyzing the responses when humans experience different emotions. In this thesis, we research into emotion analytics solutions using computer vision to detect emotions from facial expressions automatically using live video.

Considering anxiety is an emotion that can lead to more serious conditions like anxiety disorders and depression, we propose 2 hypotheses to detect anxiety from facial expressions. One hypothesis is that the complex emotion "anxiety" is a subset of the basic emotion "fear". The other hypothesis is that anxiety can be distinguished from fear by differences in head and eye motion.

We test the first hypothesis by implementing a basic emotions detector based on facial action coding system (FACS) to detect fear from videos of anxious faces. When we discover that this is not as accurate as we would like, an alternative solution based on Gabor filters is implemented. A comparison is done between the solutions and the Gabor-based solution is found to be inferior.

The second hypothesis is tested by using scatter graphs and statistical analysis of the head and eye motions of videos for fear and anxiety expressions. It is found that head pitch has significant differences between fear and anxiety.

As a conclusion to the thesis, we implement a systems software using the basic emotions detector based on FACS and evaluate the software by comparing commercials using emotions detected from facial expressions of viewers.

Acknowledgements

I would like to thank my supervisor Dr. Taehyun Rhee and my co-supervisors Dr. Harvey Ho and Prof. Neil Dodgson. It's with their guidance, support and valuable advice that I can complete my master's research and thesis.

I would also like to thank my fellow postgraduate students from the computer graphics group for sharing their research and providing valuable feedback during our weekly group meetings.

In addition, I would like to give special thanks to Auckland Bioengineering Institute for sponsoring the research grant to support my thesis.

Finally, I would like to take this opportunity to express my gratitude to my brother and parents for their support and encouragement given to me throughout my life. Without them, it would have been possible for me to pursue my interests in computer graphics and undertake this master's degree.

Contents

Glossary of Acronyms

AU	Action Unit
AEAFE	Automatic Emotions Analysis using Facial Expressions
AM-FED	Affectiva-MIT Facial Expression Dataset
CK/CK+	Cohn-Kanade (Database) / Cohn-Kanade (Database, Version 2)
FACS	Facial Action Coding System
JAFFE	Japanese Female Facial Expression
NIR	Near Infrared-Red
PCA	Principal Component Analysis
SVM	Support Vector Machines
VM	Virtual Machine
WebRTC	Web Real-Time Communication

Chapter 1

Introduction

1.1 Motivation of thesis

Emotion analytics is an area in data mining to study humans' behavior by analyzing the full spectrum of human emotions. One of the research areas under emotion analytics is automatic emotions analysis using facial expressions (AEAFE). This is a problem that has its origins in psychology in understanding how humans perceive emotions from facial expressions [1]. Early AEAFE solutions required putting physical markers on the face to track expressions and did not run in real time [2]. Subsequently, advances in AEAFE enabled applications to identify emotions in real time. This leads to the growth in the emotions analytics market.

The worldwide emotions analytics market is projected to reach USD$1.71 billion by 2022. This represents a growth of 82.9% from 2016 in the report by Research and Markets [3]. The report forecast the increased demand for emotion analytics across different industries like media and entertainment, healthcare, retail and others. Several major players mentioned are companies like Microsoft [4], Affectiva [5], Kairos [6] and Eyeris [7]. Apple is planning to enter the market when they bought the company Emotient in 2016 [8]. These companies have solutions to analyze emotions using facial expressions in real time.

Detecting basic emotions (anger, disgust, fear, happiness, sadness and surprise) using facial expression is a well-studied topic. Many of the companies mentioned above have commercial products for detecting basic emotions using images and videos. However, there is less research into the automatic detection of complex emotions. Some of these complex emotions like anxiety are no less important than basic emotions. We would like to research on the detection of anxiety from facial expression.

In 2011, the New Zealand Government has committed to a goal for New Zealand to reduce tobacco smoking to less than 5% by the year 2025 [9]. An estimated 4,500 to 5,000 people die each year in New Zealand because of smoking or second-hand smoke [10]. Research shows that the brain is wired to increased anxiety during nicotine withdrawal [11]. Smokers who are trying to quit may experience increased anxiety as a side effect. If this anxiety is unchecked, it may become depression in the long term [12].

One of the initiatives by Auckland Bioengineering Institute in 2011, is to assist smokers to quit smoking using bioengineering technologies. Among the projects proposed, one of them is to detect anxiety in smokers using computer vision. This thesis is partly funded by this initiative to lay down the emotional analysis ground work for the project.

1.2 Objectives of thesis

The objective of this thesis is to create an emotions analytics system that can detect emotions from the human face. This system should meet the following criteria:

- It should be fully automatic.
- It should be able to detect emotions without assistance from physical facial markers.

- It should only use visual data with a single camera without assistance from other non-visual sensors like electroencephalogram (EEG).
- It should work with live video.

1.3 Research Methodology

The following details the methodology of how we conducted our research for this thesis. To meet the objectives of the thesis, we tried to provide solutions for detecting (1) basic emotions and (2) anxiety.

1.3.1 Literature survey on solutions for detecting anxiety

To identify a solution for detecting anxiety from facial expressions, we did a search on existing works in computer vision as well as psychological studies from online sources. We found some research that provides us with clues on solving the problem. In the study by Harrigan et al. [13], they found that an increased in fear expression is linked to high state anxiety. In addition, the study by Steimer [14] suggests that anxiety is the emotional response to unknown threat or internal conflict while fear is the emotional response to visible or known external danger. Perkins et al. [15] suggests that anxiety corresponds to environmental scanning faces and produces a distinct facial expression from fear.

1.3.2 Propose hypothesis on facial expression of anxiety

Using the studies from Harrigan et al, Steimer and Perkins et al., we propose 2 hypotheses to identify an anxious face as follows:

- Hypothesis 1 – The basic emotion “fear” can be detected in the facial expressions of the complex emotion “anxiety”.

- Hypothesis 2 – Head poses and eye positions change more often for anxiety compared with fear.

The rationale behind hypothesis 1 is that since fear expressions is linked to high state anxiety, we think that a basic emotions detector should be able to pick up fear emotions from an anxious face.

The rationale behind hypothesis 2 is that for an anxious person, it is his belief that either there is an existential but unseen threat, or there is an impending threat. Hence, the natural response of anxiety is to scan the environment to seek out the source and direction of the threat first. In contrast, when there is a visible and known danger, the natural response of fear is to focus one's attention on the threat.

1.3.3 FACS-based solution

To build a basic emotions detector to test hypothesis 1, we surveyed existing research on automatic facial expression analysis. We selected the method based on facial action coding system (FACS) as it is the most popular method and can achieve very high accuracy. FACS is a system pioneered by Paul Ekman [1] to quantitively encode facial expressions using facial muscles known as action units (AUs). This method is a data-driven approach requiring face expression databases for training and evaluation.

By studying existing FACS-based methods, we proposed our method to build a FACS-based classifier by modelling after the work by Velusamy et al [16] as they have achieved the highest overall accuracy in the works we surveyed. However, because their work used probability studies on the occurrence frequency of AUs with the training database to improve its results, the same accuracy cannot be achieved when other databases are used for evaluation. Hence, we hope to achieve better results across different databases by introduction some tweaks to the method.

Since data-driven approach requires a database for training, we obtained three facial expressions databases, two for basic emotions and one for complex emotions.

Of the databases labelled with basic emotions, we picked 2 that are frequently used by other facial expression analysis research for training and validation. One database is the Cohn-Kanade version 2 (CK+) Database [17] which contains FACS verified facial expressions, the other is the Japanese Female Facial Expression (JAFFE) Database [18] and is not FACS verified.

The other database we obtained is the Mind Reading DVD [19] which is labelled with complex emotional words instead of basic emotions. The purpose of obtaining this database is to test hypothesis 1 by identifying the emotional words suitable to describe anxiety and select the videos labelled under these words to form the test set.

We implemented and tested our basic emotions detector using the two databases labelled with basic emotions. Implementation requires extracting AUs from the databases as features for training. Validation requires repeatedly dividing the database into training and test sets to perform cross-validation. Testing is done using the selected videos from the Mind Reading DVD.

AU data is extracted using OpenFace [20]. OpenFace is an open source tool that can extract many features like AUs from images and videos of facial expressions. The reason why OpenFace is used is because it is the only freely available tool for extracting AUs that we can find that is updated recently, plus we do not want to implement our own AUs detector or the scope will become too big.

Training and cross-validation is done using Weka [21], a popular open source data-mining software for training classifiers. The reason why we selected Weka is because not only it is free to use, it comes with many popular classifier algorithms we can use without having to implement ourselves. Due to the lack of a large expressions dataset that can be obtained easily, we are not using deep learning frameworks like Caffe, TensorFlow, Deeplearning4j. Training deep-learning networks typically requires datasets with hundreds of thousands or more data instances to achieve high accuracy.

Testing is done by detecting basic emotions for all the frames in the videos selected from the Mind Reading DVD using the trained basic emotions classifier.

We analyze the results of the FACS-based basic emotions solution to answer hypothesis 1 by doing a statistical analysis on the frame count detected as fear for each selected video from the Mind Reading DVD. Although we could not prove hypothesis 1, we did not definitively debunked it either. This motivates us to investigate an alternative solution.

1.3.4 Gabor-based solution

From our literature survey conducted earlier, we noticed some basics emotions detection solutions based on Gabor filters that do not require using FACS.

Gabor filters are based on mathematical models of the visual cortex [22]. It is first proposed by Lyons [23], who argued that this method is based on human's perception of facial expression from an observer's perspective and hence do not need to consider the actual emotion of the person (ground truth) as strictly as AU-based methods.

In the ideal situation, it should be better to match the results of the basic emotions classifier with the ground truth from the person experiencing the emotions. However, for complex emotions like anxiety, it may be difficult to obtain ground truth from the person because he may not be aware of it or he is in denial. In such conditions, it may require a trained psychologist to assess the appropriate state of the mind. Because Gabor filter is based on the perception of emotion, it may provide us with a solution.

Using a similar approach by Bashyal & Venayagomoorthy [24], we propose a solution to apply Gabor filter and select points of certain facial features for improving the performance. However, instead of manually selecting the points, we propose using OpenFace [20] to select automatically. The points we picked corresponds to the AUs associated to basic emotions. We hope that this method not only can improve the performance, it will also be closer to the "ground truth" of the facial expression.

We used Matlab [25] for extracting Gabor features and training the classifier. This is because while Matlab is not free software, it has a built-in Gabor function not available in Weka or OpenFace. This allows us to use Gabor filters without needing to implement it ourselves.

We analyze the performance of the Gabor-based solution by comparing its cross-validation results with the AU-based solution. We found that not only the overall accuracy is lower, the performance on detecting fear is worse. Moreover, the Gabor-based solution is biased towards detecting individual facial features over emotional features, resulting in very poor performance when trained with a facial expressions database that is not FACS verified like the JAFFE database. Hence, we decided that the Gabor-based solution won't be able to answer hypothesis 1 better than the FACS-based solution, hence the verdict on hypothesis 1 is still open.

1.3.5 Detecting head and eye movement

Although we could not determine whether hypothesis 1 is true, it will still be interesting if we can answer hypothesis 2. We propose using a visual method and a statistical analysis on the eye and head motions to compare between anxiety and fear expressions.

To conduct this analysis, we use the videos from the Mind Reading DVD [19] to obtain test samples of fear and anxiety. Since we have already obtained videos selected as anxiety from the earlier experiments, we repeat the same procedure to select other videos to represent fear.

By plotting scatter graphs of eye gaze direction and head pitch/yaw with a small sample of videos from fear and anxiety using Microsoft Excel, we can visually see a trend for fear to show more cluttering of the data points compared the anxiety, hence we decide that it is worth pursuing hypothesis 2 further.

For a more complete picture, we compute the standard deviation of eye gaze direction and head pitch/yaw for each video to compare the differences. In addition, we also compare the differences between the maximum and minimum values to account for scenarios where large eye and head motion occurs in only a small fraction of the video.

We found that only the head pitch shows significant differences between anxiety and fear, hence we conclude that hypothesis 2 is only partially true. However, we leave some doubts with this conclusion due to uncertainties with our dataset from the Mind Reading DVD.

1.3.6 System application for detecting emotions

While we do not have a solution to detect anxiety, we do have a good working solution to detect basic emotions using FACS. To fulfil the thesis objective, we implement the systems application to detect basic emotions as a web-based application. We showed that the application can run with live video. However, due to lack of time and resources needed to conduct a full-scale experiment to test the application live on people, we needed an alternative plan. Instead, we evaluated the application using pre-recorded videos as "live" streams. By using videos of viewers watching commercials, we can detect happy emotions and compare them to the ground truth of how much the viewers reported their enjoyment while viewing each commercial.

1.4 Structure of thesis

The remaining of this thesis is broken up into a few sections as follows.

- Chapter 2 covers the literature survey on basic emotions, facial expressions, anxiety, FACS and Gabor-based detection methods, as well as similar affective applications.
- Chapter 3 covers the description of the facial expression databases and tools we are using in the thesis.
- Chapter 4 presents the proposed solution for classifying basic emotions using FACS and its implementation. It covers how the

facial expression databases are used for the evaluation as well. The test results for hypothesis 1 is presented and the analysis and discussion on the results concludes the chapter.

- Chapter 5 presents the proposed solution for classifying basic emotions using Gabor filter and its implementation. The analysis and discussion on the validation results concludes the chapter.

- Chapter 6 covers the proposed method to test hypothesis 2. The test results are presented together with the analysis and discussion to conclude the chapter.

- Chapter 7 covers the design and implementation of the application for detecting basic emotions. The method of evaluating the application is presented together with its results. The analysis and discussion of the results concludes the chapter.

- Chapter 8 provides an overall conclusion to this thesis and discussion of possible future work.

Chapter 2

Background and Related Works

In this chapter, we present our summary of the literature surveys we have studied to understand the background knowledge and existing works required to conduct our research. These include a brief history behind the psychological study of basic emotions and facial expressions, the facial action coding system (FACS), psychological studies on anxiety and the facial expressions associated with anxiety, FACS and Gabor-based methods for automatic facial expressions analysis, as well as similar affective applications. With these information, we hope it will be helpful for the reader to follow the remaining chapters.

2.1 Basic Emotions

The earliest known literature on human emotions can be found in *Li Ji* (The Classic of Rites) under *Li Yun* (Ceremonial usages) from the collection of text on Confucianism in ancient china [26]. The text describes the seven feelings of men as *joy*, *anger*, *sadness*, *fear*, *love*, *disliking*, and *liking* and that these feelings belong to men without having to learn them.

Charles Darwin is one of the first person in recent history to conduct studies on emotions in humans. He used the photographs taken by Duchenne (see Chapter 2.2) of facial expressions of people whose faces are electrically stimulated to portray certain emotions to conduct a blind test [27]. For the test, he chose 11 photos from Duchenne's work and invited over 20 guests to his house to guess which emotion each photo represents without any help given. The result shows that his guests agreed on emotions like happiness, sadness, fear and surprise but disagreed on the others. He concluded that not all the emotions shown in Duchenne's work are universal. While Darwin's method of survey is not considered scientific today [28] since it lacked a control and the participants didn't all viewed the same set of photos, the methodology in principal is still being used in evaluating the validity of some facial expressions databases now.

In the modern psychology, Paul Ekman define emotions as "basic" when they meet certain criteria [29] like having distinctive and universal signals; having consistent response; being present in other primates; having distinctive physiology; and a few other criteria. However, there are still debates on what these basic emotions comprise of. A recent survey among psychologists [30] reveals that most of them agreed on only 5 basic emotions (Anger, Fear, Disgust, Sadness and Happiness). Another recent study by Jack et al. [31] suggests that only 4 emotions (Happy, Sad, Fear/Surprise, Disgust/Anger) are functionally common. They argue that fear is like surprise (approaching danger), while disgust is like anger (stationary danger).

For this thesis, we adopt the classical view developed by Paul Ekman [1] of 6 basic emotions (anger, disgust, happiness, sadness, fear and surprise) as it is the most well studied, and most of research into automatic emotions analysis using facial expressions (AEAFE) are following it. This would enable us to benchmark our results against other works.

2.2 Facial Action Coding System (FACS)

The studies by French neurologist Duchenne [32] in the 19^{th} century show that facial expressions are the results of the activation of different facial muscles. He conducted experiments to recreate facial expressions by using electrical probes placed on human faces to stimulate facial muscles. He showed that by stimulating certain groups of muscles, he could reproduce the facial expressions of a range of emotions without the person experiencing them. His findings and photographs of the facial expressions are published in the book *The Mechanism of Human Facial Expression* in 1862 (See Figure 2-1 for a sample photo published).

Figure 2-1: Recreation of facial expression of terror by neurologist Duchenne. The expression is created by electrical contraction of facial muscles "mm. platysma" and "mm. frontalis", plus the voluntary dropping of lower jaw. Photo taken from plate 61 in Duchenne et al. [33].

While conducting a study on nonverbal behavior, Ekman [34] with his colleague Waly Frieson used the works by Duchenne and Hjorztsjö's [35] to conduct experiments on themselves to photograph and analyze over 10,000 combinations of facial muscles movement. Using the results, they constructed a measuring technique now known as facial action coding system (FACS) [36] to quantitively measure facial muscle movements. The latest FACS manual is published in 2002 on a CD-ROM which contains example photos and videos as well as the PDF manual with the technical details on FACS.

FACS is comprised of Action Units (AUs), which are the actions of individual or groups of muscles. Each AU is measured by an intensity score of *A* for trace to *E* for maximum. With FACS, it is possible to manually code almost all anatomically possible facial expressions [37]. Table 1 summarizes the AUs that are associated for each basic emotion [38]. Table 2 shows the description, facial muscles and an example for each of these AUs.

Emotion	Action Units associated with
Anger	AU04, AU05 and/or AU07, AU22, AU23, AU24
Contempt	Unilateral AU12, Unilateral AU14
Disgust	AU09 and/or AU10, optionally AU25 or AU26
Fear	AU01, AU02, AU04, AU05, AU07, AU20, optionally AU25 or AU26
Happiness	AU06, AU12
Sadness	AU01, AU15, optionally AU04, AU17
Surprise	AU01, AU02, AU05, AU25, AU26

Table 1: The action units associated for each basic emotion. Values for each emotion extracted from Matsumoto & Ekman [38]

AU	Description	Facial Muscle	Example
1	Inner Brow Raiser	*Frontalis, pars medalis*	
2	Outer Brow Raiser	*Frontalis, pars lateralis*	
4	Brow Lowerer	Corrugator supercilii, Depressor supercilii	
5	Upper Lid Raiser	Levator palpebrae superioris	
6	Cheek Raiser	Orbicularis oculi, pars orbitalis	
7	Lid Tightener	Orbicularis oculi, pars palpebralis	
9	Nose Wrinkler	Levator labii superioris alaquae nasi	
10	Upper Lip Raiser	Levator labii superioris	
12	Lip Corner Puller	Zygomaticus major	
14	Dimpler	Buccinator	
20	Lip stretcher	Risorius with platysma	

AU	Description	Facial Muscle	Example
22	Lip Funneler	Orbicularis oris	
23	Lip Tightener	Orbicularis oris	
24	Lip Pressor	Orbicularis oris	
25	Lips parted	Depressor labii inferioris or relaxation of Mentalis, or Orbicularis oris	
26	Jaw Drop	Masseter, relaxed Temporalis and internal Pterygoid	

Table 2: List of action units, description and visual image associated with basic emotions. Table and images extracted from Cohn, Ambadar & Ekman [39].

2.3 Facial Expression of Anxiety

Anxiety is a general term for several disorders [40] like Generalized Anxiety Disorder (GAD), Panic Disorder, Phobias, Social Anxiety Disorder, Obsessive-Compulsive Disorder (OCD), Post-Traumatic Stress Disorder (PTSD) and Separation Anxiety Disorder. Severe anxiety can affect our daily lives and even cause physical symptoms. Facial expression is just one of the many physical clues that can help diagnose someone suffering from anxiety disorders. But to detect anxiety from facial expressions, we must first decide what constitute as an "anxious" face.

There are few studies that can positively identify anxiety using facial expressions alone. Harrigan et al. [13] videotaped participants who were asked to recall past stressful and non-stressful events. The facial expressions of the participants were FACS coded, and they found that there are more facial movements that are related to fear expression, and eye blinks occurred more often during high anxiety state.

This brings us to the question on how to differentiate fear from anxiety if anxiety facial expressions are related to fear. In the studies of animal behavior, some ethologists suggest that there is a clear functional distinction between anxiety and fear [14]. Anxiety is a general response to an unknown threat or internal conflict, while fear is the response to a visible or known external danger. Both emotions are signals that prepare the body for different responses or behavior.

Perkins et al. [15] recruited 40 human participants who are given a list of scenarios and images of facial expressions. They are asked to match the scenarios to the expressions and produce emotional labels for them. The list of scenarios contains situations with ambiguous threat and situations with clear threat. The photos are created using 8 separate volunteers who posed for facial expressions in response to these scenarios. Using the emotions labelled by the 40 participants, they conducted a second experiment with 18 different participants to match back the facial expression images to the emotional labels. They found that the participants can match the scenarios and emotional labels to the facial expressions. The authors concluded that the anxious face is associated to ambiguously threatening scenarios and is distinct from fear faces which are associated to scenarios with clear threat.

2.4 Detecting emotions from facial images

In this section, we present the background and origins of 2 contrasting methods we used in this thesis to detect emotions from facial images. These are facial action coding system (FACS) and Gabor filter. FACS has its origins in psychology during the research on how human face express emotions by facial muscles movement (See chapter 2.2). Gabor filter on other hand, has its origins in the physical sciences on studying how the brain processes visual information to recognize patterns and objects. For each method, we examined some recent works on how it is used to detect emotions.

2.4.1 FACS-based detection methods

Before computers can analyze facial expressions automatically, psychologists rely on the FACS manual, workshops and certification programs to train individuals to become certified FACS coders. Encoding AUs from facial expressions for emotional studies during that time was a slow and labor-intensive job.

Early computer systems [2] partially automated the AUs encoding process by using plastic dots placed on pre-defined regions on the face and automatically measuring the movement to determine the intensity of the AUs. Barlett et al. [41] is one of the first to fully automate the detection of 7 different AUs without the assistance of physical markers. Further advancements in automatic detection of AUs has led to more accurate classifiers that can detect a greater number of action units.

One of the objectives of detecting AUs from facial expression is to identify the emotions expressed by the face. Once the AUs are detected, identifying basic emotions is a relatively simple task of mapping the AUs to the rules defined in the FACS manual. Because the problem of automatically detecting AUs is a much harder task than identifying basic emotions using the detected AUs, many of the research stops at detecting AUs. However, because the AUs for each basic emotion are defined by psychologists in the FACS manual, some of the AUs not specified may still have a minor contribution to each emotion. Moreover, errors introduced in the automatic detection of AUs together with the fluctuations of AUs between successive image sequences may influence the emotions inferred. Thus, there are some research that extends to detect actual emotions and not just stopping at identifying AUs.

One of the earlier works that uses AUs to automatically classify emotions is by Pantic & Rothkrantz [42]. The system has 3 different parts. The first part generates images using 2 mounted cameras with one camera taking photo of the frontal view while the other takes photo from the side view. The second part tracks the facial features like head contour, eyebrows, eyes and mouth region. The last part tracks the AUs and infers the emotions by a rule based system. While the system can achieve overall accuracy of 90.57% in detecting emotions from a set of 256 face images, it is sensitive to inaccuracies. Moreover, if the system determines that any of the data is inaccurate, the data is discarded.

Some of the best possible accuracy of detecting emotions using AUs is found in the semi-automated method by Kotsia & Pitas [43]. For the detection to work, the user needs to manually fit a parameterized face mask known as Candide on to the first frame of the emotional image sequence. A Candide face mask consists of about 100 triangles that are shaped to model the human face that can be controlled by adjusting AU values. The nodes of the initial mask are tracked for subsequent sequence, and the new AU values are then computed from the new mask that results from the new facial expression. Using a simplified Candide grid with a modified multi-class SVM classifier, they achieved an overall accuracy of 99.7% on the Cohn-Kanade (CK) database [44]. With the exceptional high accuracy, most subsequent research hence focuses on improving the accuracy with fully automated solutions.

Velusamy et al. [16] used a Gabor-based detector to identify the AUs. However, instead of relying on standard FACS definitions of basic emotions, they establish a statistical relationship between AUs and emotions using probability. A portion of the Cohn-Kanade version 2 (CK+) Database [17] is used for learning the statistical relationship and the remaining for evaluation. They further evaluated the accuracy using other databases. They obtained overall accuracy of 97.0% with CK+ Database and between 82.0% to 94.0% with other databases.

Silva et al. [45] used a hardware solution to extract facial expressions for detecting emotions. The Intel RealSense 3D camera combines a standard camera, 2 infrared cameras and an infrared laser projector to calculate depth in scenes. Together with the Intel RealSense Software Development Kit (SDK), it can extract face location, landmarks, pose and expression in real time. Using 10 facial landmarks, the authors extract 5 distance values by computing the linear distances between these landmarks. The 5 distance values together with 7 face expressions (mapped as AUs) extracted by Intel RealSense SDK are used as features for training an SVM classifier. They built their own test and training database by asking 32 children between 6 to 9 years old and 11 adults between 18 to 30 years old by to pose for emotions in front of the Intel RealSense 3D sensor. Facial landmarks and extractions are obtained from the sensor.

During cross-validation, they obtained the result of 93.63% using radial basis function (RBF) kernel. They further evaluated the system with 14 adults aged between 18 and 49 and achieve an overall accuracy of 88.3%.

2.4.2 Gabor-based detection methods

In information theory, a new method for time-frequency analysis is invented by Dennis Gabor [46] in 1945. He proposed a method that combines both time and frequency components into the same wavelet to transmit the same information using less data [47]. We now call this as Gabor wavelet. It is the result of multiplying a sine wave with a Gaussian. The Gaussian carries the signal which changes over time, and the sine wave carries the modulation frequency. The Gaussian function $f(x)$ is defined [48] as follows:

$$f(x) = e^{-(x-x_0)^2/a^2} e^{-ik_0(x-x_0)} \tag{1}$$

where k_0 is the modulation frequency and a is the spread constant.

The visual cortex is the part of the brain that processes visual information. In 1962, Hubel & Wiesel [49] experimented on cats to understand how the visual cortex respond to different patterns of white light. Using the work of Hubel & Wiesel together with others on understanding the visual cortex of mammals, Mardelja [22] proposed a mathematical model to represent the visual cortex. He concluded that the visual cortex processes the visual information in both spatial and spatial frequency domain. By representing information in an abstract form as a set of excitation levels of different cells, he showed how the responses of the cells in the visual cortex of cats and monkey is like the Gabor wavelet. Mardelja proposed that by filtering the information in the Gabor representation in the spatial or spatial frequency domains, you could extract data about the position and orientations of lines and edges for pattern recognition.

Daugman [50] introduced 2D Gabor filters to account for the orientation selectivity as well as the 2D arrangement of the cells. Turner [51] showed how Gabor filters can be applied to different textures. He applied 2D Gabor filters on images by convolution on each pixel resulting in a 4D hyperplane. Using a set of 4x4 Gabor filters with 4 different frequencies (wavelengths of 4, 8, 16 and 32 pixels) and 4 orientations (0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$), Turner showed that different textures of an image can be separated in this hyperplane. Figure 2-2 shows a Gabor filter rendered on a 3D plot using Matlab.

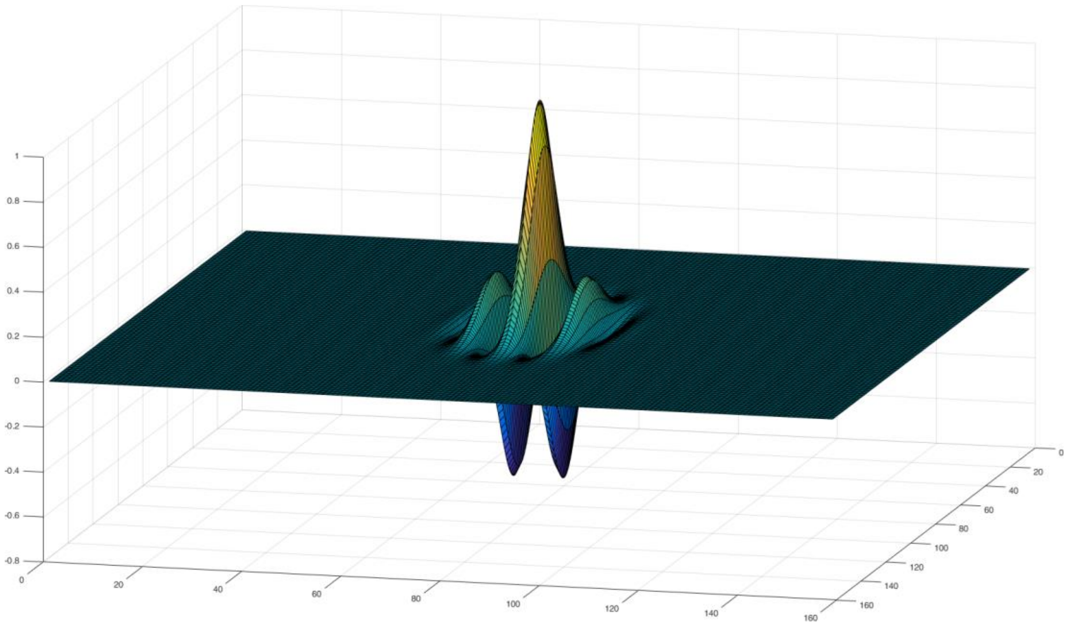


Figure 2-2 - Visual representation of a Gabor filter in 3D plotted using Matlab

Gabor filters have been used in image recognition problems successfully like thumbprint detection, characters recognition and face recognition.

Lyons [23] first proposed the method of using Gabor filters for detecting emotions using facial expressions. At that time, existing methods for detecting emotions are mainly using various heuristics like optical flow, principal component analysis and facial models. He showed that by using Gabor filters on facial expression images, an emotions classifier can be built. Lyons argues the importance of his approach compared with other methods, is that it is based on the neurobiology of how our brain perceives vision and hence has better psychological plausibility. Moreover, while there are FACS-based method that are strongly backed by psychological studies, it requires knowing the ground truth of the emotions that the subjects feel. With Gabor filters, this requirement is relaxed since it relies on the perception of emotion rather than the emotional truth.

The following are some of the more recent works using Gabor filters for detecting emotions from facial expressions.

Kumbhar et al. [52] and Bashyal & Venayagomoorthy [24] used 2D Gabor filters to extract features from facial expressions in individual images to detect 6 basic emotions. A total of 18 different Gabor filters consisting of 3 wavelengths ($\frac{\pi}{4}, \frac{\pi}{8}, \frac{\pi}{16}$) and 6 orientations (0 to 180°) were used. To extract the features from the images, 34 fiducial points were manually selected from the facial expression and convolved with each of the Gabor filters. PCA was applied to reduce the dimensionality of the features, and the trainings were done using the Japanese Female Facial

Expression (JAFFE) Database. Kumbhar et al. trained their classifier using feed forward neural network with 20 inputs and 40 to 60 hidden layers and achieved 60-70% recognition rate. Bashyal & Venayagomoorthy trained their emotions classifier using Learning Vector Quantization (LVQ) unsupervised clustering algorithm and achieved accuracy of 87.51%.

Owusu, Zhan and Mao [53] used Viola and Jones face detection [54] to detect and crop the faces from the JAFFE Database. From the cropped faces, they reduce the cropped faces to 20x20 pixels using Bessel down-sampling [55]. After applying a set of 40 Gabor Filters, a Adaboost-based algorithm is further used to reduce the dimensionality of the features. Finally, a 3-layer feed-forward neural network (MFFNN) classifier is used. With the JAFFE database, they selected 2 images for each emotion per person for training, and the rest for testing. Average recognition rate achieved is 96.83%.

Abdulrahman et al. [56] used a total of 40 Gabor Filters (5 wavelengths, 8 orientations) on the JAFFE Database [18] without down sampling of images. Instead, the dimensions are reduced by PCA and Local Binary Pattern (LBP). They compared the results between using PCA, LBP, Gabor, Gabor + PCA and Gabor + LBP, and found that Gabor + LBP has the best performance with average of 90% recognition rate.

2.5 Beyond Basic Emotions

While detecting basic emotions from facial expressions is a well-researched area with good results, detecting complex emotions is still a very difficult problem with not many studies done. The following are some of the work we have surveyed that address the research beyond basic emotions.

Ekman [57] wrote in 1987 that other than facial actions already prototypical to the basic emotions, there are no evidence of facial features that are characteristics to affective disorders (like depression). Instead, such "blue" moods and clinical depression are characteristic in showing more pronounced periods of sadness. In addition, affective disorders like anxiety is being described using fear. This suggest that in psychology, facial expressions associated to complex emotions are more difficult to analyze compared to basic emotions.

Nevertheless, there are works that attempt to detect depression from a combination of facial actions and/or voice. One of them [58] attempts to detect depression using facial expressions and vocal expressions in video recordings from AVEC2013 Dataset, which comprises of 292 subjects being recorded by webcam while performing some Human Computer Interaction tasks. All subjects are recorded between 1 to 4 times with interval of 2 weeks and each clip is between 20-50min long. A limitation of this approach is the level of depression by the subjects is self-reported using Beck Depression Index (BDI). The problem of self-reporting is that the data can be polluted by the subject's comprehension of the survey questions as well as recalling of answers [59].

Cohn et al. [60] used automated and manual systems to investigate changes over time in the facial expressions of depressed patients during the treatment. These patients are evaluated between 1 to 4 occasions at 7-weeks interval by clinical interviewer. They found that patients with high severity depression made more facial expressions associated to contempt, and their smiles are more likely to be accompanied by contempt. The authors noted several limitations of their work. One limitation is that 3 of the interview questions that the patients are asked, may have affected their emotions and behaviors subsequently.

2.6 Affective Applications

The following is a summary of some recent applications that uses emotions detected from facial expression to perform various tasks. These applications provide inspiration for us to design a way to evaluate our system application.

2.6.1 Detecting emotional stress from facial expressions for driving safety

Gao, Yuce & Thiran [61] proposed a method to detect stress in drivers to monitor their attentiveness and emotional state for safety and comfort while driving. They mounted a near-infrared (NIR) camera on the car's dashboard that tracks the face in real-time and extracts 49 facial landmarks. After normalizing the images to 200x200 pixels, they used SIFT descriptors [62] to extract blocks comprising of 32x32 pixel around these facial landmarks. The SIFT descriptors are then concatenated and PCA is used to reduce the dimensionality to form the feature vector. With the FACES database [63] and Radboud database [64], they trained a multi-class classifier using Linear SVM for each of the 6 emotions.

For evaluation, they recorded facial expressions of participants who are asked to pose 6 basic and 1 neutral expression in 2 different environments. First is in a typical office with a NIR camera mounted on a desk, the other is from inside a car with the NIR camera mounted on the dashboard. 2-minute-long videos are then recorded with the participants posing stressful look for a duration of 1 minute after 30 seconds into the recording. For the classifiers they trained, detection rates of 90.5% and 85% are achieved for the office and car scenarios respectively.

2.6.2 Video Classification and Recommendation using Emotions

Zhao, Yao & Sun [65] proposed a system for video classification and recommendation using emotions detected from facial expressions. By extracting Haar-like features from the CK Database, they trained a AdaBoost classifier to detect 6 basic emotions and 1 neutral expression. They proposed a temporal hidden conditional random fields (HCRF) algorithm that uses the results of the AdaBoost classifier to produce a sequence of emotions detected in the video sequence.

For evaluating their system, they employed a group of students to watch 100 online videos comprising of short scenes from different movies which were classified into 6 categories (comedy, tragedy, horror, moving, boring, exciting). The facial expressions of the students are recorded using video while they watched the online videos and the emotion sequences extracted for analysis. The online videos are then categorized and recommended using the following heuristics in Table 3.

	Main facial expressions	Terminal expression	Recommendation standard
Comedy	Neutral, happiness	Happiness	P(happiness) > θ_1
Tragedy	Neutral, happiness, sadness	Sadness	P(sadness) > θ_{21} & P(happiness) > θ_{22}
Horror	Neutral, fear, surprise	Fear, surprise	P(fear)+P(surprise) > θ_3
Moving	Neutral, sadness	Sadness	P(sadness) > θ_4
Boring	Neutral, fear	Neutral, fear	Do not recommend
Exciting	Neutral, happiness, surprise	Happiness	P(happiness) + P(surprise) > θ_6

Table 3: Heuristics used by Zhao, Yao & Sun [65] for video classification and recommendation for each category. θ_1 to θ_6 are some percentage values that they set. Table is reproduced from the paper.

For the video classification, they achieved classification accuracy of 90%. Overall, 80% of their subjects agreed with the video recommendation results.

2.6.3 Predicting Movie Ratings from Audience Behaviors

Navarathna et al. [66] used infrared cameras to film an audience of 5-10 people watching movies in a darkened environment. The facial and body movements of the audience are captured by the cameras. Using FACS to identify smiles and optical flow features to capture body movement, they propose a method of analyzing individual and group behavior to predict movie ratings. The prediction is compared with ratings from ratings aggregator rottentomatoes.com and with the audience self-report. Using root mean squared error (RMSE), the average RMSE of their prediction compared to audience self-report is 16.95.

2.6.4 Intelligent Advertising Billboards

In 2015, the advertisement agency M&C Saatchi demonstrated an intelligent advertisement billboard (Figure 2-3) with a built-in Microsoft Kinect camera to analyze audience reactions to watching advertisements [67]. The billboard is installed at a bus shelter on Oxford Street in July 2015, and a second one at Clapham Common in August 2015 [68]. During this period, they showed a total of 1540 ads (Figure 2-4) and collected over 42,000 interactions. The billboard is self-evolving using an algorithm to remove ads that are deemed not popular while the popular ones are retained for next round. At the end of the initial round, results show that shorter ads are more popular with heart images frequently appearing at the end.

Figure 2-3: M&C Saatchi's intelligent billboard on Oxford Street. Photograph from online article on The Guardian [67]

Figure 2-4: Some of the ad images from M&C Saatchi's intelligent billboard. Photograph from online article on The Guardian [67]

2.6.5 Predicting Online Media Effectiveness using Smile Response

McDuff et al. [69] collected 3,268 videos of facial responses to watching 3 Super Bowl commercials in Mar 2011 to determine if they can predict the "liking" and "desire to view again" using the smiles responses. Viewers were asked to answer 3 multiple-choice questions regarding whether they liked the commercials, whether they have watched them before, and whether they will watch them again. The smile classifier tracks the region around the mouth and computes Local Binary Pattern (LBP) features in that region and outputs the smile probability value for each frame. They used a few approaches to analyze the data collected: Class Priors, Naives Bayes, SVM, Hidden Markov Models (HMM), Hidden-state Conditional Random Fields (HCRF) and Latent Dynamic Conditional Random Fields (LDCRF). They achieved the best results of 0.8 and 0.78 for the area under Receiver Operating Characteristics curve using LDCRF in predicting liking and desire to watch again and concluded that it is possible to automatically determine the effectiveness of online media using their method.

Chapter 3

Databases and Tools

In this chapter, we present the relevant background and information of the facial expression databases and tools that we used in our research and systems implementation. Because these databases and tools will be mentioned frequently in the remaining chapters, we discuss the details in this chapter to facilitate a better reading flow for those chapters.

3.1 Facial Expression Databases

In this section, we provide a summary and background of the facial expression databases that we are using in this thesis. Except for the Mind-Reading DVD which is purchasable online, the rest of the databases can be obtained for free for non-commercial use. This allows anybody to replicate and validate our results. Throughout the thesis, we are using one or more databases for the training, validation and/or evaluation of our classifiers.

3.1.1 Japanese Female Facial Expression (JAFFE)

The JAFFE database [18] is created by Lyons [23] who pioneered the use of Gabor filters for detecting emotions from facial expressions (see Chapter 2.5). It consists of 213 black and white facial images posed by 10 Japanese female models. Each model has between 2 to 4 images for each of the 7 facial expressions (anger, disgust, fear, happy, sadness, surprise and neutral expression). Figure 3-1 shows a sample of images from the JAFFE Database.

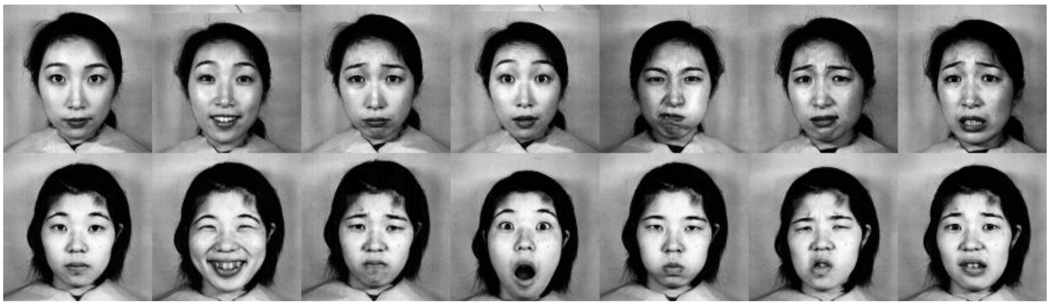


Figure 3-1: Sample images from JAFFE database (Figure 4 in [23])

Because this database is used by Lyons' work, many facial expression analysis methods that used Gabor filters have trained or tested their classifiers with JAFFE database as well. Some of these include Bashyal & Venayagamoorthy [24], Shih, Chuang & Wang [70], Owusu, Zhan & Mao [53], Gu et al. [71], Kumbhar, Jadhav & Patil [52] and Abdulrahman et al. [56]. In our thesis, we used this database in both training and evaluation of our basic emotions classifier.

3.1.2 Extended Cohn-Kanade Database (CK+)

Prior to 2000, there are limited data sets in the facial expression analysis research community where different groups can use a common data set to test and compare results. Hence, Kanade, Cohn & Tien set out to create the The Cohn-Kanade (CK) Database [44] for facial expression analysis research.

Version 1 release of the CK Database contains 486 sequences from 97 subjects between age of 18 to 50 years old from a mix of different racial groups. Subjects are asked to pose for emotions and their expressions are recorded starting from the neutral position to the peak expressions for each emotion. FACS-certified coders are employed to code each peak expression and label them with the emotions as defined by FACS.

One of the main issues with this version was that the expressions were not verified with the emotions that the subject was portraying. The issue was that subjects when asked to pose for an emotion, he/she may not be performing accordingly to the definition of the emotions outlined by FACS. This caused errors in the emotion labeling due to bad acting.

Version 2 (also known as CK+) [17] extended the original version to a total of 593 sequences from 123 subjects. In addition, the emotions labels were verified with the definition in the Emotion Prediction Table from the FACS manual. Furthermore, they checked if the expressions contain certain AUs that were not consistent with the emotion. And finally, visual judgement was made by psychologists for the facial expressions to determine if they are good representations of the emotions. A total of 327 sequences passed the emotions verification process. Figure 3-2 shows examples of verified emotions from the database.

Figure 3-2: FACS verified facial expressions from the CK+ Database [17] (Clockwise from top-left, face showing: Surprise, Happiness, Disgust, Sadness, Fear and Anger. Images used are from participants labelled S52, S55 and S106)

Because the facial expressions in the database are FACS encoded and the emotion labels certified by FACS coders, it is a very popular database used by many facial emotions analysis research. We selected this database for this thesis because it is free to use and is widely used by related works. This allow us to compare our classifier accuracy with the other works objectively. We are also able to obtain good results with this database to train our classifier. Hence, this is our database of choice for implementing the system application too.

3.1.3 Mind Reading DVD

While there are many facial expressions database available for basic emotions, there are not many that captures complex emotions. To answer our hypotheses to detect anxiety from facial expressions, we need to source for a database that is clearly labelled with the presence and absence of anxiety. However, since we could not find any such databases publicly available, the closest we could use is the Mind Reading DVD [19] developed to help people with autism who have difficulties recognizing emotions.

A team of psychologists from Cambridge University compiled the Mind Reading DVD by filming videos of actors who posed for several different emotional words. Instead of the usual 6 or 7 basic emotions agreed by most psychologists, they determined the number of emotional words by using a thesaurus to identify every word in the English language (apart from synonyms) that describes an emotional feeling. They came out with 412 dictionary distinct emotional words which were then organized into 24 related groups. Each emotional word was then portrayed by 6 actors, who learnt the definition and context of the emotional words by 6 short descriptions of scenarios that could give rise to the emotions (see Figure 3-3). We selected a subset of the database for the training and testing for our study into anxiety detection. Figure 3-4 shows a sample of some of the images we used.

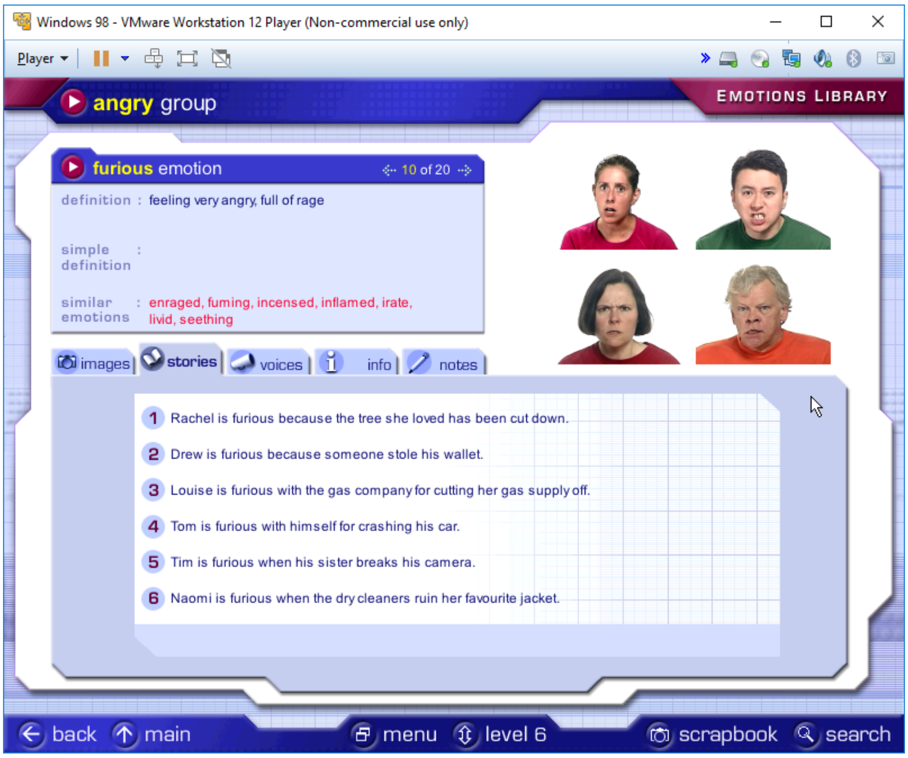


Figure 3-3: Definition and usage context of one emotional word (furious) as described in the Mind Reading DVD.

Figure 3-4: Sample images from Mind Reading DVD [19] showing the variety of emotions and subjects.

3.1.4 Affectiva-MIT Facial Expression Dataset (AM-FED)

Most facial expression databases are created with participants posing for certain facial expressions and recorded under lab conditions. This is a problem as the conditions for real-world applications are usually not ideal. The objective of the AM-FED database [72] to capture spontaneous facial expressions under real-world environment.

To create the database, Affectiva and MIT launched a website on March 2011 to record viewers' reactions to watching videos of 3 Super Bowl advertisements. (This is the same database described in Chapter 2.6.5 above earlier on.) At the end of the videos, the viewers were asked 3 questions to assess (1) if they liked the advertisement, (2) if they had watched the advertisement before and (3) if they would watch the advertisement again. The viewers were also explicitly asked for permissions to allow their facial expressions to be captured and shared for research purpose.

The dataset consists of 242 facial videos (168,359 frames) and are manually labelled for 14 AUs by at least 3 FACS trained coders. Figure 3-5 shows sample images from the database. The respective responses to the assessment questions for each video are available as well. The 3 questions asked were: (1) Did you like the video, (2) Have you seen it before and (3) Would you watch it again. There were 3 response choices for each question as follows (Table 4):

Did you like the video?	**Have you seen it before?**	**Would you see this video again?**
2 – "Heck ya! I loved it."	2 – "Yes, may times"	2 – "You bet!"
1 – "Meh! It was ok."	1 – "Once or twice"	1 – "Maybe, if it came on TV"

0 – "Na… not my thing."	0 – "Nope, first time"	0 – "Ugh. Are you kidding?"

Table 4: Response choices per question for AM-FED Database [72]

Figure 3-5: Sample image frames from AM-FED Database [72]

Because the database is not explicitly labelled using the emotional labels that our classifier used, it is not suitable for training. We used this database only to evaluate our systems application.

3.2 Tools Used

In this section, we introduce the tools and libraries that we used during our research and system application development. Other than Matlab, the rest of the tools and libraries are free and open source. There are 2 different class of tools we used for different purpose. Some tools are used directly in the implementation of our application. This include OpenFace [20], LibSVM, OpenCV and Kurento Media Server. Other tools like Matlab and Weka are purely used to speed up research only. Weka comes with open source libraries that can be used directly in applications too.

3.2.1 OpenFace

OpenFace [20] is an open-source framework developed by Baltrusaitis, Robinson & Morency that can perform several facial analysis tasks like landmark detection [73], head pose tracking, facial action unit (FACS) recognition [74] and eye gaze tracking [75]. It is the first tool that can do all 4 tasks in real time that comes with the source code and model trainer. It can extract 18 AUs (1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28 and 45), 68 points of facial landmarks, the eyes direction vector in world coordinates, location of head with respect to camera and the rotation of head around the x, y, z-axis. The application supports single images, image sequences or video from few popular formats. We are using OpenFace extensively to extract AUs, landmarks, eye gaze and head pose.

3.2.2 Weka

Weka is a popular open source data mining tool [21] written in Java produced from The University of Waikato. It contains a collection of many different algorithms for pre-processing, classification, regression, clustering and visualization. It contains a graphical interface package for users as well as Java libraries for applications developers. The Weka Explorer (Figure 3-6) is an application in the graphical interface package for users to build and evaluate new classifiers from the available algorithms that comes with the package.

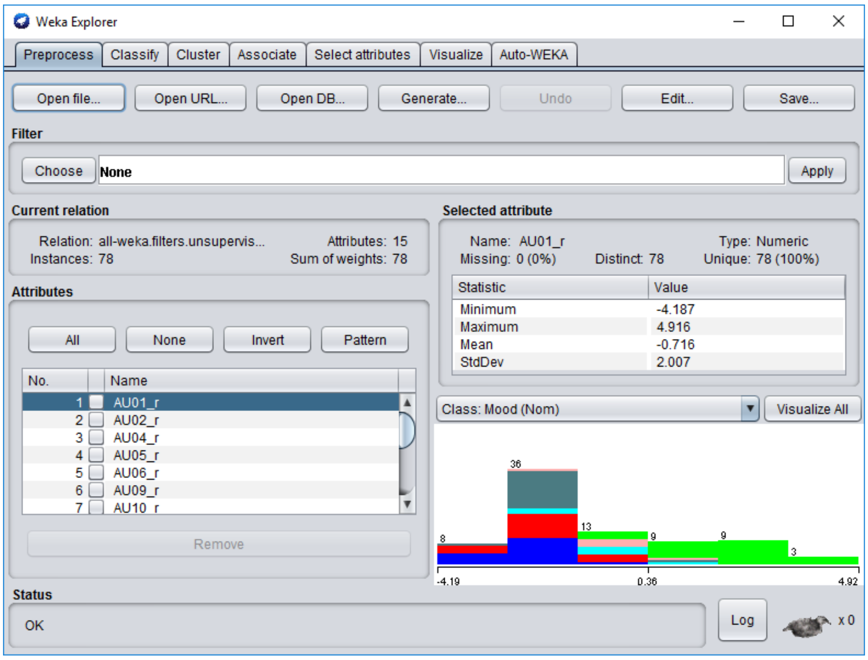


Figure 3-6: Screen capture of Weka Explorer after loading some training data.

We are using Weka extensively to determine the best combination of algorithms and features to build and evaluate our emotions classifier quickly without having to spend time writing and debugging codes.

3.2.3 Matlab

Matlab [25] is a tool from MathWorks that is widely used by engineers and scientists around the world for solving numerical problems. It is also a scripting language that is designed to perform math computations like matrices, linear algebra, numerical analysis and more. It has a rich graphical user interface for visualization and analysis of data using 2D/3D plotting functions.

We use Matlab in extracting features from facial expressions images using the built-in Gabor filter. The built-in data analysis training tool is used for finding the best classifier for detecting emotions and visualizing the results.

3.2.4 LibSVM

Support vector machines (SVM) is a supervised classification algorithm originally invented by Vladimir N. Vapnik and Alexey Ya. Chervonenkis in 1963 [76] for separating 2 data groups by mapping them into higher dimensions. A limitation with the original algorithm is that some data cannot be separated easily. It was later improved by Corinna Cortes and Vapnik in 1995 [77] by introducing soft hyperplanes to minimize the errors when separating non-separable data.

Assuming we have a line or hyperplane able to separate 2 groups of data points. Support vectors are the data points that are closest to the line or hyperplane. SVM finds the optimum line/hyperplane that minimizes the perpendicular distance of the support vectors from the line/hyperplane. Figure 3-7 show an example of an optimum linear separation of 2 sets of data points using SVM.

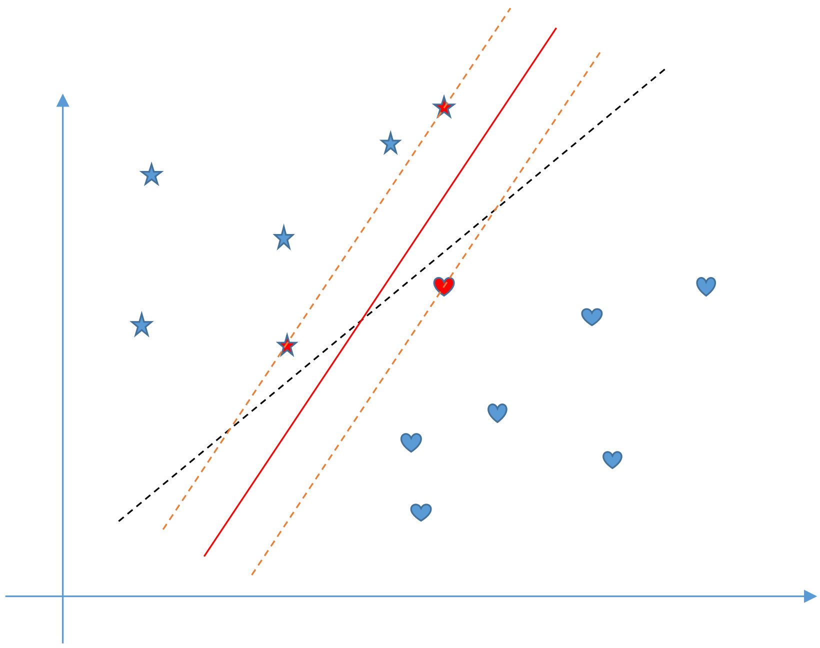


Figure 3-7: Example of linear separation of 2 sets of data points using SVM. The black dotted line shows one way to linearly separate the group of points. The red solid line shows the optimum line is identified by SVM that maximizes the perpendicular distance with the support vectors (red data points). Red dotted lines show the distance of support vectors from optimum line.

However, in many cases, data points may not be separable easily using a straight line. To solve such a problem, a kernel trick can be used to map the original points into higher dimensional spaces so that the points becomes linearly separable using a hyperplane. Figure 3-8 illustrates an example of kernel trick. A popular kernel is the Gaussian kernel.

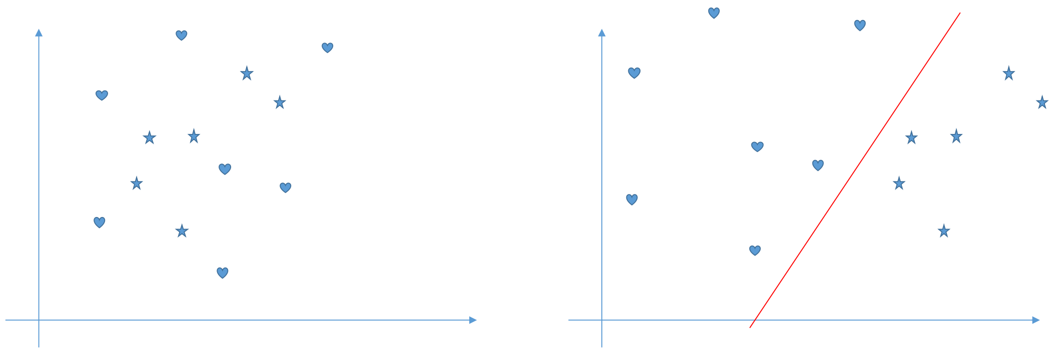


Figure 3-8: Chart on the left shows 2 groups of data points not linearly separable. Using a kernel trick, points are mapped into higher dimensions with SVM. Chart on the right shows the same 2 dimensions of these points after mapping and linear separation (red line).

LibSVM is a popular open source implementation of SVM. It is first created in 2000 in National Taiwan University. Source codes are publicly available for C++ and Java. There are now over 20 software packages maintained by other groups of people who have extended LibSVM to create software packages in other systems or platforms like Weka, Matlab, PHP and .NET.

3.2.5 WebRTC

WebRTC is an open framework originally released by Google to enable real time communications (RTC) for web browsers and applications [78]. The objective is to allow applications supporting WebRTC to seamlessly communicate using common set of protocols across any computing device. The WebRTC protocols and browser APIs are now being standardized by the Internet Engineering Task Force (IETF) [79] and the World Wide Web Consortium (W3C) [80]. The protocol is now supported on most browsers (like Chrome, Firefox and Opera) and desktop/mobile platforms. This means that most devices with the latest browsers will be able to run WebRTC-enabled web applications (like real-time video conference) without installing additional plugins.

Figure 3-9 shows the WebRTC architecture. The browsers that are compliant with WebRTC standards are expected to implement the WebRTC components and expose the functionalities via the Web API to applications.

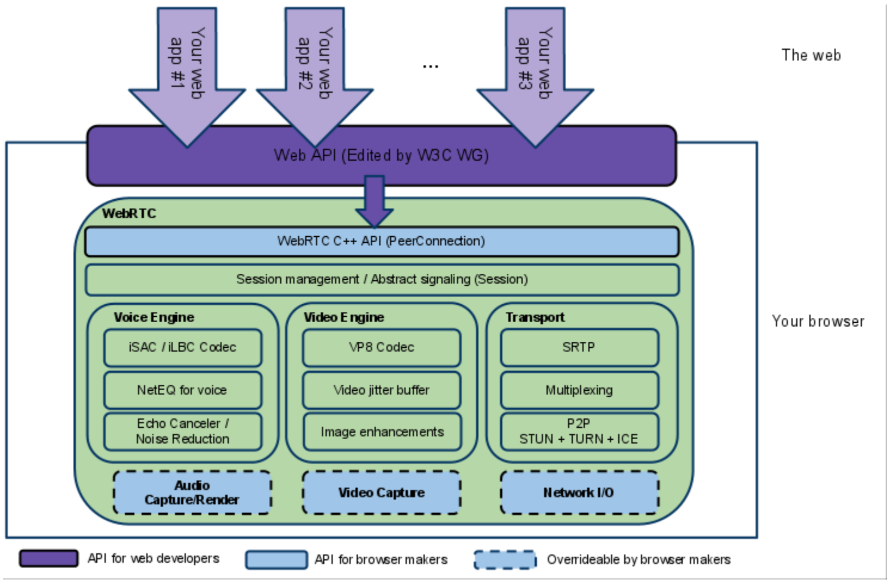


Figure 3-9: The WebRTC architecture. Image taken from https://webrtc.org/architecture/

WebRTC applications can be implemented in 2 different modes: Peer-to-peer and server-based. Peer-to-peer WebRTC applications can communicate directly between 2 devices, while server-based WebRTC applications communicates with each other through a WebRTC server (Figure 3-10).

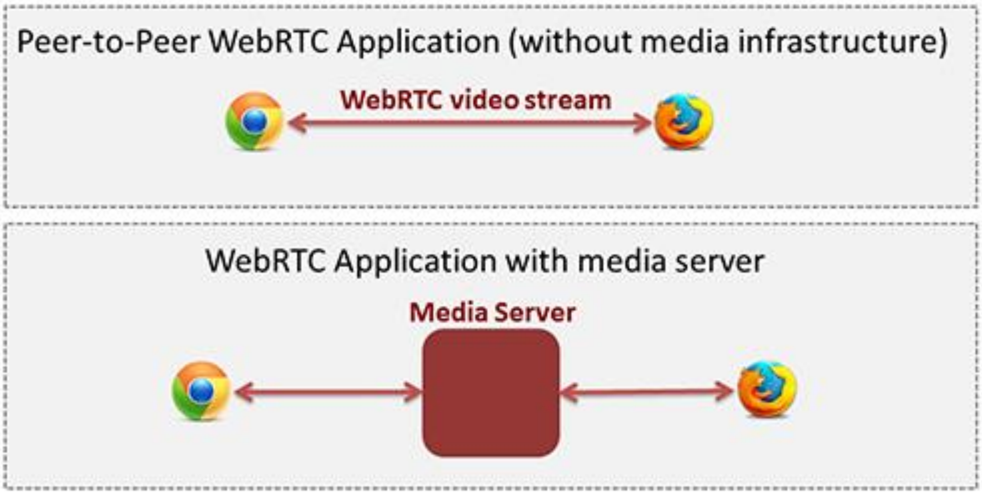


Figure 3-10: Peer-to-peer vs server-based WebRTC applications. (Image taken from Kurento documentation [81])

3.2.6 Kurento

Most WebRTC server handles the media traffic between two or more peers and provides advanced features like group communication, video transcoding and recording that are difficult to implement in a peer-to-peer model. Kurento is a WebRTC server implementation with additional built-in functionalities like augmented reality, media blending and mixing, and allows developers to provide added functionalities via custom modules [81]. It provides utilities to generate WebRTC-enabled application templates in JavaScript or Java. These built-in capabilities allow developers to easily develop WebRTC applications without needing to know about WebRTC. Figure 3-11 shows the differences between a normal WebRTC server and Kurento.

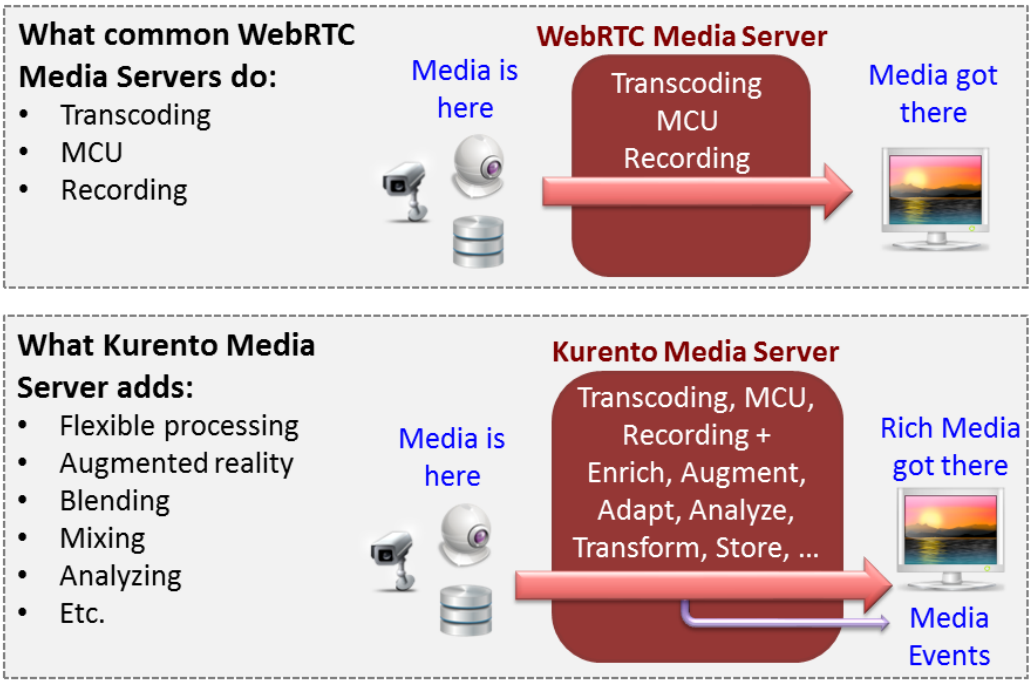


Figure 3-11: Kurento compared with normal WebRTC server (Image taken from Kurento documentation [81]))

Chapter 4

Detecting Emotions using Facial Action Units

In this chapter, we present our research into using AUs to detect basic emotions. By detecting basic emotions, we hope to answer our hypothesis 1 that fear can be detected from the facial expression of anxiety. The chapter covers the limitations of existing solution, our proposed solution, our implementation of the proposed solution, the method of evaluation, the results and the analysis.

4.1 Existing solution and its limitations

There are many works in AEAFE that focus on automatically identifying AUs but does not perform emotions classification with the AUs. In a survey of papers by Sariyanidi [82], a total of 12 papers that performed AUs detection are listed, only 2 of them continued to classify emotions using the detected AUs. The reason is because FACS is designed to code almost every facial expression possible, meaning the facial expression of basic emotions is just a subset. By detecting all the AUs present in a facial expression, it ought to be relatively straight forward to translate into emotions using the AU mapping rules defined in the FACS manual.

However, while the FACS manual contains definitions for the set of AUs to describe the facial expression for each basic emotion, it does not mean that other AUs do not have some contribution to the emotions. For example, while a typical happy expression is associated with AU06 (Cheek Raiser) and AU12 (Lip Corner Puller), there may be a statistical significant probability that the jaw may be dropped (AU26). So, instead of stopping at the classification of AUs, can we do more? Indeed, as we have seen in the related works presented in chapter 2.4.1, Velusamy et al. [16] used a statistical analysis by studying the probability of AUs detected for each emotion to achieve a very high prediction accuracy of 97.0% for the CK+ Database. They found that AU6, AU7, AU12 and AU26 have positive associations with happy emotion while AU1, AU2, AU5 and AU9 have negative associations.

While Velusamy et al. achieved a very high accuracy, the cross-database performance is not that good. When testing with JAFFE and Mind Reading DVD, they achieved accuracies of 87.5% and 82.0% respectively. One reason we speculate may contribute to this difference in performance, is because they used CK+ Database for learning the statistical significance of the AUs but tested on different databases. The statistics may be biased towards CK+ Database, hence the same performance doesn't translate to other databases. For our emotions classifier, we want to utilize as many AUs as possible for training without showing any biasness to any dataset.

4.2 Proposed Solution and Implementation

To test if additional AUs besides using the ones defined by FACS for basic emotions can improve the results, we utilize as many AUs as we can obtain and compare it with using just the emotional AUs. And to find the best classifier, we repeated the training and cross-validation process as shown in Figure 4-1. The aim is to find the best classifier algorithm and AU combination that can produce the most accurate results. The process goes as follows:

1. Using OpenFace [20], we first extract all the 18 AUs from the training database. This results in 35 features, since of the 18 AUs that OpenFace extracts (Refer to chapter 3.2.1 for the list of AUs), 17 AUs contain both intensity (decimal value between 0 to 1) and binary values (0 or 1), and the remaining AU 28 consist of only binary values.

2. Logically, intensity AUs should perform better than binary AUs since intensity AUs contains more information than binary values. However, we want to test if more AUs, even if they are in binary, can we achieve a better result. Hence, we prepare 4 different combination of AUs to test as follow:

    a. Set A – All 35 binary and intensity AUs

    b. Set B – All 17 intensity AUs

    c. Set C – All 18 binary AUs

    d. Set D - Intensity AUs associated with Emotions as indicated in Table 2. However, since OpenFace does not output AU22 and AU24, we are left with AU01, AU02, AU04, AU05, AU06, AU07, AU09, AU10, AU12, AU14, AU20, AU23, AU25 and AU26, giving us a total of 14 AUs.

3. Using Weka Explorer (See Figure 3-6), we tried training with as many classifiers as we could. These classifiers are the ones that

shows the best results: Naïve Bayes [83], LibSVM [84], Multilayer Perceptron (MLP) [85], Simple Logistics [86], Random Forest [87]. For simplicity sake, the training is first done using the default settings. Optimization on the classifiers are performed after we have identified the best set of AUs to use.

4. Evaluation is done within Weka Explorer using 10-fold cross-validation using the training data.

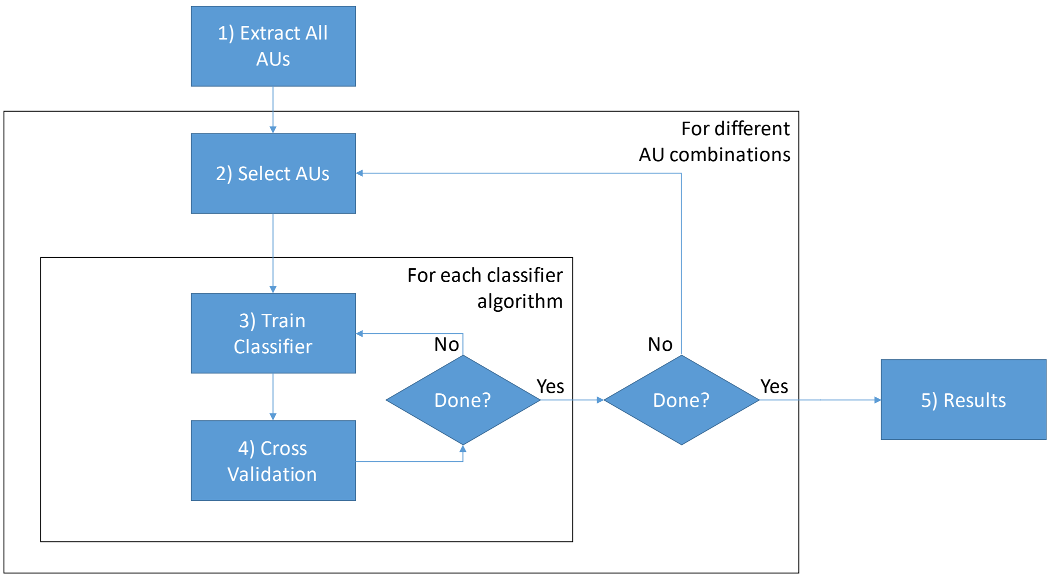


Figure 4-1: Proposed solution to test out different combinations of classifiers and AUs

We repeat the above process using CK+ [17], followed by JAFFE Database [18]. The aim is to determine which database can produce better performance using the proposed solution.

4.3 Testing Methodology

To test out hypothesis 1, we used videos from the Mind Reading DVD [19] for the test set. However, since the videos from the Mind Reading DVD are not labelled as anxiety, it is not so straight forward to select the videos for the test set. Moreover, since a video comprises of a whole sequence of images, each image can result different emotions detected. The literature we presented in chapter 2.3 does not tell us how often should the fear emotions occur during anxiety state. We need to discuss how to conclude hypothesis 1 using the detected emotions. This section describes our methodology of selecting videos for testing as well as how we interpret the results.

4.3.1 How to select the test candidates

In the Mind Reading DVD, there are a total of 412 emotional words and 6 video examples for each word, for a total of 2,412 video clips. These emotional words are grouped into 24 categories. However, anxiety is not among emotional words nor categories. Hence, we need a way to select emotional words that can fit the scenarios associated with anxiety. (Recap: we described earlier for anxiety to be the emotional response to unknown threat or internal conflict.)

El Kaliouby [88] studied on a computational model for mind reading to analyze the facial expressions of 6 complex mental states – *agreeing*, *concentrating*, *disagreeing*, *interested*, *thinking* and *unsure*. She selected them using the mental states from the contents of the Mind Reading DVD that are interesting and not comprising of any of the basic emotions. In our thesis, we selected emotional words from the same DVD as well, however instead of picking entire mental state groups, we select the relevant words using the context and re-classify them into different categories as follows.

There are 2 categories that contain emotional words that can fit the definition of anxiety. These categories are "afraid" and "bothered". Table 5 shows the list of emotional words in the 2 categories.

Emotion Group	Emotional Words
Afraid	Afraid, Consternation, Cowardly, Cowed, Daunted, Desperate, Discomforted, Disturbed, Dreading, Frantic, Intimidated, Jumpy, Nervous, Panicked, Shaken, Terrified, Threatened, Uneasy, Vulnerable, Watchful, Worried
Bothered	Bothered, Flustered, Impatient, Pestered, Restless, Ruffled, Tense

Figure 4-2 List of emotional words in the Afraid and Bothered groups from the Mind Reading DVD [19] that contains emotional words similar to "anxiety"

Each emotional word from the 2 groups is given a definition (see Table 14 in the Appendix for the whole list) as well as example stories that provide the context of usage (see Table 15 in the Appendix). With the definitions and usage context, we looked at the emotional words in these 2 categories and re-labelled them as "anxiety" or "fear" (Table 5).

To select the emotional words that we need, the notion of fear or/and anxiety must be present in the definition, and should form the dominating emotion. The words that don't have the notion of fear or/and anxiety in the definition, or if the notion of fear or/and anxiety is not a dominating emotion are categorized as “others”. For example, “shaken” is categorized as “others” because the definition describes getting “upset” as well as “worried”. “Upset” is a description more for the basic emotion “sadness”.

To differentiate the selected emotional words between “anxiety” or “fear”, we examine the context from the example stories given for each emotional word. If the source of the fear is directly related to a physical subject that is clearly visible and happening at that moment, the word is classified as “fear”. For example, “Joe feels threatened by the neighbor’s large dog”; the big dog is directly causing Joe’s fear and is visible, hence in this context “threatened” is considered as fear. If the source of the fear is indirectly related to the subject, or the fear comes from within the person, or if the source is unknown or not visible or is considered as a collective group, or if the source of the fear happened in the past, the word is classified as “anxiety”. For example, “Moana is nervous of dogs because she was bitten once”. In this example, “dogs” is a collective group and getting bitten by one dog happened in the past, hence “nervous” is considered as “anxiety”. Another example is “Ben feels consternation when the security man accuses him of stealing”. While the security man is clearly visible and accusing Ben, it’s the worry about the consequences (of stealing and getting caught) that Ben is fearful of. The fear is considered as coming from within Ben and is unknown, hence in this example “consternation” is considered as “anxiety”.

By going through this reclassification exercise, we obtained the emotional words and emotion category as shown in Table 5.

Emotion	Emotional Words
Anxiety	Afraid, Bothered, Consternation, Discomforted, Disturbed, Dreading, Flustered, Frantic, Jumpy, Nervous, Panicked, Restless, Ruffled, Shaken, Tense, Uneasy, Vulnerable, Watchful, Worried
Fear	Cowardly, Cowed, Intimidated, Terrified, Threatened
Others	Daunted, Desperate, Impatient, Pestered

Table 5: Emotional words under "afraid" and "bothered" groups from Mind Reading DVD [19] classified as "anxiety" and "fear" after re-classification exercise.

4.3.2 How to interpret the results

Since there exists a sequence of different emotions detected for every video (each image frame can show a different emotion), plus we have many videos selected for testing anxiety, to determine if fear is considered as "present" in the emotion anxiety, we will be looking at 2 aspects:

- The existential of fear emotion in a video clip. We check if there exist a single frame of fear emotion detected in the entire video clip. If fear expressions cannot be detected from any single frame of the video clip, we can conclude that the fear emotion is absent in that clip. This is the most "lenient" method to detect fear for a single video clip.

- The "dominating emotion" detected from the video clip. We count the number of frames detected for each of the 6 basic emotions in a video clip. The emotion that produces the highest frame count is the dominating emotion. If fear is the dominating emotion, we can conclude that there are increased fear expressions from that video clip.

- To answer hypothesis 1, we should detect the existential of fear in almost all the video clips (at least 95%), and a significant proportion (at least 50%) of these clips should have "fear" as the dominating emotion.

4.4 Results

The following are the results obtained. The first 2 sections show the performance differences of the different classifiers and AUs using CK+ database and JAFFE database for training and validation. The last section shows the results of detecting fear using the best classifier with the videos which are categorized as anxiety from the Mind Reading DVD.

4.4.1 Classifier performance using CK+ Database

By running the proposed solution from chapter 4.2 to train the classifiers using the CK+ Database, these are the results. With the default classifier settings, we found that Simple Logistic classifier performed the best with 90.2% overall accuracy. In addition, using all binary and intensity AUs available from OpenFace produces the best result throughout all 5 classifiers (Table 6).

Method / AUs	Naïve Bayes	MLP	LibSVM	Random Forest	SimpleLogistic
Set A	86.4%	89.9%	89.9%	89.3%	90.2%

Set B	82.5%	86.0%	84.7%	84.4%	84.7%
Set C	85.4%	84.1%	85.4%	85.7%	88.3%
Set D	81.2%	84.7%	82.8%	83.4%	86.4%

Table 6: The overall accuracy of 10-fold cross-validation by training with different classifiers in Weka using default settings on the CK+ Database. The set of AUs selected as features are mentioned in chapter 4.2

Using the results, further improvements can be made by optimizing the classifier parameters (Table 7). For Naïve Bayes classifier, turning on supervised discretization improved the accuracy to 88.6%. Multi-layer Perceptron classifier achieved 91.6% accuracy with a learning rate of 0.5, turning on decay for learning rate, and using the number of hidden layers equal to total attributes + classes. LibSVM performs the best after optimizing the parameters, setting cost to 8.0 and gamma to 0.015625, achieved overall accuracy of 92.5%. The accuracy of random forest classifier could only be improved marginally by increasing the number of iterations. But increasing the iterations increases the training time as well. Using 10,000 iterations, the accuracy improves to 90.3%. Simple Logistics classifier could be marginally improved by setting the heuristic greedy stopping to 20, achieving accuracy of 90.6%.

Method	Accuracy	Parameters used[1]
Naïve Bayes	88.6%	weka.classifiers.bayes.NaiveBayes -D
MLP	91.6%	weka.classifiers.functions.MultilayerPerceptron -L 0.5 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H t -D
LibSVM	92.5%	weka.classifiers.functions.LibSVM -S 0 -K 2 -D 3 -G 0.015625 -R 0.0 -N 0.5 -M 40.0 -C 8.0 -E 0.001 -P 0.1 -model "C:\\Program Files\\Weka-3-8" -seed 1
Random Forest	90.3%	weka.classifiers.trees.RandomForest -P 100 -I 10000 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
SimpleLogistics	90.6%	weka.classifiers.functions.SimpleLogistic -I 0 -M 500 -H 20 -W 0.0

Table 7: The best accuracy achieved for each classifier (using Set A as features) after adjusting parameters. Parameters are optimized first by randomly adjust values of available options, next by picking the options that showed significant performance differences before finetuning the values. [1]Paramaters optimized are highlighted in red. For each method:
Naïve Bayes: "-D" turns on supervised discretization
MLP: "-L" learning rate, "-D" turn on decay, "-H t" total hidden layers equal to total attributes + classes
LibSVM: "-G" gamma, "-C" cost
Random Forest: "-I" iteration count
Simple Logistics: "-H" Heuristic greedy stopping

Finally, with the best performing classifier (optimized LibSVM, using Set A), we obtained the confusion matrix as shown in Table 8.

Classified as →	Anger	Disgust	Fear	Happy	Sadness	Surprise
Anger	91.1%	6.7%	0%	0%	2.22%	0%
Disgust	1.7%	96.6%	0%	1.7%	0%	0%
Fear	0%	0%	76.0%	0%	8.0%	16.0%
Happy	0%	1.4%	0%	98.6%	0%	0%
Sadness	10.7%	0%	0%	0%	82.1%	7.1%
Surprise	0%	0%	2.4%	0%	3.7%	93.9%

Table 8: Confusion matrix for the best performing classifier (i.e. LibSVM using Set A features)

4.4.2 Classifier performance using JAFFE Database

Repeating the same procedure from chapter 4.2 on the classifiers using JAFFE Database produces worse results as shown in Table 9. The best performing classifiers are Simple Logistic and Random Forest, achieving 79.8% accuracy, and they are worse than the worst performing classifier configuration achieved using CK+ Database. With the poor performance, no further tests and optimizations are performed.

Method / AUs	Naïve Bayes	MLP	LibSVM	Random Forest	SimpleLogistic
Set A	67.8%	72.7%	73.2%	77.0%	79.8%
Set B	60.1%	76.5%	73.2%	79.8%	72.7%
Set C	61.7%	61.75%	62.8%	62.3%	62.3%
Set D	60.1%	76.0%	68.3%	78.1%	67.2%

Table 9: The accuracy of 10-fold cross-validation by training with different classifiers in Weka using default settings on the JAFFE Database. The set of AUs selected as features are mentioned in chapter 4.2

4.4.3 Detecting fear from anxiety videos

A total of 108 video clips re-labelled as anxiety are tested frame by frame using the optimized LibSVM classifier trained in Weka from chapter 4.4.1. Out of the total, 11 clips (10.2%) do not contain a single frame predicted as fear emotions. And only 23 clips (21.3%) have "fear" as the dominating emotion. (See Table 16 in the Appendix for the complete results) In fact, there are more clips with sadness (34) and surprise (25) as the dominating emotion. Since over 5% of the clips do not contain a single frame of emotions and less than 50% of the clips have "fear" as the dominating emotions, we conclude that hypothesis 1 cannot be verified.

4.5 Analysis and Discussion

We have successfully trained a good basic emotions classifier using CK+ Database and SVM classifier using LibSVM. While the overall accuracy of 92.5% does not improve on the 97.0% rate achieved by

Velusamy et al. [16] mentioned in chapter 2.4.1, it is better than the earlier work by Pantic & Rothkrantz [42] which achieved 90.56%. In the recent survey paper by Sariyanidi et al. [82], the systems listed for detecting basic emotions overall accuracy ranged from 89.9% to 95.9% with the CK+ Database.

We also showed that using the full set of AUs for training is better than using just the ones associated with emotions. This supports the idea that additional AUs may have minor contributions to the facial expressions of some emotions. Moreover, using the AUs extracted in binary form in conjunction with the intensity values performs better than using just the intensity or just the binary values for the CK+ Database. This is perhaps because the 2 sets of values help to re-enforce each other to compensate from detection errors and uncertainties.

CK+ Database proved to perform much better than using JAFFE Database. This is probably because CK+ database has gone through the additional step to verify the AUs of the expressions are present and are "valid" for the posed emotions while JAFFE does not. In a sense, the emotions are thus "encoded" using AUs within the image sequences. Hence, by using AUs as the feature set, the classifier can "retrieve" back the encoded AUs from the facial expressions and predict the emotions.

However, the classifier is unable to associate increased fear emotions with the facial expression of anxiety. There are some factors that may have contributed to this failure.

- Hypothesis 1 itself could be wrong.
- The finding by Harrigan [13] on "increased fear actions appearing in the facial expression of anxiety" by itself is vague. Our

assumption is that this translates to fear emotions dominating the facial expression may not be true. While we did find that fear is detected in more frames (25.9 frames/clip) than happy (3.39 frames/clip), anger (16.3 frames/clip) and disgust (19.1 frames/clip), it is lower than sadness (36.3 frames/clip) and surprise (30.0 frames/clip).

- Some of the emotional words describe a complex mix of emotions. Hence while fear may be present in them, so do other emotions. This means that we may not be able to use the "dominating emotion" method to conclude that there is increased in fear actions as other emotions may be dominating too.

- The Mind Reading DVD may suffer the same issues as the JAFFE Database by not being FACS coded. While the final selection of the best 6 videos to represent each emotional word is done by psychology experts, since there are no FACS definitions for each emotional word there is no way to verify if the actors performed up to the "defined" emotional word.

- Although the best classifier performed 92.5% accuracy, it's performance for fear is only 76.0% accuracy. This may have impacted its ability to identify fear expressions correctly.

With no better ways of testing out hypothesis 1 using the FACS-based classifier, we conclude that either the Mind Reading DVD is not suitable for obtaining test candidates for anxiety, or the classifier method does not work as well on the Mind Reading DVD. After all, FACS-based method relies on the ground truth to be accurate to make a good prediction. However, the videos in the Mind Reading DVD are performed by actors who may not necessarily experience the emotions they are performing, instead, they are judged by psychologists to represent the emotional words. Hence, in the next chapter, we explore the Gabor-based method that is based on the perception of emotions by humans`.

Chapter 5

Detecting Emotions using Gabor Filter

In this chapter, we present our research of using Gabor filters to detect basic emotions to determine if it can perform better than FACS based method. The chapter covers our proposed solution based on existing work, the implementation, results and the analysis on why the Gabor method did not perform better than the FACS-based method.

5.1 Proposed Solution

Many of the papers [52] [24] [56] [53] that uses Gabor filters for automatic emotions detection have a common approach as follows (Figure 5-1).

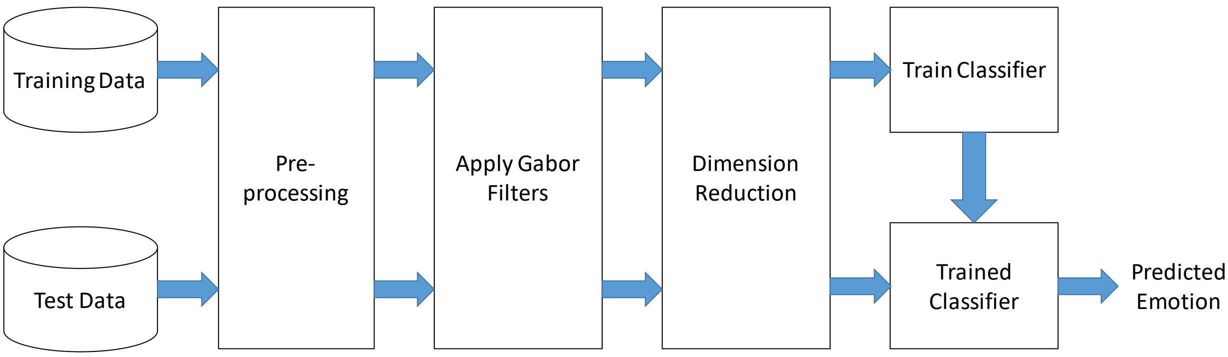


Figure 5-1: Overall procedure for training classifier using Gabor Filters and predicting emotions with the trained classifier.

The training process requires an existing image or video database comprising of actors posing for and labelled with the necessary emotions. This database is then divided into test set and training set, otherwise if the whole database is used for training, testing will require a different database.

Before Gabor filters are applied, there may be a pre-processing step. This step can involve down sampling of image, face cropping or image enhancement. A set of 2D-Gabor filters is then applied to the original image using the convolution process to produce a new image per Gabor filter. Because there is usually between 18 to 40 Gabor filters used, this results in a very large dimensionality. Hence, it is usual to perform dimensionality reduction to improve the classifier performance and to remove redundant features. Dimensionality reduction can involve one or more methods like principal component analysis (PCA), local binary patterns (LBP), selecting fiducial points etc. Finally, the reduced set of features is then used to train the classifier.

The process of testing is similar as training, except after training the final output is a trained classifier, and after testing the output is the predicted emotion.

For our proposed solution, we follow the method from Bashyal & Venayagamoorthy [24] to use selected fiducial points from the facial expression to reduce dimensionality. However, while they manually selected these points using a GUI application, we automatically detect facial features using OpenFace to these points. The idea is to select the locations of the face where the relevant facial muscles are that contributes to emotions instead of using the whole face.

5.2 Implementation

Using our proposed solution, we repeat the process shown in Figure 5-2 as follows.

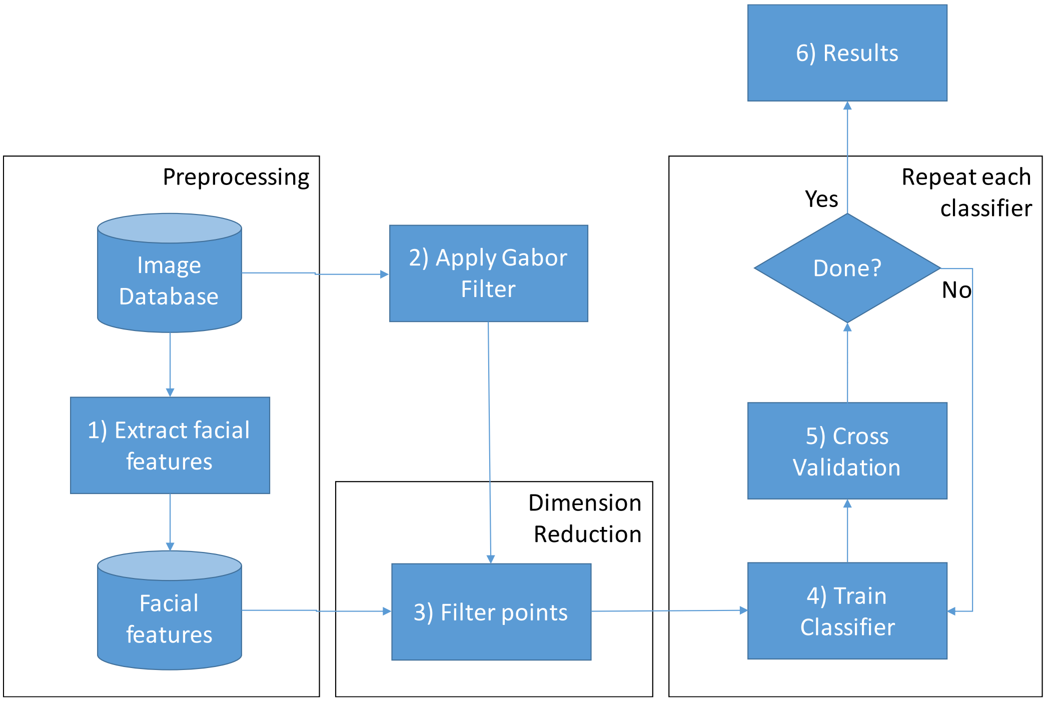


Figure 5-2: Procedure to extract Gabor features and test out different classifiers using Classification Learner application in Matlab.

Step 1 is preprocessing and it involves extracting the facial points from the databases using OpenFace and store these results in a file. The results will be needed during dimension reduction step.

Step 2 involves using the Gabor and convolution function in Matlab to process the images from the database. We used a total of 40 Gabor filters comprising of 8 orientations and 5 wavelengths. Figure 5-3 shows the Gabor filters used.

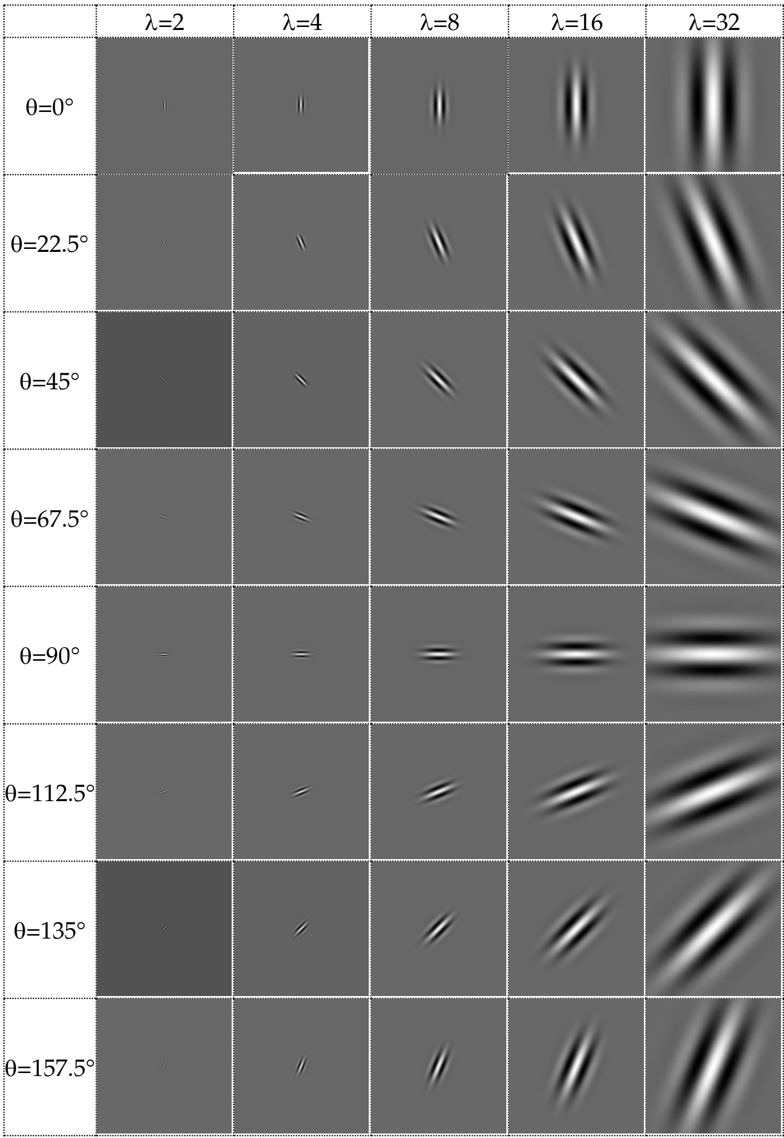


Figure 5-3: Visual representation of the 40 Gabor filters used for our implementation. 5 wavelengths of size 2, 4, 8, 16, 32, and orientations of 0, $\frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}$ *are used. Images are rendered in Matlab.*

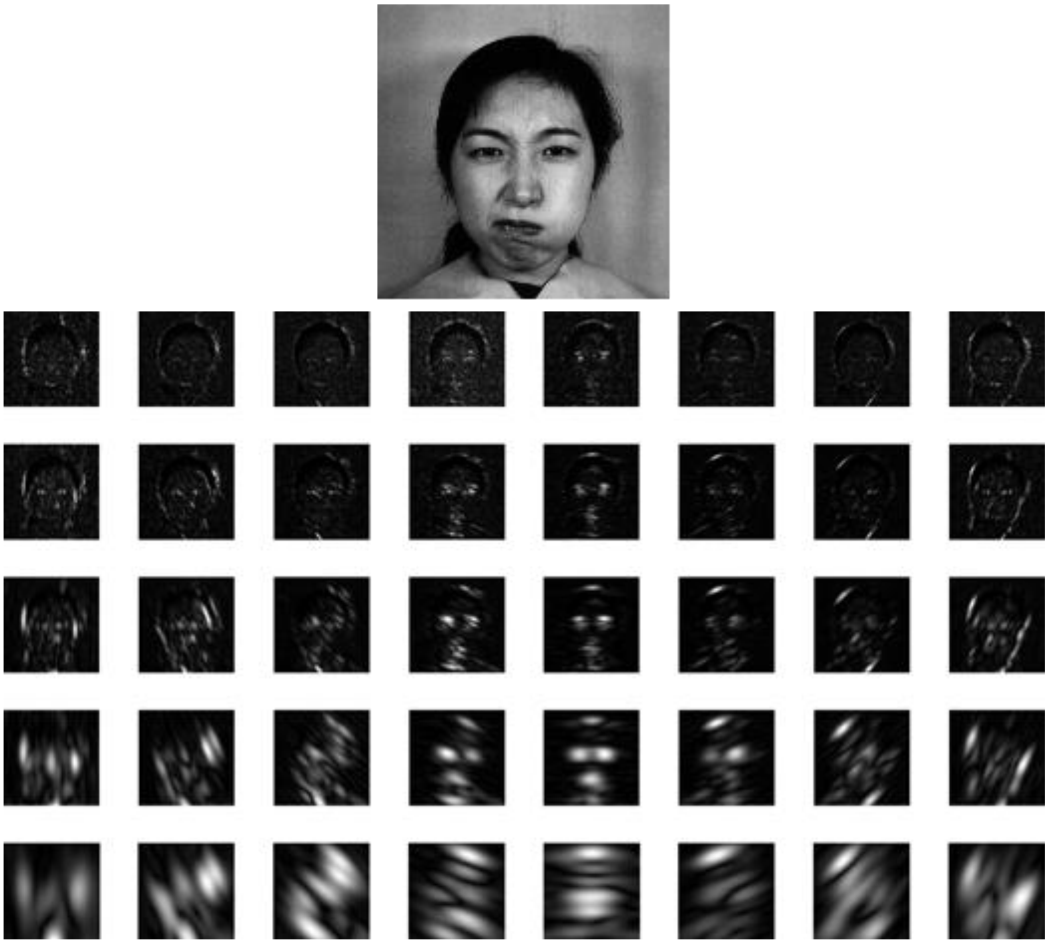


Figure 5-4: Results of applying each of the 40 Gabor filters on one image (top image) from the JAFFE Database [18] using convolution function in Matlab.

Step 3 is to reduce the dimensions by selecting points from the Gabor images. Using the facial points extracted from step 1, we select 26 points from the eye brows, eyes, nose and lips as filters and apply image mask to each of the Gabor images from step 2. Figure 5-5 shows the process of creating the image mask and applying to 1 of the Gabor image. The figure shows a 3x3 mask used for each point for illustration clarity.

However, for the actual training, only the pixel value at each point per Gabor image is used, giving us a total of 1040 features.

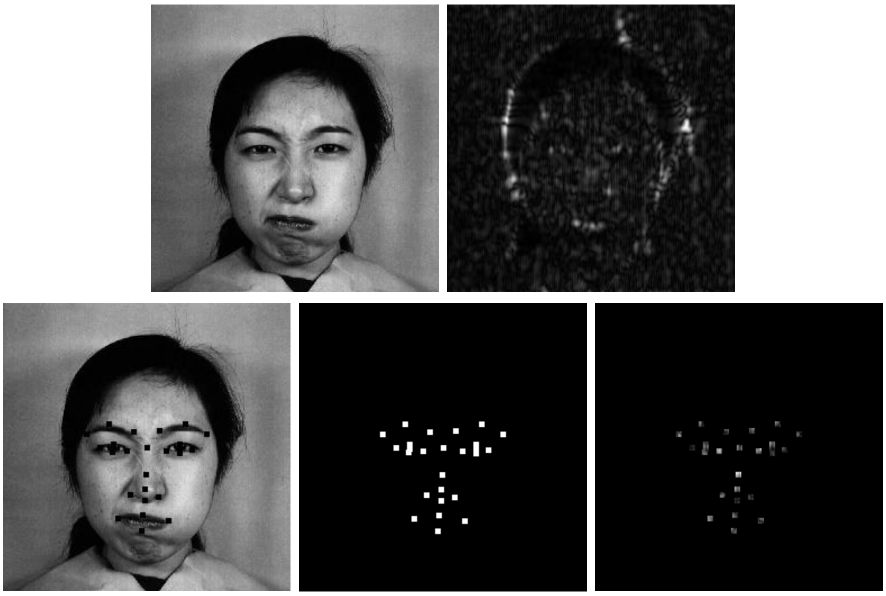


Figure 5-5: Top left shows the original image from JAFFE Database. Top right shows the convolution output of one Gabor filter on the image. Bottom left shows the facial locations marked out by OpenFace. Bottom center shows the image mask created using the facial locations. Bottom right shows the results after applying the image mask on the Gabor filtered image.

Last step is the iterative process of using the Classification Learner in Matlab to train different classifiers and perform cross validation. We repeat this process for as many different classification algorithms as we can find in Matlab. The default settings for each classifier is used initially, and the parameters are tweaked later for some of the better performing classifiers to see if improvements can be made. 5-fold cross validation is used throughout.

Previously, CK+ database performed better than JAFFE database for the FACS-based solution. This time, the JAFFE database did surprisingly well as we can see from the results section below. We considered that Gabor filters captures features globally and have been very successful in performing general face recognition, we then realized that the nature of the JAFFE database might affect the cross-validation results.

The JAFFE database consists of 3 to 4 different images for the same person with the same emotion. However, if we use the built-in cross-validation function in the Matlab Classification Learner application, there is no way we can control how we split the images into test and training sets. It is possible that the same person with the same emotion can end up in both the test and training sets. We do not want this to happen because the performance may rely on matching similar facial features that are non-emotional and unique to the individual, thus improving the results. Hence, we wrote a Matlab script that will take the best classifier that we have trained, and apply cross-validation manually. We used 2 approaches of performing manual cross-validation.

In the first approach, the script removes an entire person from the database for the test set, and used the remaining for training. Since there are 10 different persons in the database, this results in a 10-fold "minus-one-person" cross validation. This approach ensures that individual facial features will not affect the cross-validation performance.

In the second approach, the script removes one emotion from a single person from the database as the test set, and used the remaining for training. This results in a 60-fold "minus-one-emotion" cross validation. This approach tests the effects of including facial features of the individual from the test set into the training set. If the cross-validation results from this approach is different from the first approach, it means the non-emotional facial features unique to the individual do affect the cross-validation performance.

For the CK+ database, the entire set of data we used contains of only one emotion per person. Hence a random cross-validation is as good as doing a "minus-one-emotion" cross-validation.

5.3 Results

We implemented our solution for both CK+ Database and JAFFE Database. To compare our Gabor based solution against our FACS-based solution, we used the same data from the previous chapter here. The following are the results for each database.

5.3.1 CK+ Database

Using the CK+ database, we tested the range of classifiers available using Matlab with the default settings, and the best accuracy results we obtained are from SVM classifiers. Quadratic SVM obtained accuracy of 88.0%, followed by Linear SVM with 86.4%, and Cubic SVM scored 86.0%. By changing the SVM classifiers from "one-vs-one" to "one-vs-all" method, the accuracies of Quadratic SVM improved to 89.3% and Cubic SVM improved to 87.0%. Linear SVM accuracy dropped to 85.7%. Figure 5-6 shows the confusion matrix results for the Quadratic SVM classifier using "one-vs-all" method.

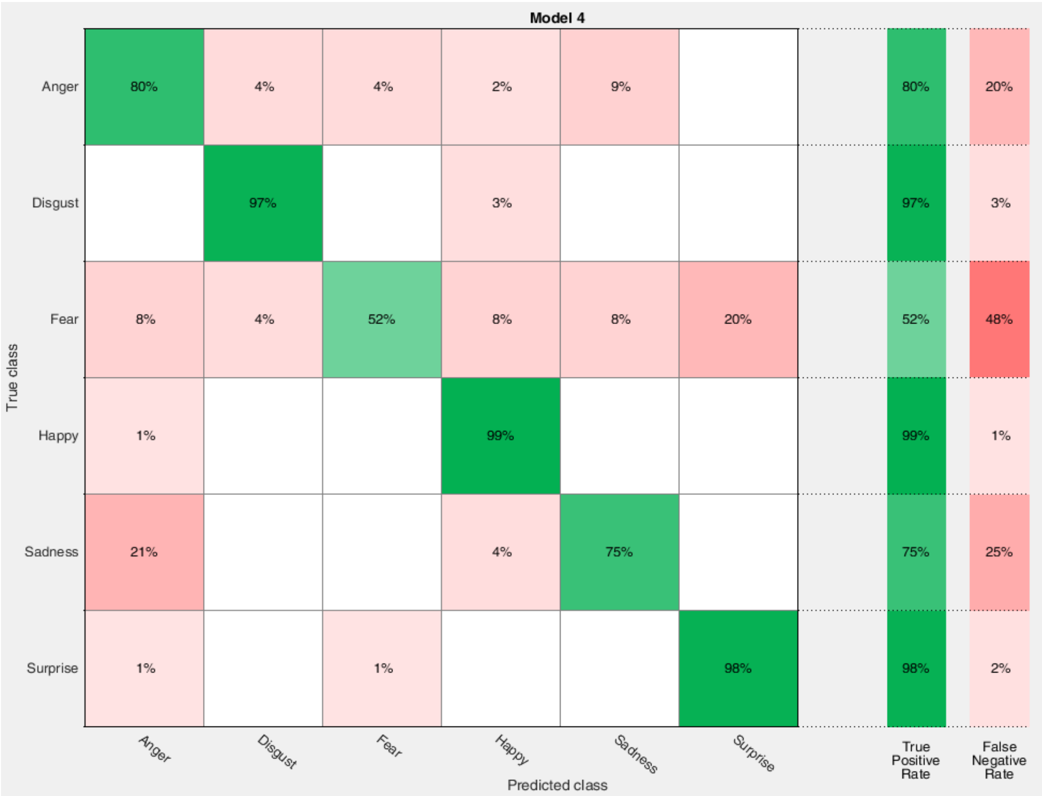


Figure 5-6: Confusion matrix for basic emotions classifier built in Matlab using Quadratic SVM and CK+ Database using "one-vs-all" method.

5.3.2 JAFFE Database

Using the JAFFE database, we tested using the default settings for the range of classifiers available in Matlab. The best accuracy results achieved are from Fine KNN (89.1%), Quadratic SVM (92.9%) and Cubic SVM with PCA at 95% variance (90.2%). Figure 5-7 shows the confusion matrix for the Quadratic SVM.

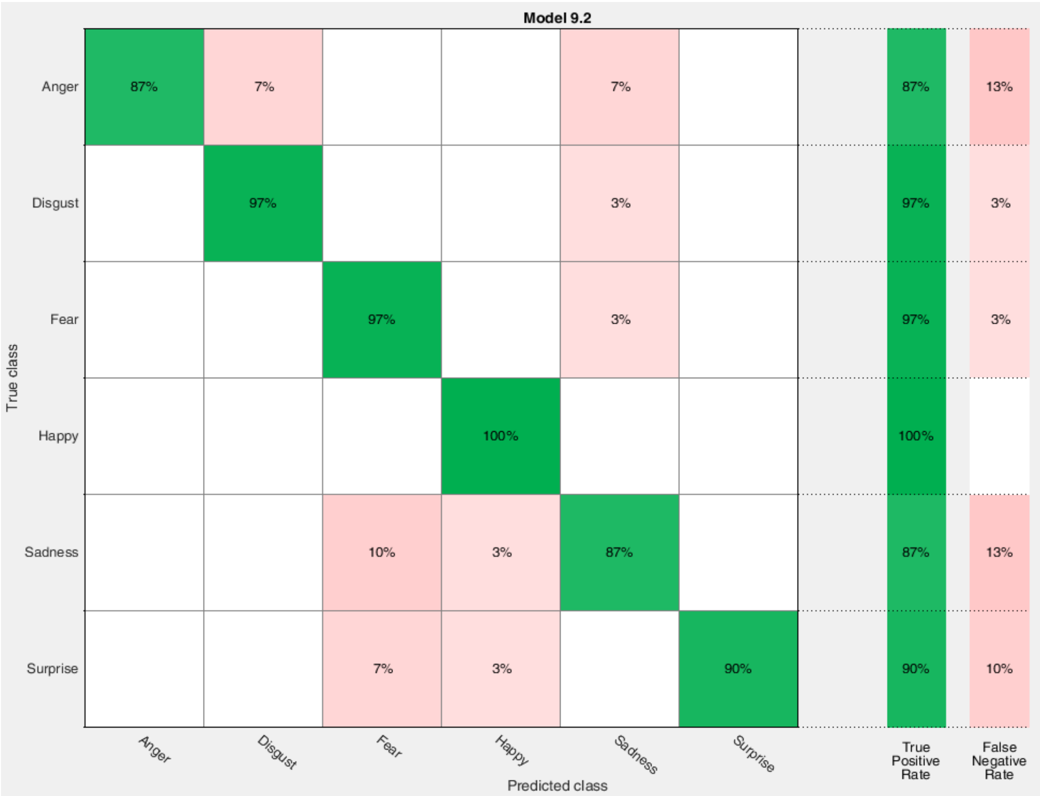


Figure 5-7: Confusion matrix for basic emotions classifier built in Matlab using Quadratic SVM and JAFFE Database

From the results, we see that best classifier for the JAFFE database outperformed the best classifier for the CK+ database by 3.6%. From our FACS-based solution in the previous chapter, the best classifier for CK+ database outperformed the best classifier for JAFFE database by 12.7%. This represents a swing of 16.3%.

To verify if the cross-validation performance of the JAFFE database is accurate, we ran our cross-validation script with the trained Quadratic SVM. For the "minus-one-person" cross-validation, the accuracy dropped to 83 correct predictions out of 183 images, or an accuracy of just 45.4%, a fall of 47.5%. For the "minus-one-emotion" cross-validation the accuracy dropped to 24 correct predictions out of 183 images, or an accuracy of 13.1%. This is lower than the accuracy of a randomized classifier which should perform around 16.7% accuracy with 6 distinct classes. This suggest that the results are not randomized and there are some misclassifications due to biasness involved.

Hence, we conclude that the best classifier using Gabor filters with CK+ Database is Quadratic SVM with an overall accuracy of 89.3% and an accuracy of 52% for fear emotion. In comparison, AU-based method with LibSVM classifier and CK+ Database achieved overall accuracy of 92.5% and 76.0% for fear emotion. With the significantly inferior results especially for the fear emotion, we decide that it is not worth testing the Gabor-based solution on the Mind Reading DVD.

5.4 Analysis and discussion

We found that under normal cross validation, the JAFFE Database performed better than CK+ Database using Gabor filters to select features. While CK+ Database can only obtain best result of 89.3% with Quadratic SVM, JAFFE Database achieved 92.9% accuracy. This result differs from using AUs as features.

However, when we use "minus-one-person" or "minus-one-emotion" method for cross validation, the accuracy of JAFFE Database dropped significantly. With the "minus-one-emotion" method, the accuracy dropped below that of a random classifier. Being worse than a random classifier suggests that this result is not random. We did not find any other research during our literature survey that reported these findings. While Gu et al. [71] did use a method equivalent to "minus-one-person" and found substantial drop in performance compared to persondependent cross-validation as well, they did not elaborate further nor provide explanations.

We think the reason this happens, is because both facial features unique to the individual and emotional features common to everyone jointly contributes to the overall expressive power of the Gabor filter. For the "minus-one-emotion" cross validation, the classifier tries to match the same emotion from the same person in the training set with the test set. However, since there are no such matches, the classifier finds a match for the same person with different emotion instead and use that emotion as the prediction. This shows that the Gabor filter is biased towards finding a facial feature match over an emotional feature match. And even though we have masked out most of the face, the Gabor filter is still able to find matching facial features.

CK+ database on the other hand, the whole dataset we used contain of only 1 emotion per person, hence a random N-fold cross-validation is as good as "minus-one-emotion" cross-validation and will not suffer the same issues as the JAFFE database. This means that 89.3% accuracy obtained using Quadratic SVM is valid and has the best performance for Gabor filters. However, while the overall accuracy is reasonably good, it is noted that Fear emotion only has a 52% accuracy rate. The performance of this classifier is mainly contributed by the high accuracies of Happy, Surprise and Disgust emotions.

These leads us to the conclusion that while Gabor filters can be used to detect emotions from facial expression given the right database, it is inferior to the FACS-based solution. In addition, any attempts to detect fear emotions from anxious expressions will have a significant amount of skepticism because of the low accuracy for detecting fear. Hence, we are unable to answer hypothesis 1.

Chapter 6

Detecting Head and Eye Motion

In this chapter, we look at analyzing the amounts of head and eye motion from facial expressions. Although we have determined from the previous chapters that we could not answer hypothesis 1, it will still be interesting to test out hypothesis 2. The first part of this chapter covers our method of studying the differences in the head and eye motion between fear and anxiety. The results are presented in the next part, and the chapter is concluded with our analysis and discussion on the results.

6.1 Method

In chapter 4.3.1, we picked the emotional words from the “afraid” and “bothered” groups from the Mind Reading DVD [19] and reclassify them under “anxiety” and “fear” using the definitions and context. Using the same videos, we extract the eye gaze and head poses using OpenFace [20].

OpenFace estimates the eye gaze and head pose by using realistic 3D face models to accurately generate large amounts of training data to learn the eye gaze and head directions [75]. The eye gaze is represented as 3D directional vectors one for each eye, and the head pose is represented as the yaw, pitch and roll in radians. Figure 6-1 shows the graphical output of the eye gaze and head pose from OpenFace.

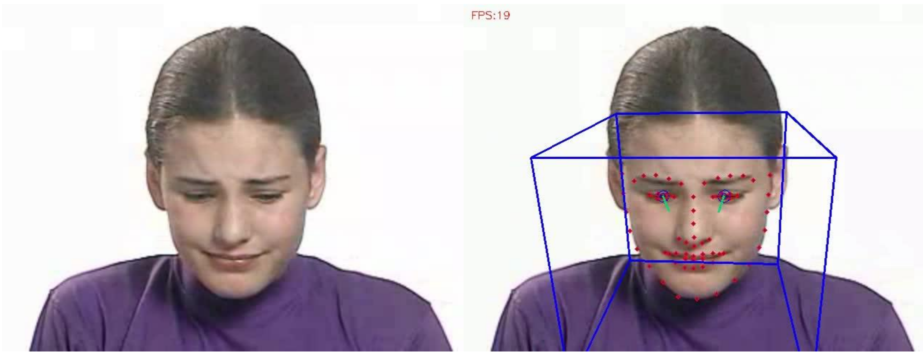


Figure 6-1: Left is the input image from Mind Reading DVD [20]. Right is the output image from OpenFace. Green lines shows the eye gaze direction, blue box shows the head orientation.

To analyze the variations of eye and head movement, we examine the x and y values for the directional vectors for the eye gaze, as well as the pitch and yaw values for the head pose. This is because we are looking at video which is in 2D, plus the actors are posing for the expressions and will naturally be looking straight into the camera, hence the z-axis and the head roll would be less important.

Hypothesis 2 predicts that fear expressions will have less eye and head movement because when there is an actual threat present, the natural response is to fix one's attention towards the source of the threat. This should result in minimal movement of the eyes and head, assuming the threat is a stationary target. Anxiety expressions on other hand, threat is unknown and unseen, hence there is no such visible target to look at. The natural response is to find the source of the threat. Hence, we should expect to find significant movement of the head and eyes in all directions so long as the threat is not identified.

We conduct a feasibility study of this method by randomly selecting 5 videos that we labelled as fear and 5 videos labelled as anxiety and extract the eye gaze and head pose information. Using Microsoft Excel, we plotted scatter graphs using x and y-axis for all the eye gaze directional vectors, and the pitch/yaw values for all the head poses. If there is relatively little eye and head movement, we should see the data points on the scatter graph to be concentrated in small areas on the scatter graph. Likewise, if there are significant amount of movement, we should see a wide spread of data points. By comparing the scatter graphs between the videos from the anxiety group vs the fear group, we should visually see some differences should hypothesis 2 be valid.

Because the feasibility study showed some positive results, we conducted a complete statistical analysis using the whole dataset. Using all 138 clips from the "anxiety" and "fear" groups, we computed the standard deviation of the x/y axis of each eye gaze, as well as the yaw and pitch of the head for each clip. The higher the standard deviation means the greater the movement for each direction. The population standard deviation S of value x for a video with N frames and mean value of x_{mean} is given by:

$$S(x) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - x_{mean})^2} \quad (2)$$

The average population standard deviation $SD_{avg}(E)$ for the emotional group E with N_E number of videos V_i is thus given by:

$$SD_{avg}(E) = \frac{1}{N_E}\sum_{i=1}^{N_E} S(V_i) \text{ where } V_i \text{ in } E \quad (3)$$

To account for video clips where the head and eye are mostly not moving but have shifted by large amounts in a small fraction of frames, thus resulting in small standard deviation, for each x/y values of eyes gaze and yaw/pitch of head pose, we compared the difference between the maximum and minimum values within each video as well. We call this the $MaxMin$ value. Thus, the average $MaxMin_{avg}(E)$ for the emotional group E with N_E number of videos V_i is given by:

$$MaxMin_{avg}(E) = \frac{1}{N_E}\sum_{i=1}^{N_E} MaxMin(V_i) \text{ where } V_i \text{ in } E \qquad (4)$$

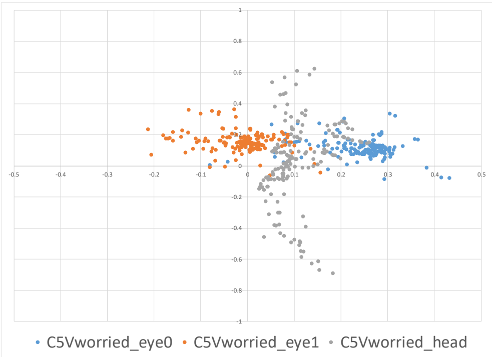
By comparing the average standard deviation and average max-min values, we should see greater values for the anxiety group compared with the fear group should hypothesis 2 be valid.

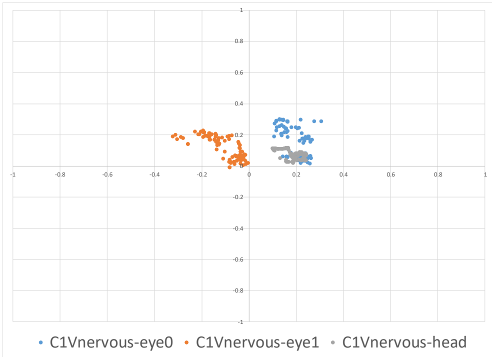
6.2 Results

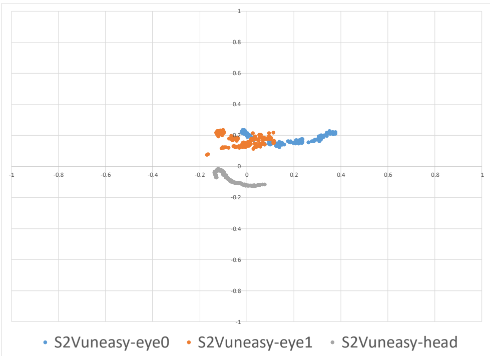
Figure 6-2 and Figure 6-3 show the scatter graphs of the eye gaze and head poses from 10 randomly selected videos from the Mind Reading DVD [19]. Each graph shows the x/y directions of both eyes and the pitch (y-axis) and yaw (x-axis) of the head pose for the same video. Grey dots are the data points for the head pose while orange and blue dots show the data points for each eye gaze direction.

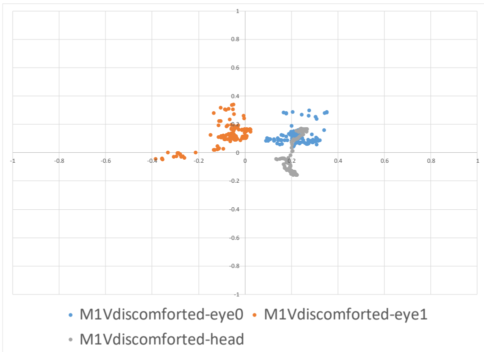
In the anxiety group (Figure 6-2), the faces in *C5Vworried* and Y3Vfrantic shows a large amount of motion in all directions, while the face in *S2Vuneasy* shows little vertical motion but significant horizontal motion. The rest of the 2 faces show moderate amount of motion in all directions. In the fear group (Figure 6-3), the faces in *Y1Vterrified* and *Y7Vthreatened* show little motion in any direction, while the rest of the faces show moderate amount of motion in all directions.

Visually from the scatter graphs, we found that hypothesis 2 may be plausible, hence we proceed to perform the statistical analysis on the whole dataset.









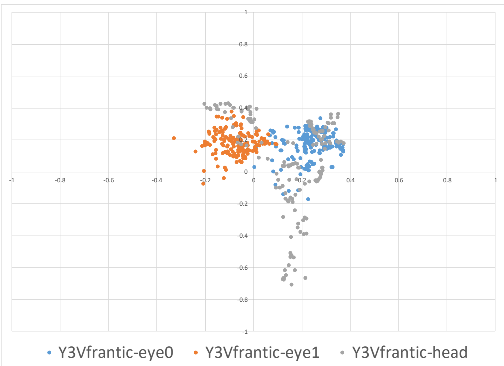


Figure 6-2: Scatter graphs of the eye gaze and head poses from 5 randomly selected videos from the anxiety group. Grey dots are the data of the head poses, orange and blue dots are data from each eye.

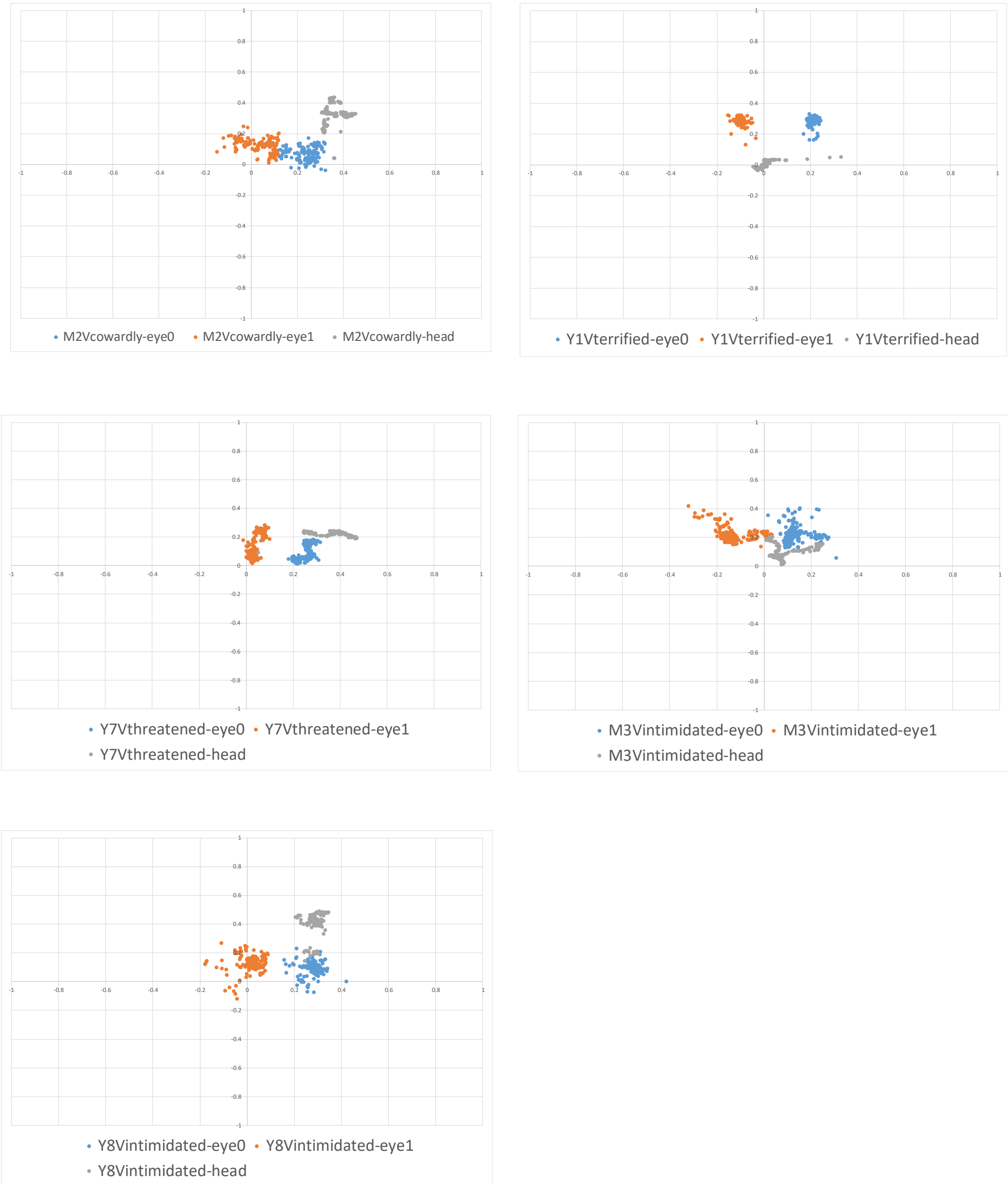


Figure 6-3: Scatter graphs of the eye gaze and head poses from 5 randomly selected videos from the fear group. Grey dots are the data of the head poses, orange and blue dots are data from each eye.

Figure 6-4 shows the average standard deviation between the anxiety and fear groups in terms of eye/head movement as well as maximum eye/head displacement. Figure 6-5 shows the average max-min values between the anxiety and fear group.

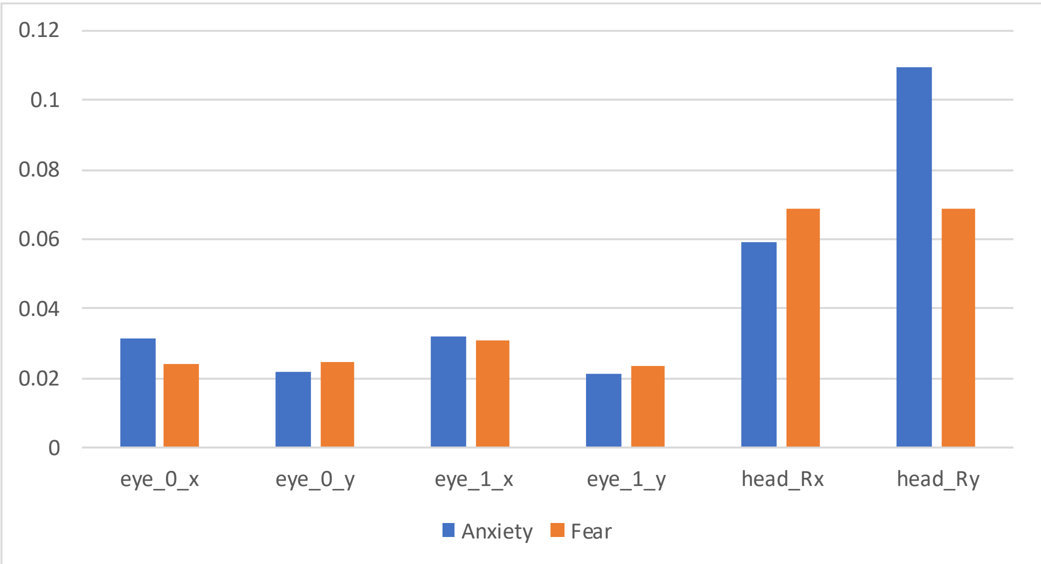


Figure 6-4: Average standard deviation comparison between Anxiety and Fear groups. The columns eye_0_x, eye_0_y, eye_1_x and eye_1_y represents x and y-axis values for each eye, head_Rx and head_Ry represents the yaw and pitch of the head respectively. Each column shows the average of the standard deviation for each video under the anxiety/fear group.

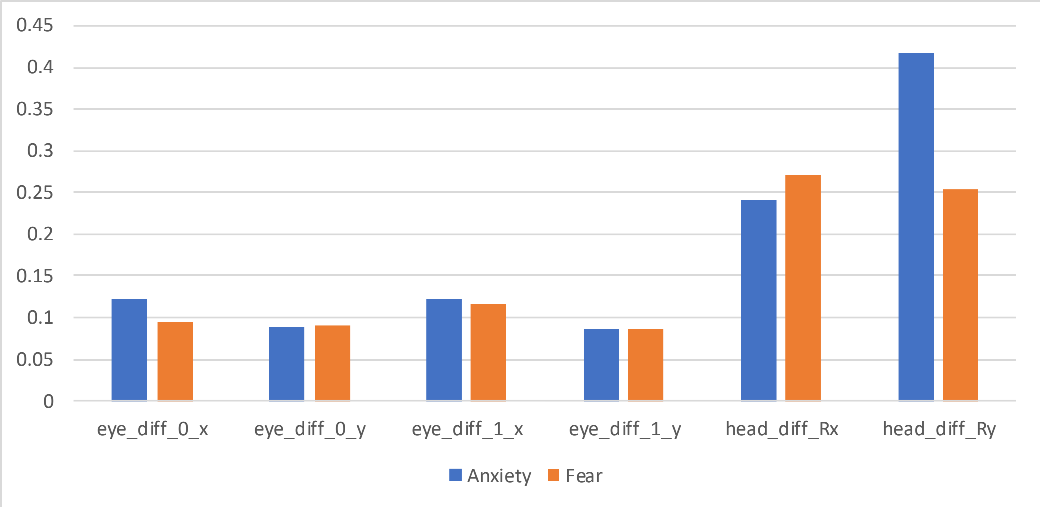


Figure 6-5: Average max-mean comparison between Anxiety and Fear groups. The columns eye_diff_0_x, eye_diff_0_y, eye_diff_1_x and eye_diff_1_y represents the difference between the maximum and minimum x and y-axis values for each eye, head_diff_Rx and head_diff_Ry represents the difference in the maximum and minimum yaw and pitch of the head respectively. Each column shows the average max-min value for each video under the anxiety/fear group.

From the results, both charts show relatively small differences between anxiety and fear for all the values (between 3.0-30.0%) other than the head pitch. For the head pitch, anxiety scored an average standard deviation of 60.1% higher than fear (0.1098 vs 0.0686). This is supported by the average max-min value where anxiety scored 64.0% higher than fear (0.417 vs 0.254). The statistics thus support the hypothesis 2 for increased head motion in anxiety but not for increased eye motion.

6.3 Analysis and discussion

In this chapter, we have done a study to test out hypothesis 2 proposed in chapter 2.3 by examining the differences in the eye gazes and head poses between anxiety and fear faces. Using OpenFace, the x/y values of the eye gaze directional vectors and pitch/yaw values of the head poses are extracted. By plotting a sample of videos on scatter graphs, we found some plausibility to the hypothesis. However, a detailed analysis using average standard deviation and average max-min values did not yield a conclusive result.

The eye/head motion scatter graphs for the faces that we sampled, 4 out of the 10 faces show visual differences in the amount of motion compared with the rest. 2 of these belong to the anxiety group and shows large amount of motion, while the other 2 belong to the fear group and show little motion. In the statistical analysis, we found 60.1% higher in average standard deviation and 64.0% higher in average max-min in the head pitch for anxiety compared with fear. However, none of the values of eye gaze directions show differences more than 30%.

While we conclude that hypothesis 2 is partially true because there are more head movement for anxiety compared with fear, there are some doubts in reaching this conclusion because there do not seems to be a reasonable explanation on why only the head pitch shows significant differences but not the yaw. And a possible reason why we obtain this strange conclusion is due to limitations of the Mind Reading DVD itself.

Due to the ambiguity of the meanings of the emotional words in the Mind Reading DVD, while doing the re-classification of the emotional words into “anxiety” or “fear” group, it isn’t so clear cut on how to differentiate them. As we have defined fear to be a visible or known danger and anxiety as unknown threat or internal conflict, some emotional words have conflicting context stories depending on interpretation. For example, (see Table 10) the emotional word "afraid" contains stories that could describe known danger, while other stories could describe unknown threat or internal conflict. When we assess the context of each story (see Table 15 in the Appendix for the whole list context stories) and attempt to classify "afraid", 3 of the stories clearly indicates "anxiety", 1 clearly indicates "fear", and the remaining 2 stories can be classified either way. The final label for "afraid" is "anxiety" because most of the stories suggest this label. With the ambiguity of the emotional words, the actors themselves may act differently depending on whether they perceive the word to be more of fear or anxiety.

Context Story	Threat Type	Reasoning
Kyle is afraid of his neighbor's dog when it barks at him. He runs away.	Known and Dangerous (i.e. Fear)	Barking dog is the threat. Dog is the source, barking indicates potential danger from angry dog.
Mike feels afraid when a stranger stops him in the street and asks him for money.	Known May or may not be dangerous (can be Fear or Anxiety)	Stranger asking for money is the source of threat. But it is unclear if his action of asking for money is perceived as dangerous or not.
Rachel is afraid when she is left alone in the house.	Unknown (i.e. Anxiety)	There is no one else around.

Context Story	Threat Type	Reasoning
Paul is afraid when he is alone at night and hear strange voices.	Unknown (i.e. Anxiety)	There is no one else around, and source of voices is unknown.
Louise is afraid when she walks alone at night.	Unknown (i.e. Anxiety)	There is on one else around.
Jessica is afraid when she hears scratching on her window late at night.	Known May or may not be dangerous (can be Fear or Anxiety)	Scratching on window gives a source of threat, but it is unclear if the scratching sound feels dangerous or not.

Table 10: The threat type categorized for the context stories of the emotional word "Afraid" from the Mind Reading DVD [19]. The first column is the context story. The middle column is the threat type that could be interpreted from the context story. The last column provides the reasoning for the interpretation.

Hence, we believe that the Mind Reading DVD is not suitable for testing anxiety without additional inputs from psychologists. However, to obtain additional resources to get psychologists input to select anxiety and fear expressions from the Mind Reading DVD, or to create our own facial expression database of anxiety is beyond the scope of the thesis. In the end, we may conclude hypothesis 2 to be partially true if and only if our dataset from the Mind Reading DVD could be validated. If we can conclusively prove hypothesis 2 to be true, it may then be easier to separate anxiety from fear. We have seen in

Chapter 7

Systems Implementation

7.1 Introduction

From our research into anxiety detection, we could not establish a working solution because hypothesis 1 failed. However, we do have working solutions for detecting basic emotions. Using our findings, we describe how we develop our systems application to identify basic emotions from the facial expression of the user in this chapter. The same application can be easily extended to detect anxiety should we find a working solution in future.

To meet our thesis objective, our application runs automatically using images from live video using the built-in camera of the device and it does not require any physical markers on the face. This chapters contain the overall architecture, details of the systems implantation, the evaluation method and results. Finally, we conclude with the analysis of the results.

7.2 System Architecture

7.2.1 Overview

The system is divided into a multi-tier architecture (Figure 7-1) for better scalability and separation of tasks. The front end consists of the web browser, which can be run from any PCs or mobile device so long as the browser supports WebRTC. The middle-tier consists of Kurento architecture, which comprise of both the Kurento Client as well as the

Media Server. The app server deals with the application logic to communicate with the client as well as with the back-end application server. The media server handles the transcoding and recording of the WebRTC video stream. The back-end consists of the application server that holds the emotions analytics module that does the actual emotions detection.

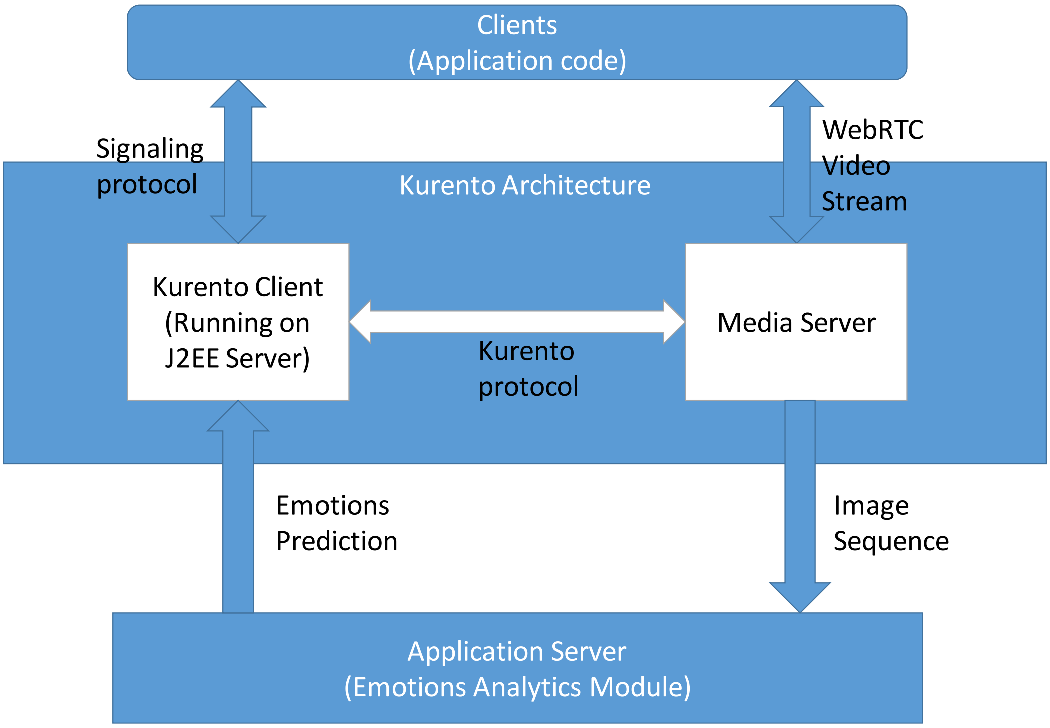


Figure 7-1: High level architecture of the emotions detection web application. The clients are the browser that runs on the end-user's device. Kurento comprises of the Kurento Client that handles the user interface and the Media Server that handles the video streaming. The Application Server hosts the emotions analytics module.

While it may be possible to implement the emotions analytics module into the Media Server as a custom module, there are 2 reasons why we separate this into an external module. One reason is due to a design consideration and the other reason is a technical issue.

Separating the emotions analytics as a separate module is a design that allows better scalability. The media server itself handles the video streaming and transcoding, hence it is both bandwidth and computationally intensive. The emotion analytics module on the other hand, while it is computationally intensive, it is not as bandwidth intensive as the media server because we do not need to extract the emotions from every frame. By running the emotion analytics module in an application server separate from the media server, it allows the media server to be optimized differently from the application server. We can scale these servers differently to cater to different demands between the emotions analytics module and media processing modules as well.

The other issue we faced when trying to implement the emotions analytics module as a custom module in the media server is a technical problem due to incompatibility between OpenCV 2 and 3. While Kurento media server provides custom modules based on OpenCV 2, OpenFace is using functions from OpenCV 3. Because we require OpenFace for extracting facial features, to integrate it as a custom module in Kurento media server means that we have dependencies for both OpenCV 2 and 3. A lot of rework have been attempted unsuccessfully to upgrade Kurento media server to support OpenCV 3, because this is not the objective of our research and we have limited time for the thesis, we abandoned this approach after a while. Instead, by separating the emotions analytics module from the media server, we could easily implement a simple interface to bridge the media server with the emotions analytics module without needing to change out OpenCV versions.

7.2.2 Kurento Client

To help develops quickly build client applications, Kurento provide 3 solutions out of the box to generate application templates to speed up development. Two of the solutions generate JavaScript based clients, and the other generates Java based client. The JavaScript client can either be client-side JavaScript code that runs directly on the web browser, or server-side JavaScript code run on top of Node.js server. Java clients can be run using J2EE Server. Figure 7-2 summarizes the 3 different client implementations. For our implementation, we create the user interface using the Java client model (see Figure 7-3).

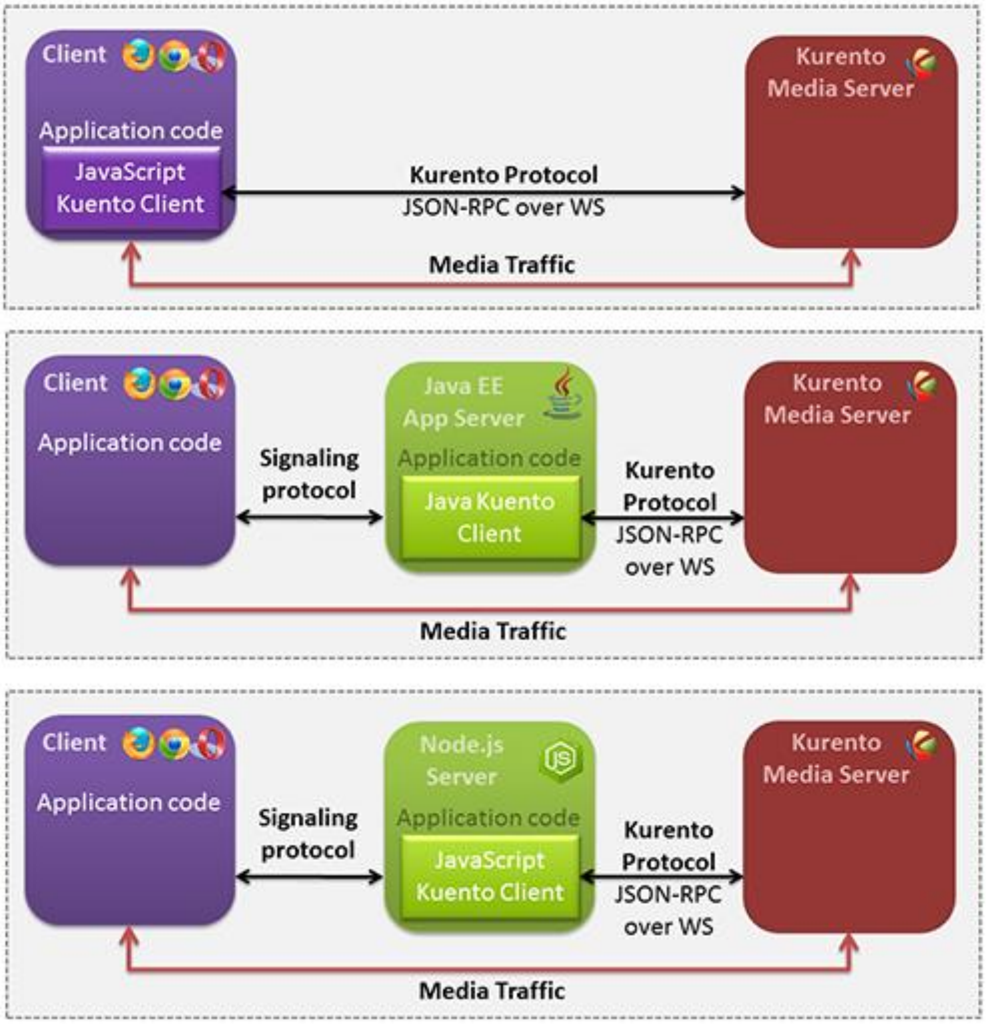


Figure 7-2: Three out-of-the-box ways to implement Kurento Clients from Kurento (image taken from Kurento documentation [81])

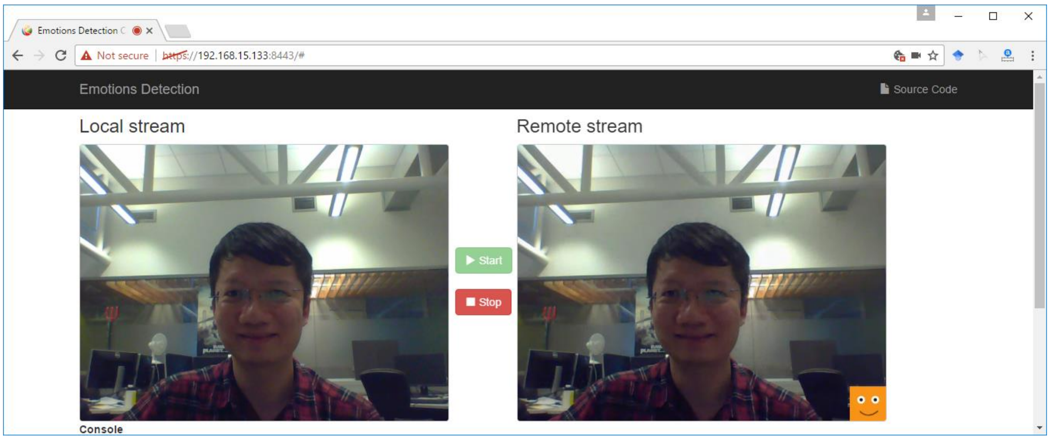


Figure 7-3: Screen capture of the Java client user interface using Chrome browser. Left screen shows the local stream from the webcam. Right screen shows the remote WebRTC stream coming from Kurento media server. The smiley face at the bottom right of screen shows the detected emotion.

7.2.3 Kurento Media Server

The media server provides utilities to generate customized GStreamer and OpenCV module templates. The GStreamer module allows applications to consume the video stream without needing to implement code for the WebRTC protocol. The OpenCV module builds on top of the GStreamer module extracts the image sequence from the video stream. We implement an OpenCV module to interface with the emotions analytics module and send the image sequences over as input. Because of the simplicity of the task, this custom module is a light-weight module and will not add much demands to the computing requirements of the media server.

7.2.4 Applications Server

This consists of a standalone C++ application that reads in image sequences and generates the prediction of emotions from the facial expression detected in the images. In chapter 3.2.1 we covered OpenFace [20] as an open source C++ application capable of several tasks, like facial landmark detection and tracking, head pose tracking, facial action unit recognition, and gaze tracking. We extended OpenFace by adding a classifier to detect 6 basic emotions (anger, disgust, fear, happiness, sadness and surprise) instead of stopping at detecting AUs. This engine is trained using LibSVM [84] classifier with the Cohn-Kanade (CK+) database [17] using the parameters we learnt from chapter 4.2. The application takes in image sequences as input and output the emotions detected as results.

7.3 Prototype

7.3.1 Specifications

For our implementation, we created a virtual machine (VM) on a single physical laptop computer to host all the server-side components. The laptop is running on 64-bit Windows 10 Pro, running on Intel Core i7-4712MQ processor with 8GB RAM and 1TB Samsung 840 EVO SSD. The VM is created using VMWare Workstation 12 Player (Build 3272444). The VM is installed with Ubuntu Desktop 14.04 configured with 2 processing cores, 4GB of RAM and 60GB disk space. Kurento Media Server 6.6.1 is installed inside the VM. We tested the client-side application on Google Chrome 56, Firefox 47 and Opera 43 outside the VM and on the same physical machine.

7.3.2 Implementation

In this prototype, we developed it to run for a single user only. The aim of this prototype is to demonstrate a working emotions detection system that can run on any web browser that supports WebRTC. The ability to function for multiple users concurrently will require an implementation using sessions, and this is outside the scope of the thesis.

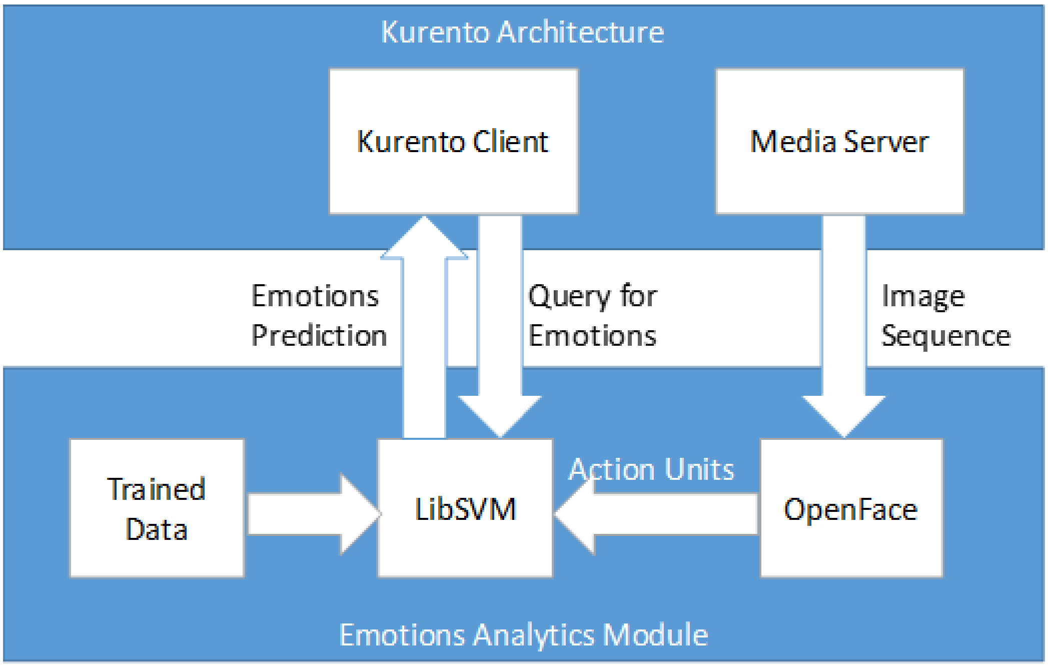


Figure 7-4: Inside the Emotions Analytics Module. It comprises of OpenFace, LibSVM and trained data of the SVM classifier.

The emotions analytics module (Figure 7-4) comprise of OpenFace, LibSVM, and data from the trained classifier. OpenFace takes in image sequences extracted by the Kurento Media Server OpenCV custom module from the WebRTC video stream to extract AUs. The basic emotions classifier is pre-loaded using the trained data by LibSVM. With the AUs from OpenFace, the trained LibSVM classifier then outputs the emotions prediction.

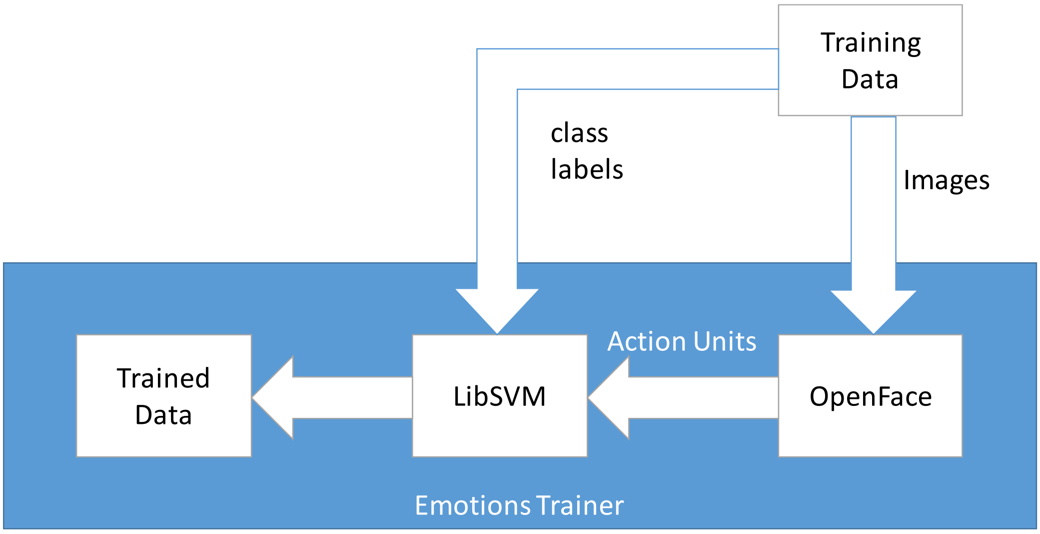


Figure 7-5: Inside the Emotions Trainer. It comprises of OpenFace and LibSVM, takes in input from training data and outputs the trained data.

For the trained data, these is created by a separate standalone Emotions Trainer (Figure 7-5) module. Images from the training data is fed into OpenFace to extract the AUs of the facial expressions, while the class labels are fed into LibSVM together with the AUs to train the classifier. The parameters of the classifier are then saved for the emotions analytics module.

7.4 Evaluation of System

To evaluate our systems application, we test its ability to detect emotions using live video automatically. To do that, we had wanted to conduct an experiment with live participants to test its accuracy. However, there was insufficient time and resources to complete this experiment.

Hence, we performed the experiment by simulating live videos by feeding in pre-recorded videos to the system as video streams. The following section details our testing procedure and results.

7.4.1 Procedure

We obtained the AM-FED Database [72] which contains the facial expression video recordings of viewers watching 3 Super Bowl commercials. (See chapter 3.1.3 for the details on the AM-FED Database) Each video from the AM-FED database comes with the results of how much the viewers liked each commercial in the form of survey questions. We use the results of questions 1 (*Did you like the video?*) and 3 (*Would you watch it again?*) to obtain the ground truth of how much they like each commercial. We then test how much the viewers like each commercial by detecting happy emotions and comparing our results with the ground truth.

To use the videos from the AM-FED database instead of live video streams from the camera, we convert the videos to produce an image sequence for each. These image sequences are then fed as video streams into the system instead of obtaining from the camera. We do this by hijacking our Kurento OpenCV custom module to discard images from the camera and replacing them with the images from AM-FED videos. In this way, the system thinks the images from the video stream are "live" from the camera.

7.4.2 Evaluation methodology

To evaluate our application, we used 2 methods. The first method obtains an emotional score of each commercial using happy emotions detected from the facial expressions of viewers and compare with the ground truth. The second method obtain the emotional score of each response type (positive, neutral, negative) using happy emotions detected and compare with the ground truth.

Method 1

To obtain the ground truth of the popularity of each commercial, we looked at the 2 questions “Did you like the video” and “Would you watch it again”. However, because there are only 1 positive option, 1 negative option and 1 “average” option to each question, it is hard to determine the level of “likeness” to the commercial. The options “Meh! It was ok.” And “Maybe, if it came on TV” are ambiguous and imply a slight preference towards “don’t like” or towards “like” depending on one’s interpretation. Hence, we are only going to consider the negative (Na… not my thing/Ugh. Are you kidding) and positive (Heck ya! I loved it/You bet!) options since there are no ambiguity for these responses.

We assign a value of +1 for each positive response and a value of -1 for each negative response. The sum of positive and negative response values will form the “ground truth” for each commercial.

We define the emotional score E_{V_i} of a video clip V_i as the percentage of frames detected for that emotion E. This can be computed with the following equation (where N_i is the number of frames in V_i):

$$E_{V_i} = \frac{1}{N_i}\sum^{N_i} E \qquad (5)$$

We define the emotional score E_C of a commercial C as the average emotional score of the video clips filmed from viewers watching that commercial. This can be computed with the following equation:

$$E_C = \frac{1}{V_j} \sum^{V_j} E_{V_j} \text{ for all } V_j \in C \tag{6}$$

We should see a direct relationship between the ground truth and the happy emotional score of each commercial if our basic emotions detector works.

Method 2

To compute the ground truth of each response type *i (positive, neutral/ambiguous, negative)*, we simply sum up the total number of each response type per question.

We calculate the happy emotional score E_{iQ} for each question Q in a similar way as how we calculate the happy emotional score for each commercial. The formula is:

$$E_{iQ} = \frac{1}{V_j} \sum^{V_j} E_{V_j} \text{ for all } V_j \in Q_i \tag{7}$$

7.4.3 Test Results

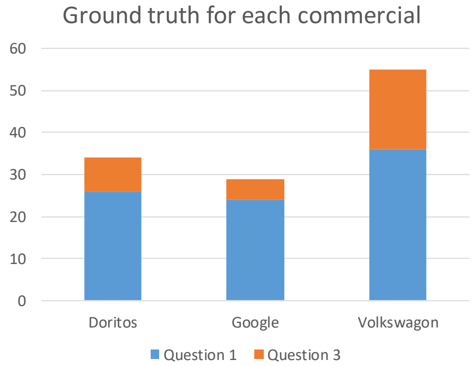
The following are the results obtained for comparing commercials.

Commercial Name	Emotional Score (Happy)
Doritos	9.54%
Google	5.24%
Volkswagon	15.3%

Table 11: (Happy) Emotional score for each commercial measured using emotions detected from viewers

Commercial Name	Positive Response	Negative Response	Ground Truth
Doritos	38	4	34
Google	42	13	29
Volkswagon	60	5	55

Table 12: Ground truth values computed using positive and negative responses to survey questions 1 and 3 for each commercial in AM-FED Database



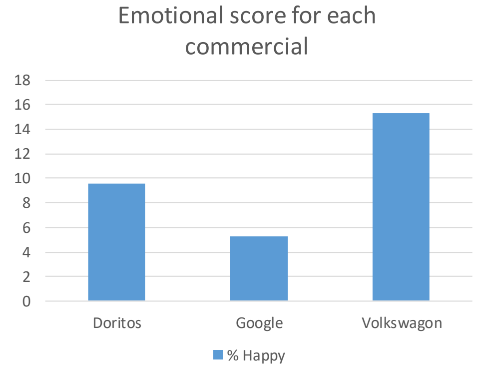


Figure 7-6: Left chart shows the "ground truth" scores for each commercial based on Table 12. Right chart shows the (happy) emotional scores for each commercial from Table 11.

From Figure 7-6, we can see as expected, the relative likeness for each commercial using the emotional score is comparable to the "ground truth" for each commercial.

The following are the results for comparing the emotional score for each response type.

Response Type	Question 1	Question 3
Positive	13.94%	15.76%
Neutral/Ambiguous	6.12%	9.77%
Negative	3.01%	1.86%

Table 13: (Happy) Emotional score for each response type per question measured using emotions detected from viewers.

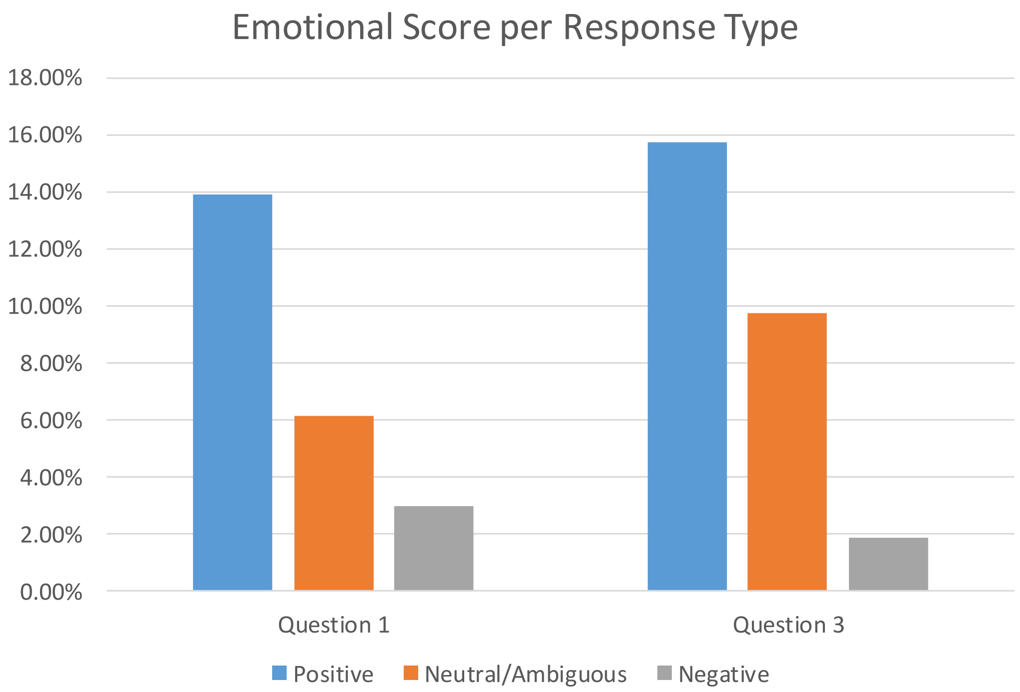


Figure 7-7: (Happy) Emotional score comparison for each response type per question based on Table 13.

From Figure 7-7, we can see that as expected, positive responses result in a higher proportion of happy emotions detected from the facial expressions than negative responses.

7.5 Analysis and discussion

In this chapter, we presented an N-tier architecture for a web-based emotions analytic application by utilizing WebRTC for the front-end client application, Kurento for the middle-tier media server and OpenFace for the back-end engine. Using this design, we developed a prototype application by training a basic emotions classifier using LibSVM and AUs extracted from the CK+ Database by OpenFace. We show that the prototype can detect emotions using live video. Using the prototype, we extracted the emotions from viewers' facial expressions when they watch commercials. By comparing the level of happy emotions detected and matching the "ground truth" of the relative likeness of the commercials obtained from the survey questions, we demonstrated that our systems application can detect basic emotion "happy" successfully from live video without using facial markers.

Chapter 8

Conclusion

In chapter 1, we proposed 2 hypotheses on how to identify anxiety from facial expressions. In this chapter, we summarize the conclusions to the hypothesis and highlight the main findings that we have discovered during our research. Limitations of our research are summarized with proposed future work that could address some of these limitations.

8.1 Summary and Findings

Hypothesis 1 suggest that basic emotion "fear" can be detected in the facial expressions of complex emotion "anxiety". In chapter 4, using the FACS-based basic emotions classifier, we could not detect fear emotion in 11 video clips of anxiety out of a total of 108 clips. And among the 97 video clips detected with fear emotion, only 23 clips have "fear" as the main emotion detected. There are more clips with sadness (34) and with surprise (25) as the main emotion detected. Hence, we are unable to make any validity claims for hypothesis 1, although we can't rule it out completely either.

Hypothesis 2 suggest that anxiety produces greater eye and head movement compared with fear. In chapter 6, we found that for anxiety compared with fear, the average standard deviation and average max-min of head pitch higher. Anxiety has average standard deviation and average max-min head pitch of 0.1098 and 0.417 respectively. Fear has average standard deviation and average max-min head pitch of 0.0686 and 0.254 respectively. For eye gaze direction, the differences between anxiety and fear for all values do not exceed 30%. This lead us to conclude that hypothesis 2 is partially true for head motion.

We have an interesting finding in chapter 5 when we try to detect basic emotions using Gabor-based solution. While this finding did not directly contribute to our main research, it came as a surprise because we did not find any other research literature having reported this finding. The finding is that using Gabor-based features for a data driven solution to detect emotions, the nature of the database used for training and the method of cross-validation can play a *greater than expected* significance regarding the accuracy of the classifier. The consequence is that, for every iteration of cross-validation, if the test set contains of *all images of one emotion* from one individual, and the training set contains the remaining data which includes *all other emotions from the same individual*, the cross-validation accuracy for the classifier can drop *below a random classifier*.

8.2 Limitations and Future Work

There are some limitations regarding our conclusion to the hypotheses. The main limitation is with the Mind Reading DVD where we obtained our test set for our experiments. The video clips in the DVD are filmed by actors posing in front of the camera while imagining scenarios to express different moods associated to each emotional word. None of these words clearly described anxiety or fear. Despite our attempts to re-classify the words, we found the task far from straight forward. This limits our ability to correctly determine which test data represent anxiety and which ones represent fear. Because of this, we have some doubts with our conclusion to both hypothesis 1 and 2.

Future work may involve repeating similar set of experiments but with collaboration with psychologists to build new facial expressions databases with actual patients suffering from anxiety disorders. A control set with people who are verified to be free from anxiety should be made as well. The whole experimental procedure should also be supported by the psychologists. With the new set of test data, we can run more experiments to attempt to test out the hypotheses again.

Appendix A

Emotional Word	**Definition**
Afraid	Unwilling to do something or worried about doing something because you are frightened about what may happen if you do it
Consternation	A feeling of anxiety or shock often caused by an unexpected event
Cowardly	Not brave enough to do something you should do or to face up to a dangerous situation
Cowed	Defeated or frightened into doing what someone else wants
Daunted	Discouraged by something that you think is going to be difficult so that you do not even attempt it
Desperate	Feeling that you have lost hope and are filled with despair
Discomforted	Made to feel uneasy, worried, or embarrassed by someone or something
Disturbed	Very worried or upset about something that you find unpleasant
Dreading	Feeling something in the future, feeling that something bad will happen
Frantic	Agitated and distracted with worry

Emotional Word	Definition
Intimidated	Feeling threatened and scared by someone, something, or a situation, so that you are frightened into submission
Jumpy	Feeling nervous and unable to relax, so tense that you are surprised by the slightest sound or movement
Nervous	Tense and worried that something might happen, so that you cannot relax
Panicked	Suddenly feeling very frightened or worried
Shaken	Made to feel upset or worried by something that has happened to disturb your peace or balance
Terrified	Extremely frightened or panicked
Threatened	Feeling anxious or afraid because someone else is behaving in an aggressive or threatening way
Uneasy	Feeling slightly worried or anxious about something
Vulnerable	Feeling unprotected and easily hurt, attacked, or struck down by illness
Watchful	Noticing everything that is going on around you, as you are frightened that something unpleasant might happen
Worried	Unsettled or anxious because you keep thinking about a problem or you think something bad is going to happen

Emotional Word	Definition
Bothered	Disturbed, worried, or upset about something
Flustered	Feeling a bit agitated or confused, often because you have too much to do or too little time
Impatient	Feeling unwilling to wait for something or someone or getting annoyed because something is not happening quickly enough
Pestered	Feeling annoyed and harassed because people keep asking you silly questions or interrupting you
Restless	Unable to stay still or remain in one place because you are bored or nervous
Ruffled	Feeling a little flustered because someone has interrupted your concentration or rest
Tense	Feeling nervous and unable to relax, often because you are worrying about something that is going to happen

Table 14: Definitions of emotional words from the "Afraid" and "Bothered" group in the Mind Reading DVD [19]

Emotional Word	Stories
Afraid	Rachel is afraid when she is left alone in the house. Kyle is afraid of his neighbour's dog when it barks at him. He runs away. Paul is afraid when he is alone at night and hear strange voices. Louise is afraid when she walks alone at night.

Emotional Word	Stories
	Mike feels afraid when a stranger stops him in the street and asks him for money. Jessica is afraid when she hears scratching on her window late at night.
Consternation	Ben feels consternation when the security man accuses him of stealing. Naomi feels consternation when she gets a credit card for several thousand pounds. Mark feels consternation when his fish start dying for no obvious reasons. Heather feels consternation about how the wrong medicine could have been given to the bird. She checked it twice before giving it to him. Peter feels consternation when his local shop closes without warning. Rose feels consternation when she hears that a hurricane is approaching
Cowardly	Ben feels cowardly when he runs away from a fight in the pub. Kim is cowardly because she won’t own up to breaking her mom’s favourite vase. Mark feels cowardly when he doesn’t say anything to the man bothering the girl on the train. Heather feels cowardly when she sees a spider and runs away.

Emotional Word	Stories
	Arthur feels cowardly and won't go on the rollercoaster with his nephew. Rose is cowardly about going to the dentist and cancels her appointment.
Cowed	John feels cowed when his boss criticizes him because customers have complained about him. Rachel feels cowed when her patient shouts at her. Drew feels cowed by the health officer and agrees to improve his hygiene standards. Mona is cowed when her son's headmaster asks if there are problems at home. Arthur feels cowed when his boss shouts at him for treating his secretary badly. Julie feels cowed when the headmaster tells her she must stop gossiping and get back to her class.
Daunted	Ben feels daunted when he realizes he has exams in two weeks and has yet to do any revision. Rachel feels daunted by her new job because she has no experience. Kyle is daunted by the work involved in building an extension to their house. Mona is daunted by the size of the hill she has to climb. Arthur is daunted by the pile of work on his desk and doesn't know if he can get through it all in time.

Emotional Word	Stories
	Julie is daunted at the prospect of cooking dinner for twenty-five people. She's never cooked for so many people before.
Desperate	Sarah feels desperate when her purse is stolen because all her money is in it. Peter feels desperate when he hears about his friend's accident. Arthur feels desperate about the mistake he made and all the trouble it has caused. Louise feels desperate to try and help. Joe feels desperate when he loses his essay. It will take hours to write again. Belinda is desperate to get away from the boy who is chasing her.
Discomforted	John feels discomforted when he has to wear a skirt and tights on a medieval night at the restaurant. Sarah is discomforted when she hears people whispering around her. Mark feels discomforted by the rowdy boys on the bus and moves towards the front. Mona feels discomforted by the man who keeps staring at her in the bar. Arthur is discomforted when a man he recently sacked joins his club. Carol is discomforted when her son finds her drinking whiskey in the middle of the afternoon.

Emotional Word	Stories
Disturbed	John feels disturbed when he hears his neighbor screaming in the night. Kim feels disturbed after watching a program about homeless people. Drew feels disturbed by the news reports of violence in the area near his children's school. Mona is disturbed when she hears a news report saying that a school bus has crashed nearby. Her son went on the bus to school today. Arthur is disturbed by the smell of gas in the street and calls the emergency gas number to report it. Rose is disturbed by the strange noises she hears in the night.
Dreading	Ben is dreading his final exams. He hadn't studied hard enough. Sarah is dreading her performance review because her boss is very strict. Peter is dreading the visit from the safety officer because he knows something always goes wrong when she comes. Sandra is dreading going back to work after her holiday. She has such a lot of work to catch up on. Paul is dreading going to the dentist because he knows he will have to have a tooth out. Julie is dreading the letter from the bank.

Emotional Word	Stories
Frantic	Ben is frantic when his girlfriend storms out after a row and doesn't call him in three days. Kim is frantic when the little girl she is babysitting disappears in the park. Peter is feeling frantic because his girlfriend is late home from work and hasn't called. Mona is frantic when the school phones to say her son has had an accident. Arthur feels frantic when his nephew borrows his car and doesn't return it. Rose is frantic when she realizes her purse has been stolen
Intimidated	John feels intimidated by his girlfriend's brother, who is a boxer. Naomi feels intimidated by the big bouncer at the nightclub. Kyle feels intimidated by his boss. Sandra feels intimidated by people in foreign countries when they talk fast and loud in a language she does not understand. Tom feels intimidated when he goes into the smart shop because he thinks the shop assistants are all staring at him. Rose is intimidated by the newspaper editor because he rejected the last article she wrote.

Emotional Word	Stories
Jumpy	Ben feels jumpy whenever he is around small dogs, as he was bitten by one once. Sarah feels jumpy when she is walking home alone at night. Kyle feels jumpy after nearly crashing his car. Mona feels jumpy when she hears strange noises during the night. Arthur feels jumpy when he goes into the dark, empty house to look for his nephew. Rose feels jumpy walking back to the hotel along the dark shadowy streets.
Nervous	Nicholas is nervous about meeting his new client. She is a famous actress. Drew is nervous about giving the best man's speech at his friend's wedding. Mona is nervous of dogs because she was bitten once. Carol feels nervous about speaking in front of lots of people. Joe is nervous when he meets his stepfather's parents for the first time. Sally is nervous when her family come to watch her netball game.
Panicked	Nicolas is panicked when he forgets something important.

Emotional Word	Stories
	Peter feels panicked when he realizes he left the front door open all night. Kim feels panicked in the swimming pool by the boys who jump in nearly on top of her. Julie is panicked when the pipe bursts and she can't think what to do. Joe feels panicked when the boss breaks an expensive glass. Belinda feels panicked when she realizes she's going to be late for the exam.
Shaken	Nicolas feels shaken when he sees the car accident. Sarah feels shaken after she sees a woman being mugged. Drew is shaken by the news that his house has been broken into. Heather feels shaken when there is fire at the house next door. Paul is shaken when he opens the letter saying he will be prosecuted. Rose is shaken when she hears that an old school friend has been murdered
Terrified	John is terrified when he hears someone break into the house. Sarah is terrified that she will drown as she could not swim.

Emotional Word	Stories
	Sandra is terrified of flying so she never travels by plane. Nicholas is terrified when his kitchen catches fire. Tim feels terrified when he sees a snake in the grass. Louise is terrified that her husband will lose his job.
Threatened	Nicolas feels threatened by the large gang on the street. Sarah feels threatened when the man holds up a knife. Kyle feels threatened by the unpleasant letter he receives. Rose feels threatened by the man following her along the street. Joe feels threatened by the neighbour's large dog. Belinda felt threatened by the big girl who pushed her.
Uneasy	John feels uneasy when he walks home alone at night. Sarah feels uneasy about walking into a crowded bar on her own. Drew is uneasy about leaving his kids with the young babysitter. Mona feels uneasy when her son is visiting his father as something always goes wrong.

Emotional Word	Stories
	Tom is uneasy about looking after his grandchildren all by himself for two weeks. Rose feels uneasy about sharing a room with the strange woman.
Vulnerable	John feels vulnerable on the train when he is surrounded by noisy football fans. Kim feels vulnerable when she is the only girl at the party. Kyle feels vulnerable when he is at the top of the ladder in the blazing house. Heather feels vulnerable when she tells her brother about her fears. Arthur feels vulnerable when he goes for a helicopter flight and they flew into fog. Naomi feels vulnerable arriving in a strange country all by herself.
Watchful	Ben is watchful when he sees a prowler next door. Kim is watchful when she is looking after the very ill patient. Drew is watchful as he walks through the dark stream at night. Heather feels watchful of the young children playing with fireworks. Paul is watchful when the boys are in the library because they've stolen books.

Emotional Word	Stories
	Julie is watchful as the students arrive at the disco. They aren't supposed to bring any alcoholic drinks.
Worried	Ben is worried that he will fail his exam. Mark is worried that his fish aren't eating properly. Sandra is worried because she can smell smoke. Carol is worried when she finds her mother looking pale and shivering. Mike is worried that his dad will find out he scratched the bible. Sally is worried that she will not be selected for the team.
Bothered	John feels bothered when he discovers there is money missing from the bill. Rachel feels bothered by her brother's swearing and wishes he would stop. Mona feels bothered by the arguments that her parents are having. Paul is bothered by his noisy neighbours at night and can't get to sleep. Mike is bothered by having such a huge pile of homework. Jessica feels bothered when her teachers ask her about her career plans. She doesn't have any but knows she should have some ideas by now.
Flustered	Ben feels flustered when the landlord turns up to look around the flat. The flat is in a terrible mess.

Emotional Word	Stories
	Kim feels flustered when she's going to be late for an important job interview. Peter feels flustered by the exam questions. He should have studied harder. Heather is flustered when her parents ask why she hasn't been round to see them recently. Paul is flustered by the very difficult problem. No-one seems to be able to help. Carol feels flustered when the dinner guests arrive half an hour early and she's still in her dressing gown.
Impatient	Rachel is impatient to get going and tells her brother to hurry up. Drew feels impatient at the doctor's because he's been waiting for so long. Arthur feels impatient when his secretary takes so long to type the letters. Julie feels impatient to hear what the boy has to say. Tim is impatient for his parents to hurry up so that can go out. Sally is impatient when she can't do her homework.
Pestered	Ben feels pestered by the children's questions when they keep asking such a lot.

Emotional Word	Stories
	Rachel feels pestered by the office junior when he won't stop asking her for a date. Mark feels pestered by stupid questions that other teachers ask him. Sandra feels pestered by her mother's questions and tells her she will speak to her tomorrow. Paul feels pestered when the women keeps ringing up to ask if he'll be on the committee even though he's said no. Julie feels pestered by the double glazing salesman who keep phoning her in the evenings.
Restless	Nicolas feels restless in his job. It might be time to look for another one. Kim feels restless because she has been studying for a week and hasn't been able to go swimming. Drew is restless to do some exercise. He hasn't played football or been to the gym for ages. Heather feels restless while waiting for news of her job application. Tom is restless and can't settle to reading or watching TV so goes for a walk. Julie feels restless about her students. They might do well in their exams but they might also do badly.
Ruffled	Ben feels ruffled when he gets an A-minus and not a straight A for his essay.

Emotional Word	Stories
	Kim feels ruffled when the man pushes her out of the way on the train. Kyle is ruffled when his wife tells him that she might leave her job. Heather feels ruffled by the silent phone calls that seems to happen every night. Tom is ruffled when the girl in front of him shouts so loudly at her little girl. Julie is ruffled when the shop assistant won't accept her credit card.
Tense	Ben feels tense before his exams. Sarah feels tense when she thinks about the creepy man at work. Drew is tense when he hears the news about the fire in his village. Louise feels tense before meeting her boss and needs at least an hour to get ready for their discussion. Paul feels tense when he waits to find out if he'll be offered the job. He can't sit still and walks up and down the corridor. Naomi always feels tense when she goes to the dentist because she hates having treatment.

Table 15: Example stories on how emotional words from "Afraid" and "Bothered" groups are used in the Mind Reading DVD.

Filename	Anger	Disgust	Fear	Happy	Sadness	Surprise	Most Frequent
0600101C4Vafraid	3	10	91	0	18	3	Fear
0600101C5Vafraid	12	10	39	0	34	46	Surprise
0600101M3Vafraid	27	10	37	0	2	49	Surprise
0600101M8Vafraid	0	0	7	0	45	43	Sadness
0600101S3Vafraid	1	0	37	0	91	4	Sadness
0600101Y6Vafraid	13	6	70	0	20	16	Fear
0600202M4Vdisturbed	4	67	12	0	2	40	Disgust
0600202M5Vdisturbed	1	0	12	0	100	12	Sadness
0600202S1Vdisturbed	12	43	33	1	0	36	Disgust
0600202S2Vdisturbed	0	0	52	0	95	0	Sadness
0600202Y1Vdisturbed	22	39	2	0	45	17	Sadness
0600202Y4Vdisturbed	24	110	19	0	0	0	Disgust
0600305C2Vworried	15	77	29	4	1	6	Disgust
0600305C5Vworried	32	7	13	25	0	58	Surprise
0600305M1Vworried	25	44	57	6	19	0	Fear
0600305M6Vworried	4	0	0	0	106	15	Sadness
0600305S4Vworried	25	1	22	0	69	17	Sadness
0600305Y3Vworried	12	0	20	0	29	77	Surprise
0600402C1Vnervous	22	0	31	0	15	17	Fear
0600402C2Vnervous	38	11	45	25	3	14	Fear
0600402M4Vnervous	19	68	10	29	0	0	Disgust
0600402M5Vnervous	16	0	23	0	60	46	Sadness
0600402S4Vnervous	13	0	59	0	24	32	Fear
0600402Y7Vnervous	1	28	42	0	30	24	Fear
0600601M1Vconsternation	67	33	17	0	8	0	Anger
0600601M2Vconsternation	43	28	27	0	0	11	Anger
0600601M7Vconsternation	0	0	0	0	81	14	Sadness
0600601S2Vconsternation	2	0	0	0	123	0	Sadness
0600601Y3Vconsternation	61	4	2	0	60	37	Anger
0600601Y8Vconsternation	50	0	14	0	75	11	Sadness
0601101M1Vdiscomforted	13	1	33	0	71	7	Sadness
0601101M4Vdiscomforted	20	51	5	0	44	5	Disgust
0601101S1Vdiscomforted	0	26	36	35	6	17	Fear
0601101S4Vdiscomforted	9	0	30	0	48	36	Sadness
0601101Y1Vdiscomforted	17	91	4	0	13	0	Disgust
0601101Y2Vdiscomforted	5	0	99	0	2	24	Fear

Filename	Anger	Disgust	Fear	Happy	Sadness	Surprise	Most Frequent
0601201M6Vdreading	0	0	0	0	119	6	Sadness
0601201M7Vdreading	40	0	0	0	11	53	Surprise
0601201S3Vdreading	31	24	9	0	15	60	Surprise
0601201S6Vdreading	64	6	56	0	29	6	Anger
0601201Y2Vdreading	6	0	60	51	0	21	Fear
0601201Y3Vdreading	28	10	13	0	42	54	Surprise
0601303M3Vjumpy	5	13	35	0	56	16	Sadness
0601303M4Vjumpy	23	5	32	0	9	68	Surprise
0601303S1Vjumpy	19	22	31	0	22	31	Fear
0601303S2Vjumpy	1	10	37	0	69	8	Sadness
0601303Y2Vjumpy	33	0	28	0	30	34	Surprise
0601303Y3Vjumpy	0	0	43	0	8	112	Surprise
0601401M4Vfrantic	12	69	35	6	0	3	Disgust
0601401M7Vfrantic	14	0	13	0	40	58	Surprise
0601401S1Vfrantic	25	14	25	3	19	63	Surprise
0601401S2Vfrantic	5	0	79	35	1	5	Fear
0601401Y3Vfrantic	8	13	17	0	26	80	Surprise
0601401Y4Vfrantic	53	48	20	0	2	17	Anger
0601501C1Vpanicked	39	7	36	0	55	21	Sadness
0601501C6Vpanicked	14	7	24	0	0	21	Fear
0601501M7Vpanicked	0	2	0	0	41	25	Sadness
0601501S6Vpanicked	10	0	10	0	116	4	Sadness
0601501Y4Vpanicked	8	12	17	46	34	8	Happy
0601501Y7Vpanicked	0	0	21	0	43	66	Surprise
0601901M4Vuneasy	2	124	6	0	0	0	Disgust
0601901M5Vuneasy	26	0	41	0	25	33	Fear
0601901S2Vuneasy	2	0	0	0	138	0	Sadness
0601901S5Vuneasy	54	141	18	0	0	7	Disgust
0601901Y1Vuneasy	0	0	33	0	19	73	Surprise
0601901Y2Vuneasy	0	0	83	0	1	67	Fear
0602001M2Vvulnerable	1	124	0	0	0	0	Disgust
0602001M3Vvulnerable	4	0	22	0	22	77	Surprise
0602001S1Vvulnerable	0	0	60	0	17	48	Fear
0602001Y1Vvulnerable	5	0	20	0	41	61	Surprise
0602001Y4Vvulnerable	9	11	1	0	37	67	Surprise

Filename	Anger	Disgust	Fear	Happy	Sadness	Surprise	Most Frequent
0602001Y8Vvulnerable	11	0	31	0	98	0	Sadness
0602102M2Vwatchful	0	55	6	0	64	0	Sadness
0602102M5Vwatchful	0	0	1	0	46	89	Surprise
0602102S3Vwatchful	0	0	59	0	64	2	Sadness
0602102S6Vwatchful	59	0	1	0	65	0	Sadness
0602102Y3Vwatchful	0	0	7	0	106	49	Sadness
0602102Y4Vwatchful	0	0	3	0	30	99	Surprise
1800201C4Vbothered	35	22	9	0	26	33	Anger
1800201C5Vbothered	13	10	1	0	39	62	Surprise
1800201M4Vbothered	4	30	3	0	64	10	Sadness
1800201S3Vbothered	13	30	13	0	58	27	Sadness
1800201Y1Vbothered	3	53	26	45	0	21	Disgust
1800201Y6Vbothered	1	0	46	2	14	11	Fear
1800501M2Vflustered	28	26	41	6	3	36	Fear
1800501M7Vflustered	46	0	0	0	12	85	Surprise
1800501S3Vflustered	12	7	23	0	35	56	Surprise
1800501S4Vflustered	5	3	75	0	3	69	Fear
1800501Y3Vflustered	2	2	10	3	31	54	Surprise
1800501Y4Vflustered	23	61	26	2	0	18	Disgust
1801301M2Vrestless	8	40	24	0	21	32	Disgust
1801301M5Vrestless	3	9	25	1	51	50	Sadness
1801301S5Vrestless	14	33	8	3	19	93	Surprise
1801301S6Vrestless	36	0	34	0	0	35	Anger
1801301Y4Vrestless	0	81	21	11	8	22	Disgust
1801301Y7Vrestless	14	14	94	10	8	31	Fear
1801401M2Vruffled	26	56	15	0	14	14	Disgust
1801401M3Vruffled	38	4	23	0	38	22	Anger
1801401S5Vruffled	47	25	26	0	60	22	Sadness
1801401S6Vruffled	12	0	18	0	102	7	Sadness
1801401Y3Vruffled	6	0	0	0	104	15	Sadness
1801401Y4Vruffled	0	73	0	0	2	28	Disgust
1801701M5Vtense	4	0	15	0	97	9	Sadness
1801701M8Vtense	39	0	42	0	13	25	Fear
1801701S3Vtense	29	0	49	0	36	16	Fear
1801701Y2Vtense	27	0	42	17	28	36	Fear
1801701Y3Vtense	8	5	19	0	61	32	Sadness

Filename	Anger	Disgust	Fear	Happy	Sadness	Surprise	Most Frequent
1801701Y8Vtense	26	0	12	0	71	16	Sadness
Average (frames/clip)	16.3	19.1	25.9	3.39	36.3	30.0	

Table 16: Results of basic emotions detection on the video clips from Mind Reading DVD [19]. A total of 108 video clips tested are those under the emotional words classified under "anxiety" listed in Table 5. Last row shows the average frames for each emotion. For the remaining rows, first column shows the filename, the remaining columns show the number of frames detected per emotion. Last column shows the dominating emotion for that file based on the emotion with the most frame count.

Bibliography

[1] P. Ekman and H. Oster, "Facial expressions of emotion," *Annual review of psychology,* vol. 30, no. 1, pp. 527-554, 1979.

[2] S. Kaiser and T. Wehrle, "Automated coding of facial behavior in human-computer interactions with FACS," *Journal of Nonverbal Behavior,* vol. 16, no. 2, pp. 67-84, 1992.

[3] "Research and Markets," Research and Markets, Dec 2016. [Online]. Available: http://www.researchandmarkets.com/research/5w8htb/worldwide_emotion. [Accessed 2 Feb 2017].

[4] "Microsoft Cognitive Service," Microsoft, [Online]. Available: https://www.microsoft.com/cognitive-services/en-us/emotion-api. [Accessed 17 Feb 2017].

[5] "Emotion Recognition Software and Analysis," Affectiva, [Online]. Available: http://www.affectiva.com. [Accessed 19 Feb 2017].

[6] "Human Analytics, Emotion Analysis & Face Recognition | Kairos," Kairos, [Online]. Available: https://www.kairos.com. [Accessed 19 Feb 2017].

[7] "EmoVu emotion recognition software," Eyeris, [Online]. Available: http://emovu.com/e/. [Accessed 19 Feb 2017].

[8] K. Kokalitcheva, "Apple Acquires Startup That Reads Emotions From Facial Expressions," Fortune, 8 Jan 2016. [Online]. Available: http://fortune.com/2016/01/07/apple-emotient-acquisition/. [Accessed 23 Feb 2017].

[9] "Smokefree 2025," Ministry of Health, 2011. [Online]. Available: http://www.health.govt.nz/our-work/preventative-health-wellness/tobacco-control/smokefree-2025. [Accessed 19 Feb 2017].

[10] N. Z. Parliament, "Inquiry into the tobacco industry in Aotearoa and the consequences of tobacco use for Māori," *Report of the Māori Affairs Select Committee. Wellington: New Zealand Parliament,* 2010.

[11] R. Zhao-Shea, S. R. DeGroot, L. Liu, M. Vallaster, X. Pang, Q. Su, G. Gao, O. J. Rando, G. E. Martin and O. George, "Increased CRF signalling in a ventral tegmental area-interpeduncular nucleus-medial habenula circuit induces anxiety during nicotine withdrawal," *Nature communications,* vol. 6, 2015.

[12] "Nicotine and Tobacco Symptoms of Withdrawal," NY Times Health, 26 Feb 2013. [Online]. Available: http://www.nytimes.com/health/guides/disease/nicotine-withdrawal/symptoms-of-withdrawal.html. [Accessed 20 Feb 2017].

[13] J. A. Harrigan and D. M. O'Connell, "How do you look when feeling anxious? Facial displays of anxiety," *Personality and Individual Differences,* vol. 21, no. 2, pp. 205-212, 1996.

[14] T. Steimer, "The biology of fear-and anxiety-related behaviors," *Dialogues in clinical neuroscience,* vol. 4, pp. 231-250, 2002.

[15] A. M. Perkins, S. L. Inchley-Mort, A. D. Pickering, P. J. Corr and A. P. Burgess, "A facial expression for anxiety.," *Journal of personality and social psychology,* vol. 102, no. 5, p. 910, 2012.

[16] S. Velusamy, H. Kannan, B. Anand, A. Sharma and B. Navathe, "A method to infer emotions from facial action units," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011.

[17] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," 2010.

[18] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba and J. Budynek, "The Japanese female facial expression (JAFFE) database," in *Proceedings of third international conference on automatic face and gesture recognition*, 1998.

[19] S. Baron-Cohen, J. Hill, O. Golan and S. Wheelwright, "Mindreading made easy," *Cambridge Medicine,* vol. 17, pp. 28-29, 2002.

[20] T. Baltrušaitis, P. Robinson and L.-P. Morency, "Openface: an open source facial behavior analysis toolkit," in *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, 2016.

[21] I. H. Witten, E. Frank, M. A. Hall and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2016.

[22] S. Marĉelja, "Mathematical description of the responses of simple cortical cells," *JOSA,* vol. 70, no. 11, pp. 1297-1300, 1980.

[23] M. Lyons, S. Akamatsu, M. Kamachi and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, 1998.

[24] S. Bashyal and G. K. Venayagamoorthy, "Recognition of facial expressions using Gabor wavelets and learning," *Engineering Applications of Artificial Intelligence,* vol. 21, no. 7, pp. 1056-1064, 2008.

[25] "MATLAB," [Online]. Available: https://www.mathworks.com/products/matlab.html. [Accessed 2 Mar 2017].

[26] "Chinese Text Project," [Online]. Available: http://ctext.org. [Accessed 22 Feb 2017].

[27] F. Jabr, "The evolution of emotion: Charles Darwin's little-known psychology experiment," Scientific American, 24 May 2010. [Online]. Available: https://blogs.scientificamerican.com/observations/the-evolution-of-emotion-charles-darwins-little-known-psychology-experiment/. [Accessed 22 Feb 2017].

[28] "Darwin Correspondence Project - Emotion Experiment," University of Cambridge, 2016. [Online]. Available: https://www.darwinproject.ac.uk/commentary/human-nature/expression-emotions/emotion-experiment. [Accessed 24 Feb 2017].

[29] P. Ekman, "An argument for basic emotions," *Cognition & emotion,* vol. 6, no. 3-4, pp. 169-200, 1992.

[30] P. Ekman, "What scientists who study emotion agree about," *Perspectives on Psychological Science,* vol. 11, no. 1, pp. 31-34, 2016.

[31] R. E. Jack, O. G. Garrod and P. G. Schyns, "Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time," *Current biology,* vol. 24, no. 2, pp. 187-192, 2014.

[32] "Wikipedia - The Free Encyclopedia," Wikipedia, 18 Jan 2017. [Online]. Available: https://en.wikipedia.org/wiki/Duchenne_de_Boulogne. [Accessed 22 Feb 2017].

[33] G.-B. Duchenne and R. A. Cuthbertson, The mechanism of human facial expression, Cambridge university press, 1990.

[34] P. Ekman, "Duchenne and facial expression of emotion.," in *The Mechanism of Human Facial Expression*, Cambridge, Cambridge University Press, 1990, pp. 270-284.

[35] C.-H. Hjorztsjö, Man's face and mimic language, Studen litteratur, 1969.

[36] P. Ekman, W. Friesen and J. Hager, *Facial Action Coding System: The Manual on CD ROM. A Human Face,* 2002.

[37] "Facial Action Coding System," Wikipedia, 22 Jan 2017. [Online]. Available: https://en.wikipedia.org/wiki/Facial_Action_Coding_System. [Accessed 24 Feb 2017].

[38] D. Matsumoto and P. Ekman, "Facial expression analysis," *Scholarpedia,* vol. 5, no. 3, p. 4237, 2008.

[39] J. F. Cohn, Z. Ambadar and P. Ekman, *The handbook of emotion elicitation and assessment,* pp. 203-221, 2007.

[40] P. Crosta, "Anxiety: Causes, Symptoms and Treatments," 03 Aug 2015. [Online]. Available: http://www.medicalnewstoday.com/info/anxiety.

[41] M. S. Bartlett, J. C. Hager, P. Ekman and T. J. Sejnowski, "Measuring facial expressions by computer image analysis," *Psychophysiology,* vol. 36, no. 2, pp. 253-263, 1999.

[42] M. Pantic and L. Rothkrantz, "An expert system for multiple emotional classification of facial expressions," in *Tools with Artificial Intelligence, 1999. Proceedings. 11th IEEE International Conference on*, IEEE, 1999, pp. 113-120.

[43] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE transactions on image processing,* vol. 19, no. 1, pp. 172-187, 2007.

[44] T. Kanade, J. F. Cohn and Y. Tian, "Comprehensive database for facial expression analysis," 2000.

[45] V. Silva, F. Soares, J. S. Esteves, J. Figueiredo, C. P. Leão, C. Santos and A. P. Pereira, "Real-time emotions recognition system," in *Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2016 8th International Congress on*, IEEE, 2016, pp. 201-206.

[46] D. Gabor, "Theory of communication. Part 1: The analysis of information," *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering,* vol. 93, no. 26, pp. 429-441, 1946.

[47] "Gabor atom," Wikipedia, 22 Dec 2016. [Online]. Available: https://en.wikipedia.org/wiki/Gabor_atom. [Accessed 1 Mar 2017].

[48] D. Gabor, "Information Theory and Coding - Lecture Notes and Exercise," [Online]. Available: http://www.cl.cam.ac.uk/teaching/1314/InfoTheory/InfoTheoryNotes2013.pdf. [Accessed 1 Mar 2017].

[49] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of physiology,* vol. 160, no. 1, pp. 106-154, 1962.

[50] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *JOSA A,* vol. 2, no. 7, pp. 1160-1169, 1985.

[51] M. R. Turner, "Texture discrimination by Gabor functions," *Biological cybernetics,* vol. 55, no. 2, pp. 71-82, 1986.

[52] M. Kumbhar, A. Jadhav and M. Patil, "Facial Expression Recognition Based on Image Feature," *International Journal of Computer and Communication Engineering,* vol. 1, no. 2, p. 117, 2012.

[53] E. Owusu, Y. Zhan and Q. R. Mao, "A neural-AdaBoost based facial expression recognition system," *Expert Systems with Applications,* vol. 41, no. 7, pp. 3383-3390, 2014.

[54] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision,* vol. 57, no. 2, pp. 137-154, 2004.

[55] P. G. Mohan, C. Prakash and S. V. Gangashetty, "Bessel transform for image resizing," in *Systems, Signals and Image Processing (IWSSIP), 2011 18th International Conference on*, 2011.

[56] M. Abdulrahman, T. R. Gwadabe, F. J. Abdu and A. Eleyan, "Gabor wavelet transform based facial expression recognition using PCA and LBP," in *2014 22nd Signal Processing and Communications Applications Conference (SIU)*, 2014.

[57] P. Ekman and A. Fridlund, "Assessment of facial behavior in affective disorders," *Depression and expressive behavior,* pp. 37-56, 1987.

[58] H. Meng, D. Huang, H. Wang, H. Yang, M. AI-Shuraifi and Y. Wang, "Depression recognition based on dynamic facial and vocal expression features using partial least square regression," in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, 2013.

[59] A. Bowling, "Mode of questionnaire administration can have serious effects on data quality," *Journal of public health,* vol. 27, no. 3, pp. 281-291, 2005.

[60] J. F. Cohn, T. S. Kruez, I. Matthews, Y. Yang, M. H. Nguyen, M. T. Padilla, F. Zhou and F. De la Torre, "Detecting depression from facial actions and vocal prosody," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 2009.

[61] H. Gao, A. Y¨uce and J.-P. Thiran, "Detecting emotional stress from facial expressions for driving safety," 2014.

[62] W.-S. Chu, F. De la Torre and J. F. Cohn, "Selective transfer machine for personalized facial action unit detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[63] N. C. Ebner, M. Riediger and U. Lindenberger, "FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation," *Behavior research methods,* vol. 42, no. 1, pp. 351-362, 2010.

[64] O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk and A. van Knippenberg, "Presentation and validation of the Radboud Faces Database," *Cognition and emotion,* vol. 24, no. 8, pp. 1377-1388, 2010.

[65] S. Zhao, H. Yao and X. Sun, "Video classification and recommendation based on affective analysis of viewers," *Neurocomputing,* vol. 119, pp. 101-110, 2013.

[66] R. Navarathna, P. Lucey, P. Carr, E. Carter, S. Sridharan and I. Matthews, "Predicting movie ratings from audience behaviors," in *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*, IEEE, 2014, pp. 1058-1065.

[67] "The Guardian," 27 Jul 2015. [Online]. Available: https://www.theguardian.com/media-network/2015/jul/27/artificial-intelligence-future-advertising-saatchi-clearchannel. [Accessed 17 Feb 2017].

[68] J. O'Gorman, "Watch: M&C Saatchi launches artificially intelligent outdoor campaign," Campaign, 24 July 2015. [Online]. Available: http://www.campaignlive.co.uk/article/watch-m-c-saatchi-launches-artificially-intelligent-outdoor-campaign/1357413. [Accessed 6 Mar 2017].

[69] D. McDuff, R. El Kaliouby, D. Demirdjian and R. Picard, "Predicting online media effectiveness based on smile responses gathered over the internet," in *Automatic*

Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, IEEE, 2013, pp. 1-7.

[70] F. Y. Shih, C.-F. Chuang and P. S. Wang, "Performance comparisons of facial expression recognition in JAFFE database," *International Journal of Pattern Recognition and Artificial Intelligence,* vol. 22, no. 3, pp. 445-459, 2008.

[71] W. Gu, C. Xiang, Y. Venkatesh, D. Huang and H. Lin, "Facial expression recognition using radial encoding of local Gabor features and classifier synthesis," *Pattern Recognition,* vol. 45, no. 1, pp. 80-91, 2012.

[72] D. McDuff, R. Kaliouby, T. Senechal, M. Amr, J. Cohn and R. Picard, "Affectiva-mit facial expression dataset (am-fed): Naturalistic and spontaneous facial expressions collected," 2013.

[73] T. Baltrušaitis, P. Robinson and L.-P. Morency, "Constrained local neural fields for robust facial landmark detection in the wild," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013.

[74] T. Baltrušaitis, M. Mahmoud and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic action unit detection," 2015.

[75] E. Wood, T. Baltrušaitis, X. Zhang, Y. Sugano, P. Robinson and A. Bulling, "Rendering of eyes for eye-shape registration and gaze estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015.

[76] "Support vector machinese," Wikipedia, 9 Dec 2016. [Online]. Available: https://en.wikipedia.org/wiki/Support_vector_machine#History. [Accessed 2 Mar 2017].

[77] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, no. 3, pp. 273-297, 1995.

[78] "WebRTC," [Online]. Available: https://webrtc.org/faq/. [Accessed 3 Mar 2017].

[79] "Real-Time Communication in WEB-browsers (rtcweb)," Internet Engineering Task Force, [Online]. Available: https://datatracker.ietf.org/wg/rtcweb/charter/. [Accessed 3 Mar 2017].

[80] "WebRTC 1.0: Real-time Communication Between Browsers," World Wide Web Consortium, 24 Nov 2016. [Online]. Available: https://www.w3.org/TR/webrtc/. [Accessed 3 Mar 2017].

[81] "Kurento," Kurento, [Online]. Available: http://www.kurento.org/about. [Accessed 3 Mar 2017].

[82] E. Sariyanidi, H. Gunes and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 37, no. 6, pp. 1113-1133, 2015.

[83] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., 1995, pp. 338-345.

[84] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST),* vol. 2, no. 3, p. 27, 2011.

[85] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals, and Systems (MCSS),* vol. 2, no. 4, pp. 303-314, 1989.

[86] N. Landwehr, M. Hall and E. Frank, "Logistic model trees," *Machine Learning,* vol. 59, no. 1-2, pp. 161-205, 2005.

[87] L. Breiman, "Random forests," *Machine learning,* vol. 45, no. 1, pp. 5-32, 2001.

[88] R. A. El Kaliouby, "Mind-reading machines: automated inference of complex mental states," University of Cambridge, 2005.

[89] W. Liao, W. Zhang, Z. Zhu and Q. Ji, "A real-time human stress monitoring system using dynamic Bayesian network," in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, 2005.