# Emotion Recognition Using Facial Expression and Electroencephalography Features with Support Vector Machine Classifier

**Student**

**Henry Candra**

**Supervisor**

**Associate Professor Steven Su**

**Professor Hung. T. Nguyen**

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy


Centre for Health Technologies

School of Electrical, Mechanical and Mechatronic Systems

Faculty of Engineering and Information Technology

May, 2017

# Certificate of Authorship/Originality

I, Henry Candra, certify that the work in this thesis has not previously been submitted for a degree, nor has it been submitted as part of the requirements of a degree, except as fully acknowledged within the text.

I also certify that this thesis has been written by me. Any help that I have received in my research work and in the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature:

Date: 8 May 2017

*“As long as we are mindful and aware, no one practice is better than another.”*

*DJKR*

# Acknowledgements

First and foremost I would like to express my gratitude to my principal supervisor, Associate Professor Steven Su for the opportunity to carry out this research and for his invaluable guidance, support, encouragement, and life teachings. He has been with me since the beginning of my PhD and has been through my ups and downs in my PhD life. Without his help and patience I will not be able to complete my PhD study.

I am also deeply grateful to my co-supervisor, Professor Hung Tan Nguyen for his continuous support, constant guidance, valuable time, inspiration and valuable advice.

My special thanks to all the members of the Centre for Health Technologies (CHT) and other school who have been a source of collaboration, friendships as well as good advice and sharing. My grateful especially to Dr. Mitchell Yuwono for his continuous support, guidance and encouragement; to Dr. Rifai Chai for his sharing, discussion and inspiration; to (Fr.) Dr. Ardi Handojoseno (RIP) for his prayer, advice, and motivation; to Adrian Johannes for the listening and heartening; to Daniel Roxby and Zhichao Sheng for our inspiring lunch discussion; to Ganesh Naik, Steve Ling, Simon Ting for the advice and support; to Kai Cao, Lin Ye, Tao Zhang, and many others whose name has not been mentioned here, through all of you I learn to raise my young spirit once more.

Finally and most importantly, my deep appreciation goes to my Mum, my wife, my son, and all my family. Their encouragement, moral support, understanding, constant love, and prayer become my inspiration, and this thesis is my gift to all of you.

# Contents

# List of Figures

# List of Tables

# Nomenclature

| | | |
|---|---|---|
| AAM | : | Active Appearance Models |
| ADCs | : | Analogue-to-Digital Converters |
| AI | : | Artificial Intelligent |
| AU | : | Action Units |
| BCI | : | Brain-Computer interface |
| CK+ | : | Extended Cohn–Kanade database |
| CoC | : | Coefficient of Correlation |
| CT | : | Computerized Tomography |
| CWT | : | Continuous Wavelet Transform |
| DEAP | : | Database for Emotion Analysis Using Physiological Signals |
| DWT | : | Discrete Wavelet Transforms |
| ECG | : | Electrocardiography |
| EEG | : | Electroencephalography |
| EGG | : | Electrogastrography |
| E-HOG | : | Edge-Histogram of Oriented Gradients |
| EMG | : | Electromyography |
| EOG | : | electrooccluography |
| ERCE | : | Ensemble Rapid Centroid Estimation |
| ERD | : | Event-Related Desynchronization |
| ERP | : | event- related potentials |
| ERS | : | event-related synchronization |
| FACS | : | Facial Action Coding System |
| FD | : | Fractal Dimension |
| FFT | : | Fast Fourier Transform |
| fMRI | : | Functional Magnetic Resonance Imaging |
| GB | : | Giga Bytes |
| GLCM | : | Gray-Level Co-occurrence Matrix |
| GMM | : | Gaussian Mixture Model |
| GSR | : | Galvanic Skin Response |
| HCI | : | Human-Computer Interaction |
| HOG | : | Histogram of Oriented Gradients |
| HOS | : | Higher Order Spectra |

| | | |
|---|---|---|
| IADS | : | International Affective Digitized Sound System |
| IAPS | : | International Affective Picture System |
| ICA | : | Independent Component Analysis |
| ID | : | Identification |
| JS | : | Jensen-Shannon |
| KL | : | Kullback-Leibler |
| KPCA | : | Kernel Principal Component Analysis |
| LBP | : | Local Binary Patterns |
| LDA | : | Linear Discriminant Analysis |
| MEG | : | Magnetoencephalography |
| MSCE | : | Magnitude Square Coherence Estimation |
| MSR | : | Manifold based Sparse Representation |
| NSI | : | Non Stationary Index |
| ORI E-HOG | : | Original Edge-Histogram of Oriented Gradients |
| OWS | : | Optimal Window Selection |
| PAD | : | Pleasure, Arousal and Dominance Emotion Scales |
| PCA | : | Principal Components Analysis |
| PET | : | Positron Emission Tomography |
| PSD | : | Power Spectral Density |
| RBF | : | Radial Basis Function |
| RCE | : | Rapid Centroid Estimation |
| RED E-HOG | : | Reduced Edge-Histogram of Oriented Gradients |
| RHMM | : | Regional Hidden Markov Model |
| SAM | : | Self-Assessment Manikin |
| SIFT | : | Scale Invariant Feature Transform |
| SMO | : | Sequential Minimal Optimization |
| SPET | : | Single Photon Emission Tomography |
| SVM | : | Support Vector Machine |
| WE | : | Wavelet Energy |
| WT | : | Wavelet Transform |

# Abstract

Recognizing emotions from facial expression and electroencephalography (EEG) emotion signals are complicated tasks that require substantial issues to be solved in order to achieve higher performance of the classifications, i.e. facial expression has to deal with features, features dimensionality, and classification processing time, while EEG emotion recognition has the concerned with features, number of channels and sub band frequency, and also non-stationary behaviour of EEG signals. This thesis addresses the aforementioned challenges.

First, a feature for facial expression recognition using a combination of Viola-Jones algorithm and improved Histogram of Oriented Gradients (HOG) descriptor termed Edge-HOG or E–HOG is proposed which has the advantage of insensitivity to lighting conditions. The issue of dimensionality and classification processing time was resolved using a combination of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) which has successfully reduced both the dimension and the classification processing time resulting in a new low dimension of feature called Reduced E–HOG (RED E–HOG).

In the case of EEG emotion recognition, a method to recognize 4 discrete emotions from arousal-valence dimensional plane using wavelet energy and entropy features was developed. The effects of EEG channel and subband selection were also addressed, which managed to reduce the channels from 32 to 18 channels and the subband from 5 to 3 bands.

To deal with the non-stationary behaviour of EEG signals, an Optimal Window Selection (OWS) method as feature-agnostic pre-processing was proposed. The main objective

of OWS is window segmentation with varying window which was applied to 7 various features to improve the classification results of 4 dimensional plane emotions, namely arousal, valence, dominance, and liking, to distinguish between the high or low state of the aforementioned emotions. The improvement of accuracy makes the OWS method a potential solution to dealing with the non-stationary behaviour of EEG signals in emotion recognition. The implementation of OWS provides the information that the EEG emotions may be appropriately localized at 4–12 seconds time segments.

In addition, a feature concatenating of both Wavelet Entropy and average Wavelet Approximation Coefficients was developed for EEG emotion recognition. The SVM classifier trained using this feature provides a higher classification result consistently compared to various different features such as: simple average, Fast Fourier Transform (FFT), and Wavelet Energy.

In all the experiments, the classification was conducted using optimized SVM with a Radial Basis Function (RBF) kernel. The RBF kernel parameters were properly optimized using a particle swarm ensemble clustering algorithm called Ensemble Rapid Centroid Estimation (ERCE). The algorithm estimates the number of clusters directly from the data using swarm intelligence and ensemble aggregation. The SVM is then trained using the optimized RBF kernel parameters and Sequential Minimal Optimization (SMO) algorithm.

# Chapter 1

# Introduction

## 1.1 Overview

Communication between two or more individuals can take place in the form of verbal or nonverbal language. Nonverbal communication involves many different aspects as well as proxemics (physical and interpersonal space: distance and territoriality), kinesics (body orientation: body posture, body motion, and gesture), appearance (physical attractiveness and clothing), haptics (touch), paralanguage (voice, tone, rate, pitch, and volume of the speaker), and facial expression (Sutter, 2010). All of these nonverbal communications need to be interpreted and they dominate more than 60 percent of the communication process (Foley & Gentile, 2010). These nonverbal communications have a significant role in communicating the feelings, attitudes and emotions.

Emotion itself is derived from the psychophysiological process stimulated by conscious and / or unconscious awareness to any event or object related with the mental state, characteristic and nature of a person (College *et al.*, 2014). Emotion has a significant part in the communication between individuals. The emotion of an individual will influence the relationship with other people such as family, relatives and friends at home, workplaces or other environments that create connection with other people. The emotion is manifested through the intonation in the voice, gesture and body posture, and most commonly facial

expression. Psychiatrists emphasize the importance of facial emotion interpretation because the interpretation of facial emotion in nonverbal communication can lead to a good or bad relation (Ekman, 2003). The psychiatrist has much guidance and many regulations to interpret the facial expression of the patients while helping them in a therapy session to establish good relations between the psychiatrists and their patients to obtain a successful therapy process (Machado *et al.*, 1999).

Psychology also divides emotions into three components: subjective feeling, motor expression, and physiological arousal (Scherer *et al.*, 2001), which can be described as follows:

a Subjective feeling: the awareness of the emotional episode expressed through various emotional lexica, by communicating the response to certain stimuli.

b Motor expression: the changes in gestures, posture, facial, and vocal expression, consciously and unconsciously. It is a communicated expression of a person, such as behaviour intentions, to other people.

c Physiological arousal: change in physiological body in response to emotional event, i.e. body temperature, skin conductivity, heartbeat alterations, breathing rates, and brain waves.

Based on those three components, it is obvious that emotions are not just about what appears, but they are more related to the responses of the brain manifested through the physiological signals. This opens the opportunities to detect emotion not just from the external appearance signs (facial expressions, voice, gesture and body posture) (Gunes & Piccardi, 2005; Metallinou *et al.*, 2010), but also from the internal physiological signals such as Electroencephalography (EEG) (Petrantonakis & Hadjileontiadis, 2012), Electrocardiography (ECG) (Agrafioti *et al.*, 2012), Electromyography (EMG), and Galvanic Skin Response (GSR) (AlZoubi *et al.*, 2012).

## 1.2 Emotion Recognition

Understanding the emotions has become the nature of humans to be successful in communicating with others. However, once in a while false interpretation of the emotions occurs between 2 or more individuals so that the communication becomes unsuccessful.

This event sometimes can lead to a worse situation which will affect the relation between the people who communicate, for example: relation with family or relative or friend. In a more official situation, it can affect the relation between the employee and the employer, the student and the teacher, and even the doctor and the patient.

These circumstances trigger the development of a system that can recognize human emotion to minimize the false interpretation of emotions. The emotion recognition system since then has become very popular in recent years and given significant contributions in many applications that are capable to interpreting human emotion based on different physiological signals.

The system has been implemented in many areas such as security, safety, and education process (Owayjan *et al.*, 2012; Chai *et al.*, 2016a; Wang & Niu, 2012). It is also applied in the health system by, for example monitoring the emotion of the student or patient while having the session with the teacher, therapist, or psychiatrist (Wang *et al.*, 2010; Liu *et al.*, 2010; Othman & Wahab, 2010). The most popular application is the Brain–Computer interface (BCI) which has been used to control devices or gadgets with an intelligence system (Pun *et al.*, 2006).

## 1.3 Application

In general the design of an emotion recognition system is the same as a biometric system. To be specific, emotion recognition system takes advantage of the physiological signal that is responsible to the elicitation of emotions. The definition of biometrics itself is automatic person identification by recognizing the person's physiological characters (Kim *et al.*, 2010). The block diagram of an emotion recognition system is shown in Figure 1.1.

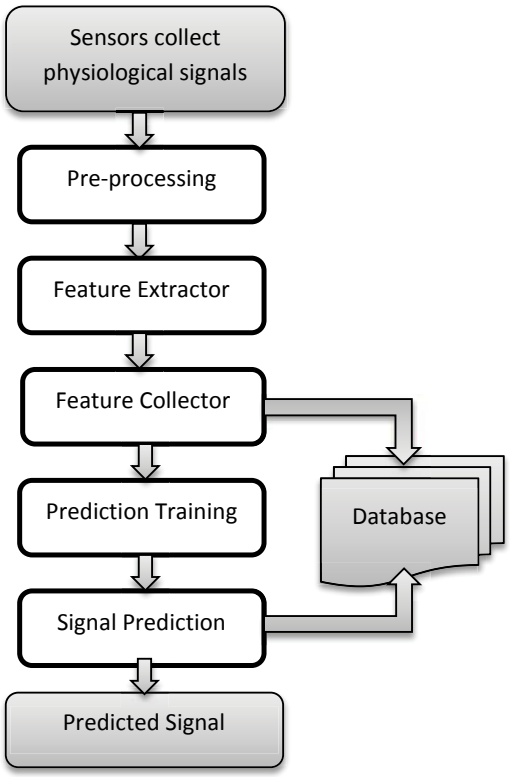



Figure 1.1: Block diagram an emotion recognition system

In Fig.1.1 the sensors are the connector between the system and the outer environment. First the sensors detect and pick up the physiological signals of an individual such as: face appearance, EEG, or ECG. The signal from the sensors undergoes pre-processing to dispose of the artifacts mixed up in the sensor and also to improve the quality of the input signal, to which sometimes a normalization method can be applied.

The feature extractor then collects sets of features from the input signal to be used as input data for the feature collector. This step is very crucial so that the correct features should be extracted in an optimal way. Next, the feature collector synthesizes the characteristics of each specific individual to get adequate recognition information.

The next most important part of a recognition system is the ability to learn and predict. When prediction training is executed in the learning process, the features of an individual are stored into the database. While in signal prediction, the features are recognized by comparing them with the information stored in the database. An emotion recognition

system provides a prediction result based on the information stored in the database.

### 1.3.1 Facial Expression Recognition System

The most favored application of emotion recognition system is the facial expression recognition system. Most of the research on facial expression refers to the work of Prof. Paul Ekman. He proposed the Facial Action Coding System (FACS) to code the facial expression extracted from thousands of photographs and tens of thousands of filmed and videotaped facial expressions (Ekman, 2003). The FACS can be extracted from the face and used as the features for facial expression classification. The implementation of a facial expression recognition system can be realized in a similar way using the biometric design. The facial expression recognition system applies more specific steps as follows:

a Detection of the face.

b Feature extraction (part of facial landmark such as eyes, mouth or the whole face).

c Expressions classification.

The first step utilizes the sensor to collect physiological signal of facial expression from the person. For that purpose, a camera can be used to produce a still image of the face. Alternatively, for continuous monitoring of facial expression, a video or web camera will be more suitable to be used for various conditions and environments.

In the facial expression recognition system, the most important part is the feature extraction process. A proper feature extraction process will produce a better recognition system with a more accurate result. The algorithms used for facial features extraction can be classified into two main groups, namely: geometric based methods which collect feature points or motion of points by tracing them from the face images and classifying the expressions from the tracked features; and, appearance based methods which collect the whole or part of the face landmarks and arrange them as one long array feature vector and apply them in the classification process. The appearance based method has more advantage compared to the geometric based method (Song *et al.*, 2010).

### 1.3.2 EEG Emotion Recognition System

Second compelling application in emotion recognition is the EEG emotion recognition system. The research on EEG emotion recognition system is mainly based on the *circumplex model of emotion* proposed by Psychologist James Russell in 1980 (Russell, 1980).

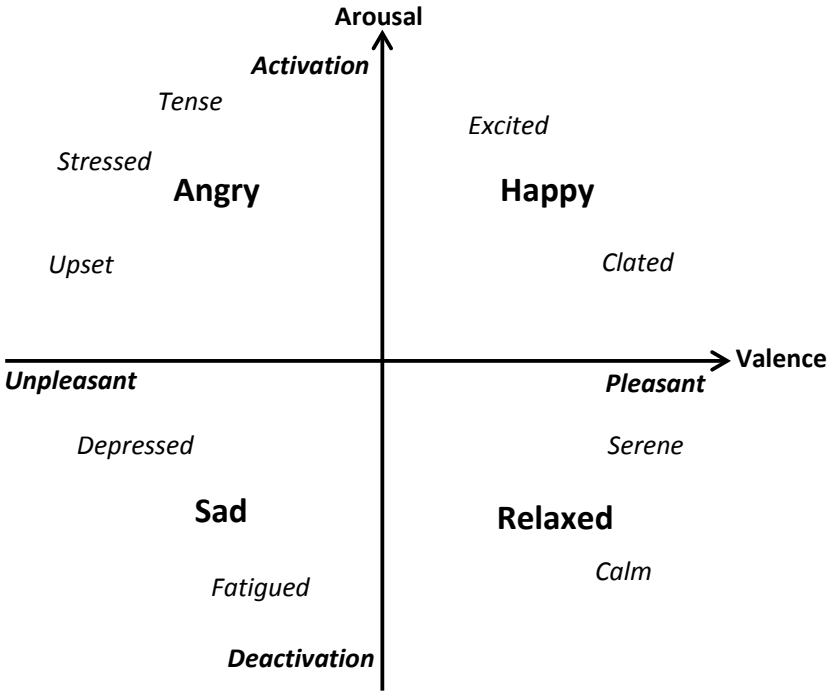



Figure 1.2: Russell's circumplex model of affect Russell (1980). Horizontal axis represents valence (pleasure); vertical axis represents arousal. Artwork is as seen in Valenza et. al (Valenza *et al.*, 2014).

The circumplex model is a conceptualized 2-dimensional continuous space where the horizontal and vertical axes correspond to the degree of valence (pleasure) and arousal, respectively. Discrete emotional states such as 'happy', 'sad', 'angry', and 'relaxed' can be inferred from the degree of valence and arousal as illustrated in Fig. 1.2. Using this model, the degree of any of the aforementioned discrete emotional states can be measured. It was further reported that the psychological condition of positive/negative arousal (activation/deactivation) and positive/negative valence (pleasant/unpleasant) can be identified from Galvanic Skin Response (GSR) and EEG signal (Torres *et al.*, 2013).

In order to map a raw EEG signal to the appropriate emotion, the following steps can be applied:

1. Preprocess the raw EEG signal to remove artifacts such as eye blinking.
2. Extract the features from the preprocessed EEG signal.
3. Classify the features with a classifier.

## 1.4 Challenges

The classification results in emotion recognition system are very impressive. Recognition of emotions from physiological signals has reached the accuracy between 70%-80%. Even the computers have successfully classified facial expressions with 80% to 90% of accuracy. However, it should be noted that the applications are still facing a few challenges. The existing challenges can be defined as follows:

1. First, challenge in facial expression recognition system. The appearance based method for facial features extraction has an issue of high dimensionality of features which is known as curse the of dimensionality. This circumstance inflicts very long classification processing time for both training and testing process (Zhao & Chellappa, 2002). More specifically, the challenge can be expressed in one sentence, that is: how to obtain reliable features in facial expression recognition using an appearance based method that has very low dimension.
2. Second, challenge in EEG emotion recognition system. A substantial factor for determining the reliability of an EEG emotion recognition systems in general is the feature selection step, including: the type of feature that is able to supply the information in both time and frequency domain, the frequency components that relate to any state of the brain represented in various subbands, and the position on the surface of the head that generates the EEG signal detected by EEG electrodes at certain position with several numbers of channels (Jenke *et al.*, 2014b). Hence, the challenge is how to obtain proper EEG emotion feature that carries useful information

in time and frequency domain which can be acquired in restricted EEG frequency subbands using limited number of EEG channels.

3. Another drawback of EEG signals are their time-varying and non-stationary characteristics.To deal with this problem, the EEG signals must be split into smaller window frames so that pattern repetitions can be extracted more easily (Picard *et al.*, 2001). In a specific length of time window, a pseudo-stationary signal has a desirable statistical property of a constant mean and variance, in which, when it is used for a prediction, it is likely yield a relatively higher predictive power (Kaplan *et al.*, 2005). Now, the challenge is how to obtain optimum window size to maximize the information gain by "zooming in" on the recurring pattern on emotion elicitation. There is one condition, the window size needs to be *just right*: A window too short will lead to incompleteness, whilst a window too long will lead to over-inclusion of non-stationary components.

4. In addition, classification using Support Vector Machine (SVM) classifier requires an optimization method to obtain a more reliable classification result (Hsu *et al.*, 2003). The challenge is what method can be used to optimize the classification with SVM.

## 1.5 Contribution of This Thesis

To answer the challenges, this thesis provides solutions to dealing with the problems with various methods and strategies which become the major contributions of the thesis that can be summarized as follows:

1. A sophisticated feature for facial expression classification using a combination of Viola-Jones algorithm and improved Histogram of Oriented Gradients (HOG) descriptor termed as Edge-HOG or E-HOG (Candra *et al.*, 2016) is proposed as appearance based features to obtain higher performance of classification results with reduced processing time compared to original HOG. This proposed E-HOG feature has been proven to be highly efficient for facial-expression recognition. A

significant improvement in processing time (by comparing the classification process applying HOG and E-HOG) with a slight reduction in classification accuracy relative to HOG can be obtained.

2. Further experiment was conducted to reduce the dimension of Original E-HOG (ORI E-HOG) features using a combination of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) which managed to reduce the dimension from thousands to tens. The new low dimension of E-HOG is called Reduced E-HOG (RED E-HOG) (Candra *et al.*, (Submitted)b). With RED E-HOG, the accuracy of facial emotion classification results was improved, while the processing time for training and testing episodes was also reduced dramatically.

3. In the case of EEG emotion recognition, a method to identify four discrete emotions namely: happy, sad, angry, and relaxed using wavelet features including energy and entropy was proposed (Candra *et al.*, 2015b). The effect of EEG channel and subband selection was also investigated, which managed to reduce the 5 subbands to become only 3 bands utlizing alpha, beta and gamma bands. The channels have also been reduced from 32 to 18 channels namely: Fp1, Fp2, AF3, AF4, F3,F4, F7, F8, FC5, FC6, T7, T8, P7, P8, P3, P4, O1, and O2, while no substantial decrease in results occurred. This gives a positive indication of the appropriateness of the selected channels and subbands.

4. To deal with the non-stationary behaviour of EEG signals in EEG emotion recognition, the effective window size was investigated to improve the classification results of EEG emotion using proposed Optimal Window Selection (OWS) (Candra *et al.*, (Submitted)a) as feature-agnostic preprocessing. Using 7 variations of features, with varying window size the information corresponding to the delta, theta, alpha, beta, and gamma bands was also calculated. These features were fed into the classification algorithm for classifying between high/low of arousal/valence/dominance/liking emotions. Using the OWS method, the localization of the information of EEG emotion can be allocated at 4–12 seconds. This makes the OWS method a potential solution to dealing with the non-stationary behaviour of EEG signals in emotion recognition.

5. A novel wavelet feature, concatenating both Wavelet Entropy and average Wavelet Approximation Coefficients (Candra *et al.*, (Submitted)a) was also proposed to identify low/high of 4 dimensional plane emotions. Based on the experiments, the classifiers trained using this novel feature consistently yield significantly higher results than those trained using other features including simple average, Fast Fourier Transform (FFT), and Wavelet Energy.

6. In all the experiments, the classification is conducted using optimized SVM trained with Sequential Minimal Optimization (SMO) algorithm. Optimization is conducted using a Radial Basis Function (RBF) kernel. The RBF kernel parameters were properly estimated in order to obtain proper learning using a particle swarm ensemble clustering algorithm called the Ensemble Rapid Centroid Estimation (ERCE) algorithm (Yuwono *et al.*, 2014). The advantage of this algorithm is its capability to estimate the number of clusters directly from the data using swarm intelligence and ensemble aggregation.

## 1.6 Outline of The Thesis

This thesis consists of 6 chapters, an appendix and a bibliography. The thesis is organized as follows: Chapter 2 reviews the literatures related to emotion measurement and emotion recognition system using the facial expression and EEG emotion signal together with the information on availability of databases, algorithms, and methods comparison to conduct the experiment in the related area, and also discussion on SVM and the optimization method for classification using SVM. Chapter 3 describes in depth the proposed facial expression recognition system using E–HOG and RED E–HOG with detail algorithm construction and discussion on the results. This is followed with Chapter 4 that provides the information on the development of discrete EEG emotion recognition in details. In Chapter 5, the Optimal Window Selection (OWS) strategy to improve classification results of EEG emotion using window segmentation strategy is explained step by step. Chapter 6 summarizes the whole discussion in all chapters in a conclusion and also provides the research limitations and possible future directions for the research.

## 1.7 List of Publications

*Published Conference Papers*

1. Candra, H., Yuwono, M., Chai, R., Nguyen, H. T., & Su, S. 2017. (In Press). *EEG Emotion Recognition Using Reduced Channel Wavelet Entropy and Average Wavelet Coefficient Features with Normal Mutual Information Method*. Proc. IEEE 39th Annu. Int. Conf. Eng. Med. Biol. Soc.

2. Guo, K., Candra, H., Yu, H., Li, H., Nguyen, H. T., & Su, S. W. 2017. (In Press). *EEG-Based Emotion Classification Using Innovative Features and Combined SVM and HMM Classifier*. Proc. IEEE 39th Annu. Int. Conf. Eng. Med. Biol. Soc.

3. Candra, H., Yuwono, M., Chai, R., Nguyen, H. T., & Su, S. 2016. *Classification of Facial-Emotion Expression in the Application of Psychotherapy using Viola- Jones and Edge-Histogram of Oriented Gradient*. Pages 423-426 of: Proc. IEEE 38th Annu. Int. Conf. Eng. Med. Biol. Soc.

4. Candra, H., Yuwono, M., Rifai, Chai, Handojoseno, A., Elamvazuthi, I., Nguyen, H. T., & Su, S. 2015a. *Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine*. Pages 7250-7253 of: Proc. IEEE 37th Annu. Int. Conf. Eng. Med. Biol. Soc.

5. Candra, H., Yuwono, M., Handojoseno, A., Chai, R., Su, S., & Nguyen, H. T. 2015b. *Recognizing emotions from EEG subbands using wavelet analysis*. Pages 6030-6033 of: Proc. IEEE 37th Annu. Int. Conf. Eng. Med. Biol. Soc.

*Submitted Journal Papers*

1. Candra, H., Yuwono, M., Chai, R., Handojoseno, A., Elamvazuthi, I., Nguyen, H. T., & Su, S. (Submitted)a. *Optimal Window Selection for Improving EEG-Emotion Classification*. Signal Processing.

2. Candra, H., Cao, K., Yuwono, M., Chai, R., Nguyen, H. T., & Su, S. (Submitted)b. *Reduced Dimension of Edge-Histogram of Oriented Gradients (RED E-HOG) Features for Facial-Emotion Recognition*. Digital Signal Processing.

# Chapter 2

# Literature Review

## 2.1 Emotion Measurement

According to the psychologists, emotional response of individuals is triggered by their own assessment. The emotion response is manifested as a specific action (motor expression), physiology signals, and subjective experience, according to a specific situation. The diagram of individuals emotional reponse is illustrated as a consensual componential model of emotion as shown in Fig. 2.1 (Mauss & Robinson, 2009).

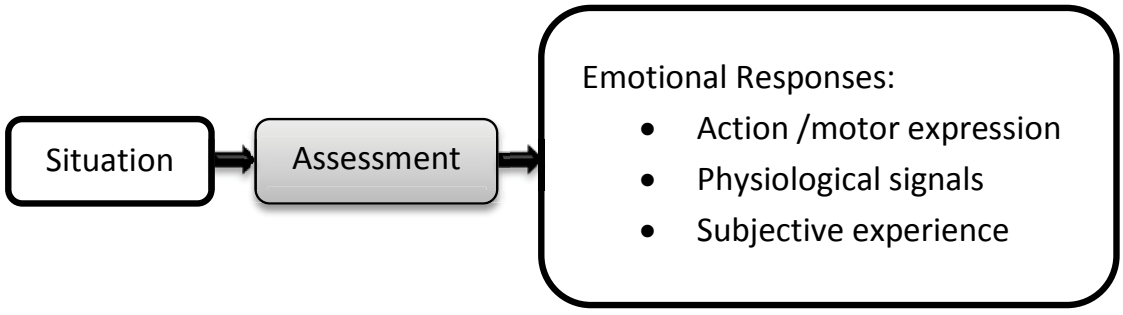



Figure 2.1: Block diagram of a consensual componential model of emotion. (Mauss & Robinson, 2009)

Psychologists also use a few different perspectives to measure the emotion response by dividing it into 3 groups as follows (Caicedo & Beuzekom, 2006):

a Discrete Emotion perspective: Each emotion corresponds to a unique profile in experience, physiology, and behaviour (Panksepp, 2007). For that purpose the emotion is divided into a few basic emotions, for example: fear, anger, joy, sadness, disgust, and surprise. According to this perspective, the basic emotions can be mixed which would create a large number variations of an individual emotional episode. The set of emotions varies depending on the theoretical background used. Ekman 6 basic facial expression is an example (Ekman, 2003).

b Dimensional perspective: The emotion is allocated between three independent dimensions: Pleasantness–unpleasantness, rest–activation, and relaxation–attention. Recently, a two-dimensional approach is adapted by allocating the emotion into pleasant–unpleasant and rest–activation dimensions as these are more sufficient to describe the emotion (Russell & Barrett, 1999).

c Componential perspective: Emotions are differentiated based on the dimensions used by the individual to evaluate an event and its affect on the individual. This perspective is more related to the assessment process.

Considering the appraisal process and emotion perspective, scientists have conducted researches to obtain reliable methods to assess emotional episodes. The methods can be classified based on the component of the emotional response that needs to be addressed (Caicedo & Beuzekom, 2006). A few different approaches that have been implemented in this thesis are discussed briefly in the following section together with their advantages and disadvantages.

### 2.1.1 Measurement of Motor Expression

Emotional response of individuals can be measured through their behaviour which is manifested in motor expressions such as facial expressions, gestures, and voice tones. Researches on the intrinsic working of facial muscles related to facial motoric action (Ekman & Friesen, 1976), as well as experiments on emotions recognition from facial expression (Bekele *et al.*, 2013) and the acoustic speech signal (Busso & Narayanan, 2007) have been conducted with different methods in this field.

Many psychologists accept the basic universal facial emotion expression concept. This concept becomes the asset of motor expression measurement and makes it possible to measure the emotion expression of individuals with cross-cultural backgrounds. The assessment can be conducted in a nonintrusive method using video cameras and microphones which can be prepared without distracting the individuals in order to minimize the affect or interfere with their reaction towards the stimuli.

The method focuses on the measurement of basic emotion which has disadvantages in measuring mixed emotions and is still facing the problem to link certain motor responses to secondary emotions. Mild emotions with little motor response are also difficult to measure. There is the possibility of falsification of facial expression related to the ability of individuals to control their motor expressions to a certain degree. Another major disadvantage is the need of expertise for the interpretation and complex instrumentation.

#### 2.1.1.1. Facial Action Coding System (FACS) for Facial Expression Measurement (Ekman & Friesen, 1976)

FACS was proposed by Paul Ekman and W.V. Friesen in 1976. It is used to identify the relation between the contraction of each facial muscle and the facial expression of an individual. The system is used to measure and describe facial behaviour based on the muscles that produce it. FACS is a descriptive tool that provides no information about the meaning or the origin of facial behaviour. A tool called EMFACS (Emotion FACS) is needed to translate the FACS into more useful information in emotion assessment. EMFACS is a limited version of FACS that provides translation of facial emotional expressions which enables the identification of basic emotions to a certain degree.

### 2.1.2 Measurement of Physiological Arousal

Measuring the physiological arousal can be conducted using specific transducers, such as electrodes, thermometers, diodes to detect the physiological changes in the body triggered by any emotion episodes. The results are represented in physiological signals such as brain waves, heart rate, blood pressure, and skin conductivity.

The main advantage of physiological measurements is the objectivity of the measurements in a way that the change of physiological signals is triggered by the body on an unconscious level of the individual. Therefore, it can be used to measure individuals from different social and cultural backgrounds.

The main drawback of the method is that the interpretation of certain physiological signals to a specific emotion is still argued by the scientists. In addition, the effect of other external factors mostly are not taken into consideration, for example, physical activity during the experiment that may affect the body temperature and heart rate of the individual whichis not correlated to the measured emotion.

Furthermore, the instrumentation attached to the participant in the experiment may create an uncomfortable feeling that affect the result of emotion measurement. The installation of the instruments also requires experts in physiology and technical engineering.

### 2.1.3 Measurement of Subjective Feeling

Subjective feelings usually are measured with self-report assessment of the participants. Using questionnaires the participants rate their emotions in a given scale or by using verbal descriptions. The method also applies pictorial models to eliminate or reduce the cultural and linguistic problem in interpreting the verbal material.

The main advantage is the accessibility to measure mixed emotions using a set of questions. It also requires very little technical background of the participants which reduces the need of technical support.

The main disadvantage is the difficulty for some participants to interpret their experiences which leads to misinterpretation of emotions (consciously or unconsciously). Besides, it is crucial to assess the emotion experience as soon as it arises. Distorted measurement may occurr if the assessment is longer than the stimulus events that trigger the emotion.

#### 2.1.3.1. The Pleasure, Arousal and Dominance (PAD) Emotion Scales (Mehrabian, 1997)

PAD is an example of a method to measure subjective feeling. The PAD method uses 3 dimensional planes to describe and measure the emotional response. Using the PAD method, eight basic emotions can be grouped and identified according to their position in the 3 dimensional planes. The 3 dimensions used in PAD methods are: Arousal (alert–not alert which defines the issue of physical activity and mental alertness); Dominance (dominance–submissiveness which assesses the individual's feeling of control on a given situation); and Valence (pleasant–unpleasant which address the affective quality of the experience). The emotions are identified from possible combinations of the levels in Arousal, Dominance, and Valence dimensions. Some example of the measured emotions are hostile, anger, defiance and insolence.

#### 2.1.3.2. Self-Assessment Manikin (SAM) (Bradley & Lang, 1994)

SAM is a pictorial graphs that replace the verbal description of emotion which is used in the self-assessment method. The graphs consist of 3 different series of images representing 3 dimensional planes of Arousal, Dominance and Valence with the scale from 1 to 9 for each dimension. The users interpret their emotions by selecting a particular image which they feel is best represents their emotion at a triggered emotion episode. Using SAM, the emotion of different cross-cultural participants can be measured.

Although self-reports of emotion using SAM are likely to be valid for current experienced emotions, however, there is a report that the users have faced difficulty to interpret and relate the images used in SAM to their emotion so that some specific emotions were interpreted as other contradictory emotions which leads to degradation of the emotion measured (Isomursu *et al.*, 2007).

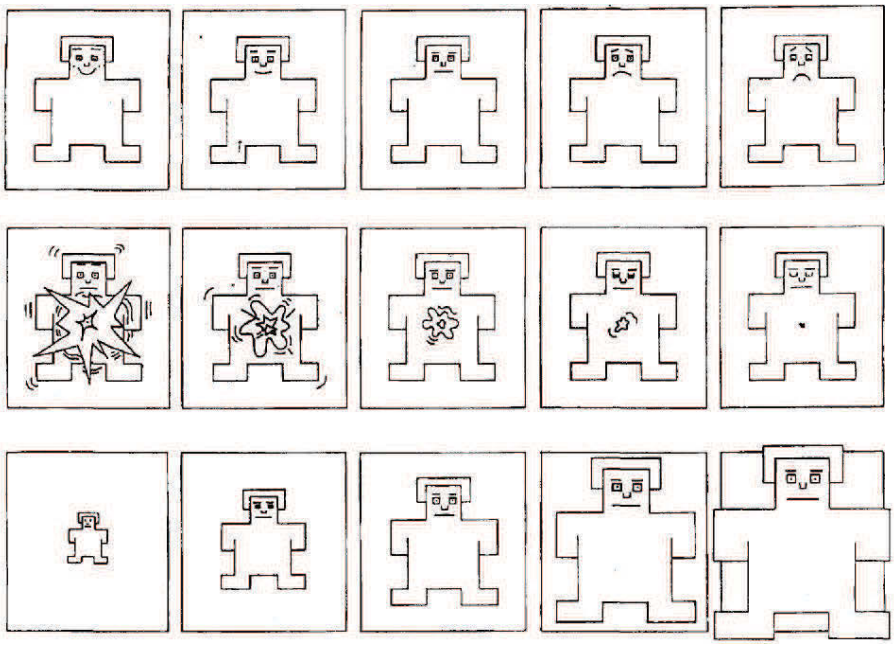


Figure 2.2: Self Assessment Manikin SAM (Bradley & Lang, 1994)

## 2.2 Face Detection and Recognition

With the rapid development of computational powers and availability of modern equipment and technologies, computers are becoming more and more intelligent. Many research projects and commercial products have shown the capability for a human computer interaction in a natural way by looking at people through cameras, listening to people through microphones, understanding these inputs, and reacting to people in a friendly manner.

### 2.2.1 Face Detection

One of the fundamental techniques that enables such natural human-computer interaction (HCI) is face detection. Face detection is the main step to all facial analysis algorithms, including face modeling, face recognition, face authentication, head pose tracking, facial expression recognition, age recognition, and many other methods.

The goal of face detection is to determine whether or not there are any faces in an image and return the image location and extent of each face (Yang *et al.*, 2002). Although this is a simple task for human beings, it is a very challenging task for computers, and has been

one of the most favoured research topics in the past few decades. The problems associated with face detection can be attributed to many variations, i.e. scale, location, orientation, poses, facial expression, lighting conditions, occlusions.

There have been numerous approaches for face detection (Yang *et al.*, 2002; Hjelms & Low, 2001). For instance, Yang et al. (Yang *et al.*, 2002) grouped the various methods into four categories: knowledge-based methods, feature invariant approaches, template matching methods, and appearance-based methods. Knowledge-based methods use pre-defined rules to determine a face based on human knowledge; feature invariant approaches aim to find face structure features that are robust to pose and lighting variations; template matching methods use pre-stored face templates to judge if an image is a face; appearance-based methods learn face models from a set of representative training face images to perform detection. In general, appearance-based methods had been showing superior performance to the others due mainly to the rapid growth in computation power and data storage.

The field of face detection has made significant progress in the past decade. In particular, the important work by Viola and Jones (Viola & Jones, 2001) has made face detection practically feasible in real world applications such as digital cameras and photo modification and organization.

#### 2.2.1.1. Viola Jones Algortihm for Face Detection

The famous face detection algorithm that has the most impact is the Viola and Jones algortihm (Viola & Jones, 2001). The advantages of the Viola–Jones algorithm is related to its robustnest with high detection rate to true-positive rate rather than false-positive rate. It is also applicable for real time processing. The algorithm detects faces from non-faces images with 4 main steps as follows:

1. Haar Feature: is sets of 3 rectangular features. First, a two-rectangle feature which is the difference between the sum of the pixels within two rectangular regions. The regions have the same size and shape and are horizontally or vertically adjacent.

Second, a three-rectangle feature which computes the sum within two outside rectangles subtracted from the sum in a center rectangle. And third, a four-rectangle feature computes the difference between diagonal pairs of adjacent rectangles. Fig. 2.3 illustrates the Haar feature.

Figure 2.3: The Haar feature (Viola & Jones, 2001)

2. Integral Image: is an algorithm for quickly and efficiently generating the sum of values of the rectangle features . The itegral image at location x, y contains the sum of the pixels above and to the left of x, y, inclusively.

3. Adaboost classifier: is a machine learning boosting algorithm capable of constructing a strong classifier through a weighted combination of weak classifiers. (A weak classifier classifies correctly in only a little bit more than half the cases.) To match this terminology to the presented theory each feature is considered to be a potential weak classifier.

4. Cascading Classifiers: is composed of stages each containing a strong classifier. The function of each stage is to determine whether a given sub-window is definitely not a face or maybe a face. When a sub-window is classified to be a non-face by a given stage it is immediately discarded. Conversely a sub-window classified as a maybe-face is passed on to the next stage in the cascade. It follows that the more stages a given sub-window passes, the higher the chance the sub-window actually contains a face.

The computation of Viola-Jones face detection algorithm in details is as follows:

1. *Computing the integral image* $\mathcal{I}$. An integral image $\mathcal{I}$ is an image whose value at $\mathcal{I}(x, y)$ is the sum of all the pixels above and to the left of $x$ and $y$, inclusive. That

is,

$$\boldsymbol{\mathcal{I}}(x, y) = \sum_{\substack{x' \leq x \\ y' \leq y}} \mathbf{I}(x', y'), \tag{2.1}$$

which can be done in a single pass recursively,

$$\begin{aligned} \boldsymbol{\mathcal{I}}(x, y) =& \mathbf{I}(x, y) + \boldsymbol{\mathcal{I}}(x - 1, y) + \dots \\ & \boldsymbol{\mathcal{I}}(x, y - 1) - \boldsymbol{\mathcal{I}}(x - 1, y - 1), \end{aligned} \tag{2.2}$$

with the base case being

$$\boldsymbol{\mathcal{I}}(x, y) = \{0 : x < 0 \vee y < 0\}. \tag{2.3}$$

2. Compute the value of each rectangle feature from the integral image $\boldsymbol{\mathcal{I}}$. This can be computed in linear time using at most nine array references (e.g. for four rectangle features).

3. *Optimizing a classification model from a given training set*. Viola and Jones propose cascading multiple sets of weak classifiers (Viola & Jones, 2001) trained using AdaBoost (Freund & Schapire, 1996). 38 layers cascaded classifier is reportedly sufficient for detecting frontal upright faces (Viola & Jones, 2001).

This thesis implemented the Viola Jones algorithm for the face detection considering the fact that Viola Jones algorithm is a lightweight algorithm to detect faces and various facial landmarks including eyes and mouth and also it has low algorithmic complexity which makes the algorithm suitable for time-critical applications.

### 2.2.2 Face Recognition

In recent years, face recognition has gained more attention from engineers and neuroscientists, since it gives many potential opportunities to develop automatic access control and computer vision applications. Face detection has an important role in face recognition algorithm as it becomes the first step of automatic face recognition development. The quest in face recognition is how to determine the identity of a person from any given input image using existing database of face images from known individuals.

Biometric-based methods have become the favourite option for person identification which can be alternatives replacing authentication and access granting methods using physical and virtual tools that use domains based, such as: key, cards, smart cards, passwords, Each method has disadvantage such as: passwords and PINs sometimesare difficult be remember or can be predicted systematically or randomly; keys and cards sometimes are lost, stolen or duplicated; and also magnetic cards may become unreadable; the biometric-based methods make use of any physiological characteristics of individuals to determine their identity which cannot be misplaced, forgotten or stolen (Parmar & Mehta, 2014).

Biometric-based technologies include identification based on physiological characteristics (such as face, fingerprints, hand geometry, palm, iris, retina, ear and voice) (Bolle & Pankanti, 1998). Face recognition appears to offer several advantages over other biometric methods, a few of which are outlined here.

All the technologies require voluntary action by the user, such as: the user needs to place his hand on a device to identify fingerprinting or hand geometry detection or has to stand in a fixed position in front of a camera for iris or retina identification. However, face recognition can be conducted passively without any explicit action or participation on the part of the user since face images can be acquired from a distance by a camera. This is particularly beneficial for security and surveillance purposes.

Furthermore, data acquisition in general is fraught with problems; for example, hands and fingers can be rendered useless if the epidermis tissue is damaged (bruised or cracked). Iris and retina identification require expensive equipment and are very sensitive to any body motion. Voice recognition is susceptible to background noises and auditory fluctuations on a phone line or tape recording. Signatures can be modified or forgotten. In addition, methods that require multiple individuals to use the same equipment to capture the biometric characteristics potentially expose the user to the transmission of germs and impurities from other users.

In contrast, facial images can be easily obtained with an inexpensive fixed camera. Good face recognition algorithms and appropriate preprocessing of the images can compensate for noise and slight variations in orientation, scale and illumination. Additionally, face

recognition is non-intrusive and does not carry any health risks (Rabia & Hamid, 2009).

The disadvantage of face recognition is the fact that in the most common form of the frontal view, faces appear to be roughly alike and the differences between them are quite subtle. Consequently, frontal face images form a very dense cluster in image space which makes it virtually impossible for traditional pattern recognition techniques to accurately discriminate among them with a high degree of success (Nastar & Mitschke, 1998). Furthermore, the human face is not a unique, rigid object. Indeed, there are numerous factors that cause the appearance of the face to vary. The sources of variation in the facial appearance can be categorized into two groups: intrinsic factors and extrinsic ones (Gong *et al.*, 2000).

Intrinsic factors are related to the physical nature of the face and are independent of the observer. These factors can be further divided into two classes: intrapersonal and interpersonal (Jebara, 1995). Intrapersonal factors are responsible for varying the facial appearance of the same person, some examples being age, facial expression and facial paraphernalia (facial hair, glasses, and cosmetics). Interpersonal factors are responsible for the differences in the facial appearance of different people, some examples being ethnicity and gender.

Extrinsic factors cause the appearance of the face to alter via the interaction of light with the face and the observer. These factors include illumination, pose, scale and imaging parameters, such as resolution, focus, imaging, noise.

Considering such challenging yet interesting factors, the face recognition methods can be classified into 3 groups as follows:

a Feature-based (structural)

b Holistic Matching

c Hybrid

#### 2.2.2.1. Feature-based (structural)

In this method local features such as eyes, nose and mouth are first of all extracted and their locations and local statistics (geometric and/or appearance) are fed into a structural classifier. A big challenge for feature extraction methods is feature "restoration"; this is when the system tries to retrieve features that are invisible due to large variations, for example, head pose when we are matching a frontal image with a profile image (Zhao *et al.*, 2003). The method can be further divided into 3 different extraction techniques: Generic methods based on edges, lines, and curves; Feature-template-based methods; Structural matching methods that take into consideration geometrical constraints on the features.

#### 2.2.2.2. Holistic Matching

Using the holistic approach, the complete face region is taken into account as input data into face catching system rather than on local features of the face. The methods can be subdivided into two groups: statistical and Artificial Intelligent (AI) approaches, while the most widely used algorithms in this method are eigenfaces (Turk & Pentland, 1991), Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA) (Delac *et al.*, 2005).

#### 2.2.2.3. Hybrid

Hybrid face recognition systems use a combination of both holistic and feature extraction methods. 3D Images are used in hybrid methods. The image of a person's face is caught in 3D, allowing the system to note the curves of the eye sockets, for example, or the shapes of the chin or forehead. Even a face in profile would serve because the system uses depth, and an axis of measurement, which gives it enough information to construct a full face (Parmar & Mehta, 2014). The 3D system usually proceeds thus: Detection, Position, Measurement, Representation and Matching.

### 2.2.3 Applications of Face Recognition

Face recognition can be implemented in many area with different purposes. Some possible applications can be summarized into the following:

a Face Identification: Face recognition systems identify people from their face images. Face recognition systems establish the presence of an authorized person and replace the use of a valid identification (ID), key, Pins, or passwords.

b Access Control: Giving the access control to a device such as office access or computer logon. In the access control, the face pictures can be caught under natural conditions, such as frontal faces and indoor illumination with high accuracy and without much co-operation from the user.

c Security: An example of a security system application is in airport protection systems that use face recognition technology; these have been implemented at many airports around the world.

d Surveillance: Like security applications in public places, surveillance by face recognition systems has a low user satisfaction level, if not lower. Free lighting conditions, face orientations and other divisors all make the deployment of face recognition systems for large scale surveillance a challenging task.

e Facial expression recognition: This is the interpretation of human facial characteristics that is read by an input sensing device such as a webcam. The facial characteristics are then interpreted using any face recognition algorithm to identify specific face expression and translated into emotional state of the individual, such as: happy, sad, angry, fear, and other emotions. Facial expression recognition can be useful in many areas, for example in medical science a doctor can be alerted when a patient is in severe pain and immediately take prompt action to help the patient. The application of facial expression recognition system is discussed further in the next section as part of the research focus of this thesis.

## 2.3 Facial Expression Recognition System

Facial expressions have been studied by clinical and social psychologists, medical practitioners, actors and artists. However in the last quarter of the 20th century, with the advances in the fields of robotics, computer graphics and computer vision, animators and computer scientists have started showing interest in the study of facial expressions.

In the past, facial expression analysis was a primary research subject for psychologists such as Ekman and Friesen, who have postulated six primary emotions. Each emotion possess a distinctive content with a specific facial expression. These prototypic emotional displays are also called basic emotions. The referred emotions are universal across human ethnicities and cultures and comprise happiness, sadness, fear, disgust, surprise and anger (Ekman & Friesen, 1976) .

Recent advances in image analysis and pattern recognition open up the possibility of automatic detection and classification of emotional and conversational facial signals. Automating facial expression analysis could bring facial expressions into man-machine interaction as a new modality and make the interaction tighter and more efficient. Such a system could also make classification of facial expressions widely accessible as a tool for research in behavioral science and medicine.

Various applications using automatic facial expression analysis can be envisaged in the near future, fostering further interest in doing research in different areas, including image understanding, psychological studies, facial nerve grading in medicine (Dulguerov *et al.*, 1999), face image compression and synthetic face animation, video-indexing, robotics as well as virtual reality.

However, facial expression recognition should not be confused with human emotion recognition as is often done in the computer vision community. While facial expression recognition deals with the classification of facial motion and facial feature deformation into abstract classes that are purely based on visual information, human emotions are a result of many different factors and their state might or might not be revealed through a number of channels such as emotional voice, pose, gestures, gaze direction and facial expressions.

Furthermore, emotions are not the only source of facial expressions. In contrast to facial expression recognition, emotion recognition is an interpretation attempt and often demands understanding of a given situation, together with the availability of full contextual information. Fig. 2.4 provides the details of this conceptual thinking (Fasel & Luettin, 2003).

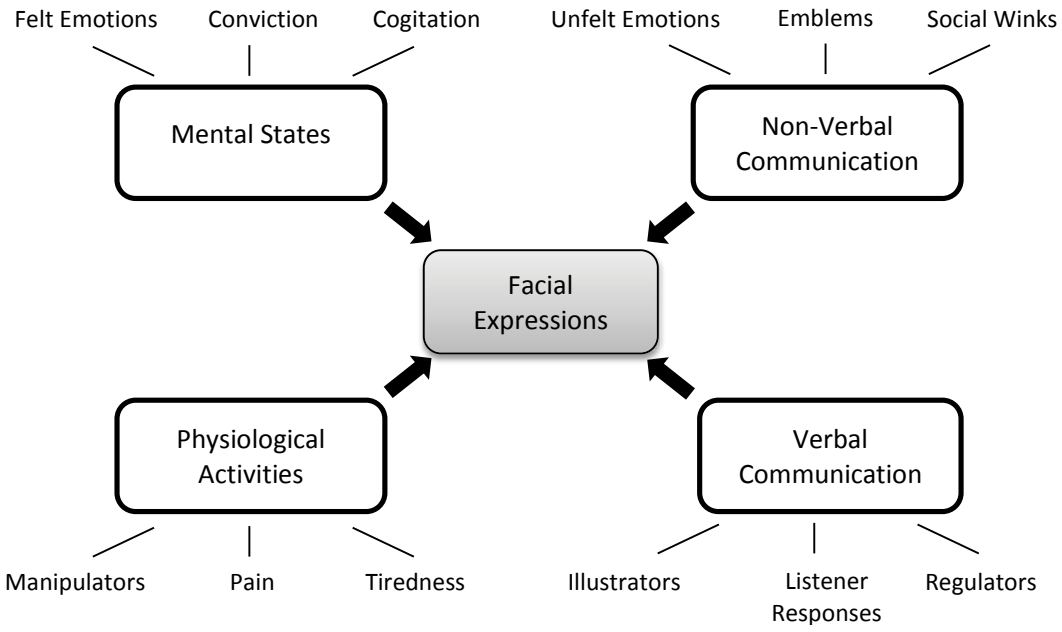



Figure 2.4: Sources of facial expressions (Fasel & Luettin, 2003)

Automatic facial expression analysis is a complex task as physiognomies of faces vary from one individual to another quite considerably due to different age, ethnicity, gender, facial hair, cosmetic products and occluding objects such as glasses and hair. Faces appear disparate because of pose and lighting changes. Variations such as these have to be addressed at different stages of an automatic facial expression analysis system (Fasel & Luettin, 2003).

In general, facial expression analysis includes both measurement of facial motion and recognition of expression. The general approach to automatic facial expression analysis consists of 3 steps: face detection, facial feature extraction, and facial emotion recognition (Tian *et al.*, 2011). Fig. 2.5 illustrates the process.

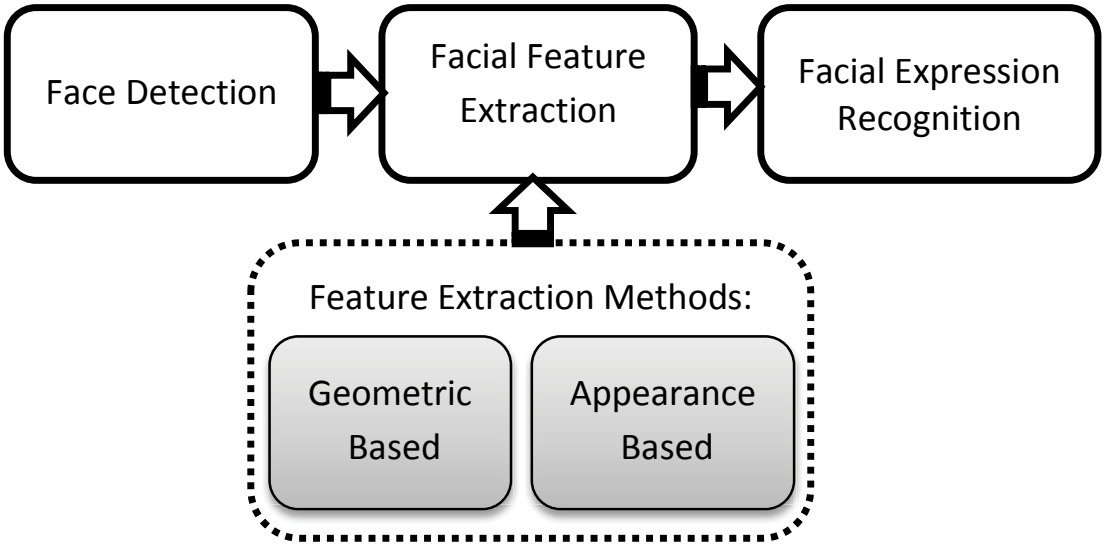



Figure 2.5: Facial expressions analysis (Tian *et al.*, 2011)

The algorithm used in facial expression recognition system is adopted from facial recognition techniques. However, the success of facial expression recognition system relies on the features used to distinguished the facial expressions. Therefore, facial emotion recognition puts a particularly strong emphasis on facial feature extraction to collect some specific features from the face. The facial features extraction method for facial emotion recognition can be categorized into 2 groups namely: Geometric Based and Appearance Based (Tariq *et al.*, 2012).

### 2.3.1 Geometric Based Method

The geometric facial features present the shape and locations of facial components (including mouth, eyes, brows, nose, etc.). The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. Geometric features present the shape and locations of facial components, which are extracted to form a feature vector that represents the face geometry.

A famous geometric feature is Active Appearance Models (AAM) (Edwards *et al.*, 1998; Zheng & Liu, 2016). However, the geometric feature-based methods usually require accurate and reliable facial feature detection and tracking, which is difficult to accommodate in many situations (Shan *et al.*, 2009). The disadvantages of the geometric method is

described in the following (Donato *et al.*, 1999):

a The approximate locations of individual face features are detected automatically in the initial frame; however, in order to carry out template based tracking, the contours of these features and components have to be adjusted manually in this frame to each individual subject.

b In cases of pose and illumination changes, the problems of robustness and difficulties emerge, while the tracking is applied on images.

c As actions and expressions tend to change both in morphological and in dynamical senses, it becomes hard to estimate general parameters for movement and displacement. Therefore, ending up with robust decisions for facial actions under these varying conditions becomes very difficult.

### 2.3.2 Appearance Based Method

Appearance-feature based approaches represent the texture of local regions of the face, which captures the intensity changes associated with different expressions, such as wrinkles, bulges and furrows. During an expression, movement of facial organs is always associated with change in appearance of the corresponding region by producing wrinkles, skin folding etc.

Some popular appearance features are Gabor descriptor (Li *et al.*, 2015) and Local Binary Patterns (LBP) (Zhao & Zhang, 2012). Lately, Histograms of Oriented Gradient (HOG) descriptors have received a lot of attention for the purpose of object detection (Dalal & Triggs, 2005).

HOG is a shape descriptor that counts occurrences of gradient orientations in localized portions of an image and that is mainly used for the purpose of object detection but that is also intuitively useful to model the shape of the facial muscles by means of an edge analysis. HOG descriptor has been applied as a tool for object recognition (Creusen *et al.*, 2010; Liang *et al.*, 2012), human detection (Zhu *et al.*, 2006; Yang *et al.*, 2012), and face recognition (Dniz *et al.*, 2011; Salhi *et al.*, 2013; Tan *et al.*, 2014; Baltrusaitis

*et al.*, 2015). However its implementation for facial expression recognition is very few because of its complexity and high dimension as an appearance-based feature.

#### 2.3.2.1. HOG descriptor for Facial Emotion Recognition feature

HOG is a robust morphological image descriptor that is insensitive to light variation. The main concept of the HOG descriptor is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. The image must be divided into small connected regions called cells, and for the pixels within each cell, a HOG directions is compiled. The descriptor is the concatenation of these histograms. For improved accuracy, the local histograms can be contrast-normalized by calculating a measure of the intensity across a larger region of the image, called a block, and then using this value to normalize all cells within the block. This normalization results in better invariance to changes in illumination and shadowing (Dalal & Triggs, 2005). The block diagram of HOG computation is illustrated in Fig. 2.6.

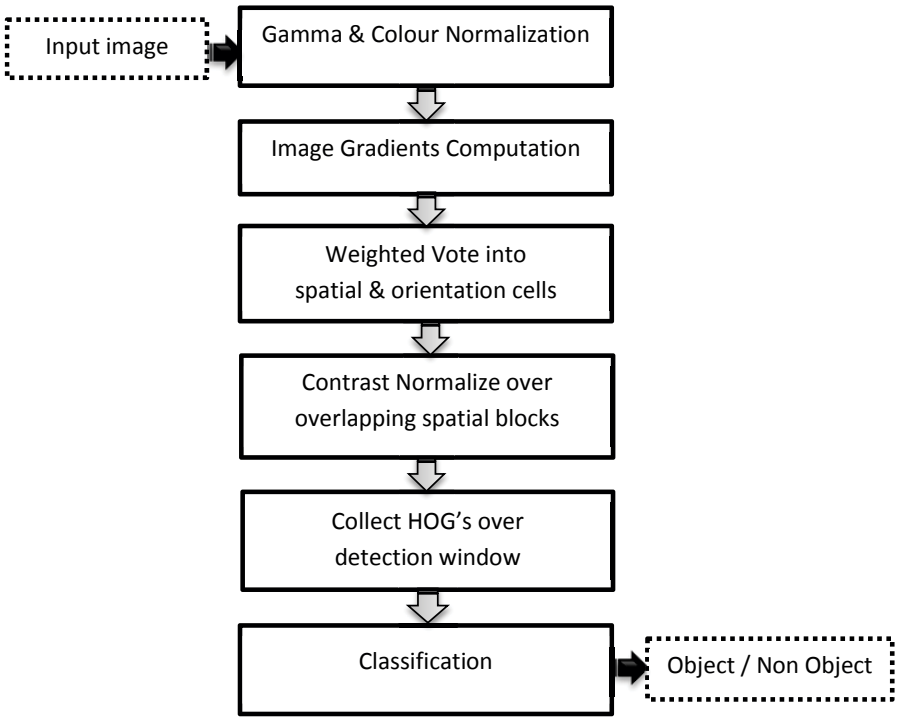



Figure 2.6: Block diagram of HOG computation (Dalal & Triggs, 2005)

The HOG descriptor has a few key advantages over other descriptors. Since it operates on

local cells, it is invariant to geometric and photometric transformations, except for object orientation. Such changes would only appear in larger spatial regions. Moreover, as Dalal and Triggs discovered, coarse spatial sampling, fine orientation sampling, and strong local photometric normalization permits the individual body movement of pedestrians to be ignored so long as they maintain a roughly upright position. The HOG descriptor is thus particularly suited for human detection in images (Dalal & Triggs, 2005).

An overview of the static HOG feature extraction computation is decribed in the following as proposed by Dalal and Triggs (Dalal & Triggs, 2005).

1. Compute the gradient of the input image E with

$$\mathbf{G}_{vert} = \mathbf{E} * [-1, 0, 1], \tag{2.4}$$

$$\mathbf{G}_{horiz} = \mathbf{E} * [-1, 0, 1]^T, \tag{2.5}$$

2. Compute the gradient magnitude $|g^E(x, y)|$ and orientation $\theta^E(x, y)$ using

$$|g(x, y)| = \sqrt{G^2_{vert}(x, y) + G^2_{horiz}(x, y)}, \tag{2.6}$$

$$\theta(x, y) = \tan^{-1}\left(\frac{G_{vert}(x, y)}{G_{horiz}(x, y)}\right), \tag{2.7}$$

3. Divide the orientation image matrix $\boldsymbol{\theta}^E$ into equally spaced cells. Each $\theta^E(x, y)$ within each cell is binned into a 9-point orientation histogram weighted by $|g^E(x, y)|$. Since $\boldsymbol{\theta}^E$ is sparse, only those cells with nonzero values needs to be computed.

4. Group the cells together into blocks and normalize the gradient strengths contained within each block,

$$v_i^{\text{normalized}} = \frac{v_i}{\sqrt{\|\boldsymbol{v}\|_2^2 + \epsilon^2}}, \tag{2.8}$$

where $\boldsymbol{v} = \{v_1, \ldots, v_9\}$ denotes a non-normalized vector containing all histograms in a given block, $\epsilon$ denotes a small constant for preventing a zero-valued denominator.

The HOG feature is implemented in this thesis considering its advantages of being relatively insensitive to the lighting condition and its robustness as facial recognition feature. Also in this thesis, an improved HOG feature is developed with the advantage of less sensitivity to lighting condition and also low dimensionality compared to the original HOG.

## 2.4 Database for Facial Expression Recognition

One of the most important aspects of developing any new recognition or detection system is the choice of the database that will be used for testing the new system. If a common database is used by all the researchers, then testing the new system, comparing it with the other state of the art systems and benchmarking the performance becomes a very easy and straightforward job. However, building such a common database that can satisfy the various requirements of the problem domain and become a standard for future research is a difficult and challenging task.

When compared to face recognition, face expression recognition poses a very unique challenge in terms of building a standardized database. This challenge is due to the fact that expressions can be posed or spontaneous which are very different in their characteristics, temporal dynamics and timings. Thus, a standardized training and testing database that contains images and video sequences of people displaying spontaneous expressions under different conditions of head posed, occlusions, and lighting conditions is required.

Some popular facial expression databases are publicly available without any cost. From many available databases, reviewing all of them will not be possible. Therefore, only those databases that have mostly been used in the past few years are presented in Table 2.1.

Table 2.1: Summary of Databases for Facial Expression Recognition

| Database | Participants Details | Description |
|---|---|---|
| Extended Cohn-Kanade (Kanade *et al.*, 2000; Lucey *et al.*, 2010). | 210 adults with 593 image sequences; Gender: 69% female; Age: 18 to 50 years; Ethnicity: 81% Euro–American, 13% Afro-Ameican, and 6% other ethnics; Each specific facial expressions are annotated with FACS Action Units. | The participants were asked to pose 23 facial expressions in either single action or combinations of action units, started and finished with neutral expression. 7 prototypic emotions (happiness, surprise, anger, fear, disgust, sadness, and contempt) were annotated by certified FACS coders. Images were taken using 2 cameras which located in front of the participants and at 30 degrees to their right side. Only the images taken from the frontal camera are provided in the database. |
| MMI Facial Expression (Pantic *et al.*, 2005) | 19 students and staffs; Gender: 44% female; Age: 19 to 62 years; Ethnicity: European, Asian, or South American; Annotated with Action Units and metadata (data format, facial view, shown AU, shown emotion, gender, age). | The database contains posed and spontaneous expressions with frontal and profile view images. The subjects were asked to display 79 series of expressions with a single AU, or a combination of a minimal number of AUs, or expressions of emotion with a short neutral expression at start at the end of each expression. The emotions were determined using an expert annotator. |
| The Japanese Female Facial Expression (JAFFEE) (Lyons *et al.*, 1998) | 10 Japanese female models; 219 images; Each image were rated on 6 emotion adjectives by 92 Japanese female undergraduates. | The database contains only posed expressions with 6 basic facial expressions (happiness, sadness, surprise, anger, disgust, fear) and a neutral face. The images have been taken under strict controlled conditions of similar lighting and with the hair tied away from the face. All the expressions are multiple AU expressions. |

Based on the comparison of the databases in Table 2.1, the Cohn-Kanade database has shown more advantages compared to others. Therefore, the Cohn-Kanade database is selected for the thesis experiment in facial expression recognition to build the dataset for training and testing of the classifier.

## 2.5 Summary of Latest Researches Methodology on Facial Expression Recognition System

The most recent application of facial expression recognitions are listed in the following Table 2.2, comparing latest reference with the method, feature used, number of emotions measured, and database used.

Table 2.2: Comparison Between Facial Expression Recognition Methods

| Method | Feature | # Emotions | Database | Reference |
|---|---|---|---|---|
| Appearance | Modified SIFT combined with DWT and CoC | 5 | JAFFEE | (Neeru & Kaur, 2016) |
| Appearance | Gabor filters and GLCM | 7 | JAFFEE | (Li *et al.*, 2015) |
| Appearance | Gabor filtering and KPCA | 7 | Cohn–Kanade | (Li & Lam, 2015) |
| Geometric | RHMM | 7 | Cohn–Kanade | (Sun & Akansu, 2014) |
| Appearance | MSR | 7 | Cohn–Kanade | (Ptucha & Savakis, 2013) |
| Geometric | AUs | - | MMI | (Pantic & Patras, 2006) |

SIFT:Scale Invariant Feature Transform; DWT: Discrete Wavelet Transform;
CoC: Coefficient of Correlation
GLCM Gray-Level Co-occurrence Matrix
KPCA: Kernel Principal Component Analysis
RHMM:Regional Hidden Markov Model
MSR: Manifold based Sparse Representation
AU: Action Units

Firstly, Table 2.2 shows that appearance-based methods are the most widely used due to their advantage compared to geometric-based method as shown in previous section. Secondly, the researchers also prefer to use Cohn-Kanade database. Lastly, the number of emotions to be detected are mostly 7 as the highest.

## 2.6 EEG Measurement and Its Applications

Early diagnosis of a variety of diseases can be obtained by acquiring signals and images from the human body by collecting the data in the form of electrobiological signals such as electroencephalography (EEG) and magnetoencephalography (MEG) from the brain, electrooccluography ( or electrooptiphy (EOG)) from eye nerves, and electrogastrography (EGG) from the stomach.

Different measurements can also be collected in the form of ultrasound or radiograph (ultrasound image), computerized tomography (CT), magnetic resonance imaging (MRI) or (fMRI), positron emission tomography (PET), and single photon emission tomography (SPET). Physiological and functional changes in the brain can be depicted by either EEG, MEG, or functional MRI (fMRI). However, application of fMRI is very limited in comparison with EEG because of its limitation of low time resolution (two frames/s) and higher cost.

The first electrical neural activities were registered using simple galvanometers. In order to magnify very fine variations of a pointer, a mirror was used to reflect the light projected to the galvanometer on the wall. The d'Arsonval galvanometer then featured a mirror mounted on a movable coil and the light focused on the mirror was reflected when a current passed the coil. The capillary electrometer was introduced by Lippmann and Marey. The string galvanometer as a very sensitive and more accurate measuring instrument was introduced by Einthoven in 1903. This became a standard instrument for a few decades and enabled photographic recording(Sanei & Chambers, 2007).

The first EEG of a human being was recorded by Hans Berger, a German neurologist as early as in 1924. An EEG can be recorded using an electrode which is displayed as an oscillating signal reflecting the electric potential from the group of neurons located in close proximity to the electrode. This signal actually shows the activity of synchronized neurons from a certain region of the brain.

In the early day, the recording of the EEG was only suitable to detect large differences arising in the signal pattern produced, such as epileptic convulsive. This is because the

fact that the rhythms of the brain produced from the low quality instrument had to be manually inspected in order to detect changes in the recording.

Nowadays, with more precise recording equipment, and the availability of sufficient computational power in modern computers, and also more empirical studies of EEG, it is possible to detect even more subtle changes in the electric potential recorded. This means that the cognitive processes such as selective attention, working memory, mental calculations, as well as specific cognitive states and different types of behaviour can be encoded and recognized from these subtle changes (Kvaale, 2012).

Next generation EEG systems combine a number of electrodes and a set of differential amplifiers (one for each channel) followed by filters, and needle (pen)-type registers. These multichannel EEGs are then plotted on plain or grid paper. Afterward, researchers looked for a computerized system with digitized and stored signals. This encouraged the emergence of digital signals EEG which required sampling, quantization, and encoding process. This was followed by the growth of the number of electrodes used which also raised the data volume, which means number of bits increases. The digital systems allow different settings like variable, stimulations, and sampling frequency together with advanced processing tools (Sanei & Chambers, 2007).

### 2.6.1 EEG Signal Processing and Signal Conditioning

In digital EEG, the conversion from analogue to digital is carried out with multichannel analogue-to-digital converters (ADCs). Effective bandwidth for EEG signals approximately is 100 Hz. Therefore, to satisfy the Nyquist criterion, a minimum sampling frequency of 200 samples/s is adequate for EEG signals. In some applications with higher resolution, sampling frequencies of up to 2000 sample/s may be used to represent brain activities in the frequency domain.

In order to maintain the diagnostic information, representation of each signal sample with up to 16 bits for the quantization of EEG signals is normally very fine. To record one hour of EEG signal with 128-electrodes and 500 samples/s of signal sample will need a memory size of $128 \times 60 \times 60 \times 500 \times 16 \approx 3.68$ Giga Bytes $\approx 0.45$ GB. Therefore, to

record large number of patients would need huge storage facilities such as large removable hard drives, and optical disks.

Furthermore, there are many and varied formats of EEG data captured from different EEG machines. However these formats are easily convertible to readable spreadsheets format and processed with most signal processing software packages such as MATLAB (Sanei & Chambers, 2007).

The raw EEG signals carry out amplitudes in the order of volts and contain variation of frequency components of up to 300 Hz. These signals are amplified before the ADC, and then filtered either before or after the ADC, to reduce noise and make the signals ready for processing and visualization. The filters are designed to reduce any change or distortion to the signals. Often, Highpass filters with a cut-off frequency less than 0.5 Hz are used to remove very low frequency disturbing components such as breathing. On the other hand, lowpass filters with a cut-off frequency of approximately 50–70 Hz are used to reduce high-frequency noise. Notch filters with a null frequency of 50 Hz are also necessary to ensure perfect rejection of the strong 50 Hz power supply. In this case the sampling frequency used can be as low as twice the bandwidth in common EEG systems. The common sampling frequencies for EEG recordings are 100, 250, 500, 1000, and 2000 samples/s.

There are 2 main artifacts for EEG signal: patient-related (physiological) and system artifacts. Often, these artifacts are highly mitigated in the preprocessing stage so that useful information can be restored. Including in the patient-related artifacts are body movement-related, EMG, ECG (and pulsation), EOG, and sweating, while the system artifacts are 50/60 Hz power supply interference, impedance fluctuation, cable defects, electrical noise from the electronic components, and unbalanced impedances of the electrodes.

### 2.6.2 EEG Electrodes and Electrodes Positioning

The EEG electrodes are utilized to acquire high quality data. There are different types of electrodes which can be used in the EEG recording systems, such as: disposable (gel-less, and pre-gelled types); reusable disc electrodes (gold, silver, stainless steel, or tin);

headbands and electrode caps; saline-based electrodes; and needle electrodes.

Electrode caps are often use for multichannel recordings with a large number of electrodes. The most commonly used are scalp electrodes consist of Ag–AgCl disks, with diameter less than 3 mm, and long flexible leads plugged into an amplifier. Needle electrodes have to be implanted under the scalp with minimal invasive operations. Distortion might occur related to high impedance between the cortex and the electrodes as well as the electrodes with high impedances, which can conceal the actual EEG signals. Commercial EEG recording systems usually are equipped with impedance monitors. The electrode impedances should be maintained to less than 5 k ohm and be balanced to within 1 kΩ of each other to get a satisfactory recording read (Sanei & Chambers, 2007).

There are various different systems for electrodes positioning (Illinois, Montreal, Aird, Cohn, Lennox, Merlis, Oastaut, Schwab, Marshall). The most widely used at present is 10/20 international system. According to this system the electrodes are positioned as follows: the common electrode is placed remote of the skull (earlobe, nose, or chin). It is counted as nasion and inion on data points. Ten percent of the data points are the prefrontal and occipital planes. The rest is divided into four equal parts of 20% each. The number of electrodes used and the position depends on the particular signal that is needed for the analysis. There are five cross-sectional planes:

- Prefrontal: Fpz
- Frontal: Fz
- Vertex: Cz
- Parietal: Pz
- Occipital: Oz (Roman-Gonzalez, 2012).

Fig. 2.7 shows the positioning of the EEG electrodes for EEG measurement using 10/20 international system.

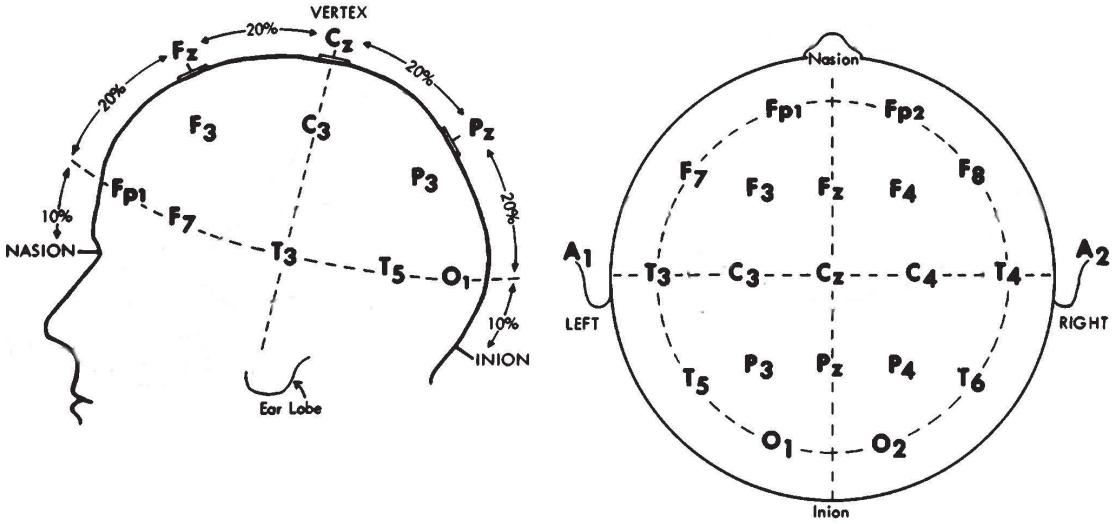



Figure 2.7: EEG electrodes positioning with 10/20 international system (Roman-Gonzalez, 2012)

### 2.6.3 Brain Rhythms

Since the first recording of EEG, the classification of different types of brain rhythms and the relation with different pathologies and functions became the concern of the researches. Brain rhythms can be divided into several frequency bands which related to different brain states, functions or pathologies (Quiroga, 1998). The division is as follows:

a. Delta ($\delta$) rhythms (0.5 – 3.5Hz): These rhythms are characteristic of deep sleep stages. Furthermore, delta oscillations with certain specific morphologies, localizations and rhythms are correlated with different pathologies.

b. Theta ($\theta$) rhythms (3.5 – 7.5Hz): These rhythms are enhanced during sleep and have an important role in infancy and childhood. In an awake adult, high theta activity is considered abnormal and it is related with different brain disorders.

c. Alpha ($\alpha$) rhythms (7.5 – 12.5Hz): These rhythms appear spontaneously in normal adults during wakefulness, while in relaxed or mental inactivity conditions. They occur when eyes are closed and are most pronounced in occipital locations.

d. Beta ($\beta$) rhythms (12.5 – 30Hz): These rhythms are best defined in central and frontal

locations, with less amplitude than alpha waves and will be enhanced upon expectancy states or tension. Traditionally the rhythms are subdivided in 1 and 2 oscillations.

e. Gamma ($\gamma$) rhythms (30 – 60Hz): These rhythms are occurring in physical and mental activity such as public appearance, game arena, panic, scared, nervous in a conscious condition.

The brainwave rhytms in different frequencies are illustrated in Fig. 2.8.

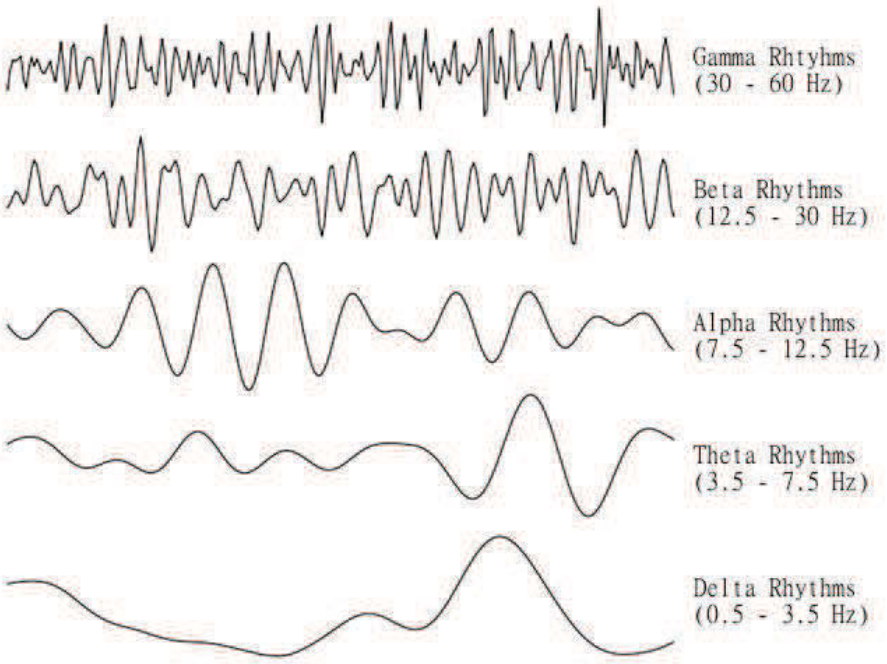



Figure 2.8: Brainwave Rhythms in various band frequencies

### 2.6.4 Applications of EEG Signals

Considering that all human activities are related to the brain both mentally or physically, including the planning, coordination, execution and controlling of the activities. These activities also give feedback to the brain as an affect or change in an alteration of brainwave pattern. Therefore, the application of EEG is very vast and open to any possibilities. The limit is only the availability of technology and knowledge to exploit and implement the methods. This also means that new possibilities require new exploration and experiment to verify the result to a certain acceptable level depending on the area of implementation.

For example, the applications in health and medical areas require a higher standard of acceptance compared to computer control applications such as Brain Computer Interface (BCI). Some existing researches in EEG applications are: EEG in sleep stages both in normal and abnormal (Colten *et al.*, 2006), abnormal EEG patterns of the ageing (Purdon *et al.*, 2015), mental disorders, for example in dementia (Signorino *et al.*, 1995), epileptic seizure (Jouny *et al.*, 2011), psychiatric disorders (Oberman *et al.*, 2008). Recently, it is also implemented for detection of freezing of gait (Handojoseno *et al.*, 2015), and driver fatigue (Chai *et al.*, 2016b). One of the the most favoured implementation is in BCI for neuroscience applications(Machado *et al.*, 2010) or gaming applications (Liao *et al.*, 2011).The application is also implemented for emotion recognition system (Wang *et al.*, 2015; Papadaniil *et al.*, 2015; Khosrowabadi *et al.*, 2014).

## 2.7 EEG emotion recognition system

To find the correlation between EEG and emotional states should be started with the methodology to define the emotional state space. There are two large categories of emotional state spaces as previously mentioned: discrete and continuous space. The discrete space describe the emotional state as various different types of discrete emotional states as well as happiness, surprise, sadness, anger disgust, contempt and fear (sometimes with more additional state space), while continuous space represents the emotional state as a multidimensional space vector. A few different emotional dimensions have been proposed and the most renowned is 2-dimensional circular space using arousal and valence dimensions which known as the circumplex model (Kim *et al.*, 2013; Russell, 2003; Barrett, 1998; Schaefer, 1959).

The study of EEG and emotional states should also consider the stimuli to induce the emotions. Usually, emotion stimuli are taken to address arousal and valence states which are presented in many modalities including visual, auditory, tactile and odor stimulation. The accuracy of the result generated by the stimulus are measured with self-rating of subjects or standard stimulus sets such as the International Affective Picture System (IAPS)

or the International Affective Digitized Sound system (IADS). IAPS provides sets of normative pictures for emotional stimuli to induce emotional changes and attention levels while IADS supplies acoustic stimuli to induce emotions (Kim *et al.*, 2013).

The study in the neuropsychological area divides the EEG features in correlation to emotional states into two domains: time domain and frequency domain. The time domain focusses on the event- related potentials (ERPs) that reflect the emotional states. The ERP components of short to middle latencies are correlated with valence, whereas the ERP components of middle to long latencies are correlated with arousal. The ERPs computation takes the averaging of EEG signals over multiple trials. The frequency domain correlates the spectral power in various frequency bands with different emotional states. For example, the alpha power is correlated to the valence state or with discrete emotions such as happiness, sadness, and fear. Also, the event-related synchronization (ERS) and desynchronization (ERD) of the gamma power are correlated to happiness and sadness (Kim *et al.*, 2013; Balconi & Mazza, 2009; Balconi & Lucchiari, 2006, 2008; Keil *et al.*, 2001; Muller *et al.*, 1999).

Although the use of EEG for emotion analysis has been conducted with various methods and algorithms by different researchers, however, there is no consensus of a standard procedure for EEG signal processing (Wichakam & Vateekul, 2014). Some examples of methods and algorithms used for EEG emotion analysis are demonstrated in the following articles (Petrantonakis & Hadjileontiadis, 2012; Conneau & Essid, 2014; Liu *et al.*, 2013).

### 2.7.1 Features, Channel Selection and Number of Electrodes for EEG Emotion Recognition

The features used in EEG emotion classification can be divided into 2 main categories: 1) time; 2) and frequency domain. The time domain feature exploits Event Related Potential (ERP), statistical computation of EEG signal, Hjorth features, Non Stationary Index (NSI), Fractal Dimensions, and Higher Order Crossing, while the frequency domain makes use of Band Power and Higher Order Spectra (HOS). There is also the implementation of a combination of time and frequency domains which exert Hilbert-Huang

Spectrum and Discrete Wavelet Transform (Jenke *et al.*, 2014a).

Other substantial factors which should be considered in EEG emotion classification are the number of electrodes used and the channel selection for EEG recording (Zheng & Lu, 2015a). For example, Jatupaiboon et.al. using 14 channels EMOTIV EEG system managed to reduce the number of electrodes used from 7 pairs (14 electrodes) to 5 pairs (10 electrodes) without significant difference in the result. They applied channel selection in pairs namely: F7–F8, AF3–AF4, F3–F4, FC5–FC6, T7–T8, and O1–O2. Using Wavelet Transform feature, the EEG signal is subdivided into 5 frequency bands of Delta (0–4Hz), Theta (4–8Hz), Alpha (8–16Hz), Beta (16–32Hz), and Gamma (32–64Hz) (Jatupaiboon *et al.*, 2013).

Other researchers reduced the number of channels from 64 to 5 channels using Synchronization Likelihood (SL) method with only a slight loss of classification. The method proposed the following selected channels: AF2, F3, F4, CP5 and CP6 respectively (Ansari-Asl *et al.*, 2007).

Moreover, Wichakam and Vateekul proposed a different channel selection. They proposed reduction of 32 to 10 channels by utilizing a band power method, and the selected channels are: FP1, FP2, F3, F4, T7, T8, P3, P4, O1, and O2 (Wichakam & Vateekul, 2014).

### 2.7.2 Wavelet Transform for EEG Feature Extraction

In recent years, the Wavelet Transform (WT) has become popular for EEG features extraction method to obtain features from EEG signal in time and frequency domain at the same time. It has been indicated that the WT is advantageous for several reasons (Rosso *et al.*, 2001b):

1 The relative wavelet energy can be naturally associated with the EEG frequency bands;

2 The relative entropy of its wavelet power can serve as a reliable estimate of the degree of similarity between different segments of the signal.

The word wavelet means a small wave with limited duration and zero average values. Wavelet is used as a function to localize another function or a set of data in time and frequency domain. The first reference to wavelets is Haar wavelets in 1909. In the mid 80s the broader concept of wavelet was introduced by Grossman and Morlet. The wavelet transform becomes famous because its effectiveness in obtaining a signal at a particular time and frequency, or extracting features at various locations in space at different scales (time frequency localization) and its capabilities for multi–rate filtering (differentiating signals with their various frequencies). Desired features from an input signal can be extracted which are characterized by certain local properties in time and space of the signal (Graps, 1995).

Wavelet analysis can be divided into 2 groups: Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). With CWT the signal to be analysed is matched and convolved with the wavelet basis function at continuous time and frequency increments. Even in CWT, the data have to be digitized. Continuous time and frequency increments indicate that data at every digitized point or increment is used. As a result, the original signal is expressed as a weighted integral of the continuous basis wavelet function.

In contrast, with DWT, the inner product of the original signal with the basis wavelet function is taken at discrete points (usually dyadic to ensure orthogonality) and the result is a weighted sum of a series of basis functions. The basis for wavelet transform is the wavelet function. Wavelet functions are families of functions satisfying prescribed conditions, such as continuity, zero mean amplitude, and finite or near finite duration. Other properties categorize wavelets, for example, orthogonality and biorthogonality, regularity (Resnikoff & Wells, 1998).

#### 2.7.2.1. EEG Feature Extraction with Discrete Wavelet Transforms (DWT)

The DWT is advantageous in a sense that it provides time-frequency localization, multi-scale zooming, and multirate filtering for detecting and characterizing transients. These advantages allow DWT to potentially extract the appropriate information from non-stationary signals such as the EEG signals (Handojoseno *et al.*, 2012).

Given any one dimensional time signal $x(t)$, the DWT decomposes $x(t)$ based on dyadic scales and positions defined as follows:

$$\text{DWT}(x(t); a, n) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{2^a}} \psi \left( \frac{t - 2^a n}{2^a} \right) dt \tag{2.9}$$

where $2^a n$ and $2^a$ are the time localization and scale respectively, while $\psi(t)$ denotes the mother wavelet function.

The DWT can be interpreted as a filtering process using a *dyadically* shifted and scaled mother wavelet. Given a sampled signal $x[n] = x(nT_s)$ where $T_s$ denotes the time between samples, the DWT can then be computed recursively for each level of decomposition by convolving $x[n]$ with a quadrature mirror filter with high-pass impulse response $g[n]$ and low-pass impulse response $h[n]$ and downsampling the resulting signals by a factor of 2. The resulting signals $x_A = (x * h)[n]$ and $x_D[n] = (x * g)[n]$ are referred to as the approximation and detail coefficients, respectively. Programmatically, the DWT can be defined as a recursive operation as follows:

$$x_{A_{a+1}}[n] \leftarrow (x_{A_a} * g) \downarrow_2 \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty} x[k] g[2n - k], \text{and} \tag{2.10}$$

$$x_{D_{a+1}}[n] \leftarrow (x_{A_a} * h) \downarrow_2 \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty} x[k] h[2n - k]. \tag{2.11}$$

where $a$ denotes the level of decomposition, while $\downarrow_2$ denotes the downsampling operator by a factor of 2.

**a. Wavelet Energy**

The wavelet energy $E(a)$ measures the energy of the wavelet coefficient localized at the $a^{th}$ level of decomposition computed as follows,

$$E(a) = \|C_a\|^2 = \sum_n C_a^2[n] \tag{2.12}$$

where $C_a$ denotes the wavelet coefficients at the $a^{th}$ decomposition level. $C_a$ can be either $x_{A_a}$ or $x_{D_a}$.

Normalizing the wavelet energy against the total wavelet energy, a probability mass function is obtained as follows,

$$p(a) = \frac{E(a)}{\sum_{k=1}^{K} E(k)}, \tag{2.13}$$

where $K$ denotes the number of discrete wavelet decompositions, $p(a) \in \{0, 1\}$ and $\sum_a p(a) = 1$. This normalized measure is also known as the *relative wavelet energy* (Rosso *et al.*, 2001b).

**b. Wavelet Entropy**

The wavelet entropy $H(a)$ measures the degree of 'unpredictability' in the energy distribution. Given the probability mass function $p(a)$ (refer to Eq. (2.13)), the wavelet entropy is calculated as follows (Shannon, 1948),

$$H(a) = -p(a) \log p(a), \tag{2.14}$$

where $K$ denotes the number of DWT decompositions.

### 2.7.3 Window Segmentation in EEG Emotion Classification

As experts already know that emotions are elicited directly from the brain, they have tried many different approach to clasify the emotion in order to obtain higher classification results (Panksepp, 2010).

One major challenge with EEG that has not been taken into consideration by most of available methods is the time-varying and non-stationary characteristics of the EEG signals. One possible approach to deal with this difficulty is to split these signals into smaller window frames so that pattern repetitions can be extracted more easily (Picard *et al.*, Oct. 2001).

During a short time window, a pseudo-stationary signal has the desirable statistical property of having a constant mean and variance. Predictions done using pseudo-stationary signals are likely to yield relatively higher predictive power (Kaplan *et al.*, 2005).

This thesis gives more attention to the investigation of appropriate window segmentation to be used in EEG emotion classification. In this thesis, an Optimal Window Selection (OWS) method is proposed to maximize the information gain by "zooming in" the recurring pattern on emotion elicitation. The window size needs to be *just right*: A window too short will lead to incompleteness, whilst a window too long will lead to over-inclusion of

non-stationary components. Fig. 2.9 illustrates the idea of window segmentation process.

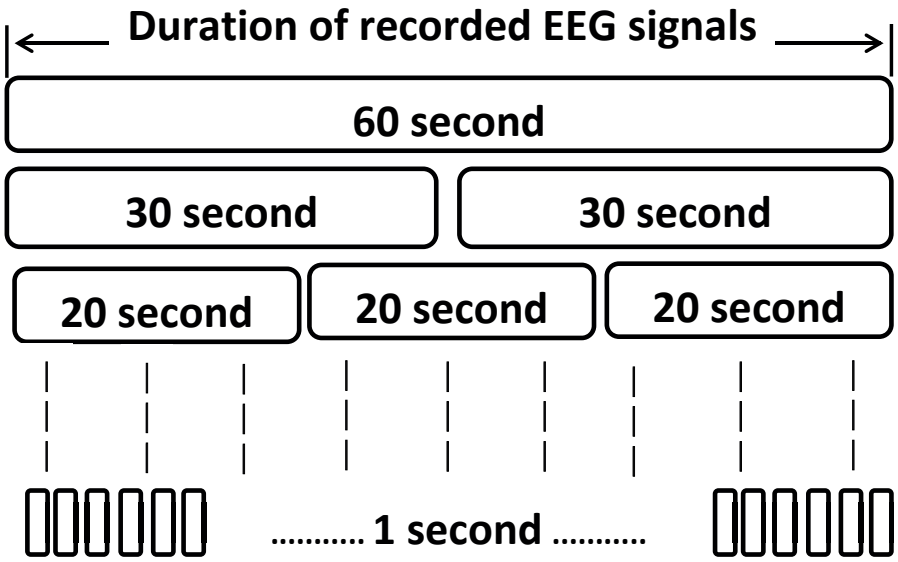



Figure 2.9: Segmentation of EEG signal with different window size. The EEG signal is segemented into different window size to obtain most effective length of EEG signal to be used in EEG emotion classification (Candra *et al.*, 2015a).

### 2.7.4 Database for EEG Emotion Classification

Similar to facial expression recognition, the demand for a reliable database to conduct the experiment in EEG emotion classification becomes a significant concern to obtain accountable results. Many researchers in EEG emotion recognition perform their studies based on a restricted number of participants which may lead to a result that is only applicable to a specific group of people with limited environment or situation.

Although a very few researches that contribute advantageous EEG emotion databases still can be found, they have become references for other researchers to conduct more profound methods in EEG emotion recognition. The databases referred to are: MAHNOB–HCI Database (Soleymani *et al.*, 2012a) and Database for Emotion Analysis Using Physiological Signals (DEAP) (Koelstra *et al.*, 2012). The description of both databases is summarized in Table 2.3.

Table 2.3: Summary of Databases for EEG Emotion Recognition

| Database | Content Summary | Description |
|---|---|---|
| DEAP (Koelstra *et al.*, 2012). | Participant: 32; video: 40; Rating scales and value: conitnuous 1–9 (Arousal, Valence, Dominance, Liking), discrete 1–5 (Familiarity); Recorded signals: 32–channel 512Hz EEG, with peripheral physiological signals, Face video for 22 participants only. | The participants were asked to watch 40 one-minute extracts of music videos while having their EEG and other physiological signals recorded: EEG (10-20 system), EOG, EMG (Zygomaticus Major and Trapezius muscles), GSR (left middle and ring fingers), respiration belt, plethysmograph (left thumb), and temperature (left pinky). Frontal face video of 22 particpants were recorded during the experiment. The participants rated each video with continuous value from 1 and 9 for valence, arousal, dominance, liking, and discrete value between 1 and 5 for familiarity. Assessment was conducted with SAM (Bradley & Lang, 1994). |
| MAHNOB–HCI (Soleymani *et al.*, 2012b) | Partiicpant: 27 (11 male, 16 female); video: 20; Rating scale and value: discrete 1–9 (arousal, valence, dominance, predictability); Recorded signals: 32-channel EEG 256Hz, face and body video using 6 cameras (60f/s), eye gaze (60Hz), audio (44.1kHz). | 30 particicpants were watching fragments of movies and pictures. The participant was asked to rate their emotive state on a scale 1 to 9 for valence and arousal. Recording was condcuted with 6 video cameras, a head-worn microphone, an eye gaze tracker, physiological sensors of ECG, EEG (32 channels), respiration amplitude, and skin temperature. Next, images/video fragments were shown with a tag containing right or wrong info at the bottom of the screen. The participant pressed a green button if they agreed with the tag, or red button if not agreed. Eye tracking and gaze locations were monitored at that time. Audio, video, gaze data and physiological data were recorded simultaneously with accurate synchronisation between sensors. |

Table 2.3 shows that DEAP dataset has more advantages compared to MAHNOB–HCI as it has more physiological signals recorded and more emotion states rated. Therefore, this thesis utilized the DEAP database to build an experimental dataset for EEG emotion recognition.

## 2.8 Sumarry of The Latest Reseaches Methodology in EEG Emotion Recognition

The latest researches in EEG emotion recognitions are listed in Table 2.4, comparing between the feature used, subband, number of channel, type of emotions, and signal window size.

Table 2.4: Comparison Between EEG Emotion Recognition Methods

| Feature | # Subband | # Channel | Emotions | Window | Reference |
|---|---|---|---|---|---|
| FD or FFT PSD | 5 | 12 | Familiarity | 4 s | (Thammasan *et al.*, 2016) |
| MSCE | 3 | 9 | Anger, Fear | 60 s | (Papadaniil *et al.*, 2015) |
| Statistical, FFT Band Power, Wavelet PSD | 4 | 32/10 | Arousal, Valence, Dominance, Liking | 60 s | (Wichakam & Vateekul, 2014) |
| GMM segment-level | 5 | 14 | Arousal, Valence, Dominance, Liking | 1 s overlapped (0.1 s step) | (Zhuang *et al.*, 2014) |
| Statitstical / FD | 5 | 32 / 14 | Arousal, Dominance, Valence | 49 s | (Liu & Sourina, 2013) |
| FFT Power Spectral | 5 | 32 | Arousal, Valence, Dominance, Liking | 1 / 2 / 4 / 8 s shifted 1 s | (Rozgi *et al.*, 2013) |

FD: Fractal Dimension; PSD: Power Spectral Density; FFT: Fast Fourier Transform
GMM: Gaussian Mixture Model
MSCE: Magnitude Square Coherence Estimation

Table 2.4 shows that Power Spectral and statistical features are often selected as the features. The number of sub bands accommodated are mostly all the 5 bands, while the number of channels depends on the EEG equipment that has been used. Some EEG equipment may only have a limited number of channels, for example:some EEG systems only equipped with 14 channels. The emotion detected is mostly based on Russel's circumplex model (Russell, 1980) that accommodates arousal, valence, and dominance emotion. The

detection of discrete emotion is usually mapped from the arousal–valence dimensional plane. Only a few researcher considered applying a different selection of window size for the signal segmentation, (Rozgi *et al.*, 2013) for example.

It should also be noticed that (Thammasan *et al.*, 2016) and (Guo *et al.*, 2017) has applied and cited the proposed window segmentation method described in (Candra *et al.*, 2015a). Thammasan et al. used Fast Fourier Transform Power Spectral Density (FFT PSD) for the feature, where the emotion to be detected is related to the familiarity to the musics, while Guo et al. used the combination of time domain and DWT feature classified with combined SVM and HMM. Their results supported the hypothesis that the implementation of window segmentation improved the classification results of EEG emotion consistently in general.

For the EEG emotion recognition, this thesis focuses on the exploration of power spectral and entropy of wavelet features while dealing with the subband and channel reduction. And also, investigation of window segment was conducted to obtain window size that provides higher performance of classification results. In addition, discrete and dimensional plane emotional states were also addressed in the recognition.

## 2.9 Support Vector Machine (SVM) Classifier for Emotion Recognition System

SVM was first introduced by Vladimir Vapnik in 1979 and first published in 1995. It is used for binary classification by discovering the hyperplane that separates the n-dimensional data into two classes. Often, the data cannot be linearly separated; hence the kernel induced feature space is introduced in SVM so that the data is moved into higher dimensional space to make it possible for separation. Moving the data into higher dimensional space raises a computational problem and overfitting. However, this is not really the case since the higher dimensional space should not be addressed directly (it is only dot product in that space). The target of SVM is developing a model from the training data to predict the target values of test data with only the test data attributes. (Joachims, 2002).

For a training set with instance-label pairs $(x_i, y_i)$, $i$ = 1, . . ., $l$ where $x_i \in R_n$ and y $\in \{1,$ -1$\}$l the SVM require the solution of the following optimization problem:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \tag{2.15}$$

is subject to

$$y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \tag{2.16}$$
$$\xi_i \geq 0, i = 1, \ldots, l.$$

Training vectors $x_i$ are mapped into a higher (maybe infinite) dimensional space by the function $\phi$. SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. C $> 0$ is the penalty parameter of the error term. Furthermore, $K(x_i, x_j) \equiv \phi(x_i)^T \ \phi(x_j)$ is called the kernel function (Hsu *et al.*, 2003). Hsu et. al. suggest to use the following four basic kernels:

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

where, $\gamma$, $r$, and $d$ are the kernel parameters.

### 2.9.1 Optimizing the SVM Classifier

SVM has several advantages, foe example, the margin maximization and the regularization term, good generalization properties (Bennett & Campbell, 2000; Jain *et al.*, 2000), insensitive to overtraining (Jain *et al.*, 2000) and to the curse-of-dimensionality (Burges, 1998; Bennett & Campbell, 2000). Having all those advantages, SVM has been shown to be a powerful classifier for the task due to its ability to operate on non-linear and high dimensional feature spaces. However, these advantages are gained at the expense of a low speed of execution (Lotte *et al.*, 2007).

Using SVM, it is possible to create nonlinear decision boundaries, with only a low increase of the classifier's complexity, by using the proper kernels. It consists in implicitly mapping the data to another space, generally of much higher dimensionality. The most common kernel used for the above mentioned purpose is RBF kernel. When using SVM with RBF kernel, there are a few hyperparameters that need to be defined manually, namely, the regularization parameter C and the RBF width $\sigma$ if using kernel 2 (Lotte *et al.*, 2007).

In this thesis, the SVM classifier is optimized using an RBF kernel by first properly estimating the RBF kernel parameters in order to obtain proper learning (Schlesinger & Hlavac, 2002) using a particle swarm ensemble clustering algorithm called the Ensemble Rapid Centroid Estimation (ERCE) algorithm (Yuwono *et al.*, 2014). The algorithm can estimate the number of clusters directly from the data using swarm intelligence and ensemble aggregation. Having the optimized RBF kernel parameters, the SVM is then trained using the Sequential Minimal Optimization (SMO) algorithm. The detaiedl steps for the optimization of the SVM classifier is presented in the following. First, calculate the RBF kernel using:

$$\mathrm{RBF}_{\mathrm{JS}}(\mathbf{x}, \boldsymbol{\mu}_k, \sigma_k) = \exp\left(-\frac{\mathrm{JS}(\mathbf{x}\|\boldsymbol{\mu}_k)}{2\sigma_k^2}\right) \tag{2.17}$$

where $k$ denotes the support vector index, $\sigma_k$ denotes the support vector radius, $\mathrm{JS}(\mathbf{x}\|\boldsymbol{\mu}_k)$ denotes the Jensen-Shannon (JS) divergence between a random vector $\mathbf{x}$ and the support vector centroid $\boldsymbol{\mu}_k$.

The JS divergence is a symmetrized and smoothed version of the Kullback-Leibler (KL) divergence (Fuglede & Topsoe, 2004). Given two discrete probability distributions $P \sim p(x)$ and $Q \sim q(x)$, the JS divergence is calculated as follows,

$$\mathrm{JS}(P\|Q) = \frac{1}{2}\mathrm{KL}(P\|M) + \frac{1}{2}\mathrm{KL}(Q\|M), \text{ where} \tag{2.18}$$

$$\mathrm{KL}(R\|S) = \sum_x R(x)\, \log \frac{R(x)}{S(x)}. \tag{2.19}$$

Here $M = (P + Q)/2$ denotes the central probability mass function and $\mathrm{KL}$ denotes KL divergence.

The next important step is the estimation of the RBF kernel parameters using a particle

swarm ensemble clustering algorithm called the ERCE algorithm (Yuwono *et al.*, 2014). It has been argued that the algorithm can estimate the number of clusters directly from the data using swarm intelligence and ensemble aggregation (Yuwono *et al.*, 2014, Jun. 2014).

Using ERCE, the support vector centroids $\boldsymbol{\mu}_k \in \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots\}$ and the kernel radius $\sigma_k \in \{\sigma_1, \sigma_2, \ldots\}$ can be inferred from the training data as follows:

1. Execute ERCE (Yuwono *et al.*, 2014) to cluster the training set $\mathbb{X}_{\text{train}} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ to an arbitrary number of cluster based on JS distance (i.e. the square root of JS divergence).

2. Aggregate the ensemble clustering results using average linkage to get the final clustered sets $\{\mathbb{C}_1, \mathbb{C}_2, \ldots, \mathbb{C}_K\}$, where $K$ is determined automatically by ERCE at ensemble aggregation. The corresponding centroid vector $\boldsymbol{\mu}_k \in \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K\}$ is computed as conditional expectation as follows,

$$\boldsymbol{\mu}_k = \mathrm{E}[\mathbf{x}|\mathbb{C}_k] = \frac{1}{|\mathbb{C}_k|} \sum_{\mathbf{x} \in \mathbb{C}_k} \mathbf{x}. \tag{2.20}$$

3. The RBF kernel radius for the $k^{th}$ support vector is taken as the square root of conditional JS divergence as follows,

$$\sigma_{\mathrm{k}}^2 = \mathrm{E}\left[\mathrm{JS}(\mathbf{x}\|\boldsymbol{\mu}_k)|\mathbb{C}_k\right] = \frac{1}{|\mathbb{C}_k|} \sum_{\mathbf{x} \in \mathbb{C}_k} \mathrm{JS}(\mathbf{x}\|\boldsymbol{\mu}_k). \tag{2.21}$$

4. Given the optimized RBF kernel parameters, the SVM is then trained using the SMO algorithm (Chang & Lin, 2011).

## 2.10 Summary

In this chapter the general concept of emotion measurement has been explained. The 3 perspectives of emotion response measurement have been discussed including: discrete, dimensional, and componential perspective. Combining the appraisal process as explained in Chapter 1 and the 3 perspectives bring about 3 possible approaches to measure the emotion. Each of the approaches is related to specific attributes. For example:

motor expression measurement is related to motoric activity of body part, and physiological arousal measurement is related to physiological signal, while subjective feeling measurement is more related to the affect.

Furthermore, the development of facial expression recognition involves much background knowledge such as: face detection and face recognition which has 3 different mechanisms: feature-based, holistic matching and hybrid methods. The facial emotion recognition gives more emphasis to facial features extraction: the geometric based method which has some disadvantages compared to appearance based.

A detailed description has been provided of the Histograms of Oriented Gradient (HOG) descriptor that has been successfully implemented for object recognition including face recognition which has condiserable potential to be used as a feature for facial expression recognition.

The availability of database for facial expression recognition should also be taken into consideration to develop the recognition system as it is required for testing the new system, comparing it with the other state of the art systems and benchmarking the performance. Such a common database will satisfy the various requirements of the problem domain and can be used as a standard for future research. The Extended Cohn–Kanade (CK+) database (Kanade *et al.*, 2000) is one good example.

In Electroencephalography (EEG) emotion recognition, it relies on the methodology, application, and equipment to measure the EEG signals. Signal processing and conditioning is also required before the information can be extracted to develop any application that uses the attributes of EEG signal as a feature. In addition, the electrodes positioning is an important factor in recording the EEG signal. The most used positioning method is the 10/20 international system. The EEG signal is usually divided into several frequency bands: delta, theta, alpha, beta, and gamma.

To build a reliable EEG emotion recognition system attention must given to the features, channel selection, and number of electrodes. Another important parameter is the window segmentation to overcome the challenge of time-varying and non-stationary characteristics of the EEG signals. Proper selection of window size will improve the performance of

the system.

A sophisticated feature to be used in EEG emotion recognition system is the Discrete Wavelet Transform (DWT) by taking its relative energy and or entropy. The advantages of the DWT are related to its capability of time-frequency localization, multiscale zooming, and multirate filtering for detecting and characterizing transients. These advantages allow DWT to potentially extract the appropriate information from non-stationary signals.

The same as in facial expression recognition, the need for a reliable database to conduct the EEG emotion recognition becomes an important issue. There are few EEG emotion databases that can meet with the expectation, and the Database for Emotion Analysis Using Physiological Signals (DEAP) (Koelstra *et al.*, 2012) becomes the most preferred one.

To complete the chapter, the Support Vector Machine (SVM) classifier has been explained together with the optimization method using a particle swarm ensemble clustering algorithm called the Ensemble Rapid Centroid Estimation (ERCE) algorithm.

# Chapter 3

# Facial Expression Recognition using E–HOG and RED E–HOG Features

In this chapter, the implementation of Edge–Histogram of Oriented Gradients (E–HOG) for facial expression recognition is discussed (Candra *et al.*, 2016). In addition, the high dimensionality issue of original E–HOG (ORI E–HOG) is resolved using a combination of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to obtain very low dimension of ORI E-HOG. Using the new feature Reduced E–HOG (RED E–HOG) (Candra *et al.*, (Submitted)b), improvement was achieved in various aspects, such as, reduced dimensions, increased accuracy, and shortened training and testing time.

The following are the primary contributions of this chapter:

the development of E-HOG feature for facial expression recognition; The implementation of dimensionality reduction using a combination of PCA and LDA to obtain a very low dimension RED E-HOG feature; The demonstration of RED E–HOG in facial expression recognition which provide improvement of accuracy and reduction of classifacation time compared to E–HOG.

This chapter is organized as follows: a quick overview on the method is explained in Section 3.1. Section 3.2 describes the experimental setup involving dataset preparation and experimental outline. In Section 3.3 the proposed E–HOG algorithm is discussed in depth, including Viola-Jones facial landmarks detection, and E–HOG computation.

Dimensionality reduction using a combination of PCA and LDA to generate RED E-HOG feature is also explained in this section, followed by the discussion on the results of E–HOG and RED E–HOG implementation in Section 3.4. The summary in Section 3.5 concludes this chapter.

## 3.1 Method Overview

For many decades scientists have shown significant interest to accurately identify human facial expressions in order to help people to better understand the human emotions. The implementations have been extended to various areas, including: human-computer interaction, advertisement, computer games, and biomedical areas, (Mohamed *et al.*, 2013; Taspinar *et al.*, 2012; Tanaka *et al.*, 2010; Ulukaya & Erdem, 2014).

Recently, the facial expression was also implemented as an assisting tool in psychotherapy. (Candra *et al.*, 2016). Using a sophisticated feature called Edge–Histogram of Oriented Gradients (E–HOG), the method was aimed to identifying patients' facial emotion expressions to provide an effective counselling session and offer the optimum treatment which is very important in psychotherapy (Foley & Gentile, 2010). Usually a therapist uses various guidance to interpret the facial expression of patients with psychological problems Machado *et al.* (1999) to provide therapeutic alliance and treatment process (Sutter, 2010). Therefore the facial expression recognition system will be very helpful for the therapists while they focus on the appropriate treatment methods.

To distinguish different types of human emotions, Ekman divides them into six basic facial expressions, namely: anger, disgust, fear, happy, sad, surprise. Contempt is then added later as another natural emotion (Ekman, 2003). These facial expressions have been implemented to Machine aided emotion recognition systems such as a lie detector for security application (Owayjan *et al.*, 2012) and autistic therapy system (Smitha & Vinod, 2015) to name a couple.

Viola Jones is renowned as a lightweight algorithm to detect faces and various facial landmarks including eyes and mouth with low algorithmic complexity which makes the

algorithm suitable for time-critical applications (Viola & Jones, 2001).

Histogram of Oriented Gradient (HOG) as a robust morphological image descriptor was first introduced by Dalal and Triggs (Dalal & Triggs, 2005). Although it has been used for various facial recognition implementations since inception (e.g.(Baltrusaitis *et al.*, 2015; Salhi *et al.*, 2013; Do & Kijak, 2012; Monzo *et al.*, 2011)) its usage in facial-emotion expression classification may not have not been adequately explored to date.

The combination of Viola-Jones algorithm and improved HOG descriptor, which is HOG feature extracted from a binary edge image is proposed to develop the E–HOG feature for facial expression recognition (Candra *et al.*, 2016). The proposed E–HOG feature can be highly efficient for facial-emotion expression recognition with high classification results and a significant improvement in processing time (by comparing the classification process applying HOG and E-HOG). The E–HOG was successfully trained to distinguish between Ekman 7 basic emotions.

In addition, the original E-HOG (ORI E-HOG) has an issue of high dimensionality which demands very long training and testing time. Therefore, the improvement of ORI E-HOG in its dimensions is necessary to make it more feasible to be implemented in a facial emotion recognition system. To obtain very low dimension of ORI E-HOG features, dimensionality reduction technique using a combination of PCA and LDA is implemented. The new feature is named Reduced dimension E-HOG (RED E-HOG). Using RED-E-HOG trained with multi-class Support Vector Machine (SVM), we achieved improvement in various aspects, such as, reduced dimensions, increased accuracy, and shortened training and testing time.

## 3.2 Experimental Setup

### 3.2.1 Dataset Prepaparation

This study used the extended Cohn-Kanade database (Lucey *et al.*, 2010). The database has 2,105 frontal faces images of 210 participants from various backgrounds, i.e., Euro-American (81%), Afro-American 13%), and other groups (6%); it consists of 31% male and 69% female with ages between 18 and 50 years. They were asked to perform 23 facial expressions in either single action or combinations of action units, starting and finishing with a neutral expression. The database was collected in an observation room that had relatively uniform lighting conditions and context. The room was equipped with a chair and 2 video cameras which were located in front of the participants and at 30 degrees to their right side. The images were recorded in 425 lines per frame and digitized by only using the odd fields (Kanade *et al.*, 2000).

The images taken from the frontal-view camera were digitized into 640 $\times$ 490 pixel with 8 bit gray-scale or 640 $\times$ 480 pixel of 24-bit colour values. From these frontal-view images, 182 subjects with 1,917 image sequences have been coded with the Facial Action Coding System (FACS).

We used the FACS coded images to construct the dataset for our experiment to identify 7 facial expression emotions including: 1. Anger (An); 2. Contempt (Co); 3. Disgust (Di); 4. Fear (Fe); 5. Happy (Ha); 6. Sad (Sa), and; 7. Surprise (Su). Each of these emotions was represented with 150 FACS coded images, giving a total number of images in the constructed dataset as much as 150 $\times$ 7 = 1,050. The dataset was then divided into 30% and 70% proportions for the training and testing purpose.

### 3.2.2 Experimental Outline

The facial expression recognition method is applied in the following steps: 1. Detection of facial landmarks using Viola Jones algorithm; 2. RED E-HOG feature extraction which is comprised of 2 consecutive steps: 2a. E–HOG Feature extraction; 2b. Dimensional reduction using a combination of PCA and LDA. This process produces 2 features

namely: E–HOG and RED E–HOG, respectively. Both features are used in the facial expression classification, which becomes the next step; 3. Facial expression recognition with Multi-class SVM. The complete steps are illustrated in the block diagram as in Fig. 3.1.

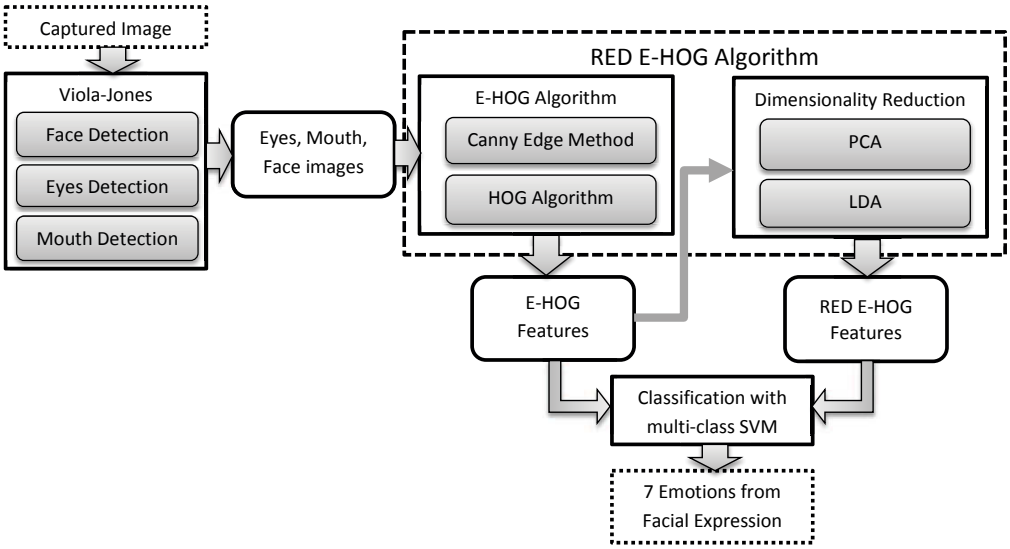



Figure 3.1: The block diagram of the proposed emotion recognition algorithm.

## 3.3 Experimental Details

In this section, we will explain the detailed process of the algorithm step by step, as referred to in Fig. 3.1. We started by performing Viola-Jones robust object detector to capture facial landmarks from the input image with the results of 3 cropped facial landmark images, i.e., face, pair of eyes, and mouth. The cropped facial landmarks were then fed to the RED E-HOG algorithm to obtain the features for facial emotions classification. The RED E-HOG algorithm itself consists of 2 steps as follows:

- First, E-HOG algorithm: Canny edge method (Canny, 1986) was applied to the cropped facial landmarks to produce binary edge images. The process was continued with the HOG algorithm to extract HOG features from the binary edge images;

- Second, Dimensionality reduction: To reduce the complexity of the E-HOG features, dimensionality reduction using PCA was implemented to obtain the projection of features vectors that best representing the distribution of features within the features space. Further dimensionality reduction was conducted using LDA to pick the maximum possible discrimination between different classes of E-HOG features by taking largest differences between them.

The output, E–HOG and RED E–HOG features were collected for facial emotion classification using multi-class SVM to distinguish 7 facial expressions. To better understand how the experiment was conducted, this section has been divided into the following 4 subsections:

1. Viola-Jones Algorithm for Facial Landmarks Detection.
2. E–HOG Feature Extraction Strategy.
3. Improving E–HOG Feature with Dimensional Reduction Technique: Constructing RED E–HOG.
4. Classifying Facial Expression using Optimized Multi-class SVM.

A more detailed algorithm on each step is discussed in the following subsections.

### 3.3.1 Viola-Jones Algorithm for Facial Landmarks Detection

Given an intensity image $\mathbf{I}$ with $x$ and $y$ denoting the row and column index, respectively and using equation 2.1 to 2.3 from Chapter 2, the Viola–Jones algorithm can be computed.

To continue the Viola–Jones algortihm, facial landmarks including face, mouth and eyes are isolated from an image with the following steps:

1. Given an image capture, localize a face entity of arbitrary size $\mathbf{F}_{m\times n}$.
2. If $\mathbf{F}_{m\times n}$ exists, standardize the image size using any interpolation method,

$$\boldsymbol{\mathcal{F}}_{r\times c} \leftarrow \texttt{imresize}\left(\mathbf{F}_{m\times n}, [r \times c]\right), \tag{3.1}$$

where $r \times c$ is a predefined standard resolution. In this experiment we find 212 $\times$ 212 pixels to be adequate for the emotion recognition purpose.

3. $\mathcal{F}$ is then divided into two parts: The upper half $\mathcal{F}^{\text{upper}}_{r/2\times c}$ — where the eyes are detected and the lower half $\mathcal{F}^{\text{lower}}_{r/2\times c}$ — where the mouth is detected. The detection process is shown in Fig. 3.2.


Figure 3.2: The facial landmarks extraction using Viola-Jones' algorithm.

### 3.3.2 E–HOG Feature Extraction Strategy

A descriptor derived from Dalal's HOG Dalal & Triggs (2005) is proposed in this thesis which is termed Edge-HOG or E-HOG. The proposed feature takes the edge image extracted using the Canny method as an input to the HOG feature extraction process. E-HOG is calculated from an input image as follows:

1. Perform edge detection using Canny method:
   (a) Normalize the intensity image.
   (b) Smoothen the image with a Gaussian mask.
   (c) Calculate the horizontal and vertical gradients of the pixels in the image, i.e.

$g_{vert}(x,y) = \partial I(x,y)/\partial x$ and $g_{horiz}(x,y) = \partial I(x,y)/\partial y$,

$$\mathbf{G}_{vert} = \mathbf{I} * [-1, 0, 1], \tag{3.2}$$

$$\mathbf{G}_{horiz} = \mathbf{I} * [-1, 0, 1]^T. \tag{3.3}$$

(d) Calculate the gradient magnitude $|g(x,y)|$ and orientation $\theta(x,y)$ as follows,

$$|g(x,y)| = \sqrt{g^2_{vert}(x,y) + g^2_{horiz}(x,y)}, \tag{3.4}$$

$$\theta(x,y) = \tan^{-1}\left(\frac{g_{vert}(x,y)}{g_{horiz}(x,y)}\right), \tag{3.5}$$

(e) Apply non-maximum suppression,

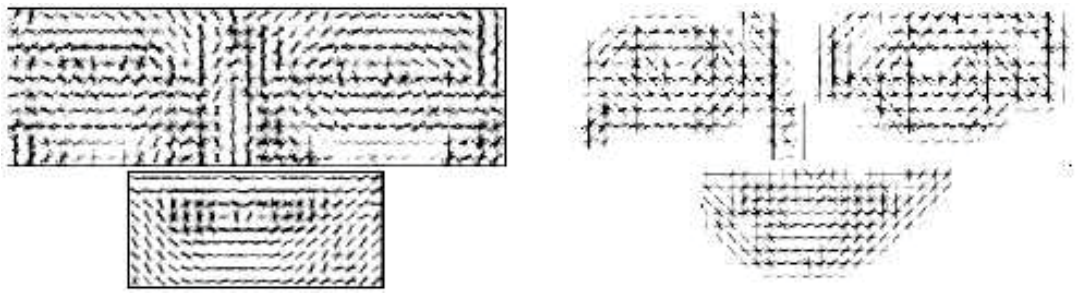
(f) Identify strong and weak edges,

(g) Track the strong edges,

(h) Return the resulting edge image **E** as an output.

2. Compute the HOG features from the edge image **E** using equation 2.4 to 2.8 from Chapter 2.

E-HOG features are relatively faster to compute than HOG due to the sparse **E** (refer to Step 2 equation 2.6 and 2.7) which potentially contributes to a computational complexity reduction. The descriptors returned by HOG and E-HOG are seen in Fig. 3.3.



(a) HOG (b) E-HOG

Figure 3.3: HOG vs E-HOG features of the eye and mouth. Sparse **E** matrix in the E-HOG computation contributes to a substantial amount of cells having zero values, yielding simpler features and a leaner extraction process.

### 3.3.3 Improving E–HOG Feature with Dimensional Reduction Technique: Constructing RED E–HOG

To reduce the complexity of E-HOG features dimension, we need to implement a dimensionality reduction algorithm. We chose a combination of PCA and LDA in our strategy due to the advantage of each method that if combined will enhance the dimensional reduction. The advantage and formulation of each method is explained in the following paragraph together with the steps for the combination.

#### 3.3.3.1. Dimension Reduction with Combination of PCA and LDA Algorithm

**A. Principal Component Analysis (PCA)**

PCA is well known as a standard technique in signal and image processing that can be used to linearly reduce the dimensionality of complex data (features) that contains redundant information into lower dimensional feature vectors. The main idea of PCA is to identify the vectors that best represent the distribution of features within the entire feature space by projecting them into the direction of highest variability (Martis *et al.*, 2013).

An E-HOG feature of $M \times N$ can be considered as a one-dimensional vector with dimension of ($M \times N$). For example, E-HOG feature of mouth with size of 80 $\times$ 36 pixels can be taken as a vector with dimension = 80 $\times$ 36 = 2,880 or equivalently points in a 2,880 dimensional space which is an ensemble of features vectors maps to a collection of points in this highly-dimensional space.

Supposing that L numbers of K dimensions E-HOG feature are arranged in matrices $K \times L$, the covariance matrices V of the E-HOG feature can be calculated with

$$V = (Y - \bar{y})(Y - \bar{y})^T, \tag{3.6}$$

where Y is the features matrices of E-HOG coefficients and $\bar{y}$ represents mean vector of Y.

The eigenvalues are the first principal component that provide the basis vector for the direction of highest variability, while eigenvectors are the basis vector for the next direction

orthogonal to the first principal component. This diagonal matrix of eigenvalues D and matrices eigenvectors E can be computed with

$$E^{-1}VE = D. \tag{3.7}$$

To obtain the new reduced dimension features matrices R, the eigenvectors E are sorted in descending order of the eigenvalues D, and then the inner product between the original features matrices and the sorted eigenvectors E are taken. Those give:

$$R = [E^T(Y - \bar{y})^T]^T. \tag{3.8}$$

**B. Linear Discriminant Analysis (LDA)**

LDA proposed the maximum possible discrimination between different classes of E-HOG features by searching for basis vectors within the features that provide the largest differences between various classes by taking the measurement of smallest within-class scatter matrices $S_w$ and the largest between-class scatter matrices $S_B$ which can be computed as follows (Zhou *et al.*, 2013):

$$S_W = \sum_{k=1}^{K} \sum_{l=1}^{L_k} (y_l^k - x_k)(y_l^k - x_k)^T, \tag{3.9}$$

$$S_B = \sum_{k=1}^{K} (x_k - x)(x_k - x)^T, \tag{3.10}$$

where $y_l^k$ represents the $l_{th}$ sample of class k, $x_k$ is the mean of class k, K is the number of classes, and $L_k$ the number of samples in class k; and *x* is the mean of all classes.

Then use the Fisher criterion to measure the discriminatory power with:

$$D(W) = \frac{\|W^T S_B W\|}{\|W^T S_W W\|}. \tag{3.11}$$

Matrix W as the optimal projection matrix can be obtained by solving the generalized eigenvalue problem:

$$S_B W = \lambda S_W W \tag{3.12}$$

The LDA coefficients that represent the reduced dimensional features of E-HOG can be obtained from the projection matrix with

$$s = W^T r, \tag{3.13}$$

where *s* is the LDA coefficients and *r* is the original E-HOG features coefficient.

**C. Combination of PCA and LDA**

PCA reduces the dimension of E-HOG features by finding basis vectors which correspond to the maximum-variance directions in the original space, while LDA aims at increasing separability of the samples in subspace by searching vectors in the underlying space that best discriminate among classes. Therefore, combination of both will provide optimized dimensionality reduction of the features (Yoo *et al.*, 2015).

The proper order of the implementation is first applying the PCA to the ORI E-HOG features to obtain maximum variance directions in space of the features. Then, the output of PCA is used as an input to LDA algorithm to gain highest discrimination among classes.

### 3.3.4 Classifying Facial Expression using Optimized Multi-class SVM

The facial expression classification is done using an optimized multi-class SVM with a Radial Basis Function (RBF) kernel as explained in Chapter 2. The kernel radius $R_{\text{SVM}}$ is inferred from the training data using Ensemble Rapid Centroid Estimation (ERCE) (Yuwono *et al.*, 2014). The SVM radius estimation step is as follows,

1. For each vectorized E-HOG descriptor $\mathbf{v}_j$ in the training set, compute the probability mass function with equally spaced bins centered at $C = \{0.000, 0.025, 0.050, \ldots, 0.500\}$ to normalize against general shape of the face to improve generalization capability by having sufficiently adequate support vector radius to avoid over or under-cluster.

$$\mathtt{p}_j = \mathtt{pmf}(\mathbf{v}_j, C), \tag{3.14}$$

   where $j$ denotes the observation index.

2. Execute ERCE (Yuwono *et al.*, 2014) to cluster $\mathbb{P} = \{\mathsf{p}_1, \mathsf{p}_2, \ldots\}$ to an arbitrary number of cluster based on Jensen-Shannon distance.

3. Aggregate the ensemble clustering results using average linkage to get the final clustered sets $\{\mathbb{C}_1 \cup \ldots \cup \mathbb{C}_K\}$, where $K$ is determined automatically by ERCE at ensemble aggregation. The corresponding centroid vectors $\{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K\}$ are computed using Eq. (2.20).

4. The SVM radius $R_{\text{SVM}}$ is taken as the average cluster radius in terms of euclidean distance and computed using Eq. (2.18), (2.19), and (2.21).

The SVM is trained using Sequential Minimal Optimization (SMO) algorithm (Chang & Lin, 2011).

## 3.4 Discussion

In order to understand the experiment results, we must first look at the steps involved prior to discussing the experiment outcome. In this section, we will first evaluate the summary of classification results with various facial landmarks of ORI E–HOG compared to original HOG (ORI HOG) together with the confusion matrix of the best trained SVM classifier using Face E–HOG. We will then discuss the impact to the classification results after the implementation of dimensionality reduction. The next step is the analysis of training and testing time by comparing the classification results before and after dimensionality reduction (ORI E–HOG Vs. RED E–HOG), followed by comparison of classification result for each facial expression using RED E–HOG Face in a confusion matrix, and finally the comparison between RED E–HOG Face and other methods. The detail for each section is discussed in the following.

### 3.4.1 Classification Results with E–HOG Feature and Comparison to Original HOG

We begin the analysis with the comparison of the classification results between E–HOG and ORI HOG as summarized in Table 3.1. A paired t-test was conducted with the null

hypothesis ($H_0$) being "*accuracy using E-HOG is not higher than ORI HOG*." $\alpha$ = 0.001.

Table 3.1: Comparison of average classification accuracy between ORI HOG and E–HOG (30 repetitions)

| Feature Source | Dimension (pixels) | Ave. Accuracy | | p-value |
|---|---|---|---|---|
| | | ORI HOG | E-HOG | |
| Eyes | 128× 40 | 0.8095 | 0.7361 | 1.74E-17 |
| Mouth | 80×36 | 0.8972 | 0.8066 | 3.32E-24 |
| Eyes & Mouth | 128×40, 80×36 | 0.9456 | 0.8936 | 2.05E-17 |
| **Face*** | **180×140** | **0.9607** | **0.9490** | **0.0016*** |

*$H_0$ could not be rejected for the face feature at $\alpha = 0.001$.

To complete the analysis, a table showing the classification processing time for each classification process utilizing ORI HOG and-E–HOG is also provided in Table 3.2. A paired t-test was also computed with the null hypothesis ($H_0$) being "*the E-HOG processing time is not smaller than ORI HOG*." $\alpha$ = 0.001.

Table 3.2: Comparison of average classification processing time between ORI HOG and E–HOG (30 repetitions)

| Feature Source | Ave. Time (s) | | improvement % | p-value |
|---|---|---|---|---|
| | ORI HOG | E–HOG | $\frac{t_{\text{ORI HOG}} - t_{\text{E–HOG}}}{t_{\text{E–HOG}}}$ | |
| Eyes | 0.98 | 0.53 | 83.07% | 0.0015 |
| Mouth | 0.75 | 0.41 | 82.20% | 0.0103 |
| Eyes and Mouth | 1.5 | 0.76 | 97.67% | 5.20E-07 |
| **Face** | **29** | **1.5** | **1833.33%** | **2.43E-17** |

Observing both Table 3.1 and 3.2, it can be noticed that the face yields the best trade-off between time complexity and accuracy as indicated with a significant improvement in classification processing time ($(t_{\text{HOG}} - t_{\text{E-HOG}})/t_{\text{E-HOG}} = 1833.33\%$, p-value = 2.43E-17) and a slight decrease in average accuracy ($acc_{\text{HOG}} - acc_{\text{E-HOG}} = 1.17\%$, p-value = 0.0016, insufficient evidence to reject $H_0$ at $\alpha$ = 0.001). Despite the high accuracy, ORI HOG (face) is significantly slower to compute compared to E-HOG (face).

The high accuracy using the face feature indicates that the key features for describing emotions are also conceived outside of the eyes and mouth. This includes skin wrinkles in the forehead and the nose.

Next, a confusion matrix showing the best trained SVM classifier using Face E-HOG feature is shown in Table 3.6. The table shows that accuracy as high as 96.4% can be achieved using the proposed feature.

Table 3.3: Confusion Matrix of the best trained classifier using Face E-HOG features

| | An | Co | Di | Fe | Ha | Sa | Su |
|---|---|---|---|---|---|---|---|
| **An** | **97.2%** | 0.0% | 0.0% | 1.4% | 0.0% | 1.4% | 0.0% |
| **Co** | 0.0% | **90.0%** | 0.0% | 8.6% | 0.0% | 1.4% | 0.0% |
| **Di** | 4.3% | 0.0% | **95.7%** | 0.0% | 0.0% | 0.0% | 0.0% |
| **Fe** | 1.4% | 0.0% | 1.4% | **94.4%** | 1.4% | 0.0% | 1.4% |
| **Ha** | 0.0% | 0.0% | 0.0% | 0.0% | **100.0%** | 0.0% | 0.0% |
| **Sa** | 1.4% | 0.0% | 0.0% | 0.0% | 0.0% | **98.6%** | 0.0% |
| **Su** | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 1.4% | **98.6%** |
| **Average Accuracy = 96.4%** | | | | | | | |

### 3.4.2 Classification Results of RED E–HOG Feature

Dimensionality reduction was performed to all facial landmarks of ORI E-HOG by first applying PCA to obtain principal components of the features that give classification results approximately equal to the results without dimensionality reduction. The results after applying PCA are provided in Fig. 3.4.

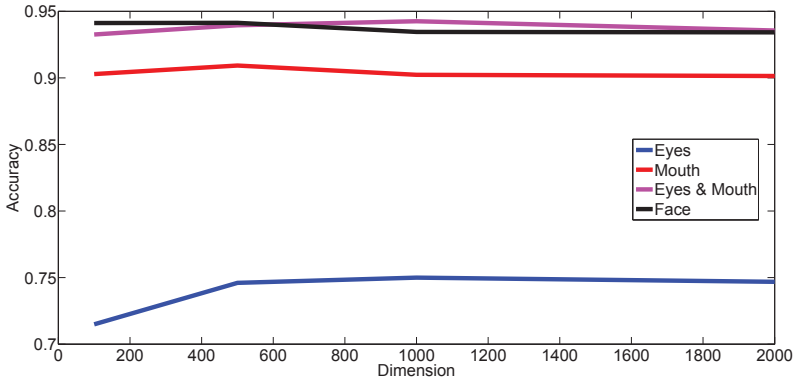



Figure 3.4: Classification results with various E–HOG features after applying PCA.

Fig. 3.4 shows that the dimension of ORI E-HOG features can be reduced from thousands to hundreds using PCA with the classification results equal to ORI E-HOG. However, reducing the dimension to lower than 100 will result in drastically reduced accuracy. Therefore, to obtain further dimensional reduction of E-HOG features, we applied LDA

by using the output from PCA algorithm that has dimension of 100 as the input to LDA algorithm. The result after LDA implementation is given in Fig. 3.5.

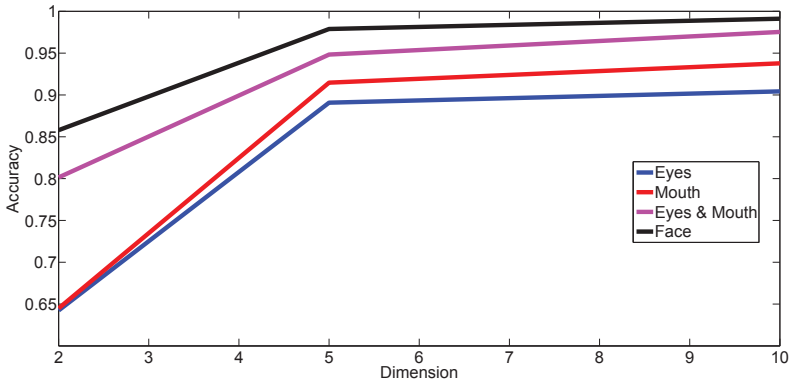



Figure 3.5: Classification result with various E-HOG features after applying a combination of PCA and LDA.

Using the combination of PCA and LDA, the dimension of E-HOG features is reduced to tens with equal classification results to ORI E-HOG. However, the accuracy dropped when the dimension was reduced to less than 5.

The observation of Fig. 3.4 and Fig. 3.5 also shows that best performance of dimensional reduction can be obtained using the combination of PCA and LDA by first implementing PCA algorithm to acquire hundreds of dimension and then continuing with the LDA algorithm to reach tens dimension.

### 3.4.3 Comparison of Classification Processing Time Between RED E-HOG vs. ORI E-HOG

Table 3.4 and Table 3.5 present the comparison of training and testing time between ORI E-HOG and 10 dimensions RED E-HOG for each classification process.

Both tables indicate that ORI E-HOG demands extensive training and testing time compared to RED E-HOG. And also, the most significant improvement of training and testing time is given by the RED E-HOG face (1.48e4% and 1.09e5%). The results suggest that dimensionality reduction applied to RED E-HOG offers time efficiency, which makes RED E-HOG more effective and feasible to be used in facial expression recognition, especially for RED E-HOG Face.

Table 3.4: Average training time of ORI E-HOG vs. 10 Dimensions RED E-HOG trained with multi-class SVM in 30 repetitions

| Feature Source | Average Time (s) | | improvement % $\frac{t_{\text{ORI}} - t_{\text{RED}}}{t_{\text{RED}}}$ |
|---|---|---|---|
| | ORI E-HOG | RED E-HOG | |
| Eyes | 4.60 | 0.90 | 411.11% |
| Mouth | 1.86 | 0.20 | 830.00% |
| Eyes & Mouth | 5.49 | 0.21 | 2,514.29% |
| **Face** | **28.22** | **0.19** | **14,752.63%** |

Table 3.5: Average testing time of ORI E-HOG vs. 10 Dimensions RED E-HOG trained with multi-class SVM in 30 repetitions

| Feature Source | Average Time (s) | | improvement % $\frac{t_{\text{ORI}} - t_{\text{RED}}}{t_{\text{RED}}}$ |
|---|---|---|---|
| | ORI E-HOG | RED E-HOG | |
| Eyes | 1.80 | 0.01 | 1.79e4% |
| Mouth | 1.00 | 0.01 | 9.90e3% |
| Eyes & Mouth | 2.93 | 0.01 | 2.92e4% |
| **Face** | **10.93** | **0.01** | **1.09e5%** |

### 3.4.4 Analysis of Classification Results for Each Emotion with Confusion Matrix of RED E-HOG Face

Using 10 dimensions RED E-HOG Face, a confusion matrix comparing the classification results for all 7 expressions is created and presented in Table 3.6. The confusion matrix denotes that using the selected features accuracy as high as 99.6% can be realized. Further analysis also reveals that anger, contempt, fear, and sad can be identified clearly with 100% accuracy. Although disgust and happy were misidentified as anger, while surprise was misinterpreted as fear, these occur in only 1.0% in each case.

Table 3.6: Confusion Matrix of the best trained SVM classifier using 10 dimensions RED E-HOG Face

| | An | Co | Di | Fe | Ha | Sa | Su |
|---|---|---|---|---|---|---|---|
| **An** | **100.0%** | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| **Co** | 0.0% | **100.0%** | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| **Di** | 1.0% | 0.0% | **99.0%** | 0.0% | 0.0% | 0.0% | 0.0% |
| **Fe** | 0.0% | 0.0% | 0.0% | **100.0%** | 0.0% | 0.0% | 0.0% |
| **Ha** | 1.0% | 0.0% | 0.0% | 0.0% | **99.0%** | 0.0% | 0.0% |
| **Sa** | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | **100.0%** | 0.0% |
| **Su** | 0.0% | 0.0% | 0.0% | 1.0% | 0.0% | 0.0% | **99.0%** |
| **Average Accuracy = 99.6%** | | | | | | | |

### 3.4.5 Analysis of The Dimensional Reduction Process with Scatter Plots Between Feature Vectors

To understand the effect of dimensionality reduction using a combination of PCA and LDA to the features, scatter plots between feature vectors are created as illustrated in Fig. 3.6. The scatter plots of the feature vectors represent the relation between each facial expression. In Fig. 3.6 (a), The ORI E-HOG with high dimensionality creates contiguous and overlapping plots between 7 facial expressions. In Fig. 3.6 (b), when the PCA was applied to the ORI E–HOG, the plot becomes convergent with clear boundaries between each facial expression, although some overlapping still occurs. Finally, when LDA was implemented after PCA, a more solid boundary with a significant distance arises between each facial expression as illustrated in Fig. 3.6 (c). Therefore, the features that represent the 7 facial expressions can be easily distinguished by the classifier.

### 3.4.6 Comparison of ORI E–HOG and RED E-HOG to Other Methods

A comparative summary showing the accuracy of ORI E–HOG Face, RED E-HOG Face with other methods that use the extended Cohn-Kanade dataset as a benchmark is presented in Table 3.7. All of the methods classified 7 facial expressions using various features. The table indicates that RED E-HOG Face outperforms the other methods for accuracy.

Table 3.7: Comparison of ORI E–HOG, RED E-HOG Faces to Other Facial Expreesion Recognition Methods

| Method | Feature | # Emotions | Accuracy (%) |
|---|---|---|---|
| CK+ (Kanade *et al.*, 2000) | Action Unit (AU) | 7 | 83.33 |
| J. Li & E. Lam (Li & Lam, 2015) | Gabor Wavelet and Kernel PCA | 7 | 91.70 |
| Ptucha et. al.(Ptucha & Savakis, 2013) | Manifoldbased Sparse Representation | 7 | 94.60 |
| H. Candra et. al. (Candra *et al.*, 2016) | ORI E-HOG Face | 7 | 94.90 |
| H Candra et.al. (Candra *et al.*, (Submitted)b) | RED E-HOG Face | 7 | **99.33** |

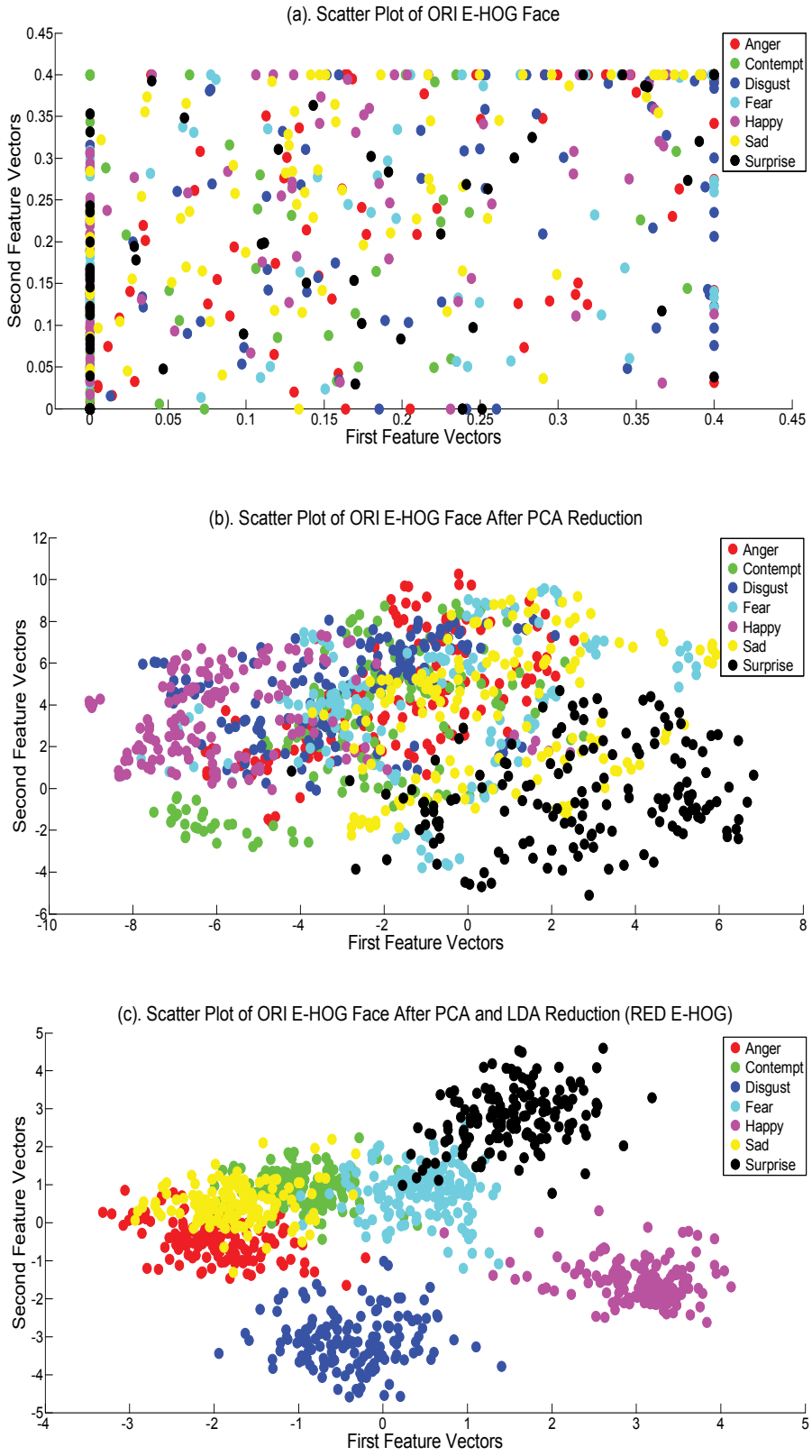



Figure 3.6: Scatter plot of ORI E-HOG features before and after dimensionality reduction using PCA and a combination of PCA and LDA.

## 3.5 Summary

The proposed Edge Histogram Oriented Gradient (E-HOG) has been shown to have very high potential as a feature for facial expression recognition; a classification result as high as 96.4% has been realized using the Extended Cohn-Kanade dataset (Kanade *et al.*, 2000).

The reduced dimension of Edge-Histogram of Oriented Gradients (RED E-HOG) is also introduced as an improved E–HOG with a very low dimension. Using a combination of PCA and LDA for dimensionality reduction technique, the dimension is reduced from thousands to tens. 7 facial expressions has been successfully recognized using 10 dimensions of RED E-HOG, as indicated by the accuracy being as high as 99.6% when trained with multi-class SVM.

The high classification result of RED E-HOG is also supported by the significant improvement in training and testing time of RED E-HOG compared to ORI E-HOG. Upon these 2 indicative improvements, the implementation of the RED E-HOG features (specifically RED E-HOG Face) becomes feasible and efficient for facial expression recognition.

Analysis of the confusion matrix indicates that anger, contempt, fear, and sad were distinguished clearly from the 7 emotions using the RED E-HOG Face features. The analysis of dimensional reduction process with scatter plots between feature vectors reveals that RED E–HOG demonstrates a simpler plot compared to ORI E–HOG. The scatter plot of RED E–HOG has a solid boundary with significant distance between 7 facial expressions so that they can be distinguished clearly by the SVM classifier.

The comparison to other methods shows that the RED E-HOG Face features outperforms. The high classification rate of RED E-HOG Face features indicates that the features to describe facial emotions include the area outside the eyes and mouth, such as skin wrinkles on the nose and forehead.

Future research in this field includes the implementation of Deep Belief Network as the classifier and application of the features in biomedical engineering to recognize facial expressions of people associated with specific emotional disorders.

# Chapter 4

# Discrete EEG-Emotions Recognition Using Wavelet Features

This chapter focus on the development of an EEG emotion recognition method to identify four discrete emotions namely: happy, sad, angry, and relaxed using wavelet features including energy and entropy. The effect of EEG channel and subband selection in relation to the classification results is also investigated to improve accuracy while also simplifying the wavelet features and reducing the number of channels. For that purpose, 2 statistical tools specifically Receiver Operating Characteristic (ROC) and Normal Mutual Information (NMI) were adopted in the analysis.

The main contribution of this chapter are: development of a discrete EEG- emotion recognition method using EEG signals and wavelet features; recommendation of subband selection using only 3 bands of wavelet energy and entropy (alpha, beta, and gamma subbands) as features for EEG discrete emotion recognition, and also; suggestion on channel selection of EEG electrodes using 18 channels (Fp1, Fp2, AF3, AF4, F3,F4, F7, F8, FC5, FC6, T7, T8, P7, P8, P3, P4, O1, and O2), without substantial decrease in the results, which gives a positive indication of the appropriateness of the selected channels and subbands.

The organization of this chapter is as follows: a quick overview on EEG emotion recognition method is discussed in Section 4.1. Section 4.2 describes the experimental setup

involving dataset resources and preprocessing, subject grouping, dataset mapping, and experimental outline. In Section 4.3 the experimental detail of the proposed discrete EEG-emotion recognition using wavelet features is explained, followed by the discussion on the results and implementation in Section 4.4. The chapter is summarized in Section 4.5.

## 4.1 Method Overview

Emotions as a part of non-verbal communication play an important role in interactions between individuals. Emotions are directly related to the brain and are manifested in the form of brain waves that affect the entire system of a person's body. Detection and study of these brain waves is known as electroencephalography (EEG) which is a type of physiological signal that has been implemented in various emotion recognition system schemes.

Psychologist James Russell proposed the *circumplex model of emotion* in 1980 (Russell, 1980). The circumplex model is a conceptualized 2-dimensional continuous space where the horizontal and vertical axes correspond to the degree of valence (pleasure) and arousal, respectively. Discrete emotional states such as 'happy', 'sad', 'angry', and 'relaxed' can be mapped to degree of valence and arousal as illustrated in Fig. 4.1. Using this model, the degree of any of the aforementioned discrete emotional states can be measured. It has been further reported that the psychological condition of positive/negative arousal (activation/deactivation) and positive/negative valence (pleasant/unpleasant) can be identified from Galvanic Skin Response (GSR) and EEG signal (Torres *et al.*, 2013).

Studies have indicated that the reliability of most EEG-based systems can be improved by proper:

1. **Feature selection** To date, there is still no consensus as to which features are most suitable for EEG emotion recognition (Jenke *et al.*, 2014b). Two of the popular methods often considered are Fast Fourier Transform (FFT) (Nie *et al.*, 2011) and Discrete Wavelet Transform (DWT) (Rizon, 2010).

2. **EEG subband selection**: EEG waveforms are generally subdivided into several

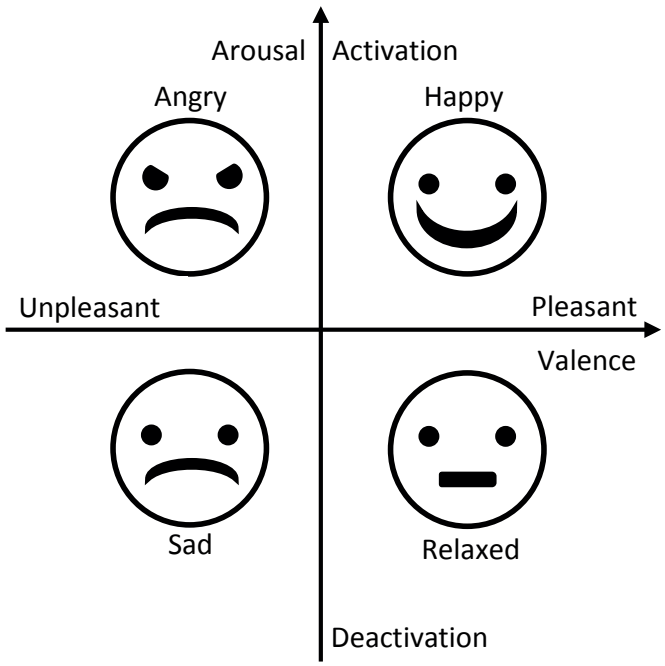



Figure 4.1: Mapping of 4 discrete emotions to Russell's circumplex model of affect (Russell, 1980). Horizontal axis represents valence (pleasure); vertical axis represents arousal.

bandwidths known as delta, theta, alpha, beta, and gamma. These bands need to be properly selected as not all of them carry the relevant information for emotion recognition. For example, Zhu et. al. discovered that EEG patterns in the Beta and Gamma bands are generally stable across emotions and subjects (Zhu *et al.*, 2014).

3. **EEG channel selection**: Jenke suggested that not all of the electrode locations carry the appropriate information for emotions (Jenke *et al.*, 2014b).

In this chapter, an EEG emotion recognition method is cosntructed to identify four discrete emotions namely: happy, sad, angry, and relaxed using wavelet features including energy and entropy. The classification algorithm used is the Multi-class Support Vector Machine (SVM) (Schlesinger & Hlavac, 2002). We also investigated the effect of EEG channel and subband selection.

On the DEAP (Dataset for Emotion Analysis using electroencephalogram, Physiological and Video Signals) (Koelstra *et al.*, 2012), our method achieved the average sensitivity and specificity of 72.3% and 83.8% using the Wavelet Entropy features. We observed an

increase of specificity from 79.8% to 82.1% using the alpha, beta, and gamma subbands. Further specificity increase to 83.8% was achieved using 18 EEG channels namely: Fp1, Fp2, AF3, AF4, F3,F4, F7, F8, FC5, FC6, T7, T8, P7, P8, P3, P4, O1, and O2 (Candra *et al.*, 2015b). We did not observe any substantial decrease in sensitivity. This implies that the channels and subbands selection are applicable for discrete EEG emotion recognition method.

## 4.2 Experimental Setup

### 4.2.1 Dataset Resource and Dataset Preprocessing

This part of the thesis uses the **Dataset for Emotion Analysis using electroencephalogram, Physiological and Video Signals (DEAP)** Koelstra *et al.* (2012) provided through the courtesy of the authors, Koelstra et. al. The dataset provides multimodal data of human affective states analysis from 32 volunteers. The volunteers were asked to watch 40 one-minute extracts of music videos while having their signals recorded. The signal sources include EEG (10-20 system), EOG, EMG (Zygomaticus Major and Trapezius muscles), GSR (left middle and ring fingers), respiration belt, plethysmograph (left thumb), and temperature (left pinky). 22 out of 32 participants had their frontal face videoed during the experiment. The participants assessed each music video with continuous real value between 1 and 9 to describe their emotions in term of valence, arousal, dominance, delight and familiarity. Assessment was conducted with the standardized Self-Assessment Manikins (SAM) (Bradley & Lang, 1994). EEG preprocessing was done according to Koelstra's Koelstra *et al.* (2012) as follows:

1. Resampling to 128Hz,
2. Removal of EOG artifacts using Blind Source Separation,
3. Filtering using 4.0 - 45.0 Hz Bandpass filter, and
4. Averaging the data to common reference.

Additionally, the dataset also provides important information to understand the dataset such as online ratings, list of videos, and participant questionnaires. The preprocessed EEG data using Koelstra's preprocessing method (Koelstra *et al.*, 2012) were also provided in Python and Matlab format. This preprocessed data was used throughout the experiment.

### 4.2.2 Subject Grouping from The Dataset

As previously mentioned in Chapter 2, the used of self-reports of emotion using SAM still faced an issue of misinterpretation of the images used in SAM to be related to certain emotions by the users. This leads to very contradicting interpretation of emotions that creates degradation of the classification reliability (Isomursu *et al.*, 2007). Therefore, to reduce the mistranslation effect, we conducted a subject grouping to properly select a group of participants for the experiment as explained in the Experimental Detail section.

### 4.2.3 Dataset Mapping

In order to extract 4 discrete emotions from the EEG signals, there is a need to map the subjective rating of arousal and valence emotions into 4 classes of discrete emotion. This can be accomplished by plotting the ratings into 4 quadrants of the circumplex model. The coordinate pairs of arousal and valence will define the position in the 4 quadrants. Each quadrant is related to a specific discrete emotion as illustrated in Fig. 4.1. The coordinate pairs are defined as follows:

- Pleasant, Activation $\Rightarrow$ Happy
- Pleasant, Deactivation $\Rightarrow$ Relaxed
- Unpleasant, Activation $\Rightarrow$ Angry
- Unpleasant, Deactivation $\Rightarrow$ Sad

### 4.2.4 Experimental Outline

The block diagram of the discrete EEG-emotion recognition is shown in Fig. 4.2.

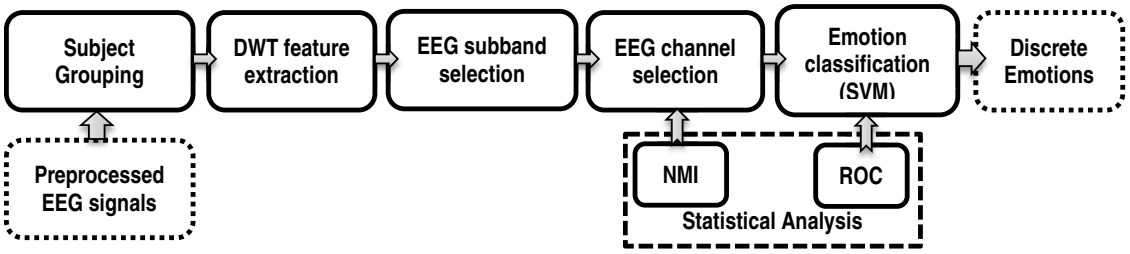



Figure 4.2: Block diagram of the Discete EEG-emotion recognition.

The method applied a few main steps to map a raw EEG signal to the appropriate emotion. The brief description of the block diagram is as follows:

- **Preprocess the raw EEG signal** to remove artifacts such as eye blinking and down-sampling the signal to 128Hz. The preprocessed DEAP dataset is readily available in the DEAP website (Koelstra *et al.*, 2012).
- **Subject grouping from the dataset** by applying the RBF kernel function (Vert *et al.*, 2004) to calculate the transformation matrix of each person's EEG features related to their own emotion rating. The result is fitted as a logistic regression. The whole results of all participants are concatenated to form the matrices for clustering. These mapping matrices are clustered based on how closely related each person is to another and put in groups with dendrogram.
- **Compute the wavelet features using DWT** from a given EEG channel and take the wavelet energy and wavelet entropy from the decomposition level that corresponds to the appropriate EEG subband.
- **Select the EEG subbands** by choosing from 5 available subbands reduced to 3 subbands only.
- **Reduce the EEG channels** by taking from 32 available electrodes subtracted into 18 selected channels.

- **Train the SVM classifier** by conducting 2 consecutive operations: dataset generation, and; optimization of the support vector centroids and radius.
- **Analyse the features and classification results** with NMI and ROC.

## 4.3 Experimental Detail

In order to better understand the process and algorithms involved in the proposed discrete EEG-emotion recognition method, this section is divided into the following subsections to provide more detailed explanation of each step:

a Subject grouping

b EEG Feature Extraction with Discrete Wavelet Transform (DWT)

c EEG subband selection

d EEG channel selection

e Optimizing the Multi-class SVM Classifier

f Receiver operating characteristic (ROC) for Classification Analysis

g Normalized Mutual Information (NMI) for EEG Channel Analysis

The details of each steps are discussed in the following paragraph.

### 4.3.1 Subject Grouping

Let $\mathbb{X}_z(s) = \{ (c_{z,1}, \mathbf{x}_1), (c_{z,2}, \mathbf{x}_2), \ldots, (c_{z,N}, \mathbf{x}_N) \}$ be defined as a set of labeled points $\mathbb{X}_z$ produced by a specific subject $s$.

For any $\mathbf{x}$, let $\phi(\mathbf{x}, \mathbf{x}')$ be defined as an RBF kernel function (Vert *et al.*, 2004),

$$\phi(x, \mathbf{x'}_m) = \exp\left(-\frac{||\mathbf{x} - \mathbf{x'}_m||^2}{2\sigma^2}\right) \tag{4.1}$$

where $\mathbf{x'}_m$ and $\sigma$ denote the center and radius of the $m^{th}$ kernel, both assumed to be reasonably optimized.

Let $\mathbf{\Phi}(\mathbf{x})$ be defined as a vector containing the outputs of the kernel function evaluated for each $\mathbf{x'}_m$, $m \in \{\, 1, \ldots, M \,\}$,

$$\mathbf{\Phi}(\mathbf{x}) = \bigcup_m \phi(\mathbf{x}, \mathbf{x'}_m). \tag{4.2}$$

For any $c_z$, let $\mathrm{logit}(c_z)$ be defined as follows,

$$\mathrm{logit}(c_z) = \log\left(\frac{c_z}{1 - c_z}\right). \tag{4.3}$$

Let $\mathbf{c}$ be defined as a vector containing the label for the four emotions,

$$\mathbf{c} = \bigcup_{z \in \mathbb{A}} c_z, \mathbb{A} \in \{\, vl, ar, do, lk \,\}. \tag{4.4}$$

The relation between $(\mathbf{c}_n, \mathbf{x}_n)$ can then be simplified as the following linear equation,

$$\mathbf{\Phi}(\mathbf{x}_n)\mathbf{P} = \mathrm{logit}(\mathbf{c}_n) \tag{4.5}$$

where $\mathbf{P}$ is a $M \times 4$ matrix explaining the linear relationship between $\mathbf{\Phi}(\mathbf{x}_n)$ and $\mathbf{c}_n$. Assuming $\mathbf{C} = \{\, \mathbf{c}_1, \ldots, \mathbf{c}_N \,\}$, $\mathbf{\Phi}(\mathbf{X}) = \{\, \mathbf{\Phi}(\mathbf{x}_1), \ldots, \mathbf{\Phi}(\mathbf{x}_N) \,\}$, and $\mathbf{\Phi}(\mathbf{X})$ is a full-column rank matrix, we can then linearly solve $\mathbf{P}$ as follows,

$$\mathbf{P} = \left(\mathbf{\Phi}^T(\mathbf{X})\mathbf{\Phi}(\mathbf{X})\right)^{-1} \mathbf{\Phi}^T(\mathbf{X})\,\mathrm{logit}(\mathbf{C}). \tag{4.6}$$

In order to compare the mapping matrices of each individual subject, we then solve $\mathbf{P}$ for every $\mathbb{X}_z(s)$. This process yields $\mathbb{P} = \{\, \mathbf{P}_1, \ldots, \mathbf{P}_S \,\}$, where $S$ is the number of subjects. We then compute pairwise euclidean distance between the $\mathbf{P}$ matrices. Subjects with similar reponse patterns will then have relatively closer distance. We then construct a dendrogram using Ward's method and divide the subjects into several subgroups according to the closeness of their matrices. We then randomly picked one of the subgroups as the subjects of interest in the experiment.

### 4.3.2 EEG Feature Extraction with Discrete Wavelet Transform (DWT)

The DEAP dataset recorded the EEG signals of 32 participants with each having 40 minutes duration. Each of the EEG signals can be divided into 40 $\times$ 1 minutes length of cropped EEG signals with each being related to a low/high state of arousal or valence emotion. Therefore, from each participant can be collected 40 different EEG signals that represent high/low state of arousal or valence emotion, which in total gives 40 EEG signals $\times$ 32 participants = 1,280 individual EEG signals.

In this experiment all of the 1,280 individual EEG signals were processed to acquire features in time-frequency domain using the DWT transform. The purpose of the DWT utilization is due to its capability to multiscale zooming, and multirate filtering for detecting and characterizing transients which allow us to extract the appropriate information from the EEG signals to be used as the feature for EEG emotion classification Handojoseno *et al.* (2012).

The DWT feature extraction of each individual preprocessed EEG signal was conducted using the following steps:

1. Given any 1 minute EEG signal $x(t)$, the DWT of $x(t)$ can be computed using Eq. (2.9). As it has already been explained in Chapter 2, the DWT can be interpreted as a filtering process using a *dyadically* shifted and scaled mother wavelet. Thus, for a given sampled EEG signal $x[n] = x(nT_s)$, the DWT can then be calculated recursively for each level of decomposition, where $T_s$ denotes the time between samples. For that purpose, the EEG signals approximation coefficients $x_A$ and detail coefficients $x_D$ can be calculated using Eq. (2.10) and (2.11), respectively.

2. The next step is to calculate the wavelet energy $E(a)$ from the EEG wavelet coefficient localized at the $a^{th}$ level of decomposition using Eq. (2.12), where $C_a$ is the EEG wavelet coefficients at the $a^{th}$ decomposition level. $C_a$ can be either $x_{A_a}$ or $x_{D_a}$ of the EEG wavelet coefficients, and then the wavelet energy is normalized against the total wavelet energy using Eq. (2.13), where $K$ denotes the number of EEG discrete wavelet decompositions, $p(a) \in \{0, 1\}$ and $\sum_a p(a) = 1$.

3. The wavelet entropy $H(a)$ of the EEG sgnal is also computed using Eq. (2.14), with $K$ as the number of EEG signal DWT decompositions.

### 4.3.3 EEG subband selection

According to Tinguely et. al, (Tinguely *et al.*, 2006) the delta and theta bands are more relevant to the sleep stage, and thus have less relevance to the emergence of emotion. For this reason, we investigated two experimental settings of EEG channel selection as follows:

$\mathrm{SB}_5$: 5 EEG frequency bands: delta, theta, alpha, beta, and gamma: $\mathrm{SB}_5 = \{\delta, \theta, \alpha, \beta, \gamma\}$, and

$\mathrm{SB}_3$: 3 EEG frequency bands: alpha, beta, and gamma: $\mathrm{SB}_3 = \{\alpha, \beta, \gamma\}$.

The results of those 2 experimental settings were then compared to inspect the effect of subband reduction to the classification results.

### 4.3.4 EEG channel selection

The effect of channels selection was also examined based on the literature review on the most commonly used EEG channels for EEG emotion recognition. We implemented the channels selection scenarios as follows:

$\mathrm{CH}_{32}$: No channel selection, i.e. using all the 32 available EEG channels.

$\mathrm{CH}_{18}$: The number of channels was reduced to the 18 channels commonly used in the EEG emotion recognition literature according to Jenke et. al (Jenke *et al.*, 2014b). The 18 channels are: Fp1, Fp2, AF3, AF4, F3,F4, F7, F8, FC5, FC6, T7, T8, P7, P8, P3, P4, O1, and O2.

### 4.3.5 Optimizing the Multi-class SVM Classifier

In order to improve the classification results of the SVM classifier, the following preparation steps were carried out :

1. *Generate the training and testing sets*:

   - Perform signal preprocessing and DWT feature extraction for each emotion for each subject from a group of n participants and assign this array of observations as the training set, so that

     $$\mathbb{X}_{\text{train}} = \{\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_n\}$$

     where $\mathbb{X}_n = \{\mathbf{x}_{1,n}, \mathbf{x}_{2,n}, \ldots\}$ denote the observation vectors extracted from participant n.

   - Similarly, the test set $\mathbb{X}_{\text{test}}$ can then be constructed from another n participants so that

     $$\mathbb{X}_{\text{test}} \cap \mathbb{X}_{\text{train}} = \emptyset.$$

2. *Optimize the support vector centroids and radius* for the RBF kernel from $\mathbb{X}_{\text{train}}$ using the Ensemble Rapid Centroid Estimation (ERCE) (Yuwono *et al.*, 2014).

3. *Train the multi-class SVM* using $\mathbb{X}_{\text{train}}$ against the class labels. Sequential Minimal Optimization (SMO) algorithm was used for the SVM training algorithm.

The emotion classification was conducted using a multi-class SVM with a Radial Basis Function (RBF) kernel which can be computed using Eq. (2.17).

The JS divergence is a symmetrized and smoothed version of the Kullback-Leibler (KL) divergence (Fuglede & Topsoe, 2004). Given two discrete probability distributions $P \sim p(x)$ and $Q \sim q(x)$, the JS divergence is calculated using Eq. (2.18) and (2.19).

Properly estimating the parameters for the RBF kernel is particularly important in order to ensure proper learning (Schlesinger & Hlavac, 2002). We used a particle swarm ensemble clustering algorithm called the ERCE algorithm for this purpose (Yuwono *et al.*, 2014). It has been argued that the algorithm can estimate the number of clusters directly from the

data using swarm intelligence and ensemble aggregation (Yuwono *et al.*, 2014; Yuwono, 2015).

Using ERCE, the support vector centroids $\boldsymbol{\mu}_k \in \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots\}$ as well as the kernel radius $\sigma_k \in \{\sigma_1, \sigma_2, \ldots\}$ can be inferred from the training data. The steps are as follows:

1. Execute ERCE (Yuwono *et al.*, 2014) to cluster the training set $\mathbb{X}_{\text{train}} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ to an arbitrary number of clusters based on JS distance (i.e. the square root of JS divergence).

2. Aggregate the ensemble clustering results using average linkage to get the final clustered sets $\{\mathbb{C}_1, \mathbb{C}_2, \ldots, \mathbb{C}_K\}$, where $K$ is determined automatically by ERCE at ensemble aggregation. The corresponding centroid vector $\boldsymbol{\mu}_k \in \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K\}$ was computed as conditional expectation using. Eq 2.20.

3. The RBF kernel radius for the $k^{th}$ support vector was taken as the square root of conditional JS divergence and computed with Eq. (2.21).

4. Given the optimized RBF kernel parameters, the SVM was then trained using the SMO algorithm (Chang & Lin, 2011).

### 4.3.6 Receiver operating characteristic (ROC) for Classification Analysis

A receiver operating characteristic (ROC) curve is a statistical tool to evaluate the quality or performance of diagnostic tests from a binary classifier system which is illustrated as a graphical plot (Bradley, 1997).

The ROC curve was first developed by electrical engineers and radar engineers during World War II for detecting enemy objects in battlefields and was soon introduced to psychology to account for perceptual detection of stimuli. ROC analysis since then has been used in medicine, radiology, biometrics, and other areas for many decades and is increasingly used in machine learning and data mining research (Hajian-Tilaki, 2013; Zou *et al.*, 2007).

Consider a two class prediction problem (binary classification) with the outcomes are labelled either as positive (p) or negative (n). It will give 4 possible outcomes as follows:

1. If the outcome from a prediction is p and the actual value is also p, then it is called a true positive (TP);
2. If the outcome from a prediction is p and the actual value is n then it is said to be a false positive (FP);
3. When both the prediction outcome and the actual value are n, then true negative (TN) is occurred;
4. When the prediction outcome is n while the actual value is p, it is called the false negative (FN).

The confusion matix of the binary classification with its possible outcomes is illustrated in Fig. 4.3.

| | | Predicted | |
|---|---|---|---|
| | | p | n |
| Actual | p | True Positive (TP) | False Negative {FN) |
| | n | False Positive (FP) | True Negative (TN) |

Figure 4.3: Confusion matrix of binary classification with 4 possible outcomes.

A ROC space is defined by FPR and TPR as x and y axes respectively, which depicts relative tradeoffs between true positive (benefits) and false positive (costs). Since TPR is

equivalent to sensitivity and FPR is equal to 1 – specificity, the ROC graph is sometimes called the sensitivity vs (1 – specificity) plot. Each prediction result or instance of a confusion matrix represents one point in the ROC space (Fawcett, 2006). The formulation of the TPR and FPR is as follows:

$$Sensitivity = TPR = \frac{TP}{P} = \frac{TP}{(TP + FN)} \tag{4.7}$$

$$Specificity = SPC = \frac{TN}{N} = \frac{TN}{(TN + FP)} \tag{4.8}$$

$$FPR = 1 - Specificity = 1 - SPC \tag{4.9}$$

### 4.3.7 Normalized Mutual Information (NMI) for EEG Channel Analysis

Mutual Information (MI) is the measurement of mutual dependence (amount of information) between a random variable X towards another random variable Y. This measurement is an indication of coherence between two distributions that generates the variables (vectors). Mutual Information has been applied for image registration over the decades based on the marginal and joint entropies (Cahill, 2010). However, its implementation has been applied in data mining to group similar data records (He *et al.*, 2008).

Recently, Normalized Mutual Information (NMI) has been proposed for feature analysis which can be used to better understand the relation between the features and their represented label (Estévez *et al.*, 2009). The NMI may inform whether the features are fully, partly or not even related to the label. The implementation of NMI algorithm in EEG signal processing offers the possibilities to analyze the relation between the features and the EEG channels. The attempt to implement NMI algorithm in EEG-emotion classification has been initiated in (Candra *et al.*, 2015b) and applied in (Candra *et al.*, 2017), using NMI algorithm to measure the relation between features in EEG channels and their represented emotions. The calculation of MI and NMI can be defined as follows (Cahill, 2010):

$$MI(X, Y) = H(X) + H(Y) - H(X, Y) \tag{4.10}$$

The MI is normalized (NMI) to obtain value between 0 (independence) and 1 (really

dependence) with the equation can be rewritten as (Strehl & Ghosh, 2003):

$$MI(X,Y) = \frac{NMI(X,Y)}{\sqrt{H(X)H(Y)}} \tag{4.11}$$

This NMI calculation is implemented in this chapter to analyse the correlation between the EEG wavelet features and the EEG channels.

## 4.4 Discussion

Prior to discussing the implementation results of the discrete EEG-emotion recognition, we divide this section into the following 4 subsections to provide the analysis of the results and its implication in depth:

a Classification Results of The Discrete EEG-Emotion Recognition

b Confusion Matrix of The Best Trained Classifier

c ROC Analysis of The Classification Results

d NMI Analysis of The EEG Channel Selection

The detailed discussion of each subsection is presented in the following paragraph in accordance with the given order.

### 4.4.1 Classification Results of The Discrete EEG-Emotion Recognition

Basically, the classification results can be divided into 2 main categories: the classification results of wavelet energy and wavelet entropy. These groups can be further subdivided into 3 more specific combinations of subband selection and channel selection, which, in total give 6 combinations of classification results, in particular: wavelet energy with 5 subbands and 32 channels; wavelet energy with 3 subbands and 32 channels; and wavelet energy with 3 subbands and 18 channels. The similar combinations can also be retrieved from the wavelet entropy.

In addition, the division of the subbands was obtained using daubechies 5 (db5) as the mother wavelet. The wavelet features were computed from the following detail coefficients to extract 5 different subbands frequency:

- $5^{th}$ detail coefficients: $x_{D_5} \approx$ Delta (3–4Hz);
- $4^{th}$ detail coefficients: $x_{D_4} \approx$ Theta (5–8Hz);
- $3^{rd}$ detail coefficients: $x_{D_3} \approx$ Alpha (9–16Hz);
- $2^{nd}$ detail coefficients: $x_{D_2} \approx$ Beta (17–32Hz); and
- $1^{st}$ detail coefficients: $x_{D_1} \approx$ Gamma (33–64Hz).

Furthermore, the channel selections were taken based on the summary of channel selections used in the references that has been adopted by Jenke et. al (Jenke *et al.*, 2014b). We selected the 18 most commonly used EEG channels which deliver the following list of channels: Fp1, Fp2, AF3, AF4, F3,F4, F7, F8, FC5, FC6, T7, T8, P7, P8, P3, P4, O1, and O2.

All the classification results of those combinations of the discrete EEG-emotion recognition are summarized in Table 4.1 which were obtained from 30 repetitions of SVM training and testing episodes using 5 subbands ($\mathrm{SB}_5$) and 3 subbands ($\mathrm{SB}_3$), as well as 32 ($\mathrm{CH}_{32}$) and 18 ($\mathrm{CH}_{18}$) EEG channels.

Using Table 4.1 we can observe diverse diagnosis. First, the emotion classification results of wavelet entropy generally pick up higher accuracy than relative wavelet energy.

Second, using only the alpha, beta, and gamma subbands ($\mathrm{SB}_3$) maintains desirable results in the overall accuracy.

Third, wavelet entropy features achieved 61.8% accuracy under the $\mathrm{CH}_{32}$, $\mathrm{SB}_3$ setting. When the number of channel is reduced to 18, a slight increase in sensitivity is observed with a substantial drop in specificity.

Table 4.1: Classification Results of The Discrete EEG-Emotion Recognition

| **Channel, Subband** | **Sensitivity*** | **Specificity*** | **Accuracy*** |
|---|---|---|---|
| **Features: Relative Wavelet Energy** | | | |
| $CH_{32}$, $SB_5$ | 72.5% ± 10.6% | 79.3% ± 7.0% | 57.3% ± 2.8% |
| $CH_{32}$, $SB_3$ | 75.4% ± 11.1% | 81.8% ± 5.9% | 59.6% ± 1.9% |
| $CH_{18}$, $SB_3$ | 73.5% ± 14.8% | 82.3% ± 6.2% | 59.6% ± 2.6% |
| **Features: Wavelet Entropy** | | | |
| $CH_{32}$, $SB_5$ | 74.6% ± 10.3% | 79.8% ± 6.5% | 59.0% ± 1.7% |
| $CH_{32}$, $SB_3$ | 75.7% ± 12.1% | 83.7% ± 6.8% | 61.8% ± 4.4% |
| $CH_{18}$, $SB_3$ | 77.4% ± 14.1% | 69.1% ± 12.8% | 60.9% ± 3.2% |

* Sensitivity, specificity and accuracy are averaged over the four emotions over 30 experiments.

### 4.4.2 Confusion Matrix of The Best Trained Classifier

Further investigation is conducted using the confusion matrix as shown in Table 4.2. We selected the wavelet entropy under $CH_{18}$, $SB_3$ setting to create the confusion matrix of the best trained classifier. From Table 4.2 it can be observed that the sad emotion was classified up to 88.89%, while the other three emotions achieved lower than 80%. The least result was attained by relaxed emotion with 64.29% of accuracy.

Table 4.2: Confusion matrix of the best trained classifier using wavelet entropy under $CH_{18}$, $SB_3$ setting.

| | | **Output class** | | | |
|---|---|---|---|---|---|
| | | **Sad** | **Relaxed** | **Angry** | **Happy** |
| **Target Class** | **Sad** | **88.89%** | 00.00% | 00.00% | 11.11% |
| | **Relaxed** | 28.57% | **64.29%** | 00.00% | 07.14% |
| | **Angry** | 09.09% | 00.00% | **72.73%** | 18.18% |
| | **Happy** | 09.09% | 20.00% | 18.18% | **72.73%** |

### 4.4.3 ROC Analysis of The Classification Results

We conducted statistical analysis of the classification results by taking the sensitivity (TPR) and 1–specificity (FPR) of the classification results for wavelet entropy with $CH_{18}$, $SB_3$ setting and expressing it as a ROC curve of an average case scenario as displayed in Fig. 4.4.

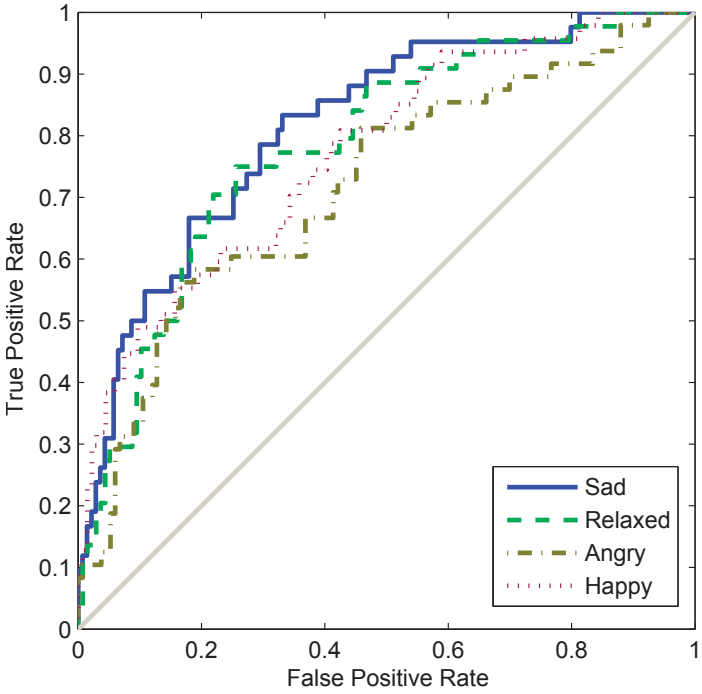



Figure 4.4: Receiver Operating Characteristic (ROC) of an average case scenario.

The ROC curve in Fig. 4.4 shows that sad emotion keeps the highest true positive rate compares to the other three emotions. However, the relaxed emotion this time occupies the second position followed by happy emotion in the third place and the angry emotion at the lowest position. These circumstances do not coincide with the confusion matrix in Table 4.2.

### 4.4.4 NMI Analysis of The EEG Channel Selection

To understand the difference between those two results, we inspect the EEG wavelet features including the EEG channel selections using the NMI method. The NMI of each

selected channel was computed by comparing the combination of 4 emotions that are represented by the feature which provide 6 possibilities, i.e. sad vs. relaxed, sad vs. angry, sad vs. happy, relaxed vs. angry, relaxed vs. happy, and angry vs. happy. The result is illustrated in a stacked bar graph with each bar representing each selected EEG channel as shown in Fig. 4.5.

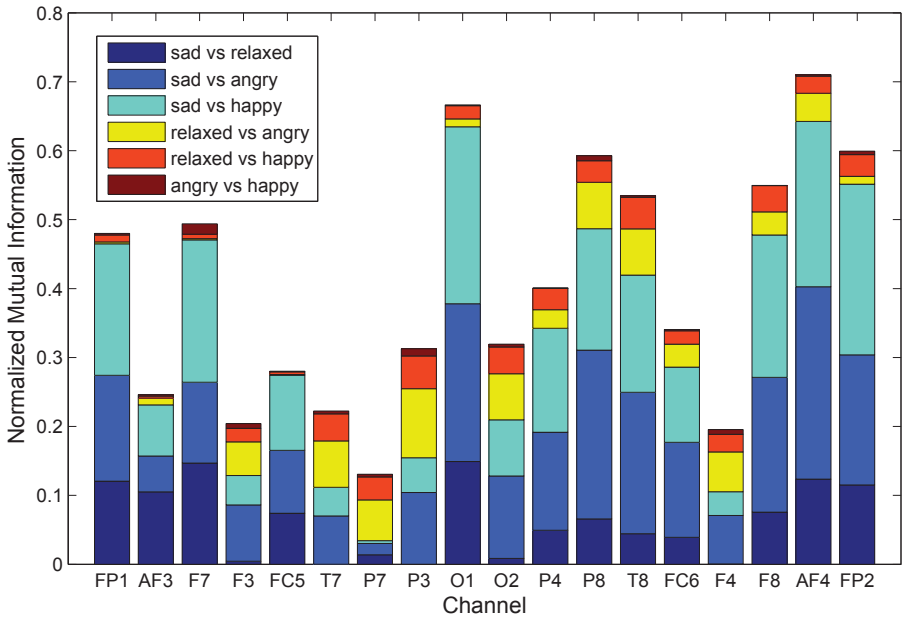



Figure 4.5: The Normalized Mutual Information between each channel and combination of emotions.

Using Fig. 4.5, the phenomena can now be revealed. The stacked bar in the graph indicates that the features are lacking the information to distinguish between angry and happy emotions. It can be seen that the wavelet entropy features were predominantly biased towards characterizing sad emotion, whereas relaxed, happy, and angry emotions were seemingly underrepresented. Therefore, the classification result of sad emotion is higher compared to other emotions. Furthermore, the relaxed emotion has less representation in each stacked bar, while angry emotion has less representation compared to the 3 other emotions on average over all channels.

## 4.5 Summary

In this chapter the discrete EEG-emotion recognition has successfully classified 4 discrete emotions (happy, sad, angry, and relaxed) from dimensional plane of arousal valence using wavelet energy and entropy.

Using the DEAP dataset (Koelstra *et al.*, 2012) we discovered that the alpha, beta, and gamma EEG subbands are particularly important for discrete EEG-emotion recognition.

Using wavelet entropy features with the $\mathrm{CH}_{18}$, $\mathrm{SB}_3$ setting, with the 18 selected EEG channels of Fp1, Fp2, AF3, AF4, F3,F4, F7, F8, FC5, FC6, T7, T8, P7, P8, P3, P4, O1, O2, our proposed approach achieved an average sensitivity and specificity of 77.4% and 69.1%, respectively.

Further analysis of the features and results using ROC curve and NMI method reveals that our method has a limitation, which is that transient patterns and quasi-stable states that may exist in the EEG are overlooked in both energy and entropy formulations. Having this in mind, potential research directions that are worth considering include incorporating EEG microstates and/or brain connectivity features.

# Chapter 5

# Improving EEG Emotion Recognition Accuracy Using OWS Method

A strategy to optimize the reliability of an EEG emotion classification system is introduced in this chapter using a simple feature-agnostic time window selection method. Using the proposed strategy, consistent inceasing of accuracy can be achieved. Analysis of time processing for training and testing is also conducted to locate the window with shorter processing time. The results provide recommendations on the optimal window size in EEG-emotion classification and its correlated time processing. To generalize the finding, 7 different types of features were extracted to identify 4 different types of dimensional state emotion.

The novel contributions of this chapter are: The demonstration of feature-agnostic preprocessing method using Optimal Window Selection (OWS) to improve the classification accuracy of an EEG emotion recognition; The generalization of the method as a potential answer to dealing with the non-stationary behavior of EEG signals using emotion recognition as an example use case; A recommendation of a novel wavelet feature, concatenating both Wavelet Entropy and segment average Wavelet Approximation Coefficients for EEG emotion recognition.

The organization of the chapter is as follows: Section 5.1. provides a brief overview of the proposed OWS method. Section 5.2 elaborates on the experimental setup for the OWS.

Section 3 describes in depth of the experimental details including subject selection algorithm, signal segmentation process step by step, details features extraction algorithms for 7 different kind of features, continuing with the training and testing protocols. Section 5.4 provides the analysis and discussion of the experimental results using OWS implementation. Section 5.5 summarize the results and future directions.

## 5.1 Method Overview

Time-varying and non-stationary characteristics of the EEG signals have become a major challenge in EEG emotion recognition. A simple and effective method to overcome the problem is by dividing the EEG signals into smaller window frames so that the signal with the similar pattern can be extracted (Picard *et al.*, 2001). The extracted signal in certain length of time (window) becomes a pseudo-stationary signal that owns a statistical property which is valuable as a feature for the classification purpose. Therefore, classifications conducted with pseudo-stationary signals provides relatively higher predictive results (Kaplan *et al.*, 2005).

In this chapter, we target the escalated information gain by "zooming in" the recurring pattern on emotion elicitation using the optimization strategy of Optimal Window Selection (OWS) method. The window size has to be *just right*: A window too short will lead to incompleteness, while a window too long will lead to over-inclusion of non-stationary components.

Previously in Chapter 2, we have described an emotion model — that described by Russel — known as Russel's Circumplex Model of Affect (Barrett & BlissMoreau, 2009) which maps a given emotion into two-dimensional planes namely arousal (activate-deactivate) and valence (pleasant-unpleasant) (Russell, 1980). The benefit of this model in emotion recognition has also been demonstrated by Nardelli et al. who shows statistically significant changes in heart rate variability (HRV) between emotion representations in the arousal-valence plane (Nardelli *et al.*, 2015).

Also in Chapter 3, we have demonstrated the use of Russel's Circumplex Model to extract

discrete emotion from the arousal – valence dimensional plane. However, we found that the classification results still can be improved. We came to the hypothesis that to improve the classification results, we first need to improve the detection of high/low of arousal and valence as they are the variable base for a dimensional plane. For that purpose, this chapter focusses on the improvement of EEG emotion classification of dimensional plane to distinguish between high/low states of arousal/valence emotion together with another two additional emotional states, namely dominance (dominant–submissive), and liking (like–dislike).

In addition, there are 2 common features used in EEG signal analysis, i.e. Fast Fourier Transform (FFT) and wavelet transform (Akin, 2002). The extensive use of wavelet transform is related to its advantages compared to FFT- such as (Rosso *et al.*, 2001a):

- The wavelet transform (WT) is both a band-pass filter and a denoiser. With WT, an EEG signal can be easily decomposed and isolated to include only the information from the desired sub-band. This characteristic is particularly important for emotion recognition as beta and gamma frequencies show a higher degree of correlation to emotion relative to other subbands (Zheng & Lu, 2015b; Candra *et al.*, 2015b).
- Wavelet entropy estimates inter-segment regularity. It can be used to identify the occurrence of a pattern.
- Wavelet approximation coefficient is a proxy for identifying the shape of a pattern.

The demonstration of WT as a powerful tool for time-critical EEG feature extraction was shown by Handojoseno et. al. for detection of freezing gait episodes (Ardi Handojoseno *et al.*, Sep. 2015), and and also by Ocak et. al. for epileptic seizures detection (Ocak, 2008).

Moreover, EEG-based emotion classifiers can be built using algorithms such as Linear Discriminant Analysis (LDA), Artificial Neural Networks (ANN), or Support Vector Machines (SVM). In particular SVM has been shown to be a powerful classifier for the task due to its ability to operate on non-linear and high dimensional feature spaces (Lotte *et al.*, 2007). In this experiment, the EEG-emotion classification was conducted using the SVM

classifier.

Previous investigation of window size to improve the classification of EEG emotion using wavelet features and SVM classifier has been conducted (Candra *et al.*, 2015a). The result is a recommendation of 3 to 12s window to be used for classification of valence and arousal emotional state with the result of improved accuracy when using wavelet entropy feature. The recommnendation is presented in the following Fig. 5.1.

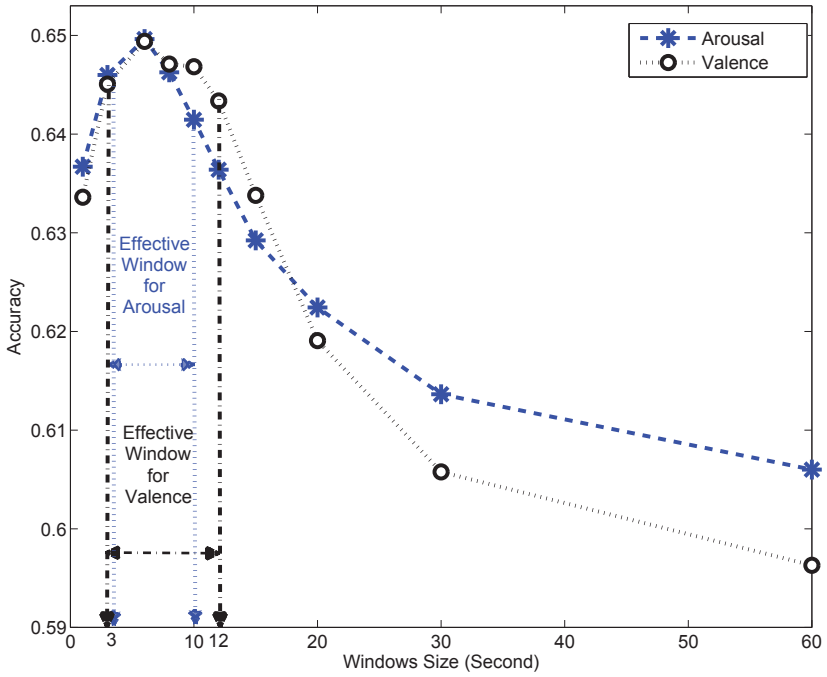



Figure 5.1: Recommendation of 3 to 12s for EEG emotion classification of arousal and valence emotional states (Candra *et al.*, 2015a).

Further investigation has been performed to generalize the finding (Candra *et al.*, (Submitted)a) by improving the investigation parameters as well as:

1. Increasing the number of investigated windows from 10 to 12,
2. Adding 6 additional features to the previous wavelet entropy,
3. Raising the number of emotional states from 2 to 4.

This chapter introduces the Optimal Window Selection (OWS) method based on the work in (Candra *et al.*, (Submitted)a) to optimize the classification results of EEG emotion

recognition in identifying high/low state of arousal/valence/dominance/liking emotions. We explored 7 types of features extracted from EEG emotion signals to demonstrate our optimization strategy with OWS. Our experiments achieved improvement of accuracy between 3% and 15% for all the experimental features. We lock SVM as the classifier of choice in the experiment to provide a fair prediction framework to cater to the objective of finding the optimum feature and window size. We also analysed the training and testing time in finding the shorter processing time. We provide recommendations of the optimal window size in EEG emotion recognition supplied with its correlated time processing.

In addition, a novel wavelet feature built from concatenated Wavelet Entropy and segment average Wavelet Approximation Coefficients is proposed. The classification results using this novel features consistently yield significantly higher accuracy compared to other features as well as simple average, Fast Fourier Transform (FFT), wavelet energy, and wavelet entropy.

## 5.2 Experimental Setup

### 5.2.1 Dataset Resource and Dataset Preparation

The experiment in this chapter was also conducted with the prepocessed DEAP dataset (Koelstra *et al.*, 2012). As previously explained, the dataset contains 32 channels EEG signals of 32 participants while watching 40 music videos with one minute duration for each video. The participants rated each video with a continuous scale 1 to 9 for arousal, valence, dominance, and liking emotions at the time they were watching the video. This experiment makes use of the ratings from the participants to label arousal/valence/dominance/liking emotions with High and Low value, respectively. The middle value of the rating is taken as the baseline.

### 5.2.2 Subject Grouping

To reduce the mistranslation of the emotions by the participants while they rate the video with the SAM method, we need to conduct subject grouping to properly select a group of participants for this experiment. The subject grouping procedure is the same as in the previous chapter. We construct a dendrogram using Ward's method and divide the subjects into several subgroups according to the closeness of their matrices. We then randomly pick one of the subgroups as the subjects of interest in the experiment.

### 5.2.3 Experimental Overview

The objective of the experiment was to find the optimal window size and feature extractor that maximizes the accuracy of an SVM emotion classifier. *Grid-search* was employed for optimizing these two parameters as illustrated in Fig. 5.2.

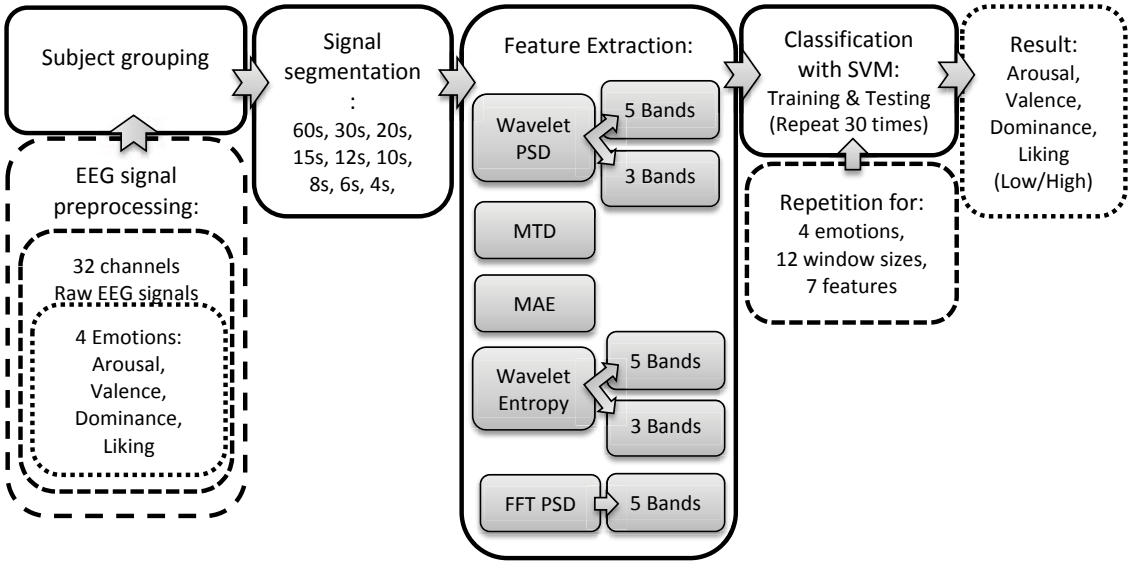



Figure 5.2: Block diagram of Optimal Window Selection (OWS) method.

The brief description of the method is as follows:

1. Subject grouping: We applied the RBF kernel function to relate EEG features of each person to their emotion rating using a transformation matrix. We fitted the result as a logistic regression. The whole results of all participants were concatenated

using mapping matrices and clustered based on how closely related each person is to another and put them in groups with a dendrogram.

2. Signal segmentation: The preprocessed EEG signal was first segmented to 12 window sizes. This parameter was upper bounded by the video length, i.e. 60s. The number of produced segments obeys the rule $N = 60/t$ (refer to Fig. 5.4).

3. Feature extraction: Let a *feature extractor* be a function that maps a preprocessed EEG signal segment $\mathbf{s}_n$ to its feature vector representation $\mathbf{x}_n$ as follows,

$$f(\mathbf{s}_n) \mapsto \mathbf{x}_n. \tag{5.1}$$

The *feature extractor* considered in this study includes:

   (a) wavelet PSD (5 bands and 3 bands)[1],

   (b) wavelet entropy (5 bands and 3 bands)[1] Candra *et al.* (2015b),

   (c) naive **M**ean of the **T**ime **D**omain signal (MTD), and

   (d) **M**ean wavelet **A**pproximation coefficients with wavelet **E**ntropy (MAE),

   (e) FFT PSD (5 bands) was added as a comparison to other features.

4. Emotion classification and evaluation

   (a) Labeled set construction

   For a given emotion[2] $z \in \mathbb{A} : \mathbb{A} = \{\, vl, ar, do, lk \,\}$, the ratings given to a video $v$ were mapped to a category label $c_{z,v}$ using the following rule:

$$c_{z,v} = \begin{cases} Lo & \text{if} \quad 1 \leq v \leq 4.5, \\ Hi & \text{if} \quad 4.5 < v \leq 9. \end{cases} \tag{5.2}$$

   This category was then mapped to the corresponding feature vectors, yielding a collection of labeled points[3], $\{\, \mathbb{X}_z : z \in \mathbb{A} \,\}$:

$$\mathbb{X}_z(v) = \{\, (c_{z,v}, \mathbf{x}_1), (c_{z,v}, \mathbf{x}_2), \ldots, (c_{z,v}, \mathbf{x}_N) \,\}, \tag{5.3}$$

[1] 5 EEG bands: delta, theta, alpha, beta, and gamma: $\mathrm{SB}_5 = \{\delta, \theta, \alpha, \beta, \gamma\}$, 3 EEG bands: alpha, beta, and gamma: $\mathrm{SB}_3 = \{\alpha, \beta, \gamma\}$

[2] $vl$: valence; $ar$: arousal; $do$: dominance; $lk$: liking

[3] a labeled point is a category-label–feature vector tuple: $(c_v^z, \mathbf{x}_n^z)$

where $N$ denotes the number of segments. The overall dataset $\mathbb{X}$ is the union of labeled points for each video:

$$\mathbb{X}_z = \bigcup_{1}^{V} \mathbb{X}_z(v) \tag{5.4}$$

(b) Training and testing

The datasets were randomly split into training set (30%) and test set (70%). SVM was trained using the training set. Classification sensitivity, specificity, and accuracy were evaluated using the test set. This random sampling and retraining step was repeated 30 times.

With 4 emotions, 12 window sizes, 7 features, and 30 re-trainings, a total of $4 \times 12 \times 7 \times 30 = 10,080$ parallel computation jobs were generated and dispatched. A more detailed process on each step will be discussed in the following section.

## 5.3 Experimental Detail

In this section, we will explain in more detail how the experiment was conducted. We started by performing subject grouping, EEG signal segmentation, continued with feature extraction employing various methods including: Discrete Wavelet Transform, statistical features, and Fast Fourier Transform. The last step of the experiment was classification with SVM. To better understand how the experiment was conducted, we have divided this section into the following 6 subsections:

a. Subject grouping

b. EEG emotion signals segmented into 12 window size

c. Discrete Wavelet Transform (DWT) feature extraction

d. Statistical features: naive **M**ean of the **T**ime **D**omain signal (MTD) and **M**ean wavelet **A**pproximation coefficients with wavelet **E**ntropy (MAE)

e. Fast Fourier Transform Power Spectral Density (FFT PSD)

f. Training and testing with SVM classifier

### 5.3.1 Subject Grouping

We implemented the subject grouping procedure the same as in Chapter 4. We then compute pairwise euclidean distance between the $\mathbf{P}$ matrices. Subjects with similar reponse patterns will then have relatively closer distance. A dendrogram using Ward's method is then constructed which result in the division of the subjects into subgroups based on the distance between their matrices. One of the subgroups was picked up randomly to be used as the dataset for the experiment. The result of the subject grouping is visualized in the following Fig. 5.3.

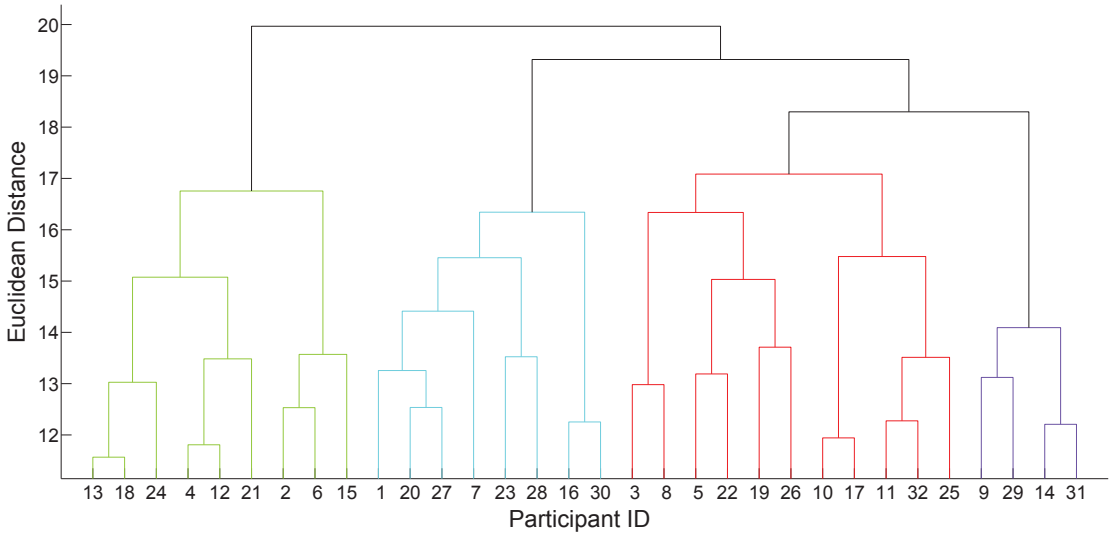



Figure 5.3: Result of subject grouping to overcome mistranslation effect of SAM rating method. The Euclidean distance reveals how closely a participant is related with another. Participants with similar response patterns will have relatively closer distance.

### 5.3.2 EEG Emotion Signals Segmented into 12 Window Sizes

The preprocessed EEG signals were segmented to prepare for feature extraction. The largest possible window size was upper bounded at the video clip length, 60s. This window was then subsequently reduced by half and continuously reduced by half from the previous window size until obtaining the smallest window of 1s which was limited by the sampling rate 128 sample/s. Additional window sizes were inserted between the previously selected sizes to subtilize the result of the investigation.

The total window sizes accommodated are 12 sizes which are: 60; 30; 20; 15; 12; 10; 8; 6; 4; 3; 2; and 1s. Fig. 5.4 illustrated the window segmentation process from largest to smallest size with a total of 12 window sizes. The results of the segmentation are stored as a different dataset for the feature extraction process. For each segmentation result, all segments were combined and used for the feature extraction which results in the multiplication of data size according to the number of segments gathered.

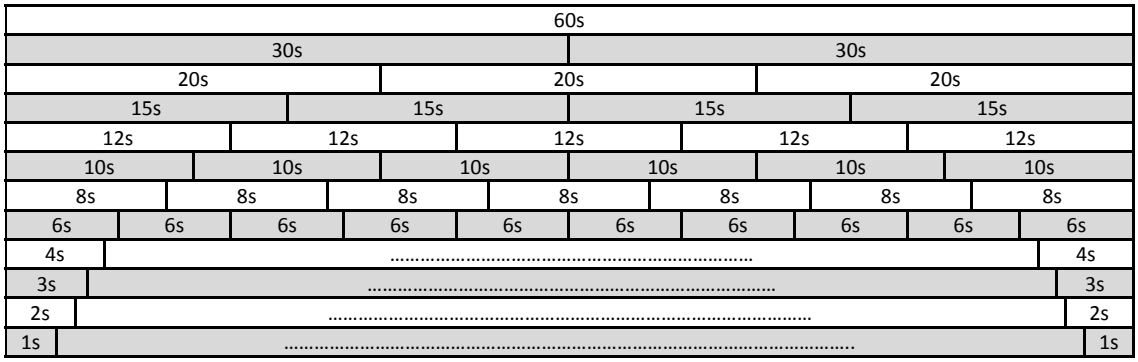



Figure 5.4: The signal segmentation scheme with 12 window sizes.

### 5.3.3 Wavelet Feature Extraction: Wavelet Energy and Wavelet Entropy

This process is needed to extract 2 types of wavelet features from the EEG signals, i.e. wavelet energy which is also called wavelet Power Spectral Density (PSD) and wavelet entropy. First, compute the wavelet coefficients $DWT$. The wavelet coefficient DWT$(x(t); a, n)$ can be calculated using Eq. (2.9).

The process is continued with the computation of wavelet PSD $E(a)$ using Eq. (2.12) to obtain the wavelet energy localized at $a^{th}$. To collect the wavelet PSD, the localized wavelet energy $E(a)$ is normalized against the total wavelet energy as probability mass function $p(a)$ and calculated using Eq. (2.13).

The wavelet entropy $H(a)$ as the degree of uncertainty in the energy distribution can be obtained from the probability mass function $p(a)$ (refer to Eq. (2.13)), using Eq. (2.14).

### 5.3.4 Creating an array of 32 channels wavelet PSD / entropy in 5 or 3 bands formation

The wavelet PSD and entropy, each is then arranged as an array of 32 channels with 5 or 3 bands formation. Each array is the representation of one segment of EEG-emotion signal. The process is illustrated in Fig. 5.5.

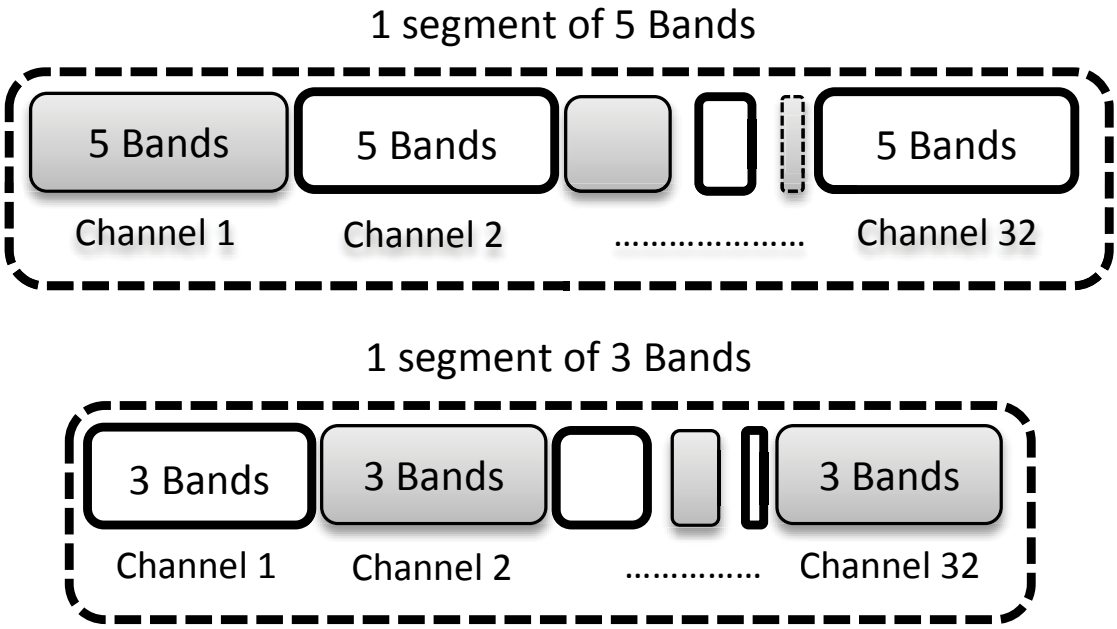



Figure 5.5: An array of 32 channels wavelet PSD / entropy in 5 or 3 bands formation as the representation of one segment EEG-emotion signal.

### 5.3.5 Statistical Features: *MTD* and *MAE*

This experiment used 2 different types of statistical features. First, the statistical feature in time domain using simple average of the EEG signals in time domain we called MTD. Second, the feature in time frequency domain using average of wavelet approximation coefficients combined with wavelet entropy we named MAE. The computation of each feature is as follows:

A. Naive **M**ean of the **T**ime **D**omain signal ***(MTD)***

Take $X_k$ as the representation of the amplitude from $k_{th}$ sample in 1 channel pre-processed EEG emotion signal, the statistical features of *MTD* can be calculated as (Picard *et al.*, 2001),

$$\mu_x = \frac{1}{K}\sum_{k=1}^{K} X_k \tag{5.5}$$

where $K = n(X)$ denotes the cardinality of $X$.

B. **M**ean wavelet **A**pproximation coefficients with wavelet **E**ntropy ***(MAE)***

Given a wavelet approximation coefficient decomposition at the $a_{th}$ level, $A_k = \text{approx}(DWT(x(t); a, k)$, the mean of wavelet approximation coefficient $\mu_A$ is computed as follows,

$$\mu_A = \frac{1}{K}\sum_{k=1}^{K} A_k \tag{5.6}$$

where $K = n(A)$ denotes the cardinality of $A$. For the *MAE* features, both $\mu_A$ and *H(a)* are concatenated in an array and ordered by the EEG channel.

### 5.3.6 Fast Fourier Transform Power Spectral Density (*FFT PSD*)

Fast Fourier Transform Power Spectral Density (FFT PSD) is a feature in frequency domain. We used it as a comparison to our proposed features in time frequency domain using the wavelet PSD, wavelet entropy and MAE. The computation of the FFT PSD is done as follows:

Given an EEG signal *x(n)*, the FFT PSD of the EEG signal can be calculated as (Van Loan, 1992)

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{kn} \tag{5.7}$$

where $W_N = e^{-j(2\pi/N)}$ with $N$ = length of the EEG signal which varies based on the window size.

Using these Fourier coefficients, the relative FFT PSD can then be obtained using Eq. (2.12) and (2.13). The FFT PSD feature is then arranged in 5 bands formation.

### 5.3.7 Recognizing the EEG Emotion with SVM Classifier

The classification of the EEG emotion is conducted with SVM classifier using Radial Basis Function (*RBF*) kernel. Ensemble Rapid Centroid Estimation (*ERCE*) is implemented to obtain the optimum kernel radius RSVM from the training data with the SVM radius estimation (Yuwono *et al.*, Jun. 2014, 2014). The computation is conducted following the steps given in Chapter 2 using the Eq. (2.17) to (2.21). The SVM is also trained using Sequential Minimal Optimization (*SMO*) algorithm (Chang & Lin, 2011).

## 5.4 Discussion

In order to understand the results of the experiment, we must first look at the steps involved prior to discussing the experiment outcome. In this section, we will first show the analysis of segmented EEG signal using wavelet PSD spectrum to show the benefit of window segmentation. We will then perform the analysis of the comparative summary of the classification results with 7 features. The next step is to analyse the wavelet features in more depth, followed by comparison of MAE to other features. We will discuss the training and testing time with 7 features in regard to locating the window size that has shorter processing time, and then explain the process on how we finally arrived at recommending the optimal window selection. This section is divided into 6 subsections as follow:

a. Analysis of Segmented EEG Signals with Wavelet PSD Spectrum: Benefit of Window Segmentation

b. Comparative Summary of The Classification Results for 7 Features

c. Anaysis for The Wavelet Features

d. Comparison of MAE to other Features

e. Analysis of Training and Testing Time: Finding Window with Less Processing Time

f. Suggested optimal window selection in EEG emotion classification with 7 features for all emotions

### 5.4.1 Analysis of Segmented EEG Signals with Wavelet PSD Spectrum: Benefit of Window Segmentation

We begin the analysis with the demonstration of how the EEG emotion segmentation affects the features and what benefit can be obtained by implementing the segmentation procedure. For that purpose, one sample of preprocessed EEG signals from a randomly selected channel was taken from one arbitrary participant. The signals are visualized in their related 5 stacked bands wavelet PSD spectrum as shown in Fig. 5.6, showing the comparison between 60s window and 15 consecutive 4s windows.

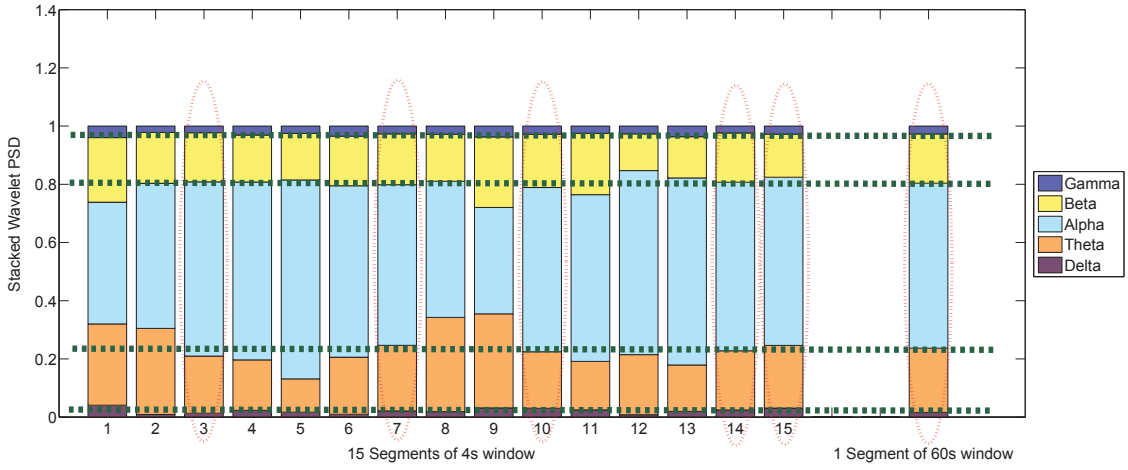



Figure 5.6: Visualization of 1 channel EEG signals in their related 5 stacked bands wavelet PSD spectrum of 60s window compared to 15 consecutive 4s window (marked by red dotted oval) showing the repetitional pattern between both windows. This is the benefit of the segmentation process.

The 60s segment in fact is the exact combination of 15 consecutive 4s segments. However the 60s segment only holds 1 stacked wavelet PSD spectrum whereas the 4s segments have 15 stacked spectrums. There are repetition of pattern in $3^{rd}$,$7^{th}$, $10^{th}$, $14^{th}$, and $15^{th}$ spectrum of 4s segment (marked by red dotted oval) which have the similar pattern as in the 60s segment.

This repetition of pattern reveals that segmented EEG signals with a certain window size carry a stationary pattern which is repeated in the segments with smaller window size. This becomes the benefit of the segmentation process as it will intensify the feature that represent the emotion state.

However, it should be noted that a non-regular pattern will also appear in the additional segments. Therefore, increasing the number of segmentations does not always provide an advantage, as it will depend on the size of the window segment selected. Consequently proper selection becomes very important to obtain an improvement of classification results that count on the segmented feature.

### 5.4.2 Comparative Summary of The Classification Results for 7 Features

The comparative summary for the classification results of 4 emotions using the 6 investigated features as well as the FFT PSD feature as a comparison are demonstrated in Fig. 5.7. The graphs in Fig. 5.7 show that MAE feature outperformed toward all other features including the comparative FFT PSD feature.

Although each emotion has a different range of accuracy, nonetheless they show similar progression. They start with very low accuracy at the widest window 60 sec, after that rise slightly between 30 to 20s window. The accuracies reached a peak between 12 to 4s window and then decreased again between 3 to 1s window.

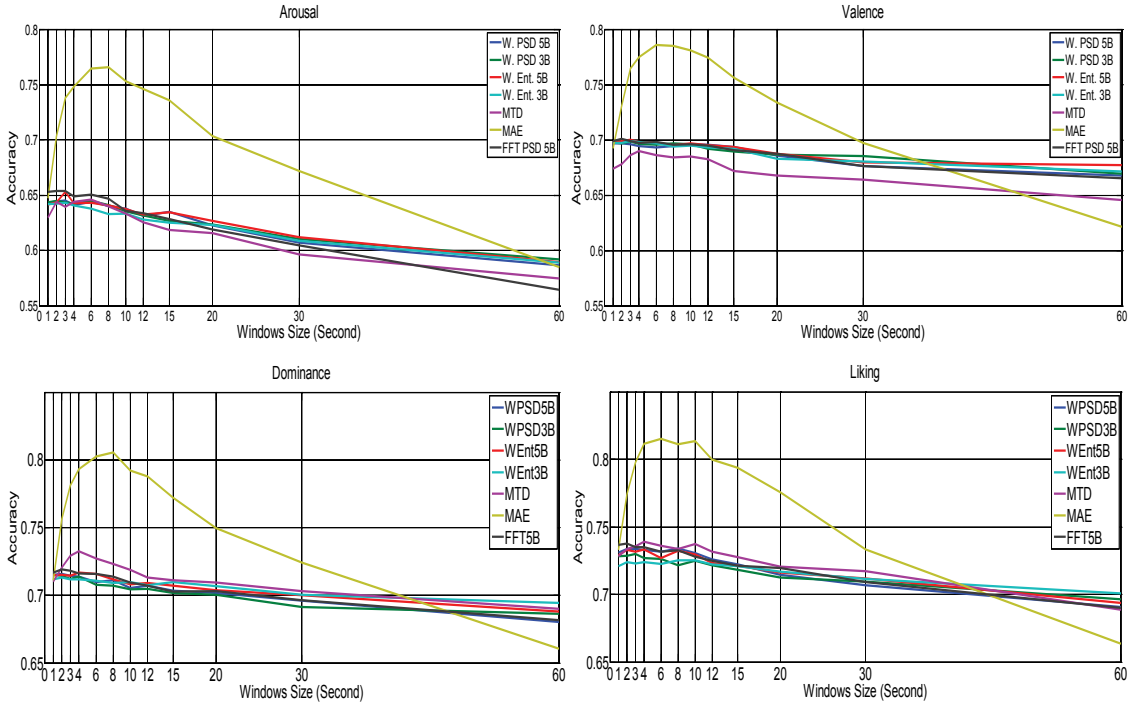



Figure 5.7: Comparative summary of classification results for 4 emotions with 7 features represented in line graphs.

By considering the similar trend of the 4 graphs that represent 4 emotions as illustrated in Fig. 5.7, we decided to simplify those 4 graphs for further investigation of the window selection method by taking the weighted average of 4 emotions for each of the 7 features and displaying them in one graph as shown in Fig 5.8.

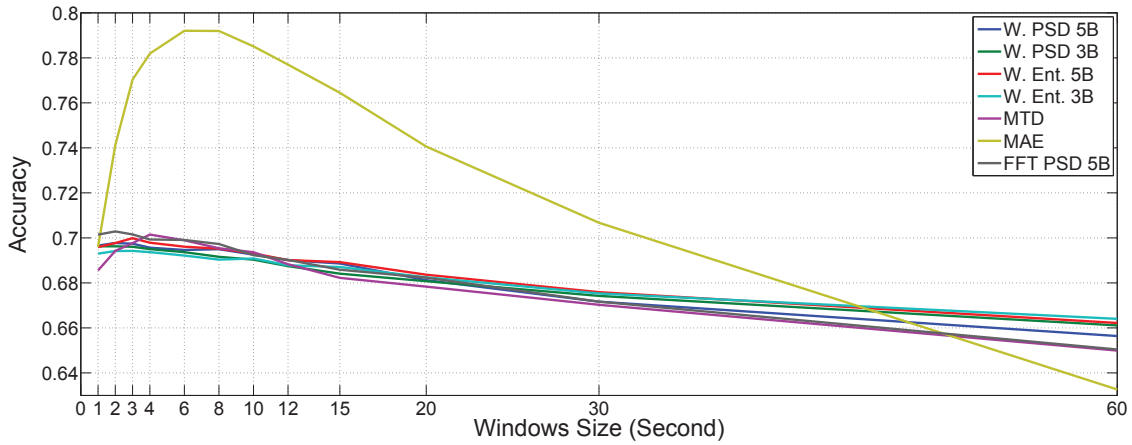



Figure 5.8: Simplified graph of 7 features using weighted average accuracy of 4 emotions.

### 5.4.3 Anaysis for The Wavelet Features

By simplifying the graphs, now it is easier for us to compare and analyse the classification results for each features using the simplified graph shown in Fig. 5.8. We begin the analysis by further investigating the relation between wavelet entropy and wavelet PSD.

The graph shows that accuracy of 5 bands wavelet entropy is slightly better than 5 bands wavelet PSD. The classification results indicate an increase within a smaller window. Significant rise started at 12 to 3s window and then reduced again. In addition, the maximum accuracy is 70%.

With the consideration that wavelet entropy has higher accuracy than wavelet PSD feature, the analysis is continued with the wavelet entropy by comparing the classification results of 5 and 3 bands wavelet entropy using the same graph in Fig. 5.8. Again, the graphs show slightly different accuracy between 5 and 3 bands wavelet entropy. Those 2 features maintain significant increase of accuracies between 12 and 3s windows.

To investigate whether the difference of accuracy between 3 and 5 bands wavelet entropy is significant or not, a t-test was conducted with the hypothesis that, "The difference of accuracy between 3 and 5 bands entropy features is not significant". The result is shown in Table 5.1 for valence emotion. The table provides *p*-value more than 0.05 for all window size which means that the difference is not significant. The *p-value* for the 3 other emotions are also consistent with no significant difference.

Table 5.1: t-test of 3 vs. 5 bands wavelet entropy features for valence emotion with hypothesis that, "The difference of accuracy between 3 and 5 bands entropy features is not significant" ($p > 0.05$)

| **Window (s)** | **Accuracy 3 bands (%±std)** | **Accuracy 5 bands (%±std)** | ***p*-value** |
|---|---|---|---|
| 60 sec | $67.1 \pm 3.0$ | $67.7 \pm 1.8$ | 0.18 |
| 30 sec | $68.1 \pm 1.7$ | $68.0 \pm 1.5$ | 0.59 |
| 20 sec | $69.3 \pm 1.3$ | $68.8 \pm 1.3$ | 0.95 |
| 15 sec | $69.2 \pm 1.1$ | $69.4 \pm 1.1$ | 0.27 |
| 12 sec | $69.3 \pm 0.8$ | $69.5 \pm 1.1$ | 0.22 |
| 10 sec | $69.5 \pm 0.9$ | $69.7 \pm 0.8$ | 0.19 |
| 8 sec | $69.4 \pm 0.8$ | $69.6 \pm 0.6$ | 0.24 |
| 6 sec | $69.7 \pm 0.8$ | $69.8 \pm 0.9$ | 0.33 |
| 4 sec | $69.8 \pm 0.6$ | $69.8 \pm 0.4$ | 0.48 |
| 3 sec | $69.7 \pm 0.5$ | $70.0 \pm 0.3$ | 0.14 |
| 2 sec | $69.7 \pm 0.4$ | $69.9 \pm 0.3$ | 0.61 |
| 1 sec | $69.7 \pm 0.3$ | $69.8 \pm 0.2$ | 0.07 |

### 5.4.4 Comparison of MAE to other Features

We continue our analysis with the MAE feature. We focus on the comparison of the MAE to other available features. We begin the analysis by observing in-depth the classification results of *MAE* compared to *MTD*, wavelet entropy, and *FFT PSD* features using the same Fig. 5.8.

The graph in Fig. 5.8 shows that *MAE* is superior to those other features. The accuracy of *MAE* is leading by about 10% at the maximum compared to other features. The accuracy of MAE reaches a peak at 79.2%.

To complete the analysis, we also give more attention to the comparison of our proposed

*MAE* feature with *FFT PSD* feature which has been selected to represent non-wavelet feature. We take the t-test with the hypothesis that, "The difference of accuracy between *MAE* and *FFT PSD* feature is significant" and present the results in Table 5.2 for the liking emotion. The results of the *p-value* supports the hypothesis that the difference is signifcant for all 4 emotions between 2 and 30s.

Table 5.2: t-test of *MAE* vs. *FFT PSD* feature for liking emotion with hypothesis that, "The difference of accuracy between *MAE* and *FFT PSD* features is significant" ($p < 0.05$)

| **Window (s)** | **Accuracy MAE (%±std)** | **Accuracy FFT PSD (%±std)** | ***p*-value** |
|---|---|---|---|
| 60 sec | 66.4 ± 2.7 | 69.0 ± 2.8 | 1.00 |
| 30 sec | 73.4 ± 2.0 | 70.9 ± 1.5 | 9.90 e-08 |
| 20 sec | 77.6 ± 1.6 | 72.0 ± 0.9 | 3.83 e-24 |
| 15 sec | 79.4 ± 1.6 | 72.2 ± 1.1 | 5.32 e-22 |
| 12 sec | 80.0 ± 1.2 | 72.5 ± 1.1 | 1.09 e-20 |
| 10 sec | 81.4 ± 1.4 | 72.8 ± 1.1 | 6.22 e-22 |
| 8 sec | 81.1 ± 1.1 | 73.3 ± 0.7 | 1.47 e-24 |
| 6 sec | 81.5 ± 0.7 | 73.2 ± 0.7 | 1.21 e-28 |
| 4 sec | 81.2 ± 0.7 | 73.5 ± 0.5 | 9.25 e-32 |
| 3 sec | 79.8 ± 0.5 | 73.5 ± 0.6 | 1.73 e-28 |
| 2 sec | 77.4 ± 0.4 | 73.8 ± 0.4 | 4.43 e-26 |
| 1 sec | 73.3 ± 0.2 | 73.7 ± 0.2 | 1.00 |

### 5.4.5 Analysis of Training and Testing Time: Finding Window with Less Processing Time

In order to provide more beneficial information to support the proposed method, we conduct additional experiments to obtain the information on the processing time for both training and testing for all 7 features using SVM classifier. The results are presented in 2 line graphs as shown in Fig. 5.9. The graphs were generated using the weighted average of training and testing time of 4 emotions for 7 features.

The first graph illustrates the training time, while the second represents the testing time. Both graphs demonstrate a similar trend for all 7 features. The first graph shows that the training time of EEG-emotion signal at 4s to 60s window is less than 100 seconds. On the

contrary, at 3s to 1s window the training time is much higher than 100s even rising up to 800s.

The testing time shows a similar trend to the training time, only that the time required is less (<< than 1s) and is not really significant compared to the training time. By considering this event, the accommodation of 4s window or above will keep the training and testing time at the shorter duration.

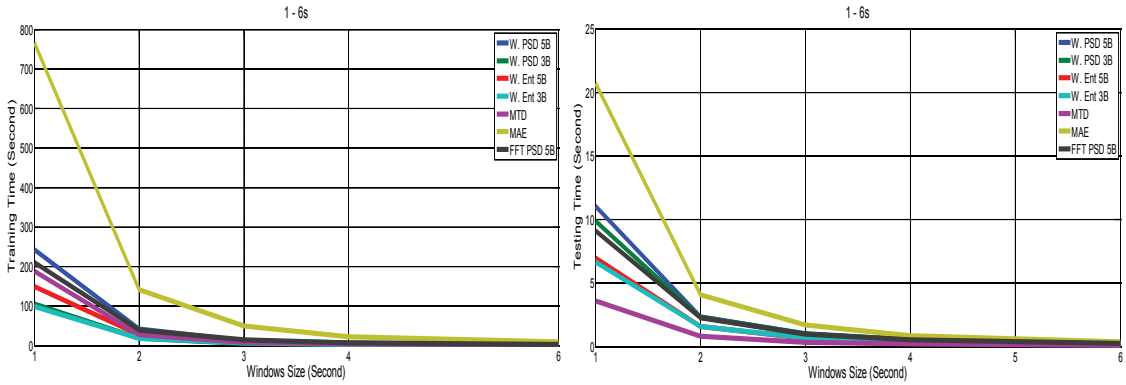



Figure 5.9: Graphichal reprentation of optimal training and testing time in EEG-emotion recognition for 7 features using weighted average of 4 emotions.

### 5.4.6 Suggested optimal window selection (OWS) in EEG emotion recognition using 7 features for all emotions

We come to the suggestion of the optimal window selection (OWS) to be used in EEG emotion recognition using 7 different features. Based on the analysis of weighted average accuracy and the training and testing time for all features, we developed a colour graph image which we named Optimal Window Selection (OWS) colour graph as illustrated in Fig. 5.10.

The OWS color graph represents the weighted average accuracy of 4 emotions for all 7 features. This OWS colour graph can be divided into 3 main regions. Using those 3 regions, the optimal window can be identified as follows:

1. The first region between 60 and 15s keeps the lowest classification results with slightly faster training and testing time.

2. The second region between 12 and 4s holds the highest classification results with shorter training and testing time. Also in this region, the classification results are improved between 3% to 15% compared to the lowest region which can be achieved using any of the features.

3. The third region between 3 and 1s, although some features have slightly higher accuracy, however, this region suffers from very long delay due to longer training and testing time.

Therefore, our recommendation for Optimal Window Selection (OWS) in EEG emotion classification is the second region between 4 to 12s (encircled in the green dotted box). Within this region, we can obtain highest classification results and at the same time achieve shorter training and testing time for all 7 features. Also in this region, the MAE demonstrates the highest classification results among other features with accuracy as high as 79.2% at 6s window.

| | W. PSD 5B | W. PSD 3B | W. Ent. 5B | W. Ent. 3B | MTD | MAE | FFT PSD 5B |
|---|---|---|---|---|---|---|---|
| 1s | 69.6579 | 69.6274 | 69.5902 | 69.2997 | 68.5549 | 69.5964 | 70.1443 |
| 2s | 69.7719 | 69.6269 | 69.7716 | 69.4175 | 69.4222 | 74.1189 | 70.2863 |
| 3s | 69.7429 | 69.6042 | 69.9862 | 69.4208 | 69.7653 | 77.0313 | 70.1546 |
| 4s | 69.566 | 69.5005 | 69.7868 | 69.3679 | 70.1471 | 78.1899 | 69.9301 |
| 6s | 69.4632 | 69.3672 | 69.6097 | 69.2143 | 69.894 | 79.2035 | 69.91 |
| 8s | 69.4968 | 69.1606 | 69.5173 | 69.0337 | 69.5392 | 79.1932 | 69.728 |
| 10s | 69.2572 | 69.0303 | 69.3093 | 69.083 | 69.3645 | 78.5082 | 69.2324 |
| 12s | 69.0088 | 68.7453 | 69.0148 | 68.7795 | 68.8298 | 77.6991 | 69.0223 |
| 15s | 68.8616 | 68.4096 | 68.923 | 68.7044 | 68.2245 | 76.45 | 68.5863 |
| 20s | 68.1225 | 68.0766 | 68.3631 | 68.2469 | 67.8336 | 74.0575 | 68.2329 |
| 30s | 67.167 | 67.4237 | 67.5893 | 67.5298 | 67.0275 | 70.6715 | 67.167 |
| 60s | 65.6399 | 66.1124 | 66.2202 | 66.4025 | 64.9926 | 63.2626 | 65.0409 |

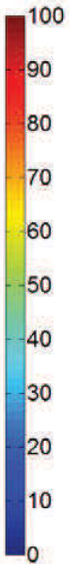



Figure 5.10: Suggested optimal window selection (OWS) color graph. The numbers in the graph reflect the weighted average accuracy of 4 emotions for each 7 features. The color heatmap expresses various levels of accuracy yield by window size between 1s and 60s for each of the 7 features, where darker colors represent higher accuracy. Referring to the graph, the optimal window can be identified.

## 5.5 Summary

Implementation of Optimal Window Selection (OWS) method in Electroencephalography (EEG) emotion recognition was performed using Support Vector Machine (SVM) to identify High/Low state of 4 dimensional plane emotions, namely: Arousal, Valence, Dominance, and Liking.

Combination of **M**ean wavelet **A**pproximation coefficients and wavelet **E**ntropy (MAE), naive **M**ean of the **T**ime **D**omain signal (MTD), wavelet Power Spectral Density (PSD) and wavelet entropy were explored as the features.

The investigation reveals that wavelet PSD and wavelet entropy features provide equal results. Frequency bands reduction can be implemented to both wavelet features using only 3 out of 5 bands which also give equal results.

Further comparison to other non wavelet features using Fast Fourier Transform Power Spectral Density (FFT PSD) reveals that the proposed *MAE* feature is outperformed.

Using optimal window selection (OWS) color graph, the optimal window can be identified. The OWS color graph suggests that the optimal window is between 4 and 12s. In this region, an increase of accuracy between 3% and 15% can be obtained depend on the features used. Also inside the region, the training and testing time can be mantained shorter. This optimal window can be applied in general to all 4 emotions: Arousal, Valence, Dominance, and Liking.

In detail, the advantage of OWS is the replication of pseudo-stationary signal within the selcted optimal window that has a desirable statistical property. This statistical property improves the classification accuracy regardless of the features used.

Further work includes the exploration of the OWS with other classifiers as well as deep belief network for optimization, and also implementation of dimensional reduction to *MAE* features.

# Chapter 6

# Conclusions and Future Direction

## 6.1 Discussion

Communication in nonverbal language involve many different aspects that have a significant role in communicating the feeling, attitudes and emotions and manifest as physical appearance such as facial expression and physiological signal including Electroencephalography (EEG) signal or brain wave (Sutter, 2010; Foley & Gentile, 2010).

The availability of methods to measure the emotions with various different perspectives open the possibilities to recognize diverse categories of emotion which can be implemented in many applications to help people better understand the human emotion in relation to another person or in dealing with certain situations and provide solutions to the problem related to them (Caicedo & Beuzekom, 2006). In this thesis, 2 appealing examples of emotion recognition method using the facial expression and EEG emotion are examined.

Facial expression recognition involves various frameworks along with face detection and recognition using a geometric based or appearance based features extraction mechanism with each having its advantages and disadvantages (Shan *et al.*, 2009; Donato *et al.*, 1999). In the case of appearance based method, a sophisticated feature known as Histograms of Oriented Gradient (HOG) (Dalal & Triggs, 2005) has been implemented for face recognition with very successful results which offers a potential solution for facial expression

recognition.

The challenges in facial emotion recognition system is related to feature, feature dimension, and classification processing time which can be elaborated in one sentence that is, 'how to obtain reliable features in facial emotion recognition using an appearance based method that has very low dimension and consumes less training and testing time'.

One important factor that should be considered in developing a facial expression recognition system is the availability of a common database that satisfies the various requirements to be used as a benchmark for the research experiment. The Extended Cohn–Kanade (CK+) database (Kanade *et al.*, 2000) is a good alternative for that purpose.

On the other hand, electroencephalography (EEG) emotion recognition requires many components and processes to collect the EEG signals, starting with the equipment preparation, signal processing and conditioning, feature extraction, number of electrodes and their positioning, as well as frequency bands allocation and selection (Jenke *et al.*, 2014b).

Another important parameter that needs attention is the time-varying and non-stationary characteristics of the EEG signals (Picard *et al.*, 2001). This issue can be handled with window segmentation strategy with proper selection of window size to improve the performance (Kaplan *et al.*, 2005).

A sophisticated feature extraction method in EEG emotion recognition is using Discrete Wavelet Transform (DWT) to collect relative energy and entropy with the advantage of time-frequency localization, multiscale zooming, and multirate filtering. These advantages make it possible to extract the useful information from non-stationary signals (Rosso *et al.*, 2001b).

The necessity for a reliable database in EEG emotion recognition is also an issue. Only a few EEG emotion databases meet the expectations, and the Database for Emotion Analysis Using Physiological Signals (DEAP)(Koelstra *et al.*, 2012) becomes the most favored.

The challenges in EEG emotion recognition system is concerned with the selection of features that is able to represent both time and frequency domain, the preference of frequency components selected from various subbands, and the positioning of EEG channel

at certain positions with a definite number of channels. More specifically, the challenge is, 'What feature is to be used to collect information in time and frequency domain allocated at specific EEG frequency subbands with finite number of EEG channels.

Another critical challenge to be settled is, 'How to resolve the non-stationary characteristics of the EEG signals by obtaining optimum window size to maximize the information extraction with a specific window size that needs to be *just right*: not too short or too long'.

An additional problem that needs solving is the optimization of classification process using Support Vector Machine (SVM) classifier. A solution using a particle swarm ensemble clustering algorithm called the Ensemble Rapid Centroid Estimation (ERCE) algorithm (Yuwono *et al.*, Jun. 2014, 2014) is offered in this thesis.

## 6.2 Conclusion

This thesis contributes a number of solutions to answer the problems as described in the discussion above. We offer the solution for each challenge by developing methods and strategies to overcome the problems. The solutions are summarized in the following paragraph.

**1. Challenges in Facial Expression Recognition**

To answer the challenge for the need of a reliable feature to be used in Facial Expression Recognition we developed an appearance based feature which we called Edge of Histogram Oriented Gradient (E-HOG) (Candra *et al.*, 2016) which has been implemented to classify the dataset acquired from the Extended Cohn-Kanade dataset (Kanade *et al.*, 2000) with classification result as up to 96.4% that makes this proposed (E-HOG) a very high potential feature for facial expression recognition.

The challenge in feature dimension reduction and time efficiency in classification processing for facial expression recognition is answered with an improved E–HOG with a very low dimension called Reduced dimension of Edge-Histogram of Oriented Gradients

(RED E-HOG) (Candra *et al.*, (Submitted)b). The dimension of the feature is reduced from thousands to tens using a combination of PCA and LDA as the dimensionality reduction technique. Using 10 dimensions of RED E-HOG trained with multi-class SVM, 7 facial expressions have been successfully recognized with the accuracy of 99.6%. The comparison toward other features reveals that RED E-HOG Face features outperforms.

This result is supported by the analysis of scatter plots for RED E–HOG features which show that RED E–HOG feature provides distinct boundary and distance for the represented facial expressions so that they can be recognized easily by the SVM classifier. Significant improvement in training and testing time was also obtained by RED E-HOG compared to ORI E-HOG. The results indicate that RED E-HOG feature (specifically RED E-HOG Face) is feasible and efficient for facial expression recognition.

**2. Challenges in EEG Emotion Recognition**

In dealing with the challenge in EEG emotion recognition we utilized Discrete Wavelet Transform (DWT) to extract wavelet feature that preserved time and frequency information represented in relative energy and entropy. We implemented the wavelet features to a discrete EEG-emotion recognition system and successfully classified 4 discrete emotions of happy, sad, angry, and relaxed from arousal–valence dimensional plane using the DEAP dataset (Koelstra *et al.*, 2012).

We also discovered that using only 3 frequency subbbands of alpha, beta, and gamma is adequate for discrete EEG-emotion recognition system, while the delta and theta subbands can be eliminated.

Besides, we also recommend the reduction of channels from 32 to 18 using only EEG channels of Fp1, Fp2, AF3, AF4, F3, F4, F7, F8, FC5, FC6, T7, T8, P7, P8, P3, P4, O1, O2 with 3 subbands wavelet entropy features for our discrete EEG-emotion recognition method, giving average sensitivity and specificity of 77.4% and 69.1%, respectively.

In summary, we provide an answer for feature with time and frequency information using the wavelet feature, together with frequency band reduction from 5 to 3 and channel reduction from 32 to 18.

To overcome the non-stationary characteristics of the EEG signals, we introduced Optimal Window Selection (OWS) method in EEG emotion recognition to identify High/Low state of 4 dimensional plane emotions, namely: of Arousal, Valence, Dominance, and Liking classified with Support Vector Machine (SVM). Using the OWS colour graph, the optimal window can be identified.

Based on the OWS colour graph we suggest that the optimal window can be allocated at 4 to 12s. Using this region, the classification results increase about 3% and 15% depending on the features used. The training and testing time can also be made shorter within the region. The optimal window is applicable to all Arousal, Valence, Dominance, and Liking emotions.

At a more advanced level, the OWS gives an advantage of the replication of pseudo - stationary signal that secures a desirable statistical property. Within the selected optimal window this statistical property improves the classification accuracy regardless of the features used.

In addition, we explored 6 various wavelet features for the OWS investigation, namely: combination of **M**ean wavelet **A**pproximation coefficients and wavelet **E**ntropy (MAE), naive **M**ean of the **T**ime **D**omain signal (MTD), wavelet Power Spectral Density (PSD), and wavelet entropy. The investigation reveals that wavelet PSD and wavelet entropy have equal results either using 3 or 5 frequency subbands. This means that subbands reduction can be implemented to both wavelet features.

The comparison of *MAE* with Fast Fourier Transform Power Spectral Density (FFT PSD) as a non-wavelet feature affirms that the proposed *MAE* feature outperforms.

**3. Support Vector Machine Optimization**

In all the experiments, the classification using optimized SVM with a RBF kernel trained with SMO algorithm provides reliable classification results which shows that the estimation of RBF kernel parameters using a particle swarm ensemble clustering algorithm called the Ensemble Rapid Centroid Estimation (ERCE) algorithm (Yuwono *et al.*, 2014) is effective. The advantage of the method is related to the capability of the algorithm

to estimate the number of clusters directly from the data using swarm intelligence and ensemble aggregation.

## 6.3 Limitation of The Research

In our discrete EEG emotion recognition method, the investigation using ROC curve and NMI method reveals that the method has a limitation of transient patterns and quasi-stable states that may exist in the EEG that has been overlooked in both energy and entropy formulations. Having this in mind, it is worth considering incorporating EEG microstates and/or brain connectivity features.

MAE feature still needs improvement in terms of its high dimensionality, which requires the implementation of dimensionality reduction to improve the dimension and classification processing time, while also considering the reduction of subbands and channels.

Both facial expression and EEG emotion recognition methods developed in this thesis depend on publicly available databases which have a few limitations such as separated databases which make the comparison or combination of the 2 methods become impractical. In addition, the availability of a real time EEG instrument which is still counted as a high end device, makes the implementation for general application in real time processing become expensive and inefficient.

Also, EEG recording requires very subtle handling of the recording process to obtain a noise free EEG signal which also mean less muscle activity especially muscles related to the face, eyes, and mouth. In contrast, facial expression needs expressive visualization of the emotion which also means very active facial muscles that create noise for the EEG. Therefore, to record the facial expression together with the EEG signal will need a powerful preprocessing algorithm for artifact removal.

Besides, the combination of both RED E–HOG and MAE in real time applications requires collection of a new database that meets the expectations such as synchronized facial expressions and EEG signals, noise free EEG signals, and expressive facial appearance, which requires complicated equipment, proper environmental settings, and an adequate

number of participants.

## 6.4 Future Directions

The methods developed for both facial expression and EEG emotion and their theoretical analysis are still undergoing improvement to tackle the aforementioned limitation which is still far from nishing. One major goal includes the implementation of other sophisticated classiers such as deep belief network which has shown promising results among other implemented classication algorithms.

The improvements of both RED EHOG for facial expression and MAE features for EEG emotion to be implemented for real time application and the comparison with the cutting-edge algorithms will be the next target of exploration. These developments will need the collection of new and more reliable datasets that can be used for these two recognition methods. The availability of such datasets will open the possibility of combining facial expressions with EEG emotions as a hybrid emotional recognition method.

The combination of both facial expression and EEG emotion recognition system will create a powerful tool for emotion recognition in a way that the EEG emotion recognition will validate the facial expression recognition and reduce its weaknesses relate to falsication of facial expressions.

In addition, scientists now have more concern in the development of applications that make more contribution to the physical and mental wellbeing across all cultures and nations. Therefore, future directions include the implementation of the facial expression and EEG emotion recognition method in biomedical engineering to identify people associated with specific emotional disorders.

# Appendix

Matlab Code for The Experiment

Code 1

```
clear all;
close all;
% Take numbers of image to be processed
for i=1:150
f= dir('.../01Anger/*.png');
ridx = i
%Read and dsiplay image
I=imread(f(ridx).name);
try
%Detect and display face
faceDetector=vision.CascadeObjectDetector('FrontalFaceLBP');
bboxes = step(faceDetector, I);
%If bboxes more than one choose the last
[c]=size(bboxes);
if c(1,1) > 1
    n=c(1,1);
    for in=1:n
    bboxes1 =bboxes(in,:);
    end
    bboxes=bboxes1;
```

```
    else
    bboxes=bboxes;
end
IFaces = insertObjectAnnotation(I, 'rectangle', ...
 bboxes, 'Face');
%Resize image based on the face detected and round up
% the detected window to 212x212
[x]=bboxes;
bboxes2=round(200/x(1,3)*bboxes);
bboxes3=[bboxes2(1,1), bboxes2(1,2),212,212];
I2=imresize(I,200/x(1,3));
IFaces2 = insertObjectAnnotation(I2, 'rectangle', ...
bboxes3, 'Face');
%Preparing image for mouth eyes and mouth detection
Faces=I2(bboxes3(1,2):bboxes3(1,2)+ bboxes3(1,4), ...
bboxes3(1,1):bboxes3(1,1)+bboxes3(1,3));
Faces3=I2(bboxes3(1,2)+22:bboxes3(1,2) ...
+ bboxes3(1,4)-11,bboxes3(1,1)+30:bboxes3(1,1) ...
+bboxes3(1,3)-43);
size (Faces3);
%Face Canny
    FaceCanny = im2single(edge(Faces3,'canny'));
  %Call Eye detection function set parameter
eyeDetector = vision.CascadeObjectDetector('EyePairBig');
    eyeDetector.MinSize = [11 45];
    eyeDetector.MergeThreshold = 1;
   %Call Mouth detection function and set parameter
    mouthDetector = vision.CascadeObjectDetector('Mouth');
    mouthDetector.MinSize = [15 25];
    mouthDetector.MergeThreshold = 16;
  %Detect Eye pairs
```

```
Faces1=Faces(40:125,1:212);
bboxEye = step(eyeDetector, Faces1); % Detect Eye Pair
   %Detect mouth
Faces2=Faces(130:212,1:212);
 bboxMouth = step(mouthDetector, Faces2); % Detect mouth
    %Display result of eyes and mouth detection
bboxEye = [bboxEye(1,1), bboxEye(1,2),128,40];
bboxMouth = [bboxMouth(1,1), bboxMouth(1,2),80,36];
  % Crop image at the eye pair
  Eyeim=Faces1(bboxEye(1,2):bboxEye(1,2) ...
  + bboxEye(1,4),bboxEye(1,1):bboxEye(1,1)+bboxEye(1,3));
  % Eye Canny
  EdEyeim = im2single(edge(Eyeim,'canny'));
   % Crop image at the mouth
  Mouthim=Faces2(bboxMouth(1,2):bboxMouth(1,2) ...
  +bboxMouth(1,4),bboxMouth(1,1):bboxMouth(1,1) ...
  +bboxMouth(1,3));
   % Mouth Canny
  EdMouthim = im2single(edge(Mouthim,'canny'));
catch
continue
end
% ------------------------------------------------
%          Compute HOG features
% ------------------------------------------------
% HOG Standard features
% Face HOG with Canny
cellSize = 4 ;
hogFaceCanny = vl_hog(FaceCanny, cellSize, 'verbose') ;
FaceimhogCanny = vl_hog('render', hogFaceCanny, 'verbose') ;
%Face HOG without canny
```

```
hogFace = vl_hog(im2single(Faces3), cellSize, 'verbose') ;
Faceimhog = vl_hog('render', hogFace, 'verbose') ;
% Eye HOG with canny
hogEyeCanny = vl_hog(EdEyeim, cellSize, 'verbose') ;
EyeimhogCanny = vl_hog('render', hogEyeCanny, 'verbose') ;
%Eye HOG without canny
hogEye = vl_hog(im2single(Eyeim), cellSize, 'verbose') ;
Eyeimhog = vl_hog('render', hogEye, 'verbose') ;
%Mouth HOG with canny
hogMouthCanny = vl_hog(EdMouthim, cellSize, 'verbose') ;
MouthimhogCanny = vl_hog('render', hogMouthCanny, ...
'verbose') ;
%Mouth HOG without canny
hogMouth = vl_hog(im2single(Mouthim), cellSize, ...
'verbose') ;
Mouthimhog = vl_hog('render', hogMouth, 'verbose') ;
%Create HOG array
%With Canny
FaceimhogmatrixCanny(:,i) = hogFaceCanny(:);
EyeimhogmatrixCanny(:,i)=hogEyeCanny(:);
MouthimhogmatrixCanny(:,i)=hogMouthCanny(:);
%Without Canny
Faceimhogmatrix(:,i) = hogFace(:);
Eyeimhogmatrix(:,i)=hogEye(:);
Mouthimhogmatrix(:,i)=hogMouth(:);
% Combine eyes and mouth with Canny
EyesandMouthhogmatrixCanny(:,i)= [hogEyeCanny(:); ...
 hogMouthCanny(:)];
% Combine eyes and mouth without Canny
EyesandMouthhogmatrix(:,i)= [hogEye(:); hogMouth(:)];
disp( ['Chosen file is: ' f(ridx).name] );
```

Appendix A.

```
lbl=1 %Anger
namafileEyeim(i)= cellstr(f(ridx).name);
label(i)= lbl;
%Create file
FileName=[namafileEyeim];
%Save file
save('CKAngernew150.mat','FileName');
save('hogmatrixCKAngernew150.mat', 'label', ...
'FaceimhogmatrixCanny', 'EyeimhogmatrixCanny', ...
 'MouthimhogmatrixCanny','EyesandMouthhogmatrixCanny', ...
'Faceimhogmatrix','Eyeimhogmatrix', 'Mouthimhogmatrix', ...
'EyesandMouthhogmatrix');
end
```

Code 2

```
clear;
clc;
close all;
%addpath('../code');
load 'EHOGFace.mat';
%d=100;
%d=500;
d=1000;
data=X';
%% Gaussian kernel PCA
DIST=distanceMatrix(data);
DIST(DIST==0)=inf;
DIST=min(DIST);
para=5*mean(DIST);
disp('Performing Gaussian kernel PCA...');
[Y3, eigVector]=kPCA(data,d,'gaussian',para);
```

Appendix A.

```
X=Y3';
save('kpcafacenewdim1000.mat','X','y');
```

Code 3

```
clear;
clc;
in_data = load('pcafaceehog100.mat');
    model = lda( in_data, 10 );
    out_data = linproj( in_data, model);
    figure; ppatterns(out_data);

    X=real(out_data.X);
    y=out_data.y;

    save('ldapcafacenewdim10.mat','X','y');
```

Code 4

```
%%
clear all
close all
% file = MTDPerson{n}.mat
n = 1;
N = 32;
varianceTarget = 97.5;
filename = 'MAEPerson';
getData = @(n) load(sprintf('%s%d.mat', filename, n));
getResults = @(n) load(sprintf ...
('Result%s%d.mat', filename, n));
pcaTransform = @(X, pcaMatrix, mu, nPC) ...
 bsxfun(@minus, X, mu) * pcaMatrix(:, 1:nPC);
%% getting the pca transformation matrix and
```

Appendix A.

```
% approximate mean
Xall = [];
for i = 1:N
    data = getData(i);
    %y = categorical(data.ya)';
    Xi = data.X';
    Xall = cat(1, Xall, Xi);
end
% compute the principal component transformation matrix
[coeff,score,latent,tsquared,explained,mu] = pca(Xall);
%[explained] = pca(Xall);
varianceExplained = cumsum(explained);
plot(varianceExplained)
nPC = 1 + find(diff(varianceExplained > varianceTarget) ...
 == 1);
disp(sprintf('the minimum number of principal components ...
required to encode %.2f%% of input variance is %d \n', ...
varianceTarget, nPC))
%% fit a logistic regression on each person and
%save the results
for i = 1:32
    data = getData(i);
    X = data.X';
    projection = pcaTransform(X, coeff, mu, nPC);
    y = [data.ya', data.yd', data.yl', data.yv'] - 1;
    y_logit = log(eps+y) - log(1-(y-eps));
    B = zeros(1 + nPC, 4);
    for j = 1:4
        b = ridge(y_logit(:, j), projection, 0.00001, 0);
        B(:, j) = b;
    end
```

```
    bias = ones(size(X,1), 1);
    y_hat = logsig(cat(2, bias, projection) * B);
    cp = classperf(y, round(y_hat));
    result.data = data;
    result.mappingMatrix = B;
    result.classperf = cp;
    save(sprintf('Result%s%d.mat', filename, i), ...
    '-struct', 'result')
end
%% for each results get the B matrix, concatenate
%them for clustering
BMatrices = zeros((nPC + 1)*4, 32);
BMatricesRaw = {};
accuracy = zeros(1, 32);
balancedAccuracy = zeros(1, 32);
for i = 1:32
    result = getResults(i);
    BMatricesRaw{i} = result.mappingMatrix;
    BMatrices(:, i) = result.mappingMatrix(:);
    cp = result.classperf;
    accuracy(i) = cp.Sensitivity * cp.prevalence ...
    + cp.specificity * (1 - cp.prevalence);
    balancedAccuracy(i) = (cp.Sensitivity ...
    + cp.specificity)/2;
end
%% cluster the mapping matrices based on how
% closely correlated they are
figure(1)
hist(accuracy)
% plot the dendrogram
figure(2)
```

Appendix A.

```
Z = linkage(tanh(BMatrices)', 'ward', 'euclidean');
th = 18;
dendrogram(Z, 32, 'colorthreshold', th)
labels = cluster(Z,'Cutoff', th, 'criterion', 'distance');
list = [13, 18, 24, 4, 12, 21, 2, 6, 15];
figure(4);
for i = 1:length(list)
    subplot(1,length(list),i); imagesc(BMatricesRaw ...
    {list(i)}, [-150 150]);
    set(gca, 'xtick', 1:4, 'xticklabel', ...
    {'a', 'd', 'l', 'v'})
    colormap hot
end
```

Code 5

```
close all;
clear all;
clc;
for y=1:9 %Number of participant
load(['s0' num2str(y) '.mat']);
for k=1:40  %Number of video
   for i=1:32 %Number of channel
    close all
data1{i}=reshape(data(k,i,1:8064),1,8064);
rdxn = data1{i};%;(1:8064);
length(rdxn)
N=length(rdxn);
n2=0:(N-1)/2; % for positiveFFT function
fs=128;
[YfreqDomainL,frequencyRange] = positiveFFT(rdxn,128);
% FFT is symmetric, throw away
```

Appendix A.

```
length(frequencyRange)
xaxis=n2*fs/N;                    % second half
length(xaxis)
K=round(N/fs)
%sampling rate 128hz, signal from 0-64hz
%d1 =            33-64hz gamma
%a1 = 0-32hz
%d2 =            17-32hz -- beta
%a2 = 0-16hz
%d3 =            9-16hz -- alpha
%a3 = 0-8hz
%d4 =             5-8hz -- theta
%a4 =0-4hz
%d5 =             3-4hz --delta
%a5 =0-2hz --delta
%gamma 33-64
%alpha 8-13
%beta 13-30
%delta 1-4
%theta 4-8
xdelta=xaxis(1:(K*4));
length(xdelta)
xtheta=xaxis((K*4+1):(K*8));
length(xtheta)
xalpha=xaxis((K*8+1):(K*13));
length(xalpha)
xbeta=xaxis((K*13+1):(K*30));
length(xbeta)
xgamma=xaxis((K*30+1):(K*64));
length(xgamma)
MXL = abs(YfreqDomainL);  % Take magnitude
```

Appendix A.

```
MX1L = 2*MXL; % Multiply by 2 to take into
length(MX1L)
figure;
%plot(n2*fs/N,MX1L);
MX1Ldelta = MX1L((1:(K*4)));
MX1Ltheta = MX1L((K*4+1):(K*8));
MX1Lalpha = MX1L((K*8+1):(K*13));
MX1Lbeta = MX1L((K*13+1):(K*30));
MX1Lgamma = MX1L((K*30+1):(K*64));
plot(xaxis,MX1L);
figure,plot(xdelta,MX1Ldelta);
figure,plot(xtheta,MX1Ltheta);
figure,plot(xalpha,MX1Lalpha);
figure,plot(xbeta,MX1Lbeta);
figure,plot(xgamma,MX1Lgamma);
% energy trapz
e_deltap=trapz((MX1Ldelta));
e_thetap=trapz((MX1Ltheta));
e_alphap=trapz((MX1Lalpha));
e_betap=trapz((MX1Lbeta));
e_gammap=trapz((MX1Lgamma));
sumenergyp=e_deltap+e_thetap+e_alphap+e_betap+e_gammap;
edeltap=e_deltap/sumenergyp;
ethetap=e_thetap/sumenergyp;
ealphap=e_alphap/sumenergyp;
ebetap=e_betap/sumenergyp;
egammap=e_gammap/sumenergyp;
Efftp=[edeltap,ethetap,ealphap,ebetap,egammap]
figure;
bar(Efftp);
Enfftrap(i,:)=Efftp;
```

Appendix A.

```
% Energy sum
e_delta=sum(abs((MX1Ldelta)).^2);
e_theta=sum(abs((MX1Ltheta)).^2);
e_alpha=sum(abs((MX1Lalpha)).^2);
e_beta=sum(abs((MX1Lbeta)).^2);
e_gamma=sum(abs((MX1Lgamma)).^2);
sumenergy=e_delta+e_theta+e_alpha+e_beta+e_gamma;
edelta=e_delta/sumenergy;
etheta=e_theta/sumenergy;
ealpha=e_alpha/sumenergy;
ebeta=e_beta/sumenergy;
egamma=e_gamma/sumenergy;
Efft=[edelta,etheta,ealpha,ebeta,egamma]
figure;
bar(Efft);
Enfft(i,:)=Efft;
[y_delta,y_theta,y_alpha,y_beta,y_gamma] ...
=wavelet_energy_decompose2modify(rdxn);
E=[y_delta,y_theta,y_alpha,y_beta,y_gamma]
figure;
bar(E);
Ewavelet(i,:)=E;
end
Enfftraps=reshape(Enfftrap,1,32*5);
Enffts=reshape(Enfft,1,32*5);
Ewavelets=reshape(Ewavelet,1,32*5);
persondir=sprintf('Person%d',y);
mkdir(persondir);
%
cd(persondir)
path=pwd
```

Appendix A.

```
%
cd ..
%
fname = [ 'person' num2str(y) 'video' num2str(k) ...
 'allchannel1mins' '.mat' ];
%%filename = get(handles.fname,'String');
fileandfolder=(fullfile(path,fname));
save(fileandfolder,'Enfftrap', 'Enfft','Ewavelet',...
'Enfftraps', 'Enffts','Ewavelets');
end
end
```

Code 6

```
clc;
clear all
for n=1:32 %Number of participant
 for m=1:2
for x=1:40 %Number of video
 load(['person' num2str(n) 'video' num2str(x) ...
 'allchannel130sec.mat']);
    len = size(reshape(Ewavelet(m,:,:),32,5));
len1 = len(1);
len2 = len(2);
Ewavelets=reshape(Ewavelet(m,:,:),32,5);
for i=1:len1
    if i==1
        Ewavelet_new= (Ewavelets(1,1:5));
    else
        Ewavelet_new=horzcat(Ewavelet_new,Ewavelets(i,1:5));
    end
end
```

Appendix A.

```
% Ewavelet_new=  Ewavelet_new';
Ewavelet_new1=log(Ewavelet_new);
Ewavelet_new2= Ewavelet_new;
Ewavelet_new3=  (-exp(Ewavelet_new1).*Ewavelet_new1 ...
./log(2))';

 X1(:,x)=Ewavelet_new3; %Entropy
 X2(:,x)=Ewavelet_new2; %Energy
 load ['labelperson' num2str(n) 'allvideo.mat'])
 y1(:,x)=labels1(x,1);
 y2(:,x)=labels2(x,1);
 y3(:,x)=labels3(x,1);
 y4(:,x)=labels4(x,1);
segmentdir=sprintf('Segment%d',m);
mkdir(segmentdir);
%
cd(segmentdir)
path=pwd
%
cd ..
fname=['Wavelet5BandsPerson' num2str(n) ...
'feature30sec2stateAllEmotionSegment' num2str(m) '.mat'];
fileandfolder=(fullfile(path,fname));
save(fileandfolder,'X1','X2','y1','y2','y3','y4')
end
end
end
```

Code 7

```
close all;
clear
```

Appendix A.

```
clc
for n=1:2
load (['ResultWavelet5BandsPerson041316263130sec2stateAll' ...
num2str(n-1) '.mat']);
llen = size(X1);
len1 = len(1);
len2 = len(2);
X1=X1;
X2=X2;
y=ya;
y=yv;
y=yd;
y=yl;
    if n==1
        X_new1= X1;
        X_new2= X2;
        ya_new= ya;
        yv_new= yv;
        yd_new= yd;
        yl_new= yl;
    else
        X_new1=horzcat(X_new1,X1);
        X_new2=horzcat(X_new2,X2);
        ya_new=horzcat(ya_new,ya);
        yv_new=horzcat(yv_new,yv);
        yd_new=horzcat(yd_new,yd);
        yl_new=horzcat(yl_new,yl);
    end
end
X1=X_new1;
X2=X_new2;
```

```
ya=ya_new;
yv=yv_new;
yd=yd_new;
yl=yl_new;
save('CombineFeatureWavelet5Bands30secAllSegmentAll.mat','...
X1','X2','ya','yv','yd','yl');
```

Code 8

```
clear
for n = 1:2
load(['Wavelet5BandsPerson4feature30sec2stateAll'' ...
num2str(n-1) '.mat']);
X12=X1;
X22=X2;
y21= y1;%valence
y22= y2;%arousal
y23= y3;%dominance
y24= y4;%liking
load(['Wavelet5BandsPerson13feature30sec2stateAl' ...
num2str(n-1) '.mat']);
X13=X1;
X23=X2;
y31= y1;
y32= y2;
y33= y3;
y34= y4;%
load(['Wavelet5BandsPerson16feature30sec2stateAll'...
 num2str(n-1) '.mat']);
X14=X1;
X24=X2;
y41= y1;
```

Appendix A.

```
y42= y2;
y43= y3;
y44= y4;
load(['Wavelet5BandsPerson26feature30sec2stateAll' ...
num2str(n-1) '.mat']);
X15=X1;
X25=X2;
y51= y1;
y52= y2;
y53= y3;
y54= y4;
load(['Wavelet5BandsPerson31feature30sec2stateAll'...
 num2str(n-1) '.mat']);
X16=X1;
X26=X2;
y61= y1;
y62= y2;
y63= y3;
y64= y4;
yv=double([y21,y31,y41,y51,y61]);%valence
ya=double([y22,y32,y42,y52,y62]);%arousal
yd=double([y23,y33,y43,y53,y63]);%dominance
yl=double([y24,y34,y44,y54,y64]);%liking
X1=double([X12,X13,X14,X15,X16])%;,X6,X7]); % ,X3,X5]);
X2=double([X22,X23,X24,X25,X26])%;,X6,X7]); % ,X3,X5]);
filename=['ResultW5BandsPerson041316263130sec2stateAl' ...
num2str(n-1) '.mat'];
save(filename,'ya','yv','yd','yl','X1','X2');
end
```

Code 9

Appendix A.

```
function [I_normalized VoI] = nmi(A, B)
% Nomalized mutual information between two matrices
assert(length(A) == length(B));
A = A(:); % make it a column vector
B = B(:); % make it a column vector
n = length(A); % number of elements
% Find the number of mixtures in A and B
A_unique = unique(A);
B_unique = unique(B);
% Calculate the marginal entropy of A and B
Ma = single(bsxfun(@eq,A,A_unique'));
Mb = single(bsxfun(@eq,B,B_unique'));
Pa = sum(Ma)/n;
Pb = sum(Mb)/n;
Ha = -sum(Pa.*log2(Pa + eps));
Hb = -sum(Pb.*log2(Pb + eps));
% Calculate the joint entropy,
Pab = Ma'*Mb/n;
Hab = -sum(Pab(:).*log2(Pab(:)+eps) );
% Calculate the mutual information
Iab = Ha + Hb - Hab;
% Normalize using Strehl-Ghosh
I_normalized = Iab/sqrt(eps + Ha*Hb);
% Variation of Information
VoI = 1 - Iab/Hab;
```

Code 10

```
clear all;
close all;
load MAESeg6sec4EmotionsPerson0107162023272830Arranged
n_ch = size(X,1);
```

Appendix A.

```
% X = 1 if X>median(X), else X = 0
X_discretized = bsxfun(@gt,median(X,2),X);
% compute the mutual information
%(from the joint probability matrix)
% between X_{0,1} against Y_{1,2,3,4}
NMI_X_against_Y = arrayfun(@(i) ...
nmi(X_discretized(i,:),y),1:n_ch);
in3band=reshape(NMI_X_against_Y',10,32)';
figure,bar(in3band)
legend('Alpha', 'Beta', 'Gamma')
xlabel('Channel')
ylabel('NMI(channel,emotion)')
% compute the mutual information
%(from the joint probability matrix)
% between X_{0,1} against Y_{1,2},
% between X_{0,1} against Y_{1,3},
% between X_{0,1} against Y_{1,4},
% between X_{0,1} against Y_{2,3},
% between X_{0,1} against Y_{2,4}, and
% between X_{0,1} against Y_{3,4}
combinations = combntns(unique(y),2)
for i = 1:length(combinations);
    classes = ismembc(y,combinations(i,:));
    NMI_X_against_Y_indv(:,i) = arrayfun(@(i) ...
    nmi(X\_discretized(i,classes),y(classes) ),1:n_ch);
        str{i} = sprintf('Emo #%d vs #%d',combinations(i,1),...
        combinations(i,2));
end
% plot the cumulative NMI, perfect mutual information
%should add up to 6
figure;
```

Appendix A.

```
bar(NMI_X_against_Y_indv(1:80,:),'stacked');
figure;
bar(NMI_X_against_Y_indv(81:160,:),'stacked');
figure;
bar(NMI_X_against_Y_indv(161:240,:),'stacked');
figure;
bar(NMI_X_against_Y_indv(241:320,:),'stacked');
figure;
bar(NMI_X_against_Y_indv(1:320,:),'stacked');
legend(str,'location','northeastoutside')
xlabel('Feature \#')
ylabel('NMI(channel,emotion)')
```

Code 11

```
clear;
clc;
for loop1=1:30
loop1
%tic
data = load('Subset2Seg30secArousal.mat'); % load data
options.solver = 'smo'; % use SMO solver
options.ker = 'rbf'; % use RBF kernel
options.arg =2.01; % kernel argument %40.51
options.C = 10.01; % regularization constant
%% Good RBF parameter is approximately the cluster radius
% For simplicity, I use shi-malik normalized cut
K = 2; % number of classes
sigma = 9.057; % RBF radius (guesstimate radius)
% calculate the Normalized Laplacian matrix
D = distmat(data.X,data.X,2);
A=exp(-D/sigma^2);
```

Appendix A.

```
%imagesc(A);
D = diag(max(1,sum(A,1)));
L_nor = eye(size(A,1)) - D^(-1/2)*A*D^(-1/2);
% Use Lanczos Algorithm to calculate the eigenvector and
%eigenvalues, arranged from smallest to largest
[eigvect,eigval] = eigs(double(L_nor),K,'SM');
in = double(eigvect(:,2:K)');
% induce K-way cut using kmeans
[classlabels]= kmeanspp(in,K);
% display the normalized mutual information (accuracy) of
%the clustering process (1 = 100%)
Normalized_mutual_information = nmi(classlabels,data.y);
disp('---clustering results ---');
fprintf('normalized mutual information ...
against ground truth labels: %.3f\n',...
Normalized_mutual_information);
centroids = arrayfun(@(i) mean(data.X(:,classlabels==i),2),...
1:max(classlabels),'uniformoutput',false);
distance_to_centroids = arrayfun(@(i) mean(sum(bsxfun ...
(@minus,data.X(:,classlabels==i),centroids{i}).^2,1)),...
1:max(classlabels));
fprintf('suggested kernel radius = ...
2*(max deviation from the centroids):
%.3f\n', sqrt(max(distance_to_centroids))*2);
%% divide to training and test set with 20\% proportion
proportion_training = 0.30; %0.2
clear train_idx;
train_idx = [];
for i = 1:max(data.y)
    I = find(data.y == i);
    N = length(I);
```

Appendix A.

```
    tr_prop = round(proportion_training*N);
    tst_prop = N - tr_prop;
    T = [zeros(1,tst_prop), ones(1,tr_prop)];
    train_idx = cat(2,train_idx, T(randperm(N)));
end
train_idx = train_idx == 1;
test_idx = ~train_idx;
data_train = struct('X',data.X(:,train_idx),'y', ...
data.y(:,train_idx));
data_test = struct('X',data.X(:,test_idx),'y', ...
data.y(:,test_idx));
tic
model = oaasvm(data_train,options ); % training
%% training & test accuracy
training_result = classperf(data_train.y, ...
svmclass(data_train.X,model))
toc
time1=toc
tic
test_result = classperf(data_test.y, ...
svmclass(data_test.X,model))
toc
time2=toc
 DT=test_result.DiagnosticTable
TrainT(loop1)=time1
TestT(loop1)=time2
 Accuracy(loop1)=test_result.CorrectRate
 models(loop1)=model;
 Sensitivity(loop1)=test_result.Sensitivity
Specificity(loop1)=test_result.Specificity
DT(:,:,loop1)=DT; %test_result.DiagnosticTable
```

```
F1(loop1)=2*DT(1)/(2*DT(1)+DT(2)+DT(3))
[ypred, dfce] = svmclass(data_test.X,model);
xtest(:,:,loop1)=data_test.X;
ytest(:,:,loop1)=data_test.y;
ypred(loop1,:)=ypred;
end
meanAcc=mean(Accuracy)
stdAcc=std(Accuracy)
meanSens=mean(Sensitivity)
stdSens=std(Sensitivity)
meanSpec=mean(Specificity)
stdSpec=std(Specificity)
meanTrainT=mean(TrainT)
stdTrainT=std(TrainT)
meanTestT=mean(TestT)
stdTestT=std(TestT)
save('TrainingResult.mat','TrainT','TestT','Accuracy', ...
'Sensitivity','Specificity','DT','F1','models','Sensitivity',...
'Specificity', 'DT','F1','ypred','xtest',' ytest','meanAcc',...
'meanSens','meanSpec','meanTrainT','meanTestT', ...
'stdAcc','stdSens','stdSpec','stdTrainT','stdTestT');
```

# Bibliography

Agrafioti, Foteini, Hatzinakos, Dimitris, & Anderson, Adam K. 2012. ECG Pattern Analysis for Emotion Detection. *IEEE Trans. Affect. Comput.*, **3**(1), 102–115.

Akin, M. 2002. Comparison of Wavelet Transform and FFT Methods in the Analysis of EEG Signals. *J. Med. Syst.*, **26**(3), 241–247.

AlZoubi, Omar, D'Mello, Sidney K, & Calvo, Rafael A. 2012. Detecting naturalistic expressions of nonbasic affect using physiological signals. *IEEE Transactions on affective computing*, **3**(3), 298–310.

Ansari-Asl, K., Chanel, G., & Pun, T. 2007. A channel selection method for EEG classification in emotion assessment based on synchronization likelihood. *Pages 1241–1245 of: 15th European Signal Processing Conference, EUSIPCO 2007*.

Ardi Handojoseno, A. M., Shine, J. M., Nguyen, T. N., Tran, Y., Lewis, S. J. G., & Nguyen, H. T. Sep. 2015. Analysis and Prediction of the Freezing of Gait Using EEG Brain Dynamics. *IEEE Trans. Neural Syst. Rehab. Eng.*, **23**(5), 887–896.

Balconi, M., & Lucchiari, C. 2006. EEG correlates (event-related desynchronization) of emotional face elaboration: a temporal analysis. *Neurosci Lett*, **392**(1-2), 118–23. Balconi, Michela Lucchiari, Claudio Comparative Study Journal Article Ireland Neurosci Lett. 2006 Jan 9;392(1-2):118-23. Epub 2005 Oct 3.

Balconi, M., & Lucchiari, C. 2008. Consciousness and arousal effects on emotional face processing as revealed by brain oscillations. A gamma band analysis. *Int J Psychophysiol*, **67**(1), 41–6. Balconi, Michela Lucchiari, Claudio Journal Article Netherlands Int J Psychophysiol. 2008 Jan;67(1):41-6. Epub 2007 Oct 13.

Balconi, M., & Mazza, G. 2009. Brain oscillations and BIS/BAS (behavioral inhibition/activation system) effects on processing masked emotional cues. ERS/ERD and coherence measures of alpha band. *Int J Psychophysiol*, **74**(2), 158–65. 1872-7697 Balconi, Michela Mazza, Guido Journal Article Netherlands Int J Psychophysiol. 2009 Nov;74(2):158-65. doi: 10.1016/j.ijpsycho.2009.08.006. Epub 2009 Aug 24.

Baltrusaitis, T., Mahmoud, M., & Robinson, P. 2015 (May). Cross-dataset learning and person-specific normalisation for automatic Action Unit detection. *Pages 1–6 of: IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2015 11th*, vol. 06.

Barrett, Lisa Feldman. 1998. Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & Emotion*, **12**(4), 579–599.

Barrett, Lisa Feldman, & BlissMoreau, Eliza. 2009. Affect as a psychological primitive. *Advances in experimental social psychology*, **41**, 167–218.

Bekele, Esubalew, Zheng, Zhi, Swanson, Amy, Crittendon, Julie, Warren, Zachary, & Sarkar, Nilanjan. 2013. Understanding how adolescents with autism respond to facial expressions in virtual reality environments. *IEEE transactions on visualization and computer graphics*, **19**(4), 711–720.

Bennett, Kristin P, & Campbell, Colin. 2000. Support vector machines: hype or hallelujah? *ACM SIGKDD Explorations Newsletter*, **2**(2), 1–13.

Bolle, Ruud, & Pankanti, Sharath. 1998. *Biometrics, Personal Identification in Networked Society: Personal Identification in Networked Society*. Norwell, MA, USA: Kluwer Academic Publishers.

Bradley, Andrew P. 1997. The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recogn.*, **30**(7), 1145–1159.

Bradley, Margaret M, & Lang, Peter J. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, **25**(1), 49–59.

Burges, Christopher J. C. 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, **2**, 121–167.

Busso, Carlos, & Narayanan, Shrikanth S. 2007. Interrelation between speech and facial gestures in emotional utterances: a single subject study. *IEEE Transactions on Audio, Speech, and Language Processing*, **15**(8), 2331–2347.

Cahill, Nathan D. 2010. Normalized measures of mutual information with general definitions of entropy for multimodal image registration. *Pages 258–268 of: Biomedical Image Registration*. Springer.

Caicedo, David Giza, & Beuzekom, Marleen Van. 2006. *How Do You Feel? An Assessment of Existing Tools for The Measurement of Emotions and Their Application in Consumer Products, Delft University of Technology.*

Candra, H., Yuwono, M., Rifai, Chai, Handojoseno, A., Elamvazuthi, I., Nguyen, H. T., & Su, S. 2015a. Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine. *Pages 7250–7253 of: Proc. IEEE 37th Annu. Int. Conf. Eng. Med. Biol. Soc.*

Candra, H., Yuwono, M., Handojoseno, A., Chai, R., Su, S., & Nguyen, H. T. 2015b. Recognizing emotions from EEG subbands using wavelet analysis. *Pages 6030–6033 of: Proc. IEEE 37th Annu. Int. Conf. Eng. Med. Biol. Soc.*

Candra, H., Yuwono, M., Chai, R., Nguyen, H. T., & Su, S. 2016 (Aug). Classification of Facial-Emotion Expression in the Application of Psychotherapy using Viola-Jones and Edge-Histogram of Oriented Gradient. *Pages 423–426 of: Proc. IEEE 38th Annu. Int. Conf. Eng. Med. Biol. Soc.*

Candra, H., Yuwono, M., Chai, R., Nguyen, H. T., & Su, S. 2017. EEG Emotion Recognition Using Reduced Channel Wavelet Entropy and Average Wavelet Coefcient Features with Normal Mutual Information Method. *In: Proc. IEEE 39th Annu. Int. Conf. Eng. Med. Biol. Soc. (In Press)*.

Candra, H., Yuwono, M., Chai, R., Handojoseno, A., Elamvazuthi, I., Nguyen, H. T., &

Su, S. (Submitted)a. Optimal Window Selection for Improving EEG-Emotion Classification. *Signal Processing*.

Candra, H., Cao, K., Yuwono, M., Chai, R., Nguyen, H. T., & Su, S. (Submitted)b. Reduced Dimension of Edge-Histogram of Oriented Gradients (RED E-HOG) Features for Facial-Emotion Recognition. *Digital Signal Processing*.

Canny, J. 1986. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, **8**(6), 679–698.

Chai, R., Naik, G., Nguyen, T. N., Ling, S., Tran, Y., Craig, A., & Nguyen, H. 2016a. Driver Fatigue Classification with Independent Component by Entropy Rate Bound Minimization Analysis in an EEG-based System. *IEEE Journal of Biomedical and Health Informatics*, **PP**(99), 1–1.

Chai, R., Naik, G., Nguyen, T. N., Ling, S., Tran, Y., Craig, A., & Nguyen, H. 2016b. Driver Fatigue Classification with Independent Component by Entropy Rate Bound Minimization Analysis in an EEG-based System. *IEEE Journal of Biomedical and Health Informatics*, **PP**(99), 1–1.

Chang, Chih-Chung, & Lin, Chih-Jen. 2011. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.*, **2**(3), 27:1–27:27.

College, OpenStax, Spielman, R.M., Dumper, K., Jenkins, W.J., Lacombe, A., Lovett, M., & Perlmutter, M. 2014. *Psychology*. OpenStax College.

Colten, Harvey R, Altevogt, Bruce M, *et al.* 2006. *Sleep disorders and sleep deprivation: an unmet public health problem*. National Academies Press.

Conneau, Anne-Claire, & Essid, Slim. 2014. Assessment of new spectral features for eeg-based emotion recognition. *Pages 4698–4702 of: Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*.

Creusen, I. M., Wijnhoven, R. G. J., Herbschleb, E., & de With, P. H. N. 2010 (Sept). Color exploitation in hog-based traffic sign detection. *Pages 2669–2672 of: 2010 IEEE International Conference on Image Processing*.

Dalal, N., & Triggs, B. 2005 (June). Histograms of oriented gradients for human detection. *Pages 886–893 vol. 1 of: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1.

Delac, Kresimir, Grgic, Mislav, & Grgic, Sonja. 2005. Independent comparative study of PCA, ICA, and LDA on the FERET data set. *International Journal of Imaging Systems and Technology*, **15**(5), 252–260.

Dniz, O., Bueno, G., Salido, J., & la Torre, F. De. 2011. Face recognition using Histograms of Oriented Gradients. *Pattern Recognition Letters*, **32**(12), 1598 – 1603.

Do, Thanh-Toan, & Kijak, Ewa. 2012. Face recognition using co-occurrence histograms of oriented gradients. *Pages 1301–1304 of: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE.

Donato, Gianluca, Bartlett, Marian Stewart, Hager, Joseph C., Ekman, Paul, & Sejnowski, Terrence J. 1999. Classifying Facial Actions. *IEEE Trans. Pattern Anal. Mach. Intell.*, **21**(10), 974–989.

Dulguerov, Pavel, Marchal, Francis, Wang, Desheng, & Gysin, Claudine. 1999. Review of objective topographic facial nerve evaluation methods. *Otology & Neurotology*, **20**(5), 672–678.

Edwards, G. J., Cootes, T. F., & Taylor, C. J. 1998. *Face recognition using active appearance models*. Berlin, Heidelberg: Springer Berlin Heidelberg. Pages 581–595.

Ekman, P. 2003. *Emotions Revealed, Recognizing Faces and Feelings to Improve Communication and Emotional Life*. New York: Henry Holt and Company LLC.

Ekman, Paul, & Friesen, Wallace V. 1976. Measuring facial movement. *Environmental psychology and nonverbal behavior*, **1**(1), 56–75.

Estévez, Pablo A, Tesmer, Michel, Perez, Claudio A, & Zurada, Jacek M. 2009. Normalized mutual information feature selection. *IEEE Trans. Neural Netw.*, **20**(2), 189–201.

Fasel, Beat, & Luettin, Juergen. 2003. Automatic facial expression analysis: a survey. *Pattern recognition*, **36**(1), 259–275.

Fawcett, Tom. 2006. An introduction to {ROC} analysis. *Pattern Recognition Letters*, **27**(8), 861 – 874. {ROC} Analysis in Pattern Recognition.

Foley, G. N., & Gentile, J. P. 2010 (jun). Nonverbal Communication in Psychotherapy. *Pages 38–44 of: Psychiatry (Edgmont)*, vol. 7.

Freund, Yoav, & Schapire, Robert E. 1996. Experiments with a New Boosting Algorithm. *Pages 148–156 of:* Saitta, Lorenza (ed), *Proceedings of the Thirteenth International Conference on Machine Learning (ICML 1996)*. Morgan Kaufmann.

Fuglede, B., & Topsoe, F. 2004 (June). Jensen-Shannon divergence and Hilbert space embedding. *Pages 31– of: Proceedings. International Symposium on Information Theory (ISIT), 2004.*

Gong, Shaogang, McKenna, Stephen J., & Psarrou, Alexandra. 2000. *Dynamic Vision: From Images to Face Recognition*. 1st edn. London, UK, UK: Imperial College Press.

Graps, A. 1995. An introduction to wavelets. *IEEE Computational Science and Engineering*, **2**(2), 50–61.

Gunes, H., & Piccardi, M. 2005 (Aug). Fusing face and body gesture for machine recognition of emotions. *Pages 306–311 of: ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*

Guo, K., Candra, H., Yu, H., Li, H., Nguyen, H. T., & Su, S. W. 2017. EEG-Based Emotion Classification Using Innovative Features and Combined SVM and HMM Classifier. *In: Proc. IEEE 39th Annu. Int. Conf. Eng. Med. Biol. Soc. (In Press)*.

Hajian-Tilaki, K. 2013. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med*, **4**(2), 627–35. 2008-6172 Hajian-Tilaki, Karimollah Babol, Iran Caspian J Intern Med. 2013 Spring;4(2):627-35.

Handojoseno, AM Ardi, Shine, James M, Nguyen, Tuan N, Tran, Yvonne, Lewis, Simon JG, & Nguyen, Hung T. 2015. Analysis and prediction of the freezing of gait using EEG brain dynamics. *IEEE transactions on neural systems and rehabilitation engineering*, **23**(5), 887–896.

Handojoseno, A.M.A., Shine, J.M., Nguyen, T.N., Tran, Y., Lewis, S.J.G., & Nguyen, H.T. 2012 (Aug). The detection of Freezing of Gait in Parkinson's disease patients using EEG signals based on Wavelet decomposition. *Pages 69–72 of: Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*.

He, Zengyou, Xu, Xiaofei, & Deng, Shengchun. 2008. k-ANMI: A mutual information based clustering algorithm for categorical data. *Information Fusion*, **9**(2), 223–233.

Hjelms, Erik, & Low, Boon Kee. 2001. Face Detection: A Survey. *Computer Vision and Image Understanding*, **83**(3), 236 – 274.

Hsu, Chih-Wei, Chang, Chih-Chung, Lin, Chih-Jen, *et al.* 2003. A practical guide to support vector classification.

Isomursu, Minna, Thti, Marika, Vinm, Soili, & Kuutti, Kari. 2007. Experimental evaluation of five methods for collecting emotions in field settings with mobile applications. *Int. Journal of Human-Computer Studies*, **65**(4), 404–418.

Jain, Anil K, Duin, Robert P. W., & Mao, Jianchang. 2000. Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence*, **22**(1), 4–37.

Jatupaiboon, N., Pan-ngum, S., & Israsena, P. 2013. Emotion classification using minimal EEG channels and frequency bands. *Pages 21–24 of: Computer Science and Software Engineering (JCSSE), 2013 10th International Joint Conference on*.

Jebara, Tony S. 1995. *3D pose estimation and normalization for face recognition*. Ph.D. thesis, McGill University.

Jenke, R., Peer, A., & Buss, M. 2014a. Feature Extraction and Selection for Emotion Recognition from EEG. *Affective Computing, IEEE Transactions on*, **PP**(99), 1–1.

Jenke, Robert, Peer, Angelika, & Buss, Martin. 2014b. Feature Extraction and Selection for Emotion Recognition from EEG. *IEEE Trans. Affective Comput.*, **5**(3), 327–339.

Joachims, Thorsten. 2002. *Introduction to support vector machines*.

Jouny, Christophe C., Franaszczuk, Piotr J., & Bergey, Gregory K. 2011. Improving early seizure detection. *Epilepsy & Behavior*, **22, Supplement 1**, S44 – S48. The Future of Automated Seizure Detection and PredictionThe production of this supplemental special issue was supported by an unrestricted educational grant from Cyberonics, Inc.

Kanade, Takeo, Tian, Yingli, & Cohn, Jeffrey F. 2000. Comprehensive Database for Facial Expression Analysis. *Pages 46– of: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*. FG '00. Washington, DC, USA: IEEE Computer Society.

Kaplan, Alexander Ya., Fingelkurts, Andrew A., Fingelkurts, Alexander A., Borisov, Sergei V., & Darkhovsky, Boris S. 2005. Nonstationary Nature of the Brain Activity As Revealed by EEG/MEG: Methodological, Practical and Conceptual Challenges. *Signal Process.*, **85**(11), 2190–2212.

Keil, A., Muller, M. M., Gruber, T., Wienbruch, C., Stolarova, M., & Elbert, T. 2001. Effects of emotional arousal in the cerebral hemispheres: a study of oscillatory brain activity and event-related potentials. *Clin Neurophysiol*, **112**(11), 2057–68. Keil, A Muller, M M Gruber, T Wienbruch, C Stolarova, M Elbert, T Journal Article Research Support, Non-U.S. Gov't Netherlands Clin Neurophysiol. 2001 Nov;112(11):2057-68.

Khosrowabadi, R., Quek, C., Ang, K. K., & Wahab, A. 2014. ERNN: A Biologically Inspired Feedforward Neural Network to Discriminate Emotion From EEG Signal. *IEEE Transactions on Neural Networks and Learning Systems*, **25**(3), 609–620.

Kim, Min-Ki, Kim, Miyoung, Oh, Eunmi, & Kim, Sung-Phil. 2013. A Review on the Computational Methods for Emotional State Estimation from the Human EEG. *Computational and Mathematical Methods in Medicine*, **2013**, 13.

Kim, T., Shin, D., Shin, D., Kim, S., & Lee, M. 2010 (Oct). Design and development of multimodal analysis system based on biometric signals. *Pages 853–857 of: 2010 3rd International Conference on Biomedical Engineering and Informatics*, vol. 2.

Koelstra, Sander, Muhl, Christian, Soleymani, Mohammad, Lee, Jong-Seok, Yazdani, Ashkan, Ebrahimi, Touradj, Pun, Thierry, Nijholt, Anton, & Patras, Ioannis. 2012.

DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *IEEE Trans. Affect. Comput.*, **3**(1), 18–31.

Kvaale, Stian Pedersen. 2012. *Emotion Recognition in EEG, A neuroevolutionary approach*. Master Theses.

Li, Junnan, & Lam, E.Y. 2015 (Sept). Facial expression recognition using deep neural networks. *Pages 1–6 of: IEEE International Conference on Imaging Systems and Techniques (IST), 2015*.

Li, R., Liu, P., Jia, K., & Wu, Q. 2015 (Dec). Facial Expression Recognition under Partial Occlusion Based on Gabor Filter and Gray-Level Cooccurrence Matrix. *Pages 347–351 of: 2015 International Conference on Computational Intelligence and Communication Networks (CICN)*.

Liang, Pengpeng, Teodoro, Gregory, Ling, Haibin, Blasch, Erik, Chen, Genshe, & Bai, Li. 2012. Multiple kernel learning for vehicle detection in wide area motion imagery. *Pages 1629–1636 of: Information Fusion (FUSION), 2012 15th International Conference on*. IEEE.

Liao, Lun-De, Chen, Chi-Yu, Wang, I-Jan, Chen, Sheng-Fu, Li, Shih-Yu, Chen, Bo-Wei, Chang, Jyh-Yeong, & Lin, Chin-Teng. 2011. Gaming control using a wearable and wireless EEG-based brain-computer interface device with novel dry foam-based sensors. *In: Journal of NeuroEngineering and Rehabilitation*.

Liu, Y., Sourina, O., & Hafiyyandi, M. R. 2013 (Sept). EEG-Based Emotion-Adaptive Advertising. *Pages 843–848 of: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*.

Liu, Yin, Wang, Wansen, Liu, Dian, & Liu, Shuai. 2010. The Study of Learners' Emotional Features in the E-learning System. *Pages 423–426 of: Proceedings of the 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing - Volume 01*. NSWCTC '10. Washington, DC, USA: IEEE Computer Society.

Liu, Yisi, & Sourina, Olga. 2013. Real-time fractal-based valence level recognition from EEG. *Pages 101–120 of: Transactions on Computational Science XVIII*. Springer.

Lotte, Fabien, Congedo, Marco, Lcuyer, Anatole, Lamarche, Fabrice, & Arnaldi, Bruno. 2007. A review of classification algorithms for EEG-based braincomputer interfaces. *Journal of Neural Engineering*, **4**, 24.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. 2010 (June). The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. *Pages 94–101 of: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*.

Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. 1998. Coding Facial Expressions with Gabor Wavelets. *Pages 200– of: Proceedings of the 3rd. International Conference on Face & Gesture Recognition*. FG '98. Washington, DC, USA: IEEE Computer Society.

Machado, P. P., Beutler, L. E., & Greenberg, L. S. 1999. Emotion recognition in psychotherapy: impact of therapist level of experience and emotional awareness. *J Clin Psychol*, **55**(jan), 39–57.

Machado, Sergio, Araújo, Fernanda, Paes, Flávia, Velasques, Bruna, Cunha, Mario, Budde, Henning, Basile, Luis F, Anghinah, Renato, Arias-Carrión, Oscar, Cagy, Mauricio, *et al.* 2010. EEG-based brain-computer interfaces: an overview of basic concepts and clinical applications in neurorehabilitation. *Reviews in the neurosciences*, **21**(6), 451–468.

Martis, Roshan Joy, Acharya, U Rajendra, & Min, Lim Choo. 2013. ECG beat classification using PCA, LDA, ICA and discrete wavelet transform. *Biomedical Signal Processing and Control*, **8**(5), 437–448.

Mauss, Iris B., & Robinson, Michael D. 2009. Measures of emotion: A review. *Cognition & emotion*, **23**(2), 209–237. 19809584[pmid] Cogn Emot.

Mehrabian, Albert. 1997. Comparison of the PAD and PANAS as models for describing

emotions and for differentiating anxiety from depression. *Journal of Psychopathology and Behavioral Assessment*, **19**(4), 331–357.

Metallinou, Angeliki, Lee, Sungbok, & Narayanan, Shrikanth. 2010. Decision level combination of multiple modalities for recognition and analysis of emotional expression. *Pages 2462–2465 of: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE.

Mohamed, Abdallah A., Gavrilova, Marina L., & Yampolskiy, Roman V. 2013. *Recognizing Avatar Faces Using Wavelet-Based Adaptive Local Binary Patterns with Directional Statistical Features*. Berlin, Heidelberg: Springer Berlin Heidelberg. Pages 137–154.

Monzo, David, Albiol, Alberto, Albiol, Antonio, & Mossi, Jose M. 2011. Color HOG-EBGM for face recognition. *Pages 785–788 of: 2011 18th IEEE International Conference on Image Processing*. IEEE.

Muller, M. M., Keil, A., Gruber, T., & Elbert, T. 1999. Processing of affective pictures modulates right-hemispheric gamma band EEG activity. *Clin Neurophysiol*, **110**(11), 1913–20. Muller, M M Keil, A Gruber, T Elbert, T Clinical Trial Journal Article Research Support, Non-U.S. Gov't Netherlands Clin Neurophysiol. 1999 Nov;110(11):1913-20.

Nardelli, Mimma, Valenza, Gaetano, Greco, Alberto, Lanata, Antonio, & Scilingo, Enzo Pasquale. 2015. Recognizing Emotions Induced by Affective Sounds through Heart Rate Variability. *IEEE Trans. Affective Comput.*, **6**(4), 385–394.

Nastar, C., & Mitschke, M. 1998. Real-Time Face Recognition Using Feature Combination. *Pages 312– of: Proceedings of the 3rd. International Conference on Face & Gesture Recognition*. FG '98. Washington, DC, USA: IEEE Computer Society.

Neeru, Nirvair, & Kaur, Lakhwinder. 2016. Modified SIFT Descriptors for Face Recognition under Different Emotions. *Journal of Engineering*, **2016**, 12.

Nie, Dan, Wang, Xiao-Wei, Shi, Li-Chen, & Lu, Bao-Liang. 2011 (April). EEG-based

emotion recognition during watching movies. *Pages 667–670 of: Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on*.

Oberman, Lindsay M., Ramachandran, Vilayanur S., & Pineda, Jaime A. 2008. Modulation of mu suppression in children with autism spectrum disorders in response to familiar or unfamiliar stimuli: The mirror neuron hypothesis. *Neuropsychologia*, **46**(5), 1558 – 1565.

Ocak, Hasan. 2008. Optimal Classification of Epileptic Seizures in EEG Using Wavelet Analysis and Genetic Algorithm. *Signal Process.*, **88**(7), 1858–1867.

Othman, Marini, & Wahab, Abdul. 2010. Affective face processing analysis in autism using electroencephalogram. *Pages E23–E27 of: Information and Communication Technology for the Muslim World (ICT4M), 2010 International Conference on*. IEEE.

Owayjan, M., Kashour, A., Haddad, N. Al, Fadel, M., & Souki, G. Al. 2012 (Dec). The design and development of a Lie Detection System using facial micro-expressions. *Pages 33–38 of: Advances in Computational Tools for Engineering Applications (ACTEA), 2012 2nd International Conference on*.

Panksepp, J. 2010. Affective neuroscience of the emotional BrainMind: evolutionary perspectives and implications for understanding depression. *Dialogues Clin Neurosci*, **12**(4), 533–45.

Panksepp, Jaak. 2007. Neurologizing the Psychology of Affects: How Appraisal-Based Constructivism and Basic Emotion Theory Can Coexist. *Perspectives on Psychological Science*, **2**(3), 281–296.

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

Pantic, M., Valstar, M., Rademaker, R., & Maat, L. 2005 (July). Web-based database for facial expression analysis. *Pages 5 pp.– of: 2005 IEEE International Conference on Multimedia and Expo*.

Papadaniil, C., Kosmidou, V., Tsolaki, A., Hadjileontiadis, L., Tsolaki, M., & Kompatsiaris, I. 2015. Age Effect in Human Brain Responses to Emotion Arousing Images: The EEG 3D-Vector Field Tomography Modeling Approach. *Autonomous Mental Development, IEEE Transactions on*, **PP**(99), 1–1.

Parmar, Divyarajsinh N., & Mehta, Brijesh B. 2014. Face Recognition Methods & Applications. *CoRR*, **abs/1403.0485**.

Petrantonakis, Panagiotis C., & Hadjileontiadis, Leontios J. 2012. Adaptive Emotional Information Retrieval From EEG Signals in the Time-Frequency Domain. *Trans. Sig. Proc.*, **60**(5), 2604–2616.

Picard, R. W., Vyzas, E., & Healey, J. Oct. 2001. Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Trans. Pattern Anal. Machine Intell.*, **23**(10), 1175–1191.

Picard, Rosalind W, Vyzas, Elias, & Healey, Jennifer. 2001. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Trans. Pattern Anal. Mach. Intell.*, **23**(10), 1175–1191.

Ptucha, Raymond, & Savakis, Andreas. 2013. Manifold Based Sparse Representation for Facial Understanding in Natural Images. *Image Vision Comput.*, **31**(5), 365–378.

Pun, Thierry, Alecu, Teodor Iulian, Chanel, Guillaume, Kronegg, Julien, & Voloshynovskiy, Sviatoslav. 2006. Brain-computer interaction research at the Computer Vision and Multimedia Laboratory, University of Geneva. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **14**(2), 210–213.

Purdon, PL, Pavone, KJ, Akeju, O, Smith, AC, Sampson, AL, Lee, J, Zhou, DW, Solt, K, & Brown, EN. 2015. The Ageing Brain: Age-dependent changes in the electroencephalogram during propofol and sevoflurane general anaesthesia. *British journal of anaesthesia*, **115**(suppl 1), i46–i57.

Quiroga, Rodrigo Quian. 1998. Quantitative analysis of EEG signals: time-frequency methods and chaos theory. *Institute of Physiology-Medical University Lubeck and Institute of Signal Processing-Medical University Lubeck*.

Rabia, Jafri, & Hamid, R. Arabnia. 2009. A Survey of Face Recognition Techniques. *Journal of Information Processing Systems*, **5**(2), 41–68.

Resnikoff, Howard L., & Wells, Jr., Raymond O. 1998. *Wavelet Analysis: The Scalable Structure of Information*. New York, NY, USA: Springer-Verlag New York, Inc.

Rizon. 2010. Discrete Wavelet Transform Based Classification of Human Emotions Using Electroencephalogram Signals. *American Journal of Applied Sciences*, **7**(7), 878–885.

Roman-Gonzalez, Avid. 2012. Eeg signal processing for bci applications. *Pages 571–591 of: Human–Computer Systems Interaction: Backgrounds and Applications 2*. Springer.

Rosso, Osvaldo, Blanco, Susana, Yordanova, Juliana, Kolev, Vasil, Figliola, Alejandra, Schrmann, Martin, & Basar, Erol. 2001a. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of Neuroscience Methods*, **105**(1), 65–75.

Rosso, Osvaldo A., Blanco, Susana, Yordanova, Juliana, Kolev, Vasil, Figliola, Alejandra, Schürmann, Martin, & Başar, Erol. 2001b. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of Neuroscience Methods*, **105**(1), 65–75.

Rozgi, V., Vitaladevuni, S. N., & Prasad, R. 2013 (May). Robust EEG emotion classification using segment level decision fusion. *Pages 1286–1290 of: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*.

Russell, J. A., & Barrett, L. F. 1999. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *J Pers Soc Psychol*, **76**(5), 805–19.

Russell, James A. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, **39**(6), 1161–1178.

Russell, James A. 2003. Core affect and the psychological construction of emotion. *Psychological review*, **110**(1), 145.

Salhi, Abdel Ilah, Kardouchi, Mustapha, & Belacel, Nabil. 2013. Histograms of fuzzy oriented gradients for face recognition. *Pages 1–5 of: Computer Applications Technology (ICCAT), 2013 International Conference on*. IEEE.

Sanei, Saeid, & Chambers, J.A. 2007. *EEG signal processing*. 2nd edition edn. West Sussex, England: John Wiley & Sons Ltd.

Schaefer, Earl S. 1959. A circumplex model for maternal behavior. *The Journal of Abnormal and Social Psychology*, **59**(2), 226–235.

Scherer, Klaus R., Schorr, Angela, & Johnstone, Tom. 2001. *Appraisal processes in emotion : theory, methods, research*. Oxford [etc.]: Oxford University Press.

Schlesinger, Michail I, & Hlavac, Vaclav. 2002. *Ten lectures on statistical and structural pattern recognition*. Kluwer Academic.

Shan, Caifeng, Gong, Shaogang, & McOwan, Peter W. 2009. Facial expression recognition based on Local Binary Patterns: A comprehensive study. *Image and Vision Computing*, **27**(6), 803 – 816.

Shannon, C. E. 1948. A mathematical theory of communication. *Bell system technical journal*, **27**.

Signorino, M., Pucci, E., Belardinelli, N., Nolfe, G., & Angeleri, F. 1995. {EEG} spectral analysis in vascular and Alzheimer dementia. *Electroencephalography and Clinical Neurophysiology*, **94**(5), 313 – 325.

Smitha, KG, & Vinod, AP. 2015. Facial emotion recognition system for autistic children: a feasible study based on FPGA implementation. *Medical & biological engineering & computing*, **53**(11), 1221–1229.

Soleymani, Mohammad, Lichtenauer, Jeroen, Pun, Thierry, & Pantic, Maja. 2012a. A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE Trans. Affect. Comput.*, **3**(1), 42–55.

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

Song, Mingli, Tao, Dacheng, Liu, Zicheng, Li, Xuelong, & Zhou, Mengchu. 2010. Image Ratio Features for Facial Expression Recognition Application. *Trans. Sys. Man Cyber. Part B*, **40**(3), 779–788.

Strehl, Alexander, & Ghosh, Joydeep. 2003. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *The Journal of Machine Learning Research*, **3**, 583–617.

Sun, Y., & Akansu, A. N. 2014. Facial expression recognition with regional hidden Markov models. *Electronics Letters*, **50**(9), 671–673.

Sutter, J. V. 2010. *Assessing Impact Of Affect Recognition On Therapeutic Relationship*. M.Phil. thesis, University of Kentucky.

Tan, Hengliang, Yang, Bing, & Ma, Zhengming. 2014. Face recognition based on the fusion of global and local HOG features of face images. *IET computer vision*, **8**(3), 224–234.

Tanaka, James W., Wolf, Julie M., Klaiman, Cheryl, Koenig, Kathleen, Cockburn, Jeffrey, Herlihy, Lauren, Brown, Carla, Stahl, Sherin, Kaiser, Martha D., & Schultz, Robert T. 2010. Using Computerized Games to Teach Face Recognition Skills to Children with Autism Spectrum Disorder: The "Let's Face It!" Program. *Journal of Child Psychology and Psychiatry*, **51**(8), 944–952.

Tariq, U., Lin, Kai-Hsiang, Li, Zhen, Zhou, Xi, Wang, Zhaowen, Le, Vuong, Huang, T.S., Lv, Xutao, & Han, T.X. 2012. Recognizing Emotions From an Ensemble of Features. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, **42**(4), 1017–1026.

Taspinar, M., Naskali, A. T., Eren, G., & Kurt, M. 2012 (May). The importance of customized advertisement delivery using 3D tracking and facial recognition. *Pages 520–524 of: Second International Conference on Digital Information and Communication Technology and it's Applications (DICTAP), 2012*.

Thammasan, Nattapong, Moriyama, Koichi, Fukui, Ken-ichi, & Numao, Masayuki. 2016. Familiarity effects in EEG-based emotion recognition. *Brain Informatics*, 1–12.

Tian, Yingli, Kanade, Takeo, & Cohn, Jeffrey F. 2011. Facial expression recognition. *Pages 487–519 of: Handbook of face recognition*. Springer.

Tinguely, Gilberte, Finelli, Luca A., Landolt, Hans-Peter, Borbély, Alexander A., & Achermann, Peter. 2006. Functional EEG topography in sleep and waking: State-dependent and state-independent features. *NeuroImage*, **32**(1), 283–292.

Torres, C.A., Orozco, A.A., & Alvarez, M.A. 2013 (July). Feature selection for multimodal emotion recognition in the arousal-valence space. *Pages 4330–4333 of: Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*.

Turk, Matthew, & Pentland, Alex. 1991. Eigenfaces for Recognition. *J. Cognitive Neuroscience*, **3**(1), 71–86.

Ulukaya, Sezer, & Erdem, Cigdem Eroglu. 2014. Gaussian mixture model based estimation of the neutral face shape for emotion recognition. *Digital Signal Processing*, **32**, 11 – 23.

Valenza, Gaetano, Citi, Luca, Lanatá, Antonio, Scilingo, Enzo Pasquale, & Barbieri, Riccardo. 2014. Revealing Real-Time Emotional Responses: a Personalized Assessment based on Heartbeat Dynamics. *Scientific Reports*, **4**.

Van Loan, Charles. 1992. *Computational Frameworks for the Fast Fourier Transform*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.

Vert, Jean-Philippe, Tsuda, Koji, & Schölkopf, Bernhard. 2004. A primer on kernel methods. *Kernel Methods in Computational Biology*, 35–70.

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

Wang, S., Zhu, Y., Yue, L., & Ji, Q. 2015. Emotion recognition with the help of privileged information. *Autonomous Mental Development, IEEE Transactions on*, **PP**(99), 1–1.

Wang, W., & Niu, H. 2012 (Oct). Face detection based on improved AdaBoost algorithm in E-Learning. *Pages 924–927 of: 2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems*, vol. 02.

Wang, Zhiliang, Gu, Xuejing, He, Jie, Zheng, Siyi, & Wang, Wei. 2010. Design and implementation of an intelligent tutoring system for English instruction. *Pages 166–169 of: Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on*, vol. 2. IEEE.

Wichakam, I., & Vateekul, P. 2014. An evaluation of feature extraction in EEG-based emotion prediction with support vector machines. *Pages 106–110 of: Computer Science and Software Engineering (JCSSE), 2014 11th International Joint Conference on*.

Yang, Ming-Hsuan, Kriegman, David J., & Ahuja, Narendra. 2002. Detecting Faces in Images: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, **24**(1), 34–58.

Yang, Wei, Song, Zhan, & Wu, Xinyu. 2012. Histogram of silhouette direction code: An efficient HOG-based descriptor for accurate human detection. *Pages 330–335 of: Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on*. IEEE.

Yoo, Sung-Hoon, Oh, Sung-Kwun, & Pedrycz, Witold. 2015. Optimized Face Recognition Algorithm Using Radial Basis Function Neural Networks and Its Practical Applications. *Neural Netw.*, **69**(C), 111–125.

Yuwono, M., Su, S.W., Moulton, B.D., Guo, Ying, & Nguyen, H.T. 2014 (July). An algorithm for scalable clustering: Ensemble Rapid Centroid Estimation. *Pages 1250–1257 of: Evolutionary Computation (CEC), 2014 IEEE Congress on*.

Yuwono, M., Su, S. W., Moulton, B. D., & Nguyen, H. T. Jun. 2014. Data Clustering Using Variants of Rapid Centroid Estimation. *IEEE Trans. Evol. Comput.*, **18**(3), 366–377.

Yuwono, Mitchell. 2015 (March). *Ensemble Rapid Centroid Estimation: A Semi-Stochastic Consensus Particle Swarm Approach for Large Scale Cluster Optimization*. Ph.D. thesis, University of Technology, Sydney (UTS), Australia.

Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. 2003. Face Recognition: A Literature Survey. *ACM Comput. Surv.*, **35**(4), 399–458.

Zhao, WenYi, & Chellappa, Rama. 2002. Image-based face recognition: Issues and methods. *Optical Engineering-New York-Marcel Dekker Incorporated-*, **78**, 375–402.

Zhao, Xiaoming, & Zhang, Shiqing. 2012. Facial expression recognition using local binary patterns and discriminant kernel locally linear embedding. *EURASIP Journal on Advances in Signal Processing*, **2012**(1), 20.

Zheng, W., & Lu, B. 2015a. Investigating Critical Frequency Bands and Channels for EEG-based Emotion Recognition with Deep Neural Networks. *Autonomous Mental Development, IEEE Transactions on*, **PP**(99), 1–1.

Zheng, W. L., & Lu, B. L. 2015b. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Trans. Auton. Ment. Dev.*, **7**(3), 162–175.

Zheng, Wenchao, & Liu, Cuicui. 2016 (May). Facial expression recognition based on texture and shape. *Pages 1–5 of: 2016 25th Wireless and Optical Communication Conference (WOCC)*.

Zhou, Changjun, Wang, Lan, Zhang, Qiang, & Wei, Xiaopeng. 2013. Face recognition based on PCA image reconstruction and LDA. *Optik-International Journal for Light and Electron Optics*, **124**(22), 5599–5603.

Zhu, Jia-Yi, Zheng, Wei-Long, Peng, Yong, Duan, Ruo-Nan, & Lu, Bao-Liang. 2014 (July). EEG-based emotion recognition using discriminative graph regularized extreme

learning machine. *Pages 525–532 of: Neural Networks (IJCNN), 2014 International Joint Conference on*.

Zhu, Qiang, Yeh, Mei-Chen, Cheng, Kwang-Ting, & Avidan, S. 2006. Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. *Pages 1491–1498 of: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2.

Zhuang, X., Rozgi, V., & Crystal, M. 2014 (June). Compact unsupervised EEG response representation for emotion recognition. *Pages 736–739 of: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*.

Zou, Kelly H, OMalley, A James, & Mauri, Laura. 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, **115**(5), 654–657.