Faculty of Computer Science and Information Technology

Malaysia Ethnicity-based Facial Expression Classification and Emotion Mapping

Izzah Nilamsyukriyah Binti Buang

Master of Science
2021

Malaysia Ethnicity-based Facial Expression Classification and Emotion Mapping

Izzah Nilamsyukriyah Binti Buang

A thesis submitted

In fulfillment of the requirements for the degree of Master of Science

(Computer Science)

Faculty of Computer Science and Information Technology
UNIVERSITI MALAYSIA SARAWAK

2021

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

………………………………

Signature

Name: Izzah Nilamsyukriyah Binti Buang

Matric No.: 17020132

Faculty of Computer Science and Information Technology

Universiti Malaysia Sarawak

Date:

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful.

Alhamdulillah, all praises to Allah for the strengths and His blessing in completing this thesis. This thesis has been kept on track and been seen through to completion with the support and encouragement of numerous people including family and friends.

Special appreciation goes to my supervisor, Dr. Hamimah Binti Ujir, for her supervision and constant support, the invaluable help of constructive comments and suggestions throughout the making to the success of this research. My deepest gratitude goes to my beloved parents Mr. Buang Bin Asin and Mrs. Maimunah Binti Buang and also to my brothers for their endless prayers and encouragement. Sincere thanks to my strong supporters and friends for their moral support, kindness and being there along the struggles, also bring joys and laughter. Thanks for the friendship and memories. Finally, thank all those who have helped me directly or indirectly in the successful of the thesis. JazakAllahu khair. May Allah reward you for the good.

ABSTRACT

Human are commonly good in notifying several emotions via facial expression. In daily human communication, it is crucial for each person to be able to convey his emotions and perceive others respectively using speech, facial expressions and body movements. The computer vision experts are continuously learning on how to achieve high performance in analysing faces, especially which occur spontaneously. Malaysian facial database and analysis are still inconspicuous, especially for local ethnicity studies. Hence, this thesis developed MUA Database, the first Malaysian ethnicity facial database, which consists of data from non-actor subjects from 4 local ethnicities that are Chinese, Iban, Indian and Malay. During the data collection, the subjects are encouraged to express facial expressions spontaneously. Facial expressions analyses are done using the database and facial deformation for each ethnicity is evaluated. From the experiments, the performance of HOG, LBP and SIFT are compared for feature extraction, and SVM, Decision Tree and KNN performance are evaluated as classifier. Results show that the combination of HOG features and KNN classifiers are the best pair for ethnic recognition with 96.90% accuracy, whereas HOG features and SVM classifier combination shows the best pair for emotion recognition with 59.10% accuracy. Indian appeared to be the most recognisable among other ethnicities. As for emotion, “happy” appear to be the most conspicuous emotion, whereas “fear” is the least visible among all tested emotion.

Keywords: Ethnic facial analysis, face recognition, face database, ethnicity classification, ethnic face database

Klasifikasi Ekspresi Muka Berdasarkan Etnik Malaysia dan Pemetaan Emosi

ABSTRAK

Manusia lazimnya boleh mengesan beberapa emosi menerusi ekspresi muka dengan mudah. Penting bagi setiap orang untuk menyampaikan dan memahami emosi orang lain melalui nada dan intonasi percakapan, ekspresi muka, dan bahasa badan dalam komunikasi harian. Pakar visi komputer sehingga kini berterusan mengkaji untuk mendapat pencapaian yang tinggi dalam proses menganalisa muka, terutamanya yang timbul secara spontan. Pangkalan data dari Malaysia masih belum kukuh, khususnya dalam kajian mengenai etnik tempatan. Justeru itu, tesis ini memperkenalkan Pangkalan Data MUA, pangkalan data wajah ethnic Malaysia yang pertama, merangkumi data bagi peserta bukan artis daripada 4 etnik utama yang terdiri daripada Cina, Iban, India dan Melayu. Peserta digalakkan untuk mengekspresikan riak muka secara spontan semasa proses pengumpulan data. Analisa bagi ekspresi muka juga dilakukan menggunakan pangkalan data tersebut sekaligus mengkaji dan menilai perubahan wajah pada setiap etnik. Tambahan lagi, melalui ini turut membandingkan prestasi HOG, LBP dan SIFT bagi proses pengekstrakan ciri serta menilai prestasi SVM, Decision Tree dan KNN sebagai pengelas. Hasil menunjukkan ciri HOG dan pengelas KNN merupakan kombinasi terbaik bagi mengenalpasti etnik dengan 96.90% ketepatan, manakala ciri HOG dan pengelas SVM adalah gabungan terbaik bagi mengenalpasti emosi dengan 59.10% ketepatan. Keputusan menunjukkan India merupakan etnik yang paling mudah dikenali berbanding etnik lain. Emosi "Gembira" merupakan emosi yang paling menyerlah, manakala emosi "Takut" paling sukar dikesan.

Kata kunci: *Analisis wajah mengikut etnik, mengenalpasti wajah, pangkalan data muka, pengelasan etnik, pangkalan data muka mengikut etnik.*

TABLE OF CONTENTS

LIST OF TABLES

TABLE OF FIGURES

LIST OF ABBREVIATIONS

AAM	Active Appearance Models
AFEW	Acted Facial Expression in the Wild Database
AI	Artificial Intelligence
ANN	Artificial Neural Network
AOML	Adaptive Online Metric Learning
AU	Action unit
BU-3DFE	Binghamton University 3d Facial Expression
CAS(ME)2	Chinese Academy of Sciences Macro-Expressions and Micro-Expressions Database
CLM	Constrained Local Model
CNN	Convolutional Neural Network
CV	Confidence value
DBN	Dynamic Bayesian Network
DISFA	Denver Intensity of Spontaneous Facial Action
DL	Dictionary Learning
EUC	Euclidean distance metric
FACS	Facial Action Coding System
FEED	Facial Expression and Emotion Database
FLIR	Forward Looking Infrared thermal camera
HCI	Human Computer Interaction
HOF	Histograms of Optic Flow
HOG	Histogram of Oriented Gradient

ICA	Independent Component Analysis
KNN	K-Nearest Neighbor
LBP	Local Binary Pattern
LBP-TOP	Local Binary Pattern histograms from Three Orthogonal Planes
LDA	Linear Discriminant Analysis
LDCRF	Latent-Dynamic Conditional Random Field
LGBP	Local Gabor Binary Pattern
LLE	Locally Linear Embedding
LoG	Laplacian of Gaussian
LPQ	Local Phase Quantization
MFCC	Mel-Frequency Cepstral Coefficient
MFP	Moving Faces and People database
NNC	Nearest Neighbor Classification
nrML	Non-robust version of metric learning
OEE	Other-ethnicity-effect
p.s.d.	Positive semidefinite
PCA	Principal Component Analysis
PLP	Perceptual Linear Prediction
PLS	Partial Least Square regression
POEM	Patterns of Oriented Edge Magnitudes
RASTA	Relative Spectral
RobustML	Robust version of metric learning
SIFT	Scale-invariant Feature Transform
STASM	Stacked Active Shape Model facial landmark

SVM	Support Vector Machine
UNBC-McMaster	Mc Master University and University of Northern British Columbia
USTC-NVIE	University of Science and Technology China Natural Visible and Infrared facial Expression database

CHAPTER 1

INTRODUCTION

1.1 Introduction

Human facial analyses are often in deliberation for various applications nowadays, starting with scientific to industrial communities competing to create consecutive value to human daily life. The utilization of facial analysis is no longer limited to surveillance and security (Chang, 2004; Beghdadi et al., 2018), but also contribute in cognitive studies (Palestra, 2018; Yu, 2018), biometric (McLain and Kefallonitis, 2019; Wouters et al., 2019), virtual reality (Briggs et al., 2018; Souto et al., 2019), multimedia (Tao et al., 2018; Chen & Nahrstedt, 2019), computer entertainment (Braun et al., 2018), and have significant part in artificial intelligence (AI) applications (Sharma et al., 2019; Wang et al., 2019). In order to create machines that can interact efficiently with human, the development of human computer interaction (HCI) nowadays is no longer restricted to constructing the groundwork of user interface. Various applications have included human cognitive and behavioural model that are producing user-centric design concept which linked to human life in order for the interaction to be more effective and efficient (Wang et al., 2019), as for example the progressive development of mobile phone interface (Lalji & Good, 2008). Therefore, the studies of human faces are crucial in order to develop machine that is able to respond to facial information rather than to depend on input commands from the user.

This thesis will focus on the challenges encountered in analysing facial expressions which are still unsolved, measure the accuracy of existing solutions and also suggest solution for the problems. We also focused to bring the problems locally, so that the specific information and studies can be obtained for the selected geographical area.

1.2 Motivation

The field of computer vision have many applications in various ways nowadays, where machine learning is performed to interpret visual information with an intention to improve the quality of life. It helps in detecting and describing knowledge in images, interpreting and even learning solutions by using the computer (Szeliski, 2010). Face analysis is one of essential process of computer vision. The information from this process enable us to learn a lot more about people, from their identity, ethnicity, age, gender, expression and emotions. These practices are then implemented to various applications, such as in HCI, human evaluation and cognitive, security and a lot more.

The most interesting part to analyse from faces is the emotion. Emotion recognition a primary framework for of emotion regulation, which both important in human relationships as the ground in reasoning, problem solving and enhancing cognitive activities (Mayer et al., 2001). This aspect is essential especially for multicultural country like Malaysia as it can be used to improve the intercultural adjustment between ethnicities (Yoo et al., 2006). In order to classify emotion from faces, a lot of complex algorithms will be involved, which some of them will be discussed in Chapter 2. Principally human display their emotion through facial expressions (Hupont et al., 2010). One of major face recognition and classification problem is usually occurred when large target set involved. Thus, specification of ethnicity, age and gender may increase the accuracy of the identification process. Human cognitively perceive ethnicity before age, gender and expression from face (Wang et al., 2019). Therefore, this research will concentrate on ethnicity-based facial expression analysis. Ethnicity defined as group of people in a region with shared culture, language and some with particular skin colour (Braun et al., 2013). Malaysian facial database and analysis are still inconspicuous, especially for local ethnicity studies. Moreover, there is no existing Malaysia ethnics' database and facial emotion analysis available. So, this research will analyse facial expressions of the main ethnicities in Malaysia which are Chinese, India and Malay, and also Iban, which is one of major ethnic in Sarawak, Malaysia.

The computer vision experts are continuously learning on how to achieve higher performance in analysing faces, especially which occur spontaneously (Saha et al., 2019). Since the study of spontaneous face is imperative and complex (Liu & Yin, 2017), as it has more variations in term of physical peculiarity and class compared to the basic facial expression (Ekman, 2001). Generally, psychologist represent facial emotion in discrete division of six universal emotions, which are "happy", "sad", "fear", "anger", "disgust" and "surprise" (Ekman et al., 1999). However, human daily interaction involves broad range and intensity of emotion other than the six universal emotions (Hupont et al., 2010). Therefore, for this research, other non-basic affective states in describing emotion will be identified using emotional mapping.

1.3 Problem Statement

Humans are commonly good in recognizing several emotion facial expressions. "Happiness" and "surprises" can be easily perceived compared to "anger" and "sadness", and even worse for "fear" and "disgust" (Martinez & Du, 2012). This is probably because the emotions of "happiness" and "surprise" are involving wider face transformations than the rest (Saha et al., 2019). In daily human communication, it is crucial for each person to be able to convey his emotions and perceive others respectively using speech, facial expressions and body movements. These processes are usually dynamic and spontaneous.

Spontaneous expressions give a huge challenge as they are exempt from any intentional attempt, not always noticeable and not fully expressed (Ekman, 1997) due to the coexistence of other basic expressions of emotion (Reisenzein et al., 2013). It is more complicated and the changes are more gradually than the acted one, directing to subtle sequence of expressions (Ekman & Friesen, 1976). This causes fuzzy distinction between different emotions. Each individual also portrays their emotions in various manners, which resulting diverse and confusing information of similar emotions due. For example, Tarnowski et al. (2017) have able to recognised discrete basic emotion using facial expression and shows the most recognition mistakes occurs between each pairs of expression, but failed to represent emotion based on the presence of each expression occurred from a spontaneous expression. As mentioned by Ekman and Friesen (1976), the existence of only one specific basic facial expression of emotion is rarely to appear. Thus, the ability to only detect 6 basic emotion facial expressions distinctively are not enough in order to describe spontaneous expressions due to intra-class diversity (Gazizullina & Mazzara, 2019).

An ideal emotion detection system should be able to recognized expressions regardless of their gender, age and ethnicity besides the ability to be consistent through various diversions including illumination, lightening conditions, glasses, facial hair and hairstyles (Sebe et al., 2004). Izard (2009) questioned whether some emotional expressions are universal, which means the individuals of distinct cultures resulting in similar facial muscle movements when expressing some emotion. In developing facial recognition system, ethnic identification helps on identity-related features, and narrows down the search in large database which will increase the search speed and efficiency of the system. This also applies to demographic statistics in many social applications (Lu & Jain, 2004). Due to ambiguous physical and psychological nature of ethnical group, it is a comprehensive problem in differencing emotions using facial expression because of the existence of similar facial features and characteristics among the ethnicities (Wang et al., 2019). Ma et al. (2019) had developed a facial expression database of Chinese ethnicity, but has not mention any distinction of ethnicity for each facial expression. Thus, there is a compelling need to have a facial analysis on different ethnicities (Wang et al., 2019).

1.4 Objective

The primary objectives of this research are as follows:

a. To determine the association between facial deformities and ethnicity in expressing different type of facial expression
b. To classify the affective values of spontaneous expression from MUA Database through emotion mapping

In this research, a Malaysian ethnicity face database named MUA Database is developed, which also include spontaneous facial expression dataset. From this, it should able to classify spontaneous facial expression images of the database through emotion mapping. The hypothesis for this thesis is the ethnicity categorization may improve the facial expression classification using FACS. Other hypothesis is that the deformation of facial expression using FACS can improved ethnicity classification.

1.5 Scope of Research

The main focus of the project is the study of facial expression classification between Malaysian ethnicity and emotion mapping of spontaneous facial expression using Whissell Space. The study is limited to 4 major ethnicities in Malaysia, which are Chinese, Iban, India and Malay. The development of MUA Database include 200 subjects and this research involve data training, testing and analysis from the database. The emotion mapping from the research also only using spontaneous facial expression from the database. In this research will only compare the performance of 3 facial features which are HOG, SIFT and LBP, and 3 classifiers that are SVM, Decision Tree and KNN.

1.6 Significance of the Study

This research helps to investigate the emotion facial recognition and analysis between ethnicity in Malaysia. The findings of this study will directly benefit to improve local face recognition system and intercultural studies, either cognitively or psychologically, between ethnicities in Malaysia. Main contribution of the thesis is the development of the first Malaysian ethnicity database, MUA Database, and facial expression analysis. This research also contributes in publishing a paper that discuss on Malaysian ethnicity recognition (Buang & Ujir, 2019).

1.7 Thesis Outline

The thesis is organized into five chapters. Chapter 1 introduces the research work. Chapter 2 presents the literature review and current research that bring out the problem and restate the need of the study of facial ethnicity, expression and recognition. Chapter 3 introduces the database that we developed for this research. The chapter also explains experimental procedure and methodology in analysing the database including the facial features and classifier that are used for the experiment. Result and analysis from the experiment are discussed in Chapter 4. Chapter 5 concludes the thesis with a thought on limitations, potential improvements and forthcoming research considerations.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This part of the thesis explains the history and theory required for better understand the purpose of this thesis. Firstly, Section 2.2 explains the differences between spontaneous and non-spontaneous facial expressions. It also includes the groundwork and benefits of studying spontaneous facial expressions. Thereafter, existing spontaneous facial expression database and its analysis were stated. Next, it explained the history of researches including the existing approaches for facial expressions analysis. Then, current studies of ethnic classification and explained the differences between ethnicity using facial expressions were discussed. After that, the comparation of existing ethnicity facial database and their analysis was made. Lastly, Section 2.7 explained the theory of Facial Action Coding System (FACS) and its application in the deformation of facial expressions.

2.2 Existing Works on Facial Expression

Since spontaneous facial expression recognition was naturally expected a greater challenge than posed expression (Wan & Aggarwal, 2014), deep studies have been done to obtained better recognition results. Adaptive Online Metric Learning (AOML) algorithm, a new metric space of facial expression was introduced by Wan and Aggarwal (2014). The precision of annotation is defined by assigning a gold standard label to each expression based on majority voting of annotators. The experiment was done using existing Moving Faces and People (MFP) dataset (O'Toole et al., 2005). Landmark points were automatically placed using Constrained Local Model (CLM) (Saragih et al., 2009) before proceed to feature extraction using Gabor filters. The recognition process was then proceed using a Robust Metric Learning approach (RobustML) and Adaptive Online Metric Learning algorithm, which helped obtaining faster convergence by adjusting the gradient adequate step size and decreasing computational tasks via substitution of eigen-decomposition based positive semidefinite (p.s.d) with much simple two-case function. The recognition accuracy method was then compared with Euclidean distance metric (EUC), Isomap (Tenenbaum et al., 2000), Locally Linear Embedding (LLE) (Saul & Roweis, 2000), SVM, Latent-Dynamic Conditional Random Field (LDCRF) (Roweis & Saul, 2000) and non-robust version of metric learning (nrML). From the comparison, RobustML and nrML were the most outstanding in differentiating the "neutral", "fear" and "sad" expression, whereas LDCRF resulted the highest recognition accuracy.

Research by El Meguid et al. (2014) deliberated the framework for fully automated spontaneous facial expression from videos using PittPatt face detector (Pittsburgh, 2011) and Random Forest tree paired with SVM classifier. The experiment done using 7 expressions of Binghamton University 3D Facial Expression (BU-3DFE) database for training and tested with few labelled spontaneous databases (BU-4DFE (Yin et al., 2008), FEED (Wallhoff, 2006) and AFEW database (Dhall et al., 2011). For this experiment, BU-4DFE obtained the best results because it originated in the same set up as the training database. Most of the confusions from the experiment were oriented by the low-intensity of the expressions. FEED database obtained lower result from incorrect frame labelling for certain video clip and also from low intensity expressions. AFEW database that contained acted spontaneous expressions resulting lowest result among all. This may due to the posed expression from actors may not represent actual human emotion.

Liu and Yin (2017) proposed analysing spontaneous facial expression using thermal data. From this experiment, an infra-red thermal video descriptor was introduced. USTC-NVIE (Wang et al., 2010) database was used for this experiment aside from their own database captured using Forward Looking Infrared (FLIR) thermal camera consist of 77 subjects. The facial data were then extracted using scale-invariant feature transform (SIFT) method. Thermal and motion video words were obtained by using K-means clustering algorithm which represented the average pooling function of the clip before the classification process via SVM. SIFT method had been compared with other descriptors such as HOG, histograms of optic flow (HOF) and Cuboid descriptor, which obtained highest accuracy and precision among all.

Next research is done by Li et al. (2015) that focused on modelling dynamic and semantic connection among AU intensities using Dynamic Bayesian Network (DBN). The experiment firstly registers 66 facial points of the DISFA database before the process of facial extraction using HOG and Localized Gabor features. The features were then processed into manifold learning, which helped preserve local information of facial appearances for classification process of facial expressions. The AU intensities are then classified using SVM and using DBN to depicted relationships among each AUs with its intensities.

Mohammadi et al. (2015) suggested a method in acknowledging the existence and absence of AUs by using sparse representation and estimated their intensities via dictionary learning (DL) approach for sparse image regression. The experiment was done using DISFA and UNBC-McMaster (Lucey et al., 2011) database. The approach emphasised the ability to joint learning of all AUs, in which notify coexistence of AUs that indicate the presence of both emotional expression and non-emotion expressions which allow the filtering of the non-emotion facial movements.

In summary, list of all revised studies on methodology in analysing spontaneous facial expression are listed in Table 2.1.

2.3 Spontaneous Facial Expression Database

There are researches working on spontaneous expression database that will be discussed in this section. Mavadati et al. (2013) proposed the Denver Intensity of Spontaneous Facial Action (DISFA) database which included labelled video with the ground truth. The database contains data from 27 young adults, in which their spontaneous expressions were video recorded via stereo camera whilst watching video clips proposed to evoke their emotions. The intensity of facial action units (AUs) were manually coded according to FACS before extracted using active appearance models (AAM) and the AUs intensity measurement were recorded. The feature extraction was done using Local Binary Pattern (LBP), Histogram of oriented gradient (HOG) and Gabor features, and reduced using spectral regression in which each of the features performance were compared after classified using Support Vector Machine (SVM).

In the following year, Yan et al. (2014) came out with micro-expression database with 200 frame per second (fps) temporal resolution and 280 x 340 pixels' spatial resolution on facial area. 247 micro-expressions with AUs and emotions were picked and marked for the database. Feature extraction and classification were done using LBP histograms from Three Orthogonal Planes (LBP-TOP) and SVM respectively. The improvement to this database was published in Qu et al. (2017) which have 2 parts, one consists of 87 long videos

Table 2.1: Comparison of Methodology in Analysing Spontaneous Facial Expression

Author & Year	Tested Database	Methodology	Feature Extraction	Classifier	Result
Wan and Aggarwal (2014)	Dynamic Facial Expressions recordings of MFP consist of 76 males and 208 females students of The University of Texas	Introducing Adaptive Online Metric Learning algorithm, a new metric space of facial expression and assigning a gold standard label to each expression based on majority voting of annotators	Gabor filters	Robust Metric Learning approach and Adaptive Online Metric Learning algorithm	Robust Metric Learning method shows better results in detecting spontaneous expressions compared to EUC, Isomap, LLE, LDA, SVM, LDCRF, and nrML
El Meguid and Levine (2014)	Training: Total 3829 images of BU-3DFE Testing:	Deliberated the framework from videos using PittPatt face detector and Random Forest tree paired with SVM classifier	PittPatt face detector	Random Forest tree paired with SVM classifier	BU-4DFE obtained the best results because originated in the same set up as the training database

	Still images and videos of spontaneous BU-4DFE, FEED and AFEW database				

Table 2.1 continued

Li et al (2015)	DISFA contained 27 young adults, included labelled video with ground truth	Modelling dynamic and semantic connection among AU intensities using DBN	HOG and Localized Gabor feature	SVM	Gabor features showed higher accuracy compared to HOG
Mohammadi et al. (2016)	DISFA and UNBC-McMaster (consist of shoulder pain expression from 129 participants)	Detecting existence and absence of AUs by using sparse representation and estimated their intensities via DL approach for sparse image regression	Sparse representation	DL	Jointly learned regression method showed better result than person-independent intensity estimation

Liu and Yin (2017)	USTC-NVIE with posed and spontaneous expressions of 215 subjects	Introducing infra-red thermal video descriptor for spontaneous facial analysis	SIFT	SVM	Obtained highest accuracy when compared with HOG, HOF and Cuboid descriptor

of spontaneous expressions and micro-expressions, and the other contains 300 cropped spontaneous expressions and 57 micro-expressions samples. The participants were asked to watch their recorded facial expressions and give their own annotation for each expression. Similarly, the feature extractions and classification were done using LBP features and SVM classifier.

Happy et al. (2015) introduced spontaneous facial expressions database of Indian origin which involves 50 participants, producing 428 fragmented video clips of emotions elicited using emotional videos and self-ratings. Facial expressions from the clips were remarked by trained decoders to obtain the peak frames before further evaluation. The experiment using grey intensities, LBP, Gabor filter, HOG, and Local Gabor Binary Pattern (LGBP) for feature extraction. In this research, Linear Discriminant Analysis (LDA) performed well as the classifier compared to the SVM, Adaboost, K-Nearest Neighbor (KNN) and Naïve Bayes.

Other spontaneous databases developed to not only include facial data. Zhalehpour et al. (2016) had emphasized the importance of including audio messages in addition to the face of the spontaneous data to recognize emotional expressions. The database contained recordings from 31 subjects of Turkish native speakers expressing the 6 basic emotions ("happy", "anger", "sad", "disgust", "fear" and "surprise") along with boredom and contempt affective. Apart from that, the database also included several mental states such as unsure (confused, undecided), thinking, concentrating and bothered from watching stimuli videos. The videos were remarked by using GTrace tool for detecting facial features, which then extracted using Local Phase Quantization (LPQ) features and Patterns of Oriented Edge Magnitudes (POEM) features. The speech features were extracted using the Mel-Frequency Cepstral Coefficients (MFCC) and relative spectral features (RASTA) based on perceptual linear prediction (PLP). Both of the face and speech features were then passed to SVM for classification.

On the other hand, Girard et al. (2017) claimed to be the first to introduced facial expression database of multiple interacting participants. The database consists of 172800 video frames of 96 participants forming three-person groups, which were seated around a circular table, and alluring in a collection of cognitive tasks. Five AUs (AU 1, 6, 10, 12 and 14) were selected for intensity coding to help in estimating the emotion. Feature extractions were done using HOG descriptors and trained using SVM. Table 2.2 shows the summary for comparison of spontaneous facial expression database that were discussed in Section 2.3.

2.3.1 Spontaneous and Non-spontaneous (Posed) Facial Expression

The origins of facial expression works are triggered from a century ago where in Darwin (1872) raised questions and observations in variations of emotions and also distinguished family-related emotions such as “rage”, “anger”, “hatred”, “defiance” and “indignation”. Darwin (1872) stated that it is possible to differentiate photographs of enjoyment and non-enjoyment smiles without any conscious process of analysis on our part. Darwin (1872) also initiated the theory that facial expressions are universal and justified the universality for some facial expressions via the answers of questions to Englishmen living in eight parts of the world, which are Africa, America, Australia, Borneo, China, India, Malaysia and New Zealand, whether or not they saw the same expressions of emotions from the countries. Though the theory is still questionable, these studies has motivated the ideas in Ekman (1973) to find the relation of facial expression across culture and determined to find the involvement of particular facial muscle movement for facial expressions. In Ekman and Friesen (1971) found that when evaluating people of certain cultures that they have never

Table 2.2: Comparison of Spontaneous Facial Expression Database

Author & Year	**Database**	**Methodology/ Experiment**	**Feature Extraction**	**Classifier**	**Result**
Mavadati (2013)	DISFA database. 27 young adults, included labelled video with ground truth	Video recorded via stereo camera whilst watching video clips proposed to evoke their emotions. AUs manually coded according to FACS	LBP, HOG and Gabor features and reduced using spectral regression	SVM	Gabor features yielded the best result among the three facial representations with 86% accuracy of a single AU.
Yan (2014)	CASME II database. 247 micro-expressions with AUs and emotions were picked and marked	Micro-expression database with 200 fps temporal resolution and 280 x 340 pixels spatial resolution on facial area.	LBP-TOP	SVM	The best performance were 63.41% in detecting emotion based on AU combination

Table 2.2 continued

Qu (2017)	CAS(ME)2 database. 2 parts; 1 consist 87 long videos of spontaneous and micro-expressions, another contains 300 cropped spontaneous expressions and 57 micro-expressions samples	The participants were asked to watch their recorded facial expressions and give their own annotation for each expressions	LBP	SVM	Best accuracy were 40.95% and lowest were 28.09% in detecting micro-expressions
Happy (2015)	Indian database involving 50 participants, producing 428 fragmented video clips of emotions	Emotions elicit using emotional videos and self-ratings were collected for validation.Facial expressions from the clips were remarked by trained decoders to obtain the peak frames	Grey intensities, LBP, Gabor filter, HOG, and LGBP	LDA in comparison with SVM, Adaboost, KNN and Naïve Bayes	Combination of LBP and LDA shows the highest recognition accuracy with 86.46%

Table 2.2 continued

Zhalehpour et. al (2016)	Recordings with voice from 31 subjects of Turkish native speakers expressing the 6 basic emotions along with boredom and contempt affective. Also included several mental states such as unsure (confused, undecided), thinking, concentrating and bothered	Using GTrace tool for detecting facial features	Facial feature: LPQ and POEM Speech feature: MFCC and RASTA	SVM	Emotion recognition accuracies for visual and audio separate modalities were 42.16% and 72.95% respectively. The experiment for combining both modalities achieved 77.02% accuracy.
Girard et al. (2017)	Facial expression database of multiple interacting participant, consist of 172800 video frames of 96	5 AUs (AU 1, 6, 10, 12 and 14) were selected for intensity coding to help in estimating the emotion	HOG	SVM	A-level occurred 42% and B-level 39%, whereas rarely to occurred C-level with 15%, D-level

Table 2.2 continued

	participants forming three-person groups				occurred 13% and E-level with 1% based on intensity category.

exposed to, there are probability that they could not recognize certain facial expressions especially “fear” and “surprise”. Mead (1975) challenged the theory and stated that the universal expressions are only for posed expressions and not spontaneous. This statement then widens the studies and findings in learning the spontaneous facial expressions.

Woodworth (1938) has impugn the accuracy in describing spontaneous as well as posed facial expression. In respond to this, Ekman and Friesen (1978) come out with the idea that facial expressions can be measured via movement of facial features and issued solution to objectively measure the abundance and complexity of facial expressions using the FACS, which are discussed further in Section 2.7. With the existence of standard coding system, the research on both posed and spontaneous expressions are expanded.

Spontaneous facial expressions are different from posed expressions, in term of their psychophysical characteristics (Ekman, 2001) to the experimental approaches in determining the expression (Bartlett et al., 2005). Subject tends to show different facial movements when asked to pose emotion expressions. For example, their pose for fear are different compared with the one portrayed when they actually enduring fear. Although it is more precise to describe emotions using spontaneous expressions, but it is quite difficult to label and describe. This is because of the coexistence of other basic expressions of emotions (Reisenzein et al., 2013) which support the statement by Ekman and Friesen (1976). Most of the existing recognition systems are using posed facial expressions, but real-life situations commonly encounter with spontaneous facial expressions (Mavadati et al., 2016).

2.4 Ethnic-specific Classification

According to Jilani et al. (2017), ethnicity is one of the most notable face identities. Nevertheless, the analysis of ethnicity-specific for faces is a challenging puzzle for computer-based experiments. For that, a study of classifying British Pakistani ethnicity with other was conducted using total of 75 face images, which are not limited to frontal image of 135 multi-ethnic subjects. The approach of the experiment was using three fundamental components which are geometric feature-based extraction, reduction using PCA and Partial Least Square Regression (PLS), and classification using SVM. The experiment achieved promising results but limited specifically for British Pakistani facial images.

The research conducted by Zuo et al. (2017) have establish face database with 254 individuals' faces of Uighur and Kazakhs of the Xinjiang province of China which contains five expressions which are "neutral", "smile", "angry", "surprised" and "closed eyes", with four different angles, three illumination variations, wearing glasses and mask. Experiments were separately conducted using PCA, Independent Component Analysis (ICA), SIFT and LGBP. However, the database has not been tested for ethnicity and emotion recognition. Furthermore, the facial expressions are limited to "neutral", "smile", "angry" and "surprised" expression.

The study by Masood et al. (2018) solved problem of ethnicity classification using three major ethnicities which were Mongolian, Caucasian and Negro from 447 images of FERET database using feature-based approach and differentiate the efficiency using Artificial Neural Network (ANN) and Convolutional Neural Network (CNN). From the experiment stated that CNN produced much better result but cost more time for feature extraction and training of the network. Yet, for this experiment, the face images obtained from FERET database have not indicated any standardization in term of their expression. The images may be assumed to acquire neutral expression but still having the smiling faces for certain images. This may affect the studies that initially want to see the different level of expression intensity between features for different races.

In 2019, Wang et al. (2019) performed facial analysis experiment and established ethnical group dataset of Chinese Uyghur, Tibetan and Korean. 100 students from each ethnicity were selected and their frontal facial images were captured with different poses and expressions. However, the databases have not standardized any type of expressions. The ethnicity recognition experiment was performed using the KNN-based fast sparse sensing. The whole training set for each group are used by the KNN to classify a sample with the help of super resolution approach. Then, the local facial features were analysed using Stacked Active Shape Model (STASM) facial landmark detector which helps to extract 77 landmarks in each facial image to determine the ethnicity. However, the result obtained was quite low (63%) for face recognition but still suitable for ethnicity recognition. In summary, a list of all revised ethnicity databases is as listed in Table 2.3.

The studies in psychology prove that when observing human face usually will trigger three conscious neural evaluations which are ethnic, gender and age (Fu et al., 2014). Among them, ethnic is said to be the most outstanding attribute to be conceived by series of social cognitive and perceptual task (Calder et al., 2011). However, the bottom line of the problem rises as the computational mechanism facing complexity to classify an ethnic's face (Fu et al., 2014). Researches have been carried out to deal with the problem. For example, researches in psychology have deliberated behaviour interrelations of race perception such as other-ethnicity-effect (OEE) and attention model in Calder et al. (2011), Rossion and Jacques (2002), Freeman et al. (2010), Eberhardt et al. (2003) and Bentin and Deouell (2000) which portrayed the existence of racially-discriminative facial features. Computational neuroscientists also have produced models to trigger and describe race perception in Bowyer et al. (2006), Chang et al. (2005), Phillips et al. (2005), Phillips et al. (2010) and Kurzban et al. (2001). Experiments by

Table 2.3: Comparison of Ethnicity Database

Author & Year	**Database**	**Methodology/Experiment**	**Feature Extraction**	**Classifier**	**Result**
Jilani et al. (2017)	The database using total of 75 face images, which are not limited to frontal image of 135 multi-ethnic subjects	Classifying British Pakistani ethnicity using geometric feature based extraction, reduction using PCA and PLS, and classify using SVM	Geometric feature and reduced using PCA and PLS	SVM	PCA obtained 71.11% accuracy whereas PLS have higher accuracy with 76.03%
Zuo et al. (2017)	254 individuals' faces of Uighur and Kazak. Contains five expressions (neutral, smile, angry, surprised and close eyes), with four different angles, three illumination variations, wearing glasses and mask	Separately conduct experiment using PCA, ICA, SIFT and LGBP.	PCA, ICA, SIFT and LGBP.	KNN	LGBP-based method showed the best overall performances compared to other descriptor in detecting expressions for all individuals

Table 2.3 continued

Masood et al. (2018)	447 images of FERET database that consist of Mongolian, Caucasian and Negro	Using feature-based approach and differentiate the efficiency using ANN and CNN	Geometric feature detected using Viola Jones	ANN and CNN	CNN give the best performance of 98.6% accuracy compared to ANN (82.4%)
Wang et al. (2019)	100 students for each ethnicity of Chinese Uyghur, Tibetan and Korean. Their frontal facial images were captured with different poses and expressions	Experimented using the KNN-based fast sparse sensing and super resolution approach, with help of STASM facial landmark detector	STASM facial landmark detector	KNN and super resolution approach	T regions results better than O region on ethnicity recognition (with 78% accuracy) compared to face recognition, which only obtained 63% accuracy while O region obtained 90%.

Levine (1996), Levine (2000) and Levine and Angelone (2002) also revealed a notable visual factor of racial features.

As for the process of classification of ethnicity, it should begin with the elemental clarification and perception of ethnicity itself. Ethnicity defined as group of humans with similar gene, culture and language in a certain geographical locality (Shiyuan, 2002). Yet, the ethnic classification often defined as the grouping via skin colour or tone, physical attribute such as hair shaft peculiar, and also human phenotypic feature (Berardesca et al., 2006). However, such aspect was not applicable for computer vision methodology.

Nonetheless, the implementation of basic algorithm for ethnic classification tends to be varied and sophisticated (Fu et al., 2014). Firstly, the classification of ethnicity is still puzzling by variety of perspectives, which lead to ambiguity in formulation and methodology (Mays et al., 2003). Secondly, in order to construct competitive automatic race recognition, a large-scale database is needed to be trained and established (Fu et al., 2014).

In 1994, Farkas (1994) adopt 25 measurements from head and facial landmarks to differentiate ethnic morphology between three groups, which are North-American Caucasians, African-Americans and Chinese. The study suggested the used of information of distinctive characteristics for each ethnic to enhance recognition accuracy. Lu et al. (2006) introduced a multimodal method to identify ethnicity using SVM using depth and texture facial data. The results showed that depth approaches were more descriptive in comparison to intensity modality for ethnicity classification. Zhong et al. (2009) proposed a fuzzy 3D face ethnic classification algorithm using Gabor filter to differentiate between Eastern and Western people. The algorithm showed more than 74% accuracy in categorizing both Eastern and Western individuals.

Tin and Sein (2011) focused on ethnic recognition based on facial images using Principal Component Analysis (PCA) and Nearest Neighbor Classification (NNC). The average accuracy rate for the approach was 96.4%. Berretti et al. (2012) then come out with a local approach to 3D face recognition based on iso-geodesic stripes (Berretti et al., 2006) combined with minimal-redundancy maximal-relevance feature selection model in order to analyse the differences of relative relevance of the facial stripes from ethnic different races. The results showed 90.3% accuracy in determined Asian and Caucasian individuals.

2.5 Facial Action Coding System

Ekman and Friesen (1978) initially introducing Facial Action Coding System (FACS) with the idea to measure facial movements that are observable on the face with the foundation of 44 unique AUs together with the movements of head and eye positions. From the 44 AUs, 30 of the AUs are anatomically linked to human facial muscle contractions, in which 12 are at the upper face and 18 at the lower face. The AUs can be coded into five-point intensity scale.

Originally, the use of FACS are not limited to measure movements that are related to emotions, but also helps the study of emotions by distinguish emotional and non-emotional facial expressions (Ekman, 1997). The combination of AUs helps to describe the details of facial expressions (Tian et al., 2001). These facial movements resulting changes in the vector of skin surfaces, thus highlight the importance of tracking the facial features (Ekman, 1997). The whole lists of AUs are listed in Appendix A (Ekman et al, 2002).

There are several developing researches on describing emotion based on AUs movements. Du et al. (2014) and Calvo et al. (2018) have categorized emotion expressions with related AUs. Both of the researches agreed that “anger” involves AU4 and AU7. In addition to that, Du et al. (2014) suggested involvement of AU6 as variant AU. As for “disgust”, both mentioned the involvement of AU4, AU9 and AU10, other than that Du et al. (2014) included AU17 and AU24 while Calvo et al. (2018) suggested existence of AU6 and AU7. AU1 and AU5 had significantly appeared in fear, although Du et al. (2014) also suggested the presence of AU2 AU4, AU20, AU25 and AU26. Both of the researches also agreed with the presence of AU6, AU12 and AU25 in “happy”. For “sad” emotion, AU4 and AU15 are mentioned in both studies, while Du et al. (2014) also mentioned the existence of AU6, AU11 and AU17 whereas Calvo et al. (2018) mentioned AU1 and AU24. Finally, both studies agreed with the presence of AU1, AU2, AU5, AU25 and AU26 for “surprise” expression.

In summary, the representation of AUs according to emotions is summarised in the Table 2.4, with evidence scores of the AUs existence.

Table 2.4: Representation of AUs According to Facial Expression of Emotion

Basic Facial Emotions	**Du et al.** (2014)	**Calvo et al.** (2018)
Anger	AU4, AU6 (51%), AU7 and AU24	AU4 (+1.86) and AU7 (+0.89)
Disgust	AU4 (31%), AU9, AU10, AU17 and AU24 (26%)	AU4 (+1.70), AU6 (+1.02), AU7 (+1.28), AU9 (+3.48), and AU10 (+3.55)
Fear	AU1, AU2 (57%), AU4, AU5 (63%) AU20, AU25 and AU26 (33%)	AU1 (+1.58), AU5 (+1.26) and AU25 (+1.48)
Happy	AU6 (51%), AU12 and AU25	AU6 (+2.88), AU12 (+4.06) and AU25 (+2.07)
Sad	AU4, AU6 (50%), AU11 (26%), AU15 and AU17 (67%)	AU1 (+1.38), AU4 (+1.55), AU15 (+0.98)

Table 2.4 continued

Surprise	AU1, AU2, AU5 (66%), AU25 and AU26	AU1 (+1.51), AU2 (+1.93), AU5 (+1.99), AU25 (+2.58) and AU26 (+2.27)

2.6 Summary

There are several existing works developing on spontaneous and ethnicity database separately, as well as in studying the efficiency of methodology on analysing spontaneous facial expression. From the researches reviewed, there is still an obvious gap to study different expressions for different ethnicity. Moreover, ethnicity database specifically for Malaysian faces do not exist. It also can be said that there is still a need in analysing spontaneous expression of an ethnicity database.

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter explains the methodology in MUA Database creation and facial analysis processes. Firstly, Section 3.2 elucidated the creation of MUA Database. Next, Section 3.3 explains the facial analysis process which initiated with the facial feature extraction. It describes the theory involved in three descriptors involved, which are HOG, LBP and SIFT. Section 3.4 described the classification process, which includes the explanations on three classifiers involved, SVM, Decision Tree and KNN. Section 3.4 discuss the mapping of the spontaneous expression using Whissel Space. In a nutshell, the overall processes of the proposed approach are explained in Figure 3.1.

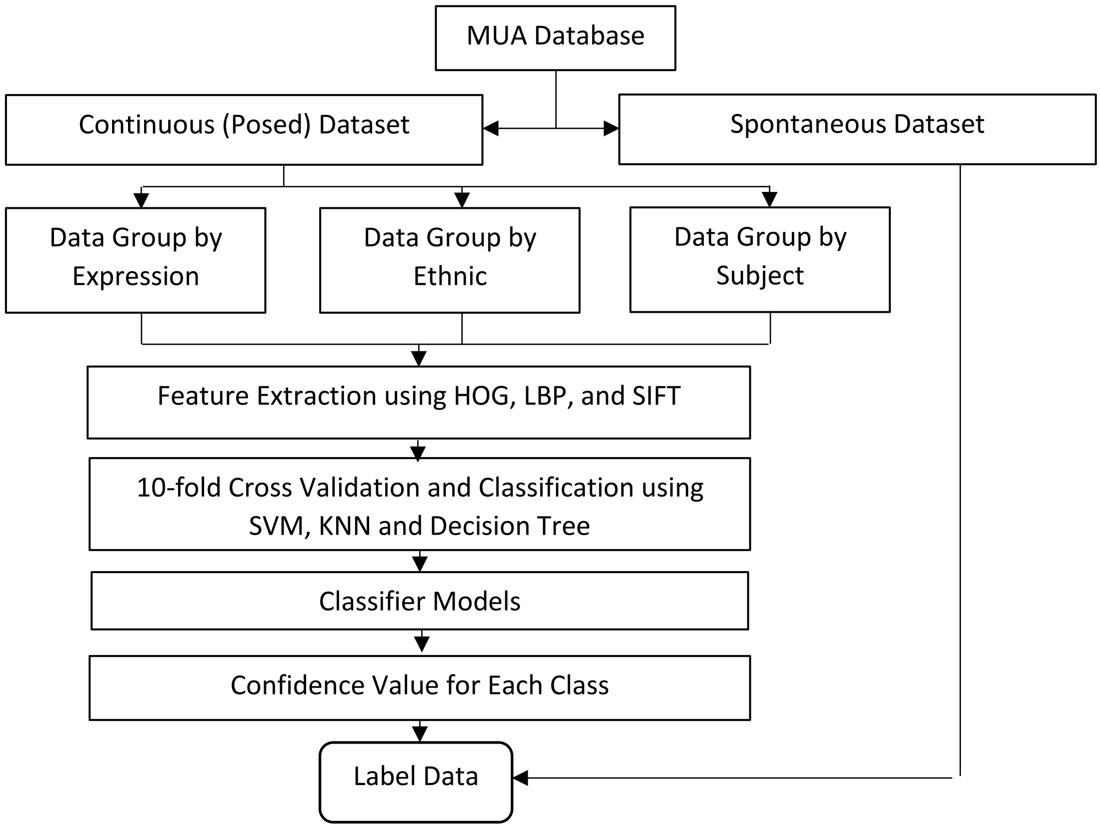


Figure 3.1: Overview of Proposed Approach

3.2 The Database Creation

This research has contributed in developing new database, called MUA database, gathering Malaysian ethnicity face database which include major ethnic in Malaysia, Malay, Chinese, Indian and also Iban, comprise the largest population now in Sarawak. “MUA” is apt from Iban language with the meaning of face. Malaysia as being multiracial country, we believe the availability of the database will bring benefits for social cognition research, especially in Malaysia.

MUA database is consist of 200 subjects where 50 Malays, 50 Chinese, 50 Iban, 50 Indian and an additional of 5 subjects from other races for outliers. The total subjects include 75 males and 125 females. Most of the subjects are students and working adults with approximation age range of 19 to 45 years old.

3.2.1 Data Collection

The subjects are not professional actors and are not trained for displaying each specific emotions in order to obtained natural emotion expressions from the experiment. The subjects had to involve in two type of data collection process: (i) spontaneous and (ii) continuous. For spontaneous experiment, 10 videos to have been shown to the subjects and the facial expressions recorded by using Kinect sensor. The videos are containing “neutral”, “angry”, “disgust”, “fear”, “happy”, “surprise” and “sad” contents. The subjects reported their intensity of emotions after each session of spontaneous experiments to record the ground truth of their emotion. They were given a list of six basic emotions (“angry”, “disgust”, “fear”, “happy”, “surprise” and “sad”) with low, medium and high intensity and also natural.

Continuous experiment has been conducted by recording the seven types of expressions ("neutral", "angry", "disgust", "fear", "happy", "surprise" and "sad") from each subjects. For each expression, 3 takes have been recorded which contains from low to high intensity of the facial expressions. In order to avoid expressions overlapping, a demo video for each expression has been shown to the subjects before recording the specific expressions. Moreover, for immediate visual feedback a mirror was placed in front of the subjects. Some videos recorded with subjects wearing glasses, partially occluded by hair, wearing scarf, and cap. Seven types of facial expressions including "neutral", "angry", "disgust", "fear", "happy", "surprise" and "sad" have been recorded from all of the subjects.

3.2.2 Data Collection Using Kinect

Kinect sensor able to acquire both colour and dense depth images. The sensor integrates structured light with depth from focus and depth from stereo. It also includes infrared laser-based IR emitter, an infrared camera and a RGB camera. The IR camera and projector compose a stereo pair with approximately 75 mm standard, whereas IR laser emitter transmit a known pattern of dots. Images from the camera correlate with different camera position since there is a distance between laser and sensor which allows the operation of stereo triangulation to compute each spec (no short format) depth. The depth and colour images are captured simultaneously at a frame rate of about 30 fps. The RGB frame has size 640×480 and 8-bit for each channel, whereas the depth frame with 11-bit depth. Depth is favourable in face detection and tracking considering that face may not have consistent colour and texture. In the meantime, it able to occupy an integrated region in space. The camera has 57^{o} horizontally and 43^{o} vertically range of view, with the minimum measurement range of 0.6 meter (m) and maximum range is somewhere between 4-5 m.

Thus, for the experiment, the camera is set up at range 1 m in front of the subjects as shown in Figure 3.2.

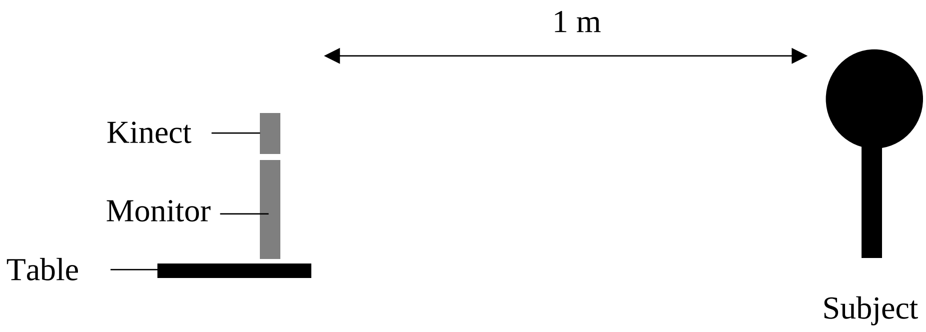


Figure 3.2: The Experimental Setup

With the help of the sensor, Windows SDK and the Face Tracking SDK, the Kinect enable to developed applications to track human faces in real-time. The face tracking engine determines 3D positions of semantic facial feature points as well as 3D head pose. It tracks the 3D location of 121 points. Additionally, the Face Tracking SDK fits a 3D mask to the face. The 3D model is based on the Candide 3 model (Universitet, 2012), which is a parameterized 3D face mesh specifically developed for model-based coding of human faces. This 3D model is widely used in head pose tracking (Saeed & Al-Hamadi, 2015).

As already indicated, Kinect sensor allows low cost sensing with high capture speed. However, the 3D maps provided by Kinect are very noisy and have relatively low resolution in comparison to typical devices, such as thermal cameras utilized in facial expression recognition. In consequence, many important fiducial points such as eye and mouth corners are not precisely locatable. Even more, some fiducial markers undergo occlusion, particularly the points that are located closed to the nose.

3.2.3 Database Structure

The database contained of full RGB images, face RGB images and 2D facial points retrieved from the experiment. Each of the datasets was arranged by (i) subjects, (ii) emotions and (iii) ethnics accordingly. The hierarchy of the database is best explained via Figure 3.3.

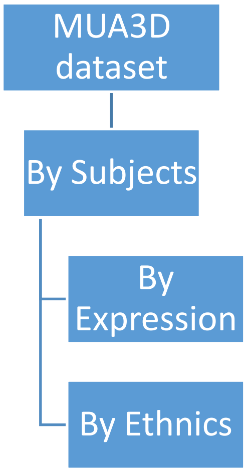


Figure 3.3: Hierarchy of MUA3D Database

Figure 3.4: Facial Images for 7 Basic Expressions (Angry, Disgust, Fear, Happy, Neutral, Sad and Surprise)

Figure 3.5: Facial Images for 4 Different Ethnicity (Iban, Malay, Indian and Chinese)

3.3 Feature Extraction

First step in analysing the faces is to go through the images or data and extract the pixels that can be used to study the faces. Without much information, our machine may not able to decide the desired characteristic precisely. However, too much of information also can lead to confusion thus increase the false positive decisions. The fundamental thought is to get information which is significant to the classification task which includes edges, points or even objects. Therefore, this research proposed to compare several techniques in order to extract all relevant shape features for the classifier to train, which are HOG, LBP and SIFT.

3.3.1 Histogram of Oriented Gradients

Histogram of Oriented Gradients (HOG) is one of a global descriptor in which it describes and generalizes the image as altogether. HOG act by a sliding window through the image which initiated on a square cell, as for example 8x8 cells that to find the local intensity gradients. Next, the unspecified gradient vector in each 8x8 cells is calculated, producing size 64 gradient vectors. Each magnitude from the vector then split between neighbouring bins of histogram, depending on the angle of the histogram. The stronger gradient vector has more affect to the histogram. Figure 3.6 shows an example of the HOG process in an image window (Shu et al., 2011). After going through all cells in the image, the histogram then normalized in blocks with its neighbouring histogram to rule out sensitive effect from the illumination and variation of contrast. The same formula of normalization applied, in which the bins in 4 neighbouring histograms are divided by the total magnitude. The feature vectors are then used for training process (Dalal & Triggs, 2005).

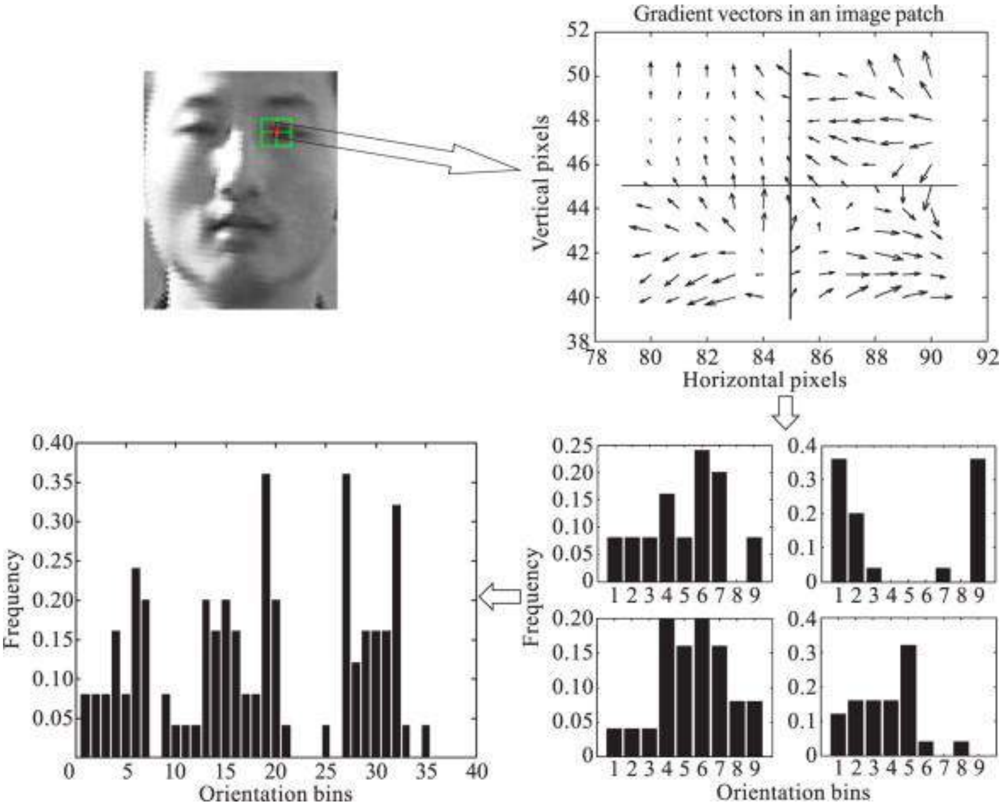


Figure 3.6: The HOG Operation (Shu et al., 2011)

3.3.2 Local Binary Pattern

Local Binary Pattern (LBP) is one of the best performing pattern descriptors and is extensively applied in various fields. It labelled every pixel of an image by threshold the centre pixel of 3x3 neighbour pixel with the value 1 if the intensity is lower than the centre pixel and 0 if the intensity is higher than the centre pixel for entire image. The 8-bit values are then converted into decimal as shown in Figure 3.7 (Ahonen et al., 2006).

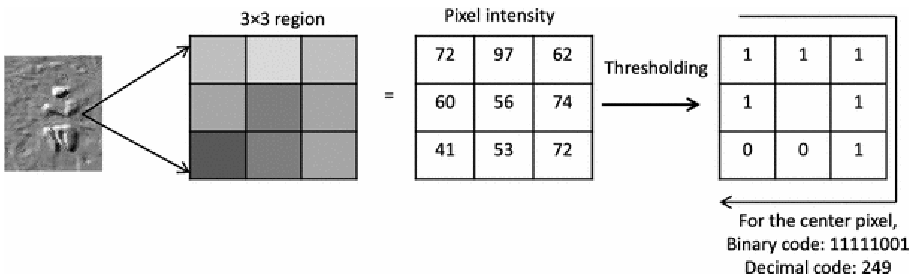


Figure 3.7: The LBP Operation (Ahonen et al., 2006)

This then formed the histogram with 8 bins of the labelled image, which contains the information of the intensity pattern for the image.

3.3.3 Scale Invariant Feature Transform

Lowe (2004) has introduced Scale Invariant Feature Transform (SIFT) algorithm that able to extract distinctive invariant features from images that either in different scale and rotation, and also changes in viewpoint and illumination. There are four fundamental steps for the algorithm. Firstly, the Laplacian of Gaussian (LoG) process, where the image will be applied for blurry effect to blur out the noises and the Laplacian derivative are calculated. From this, the edges and corners of the images are located. However, since Laplace can be intensively computed, this algorithm uses the difference between two consecutive scale space called Difference of Gaussian as shown in Figure 3.8.

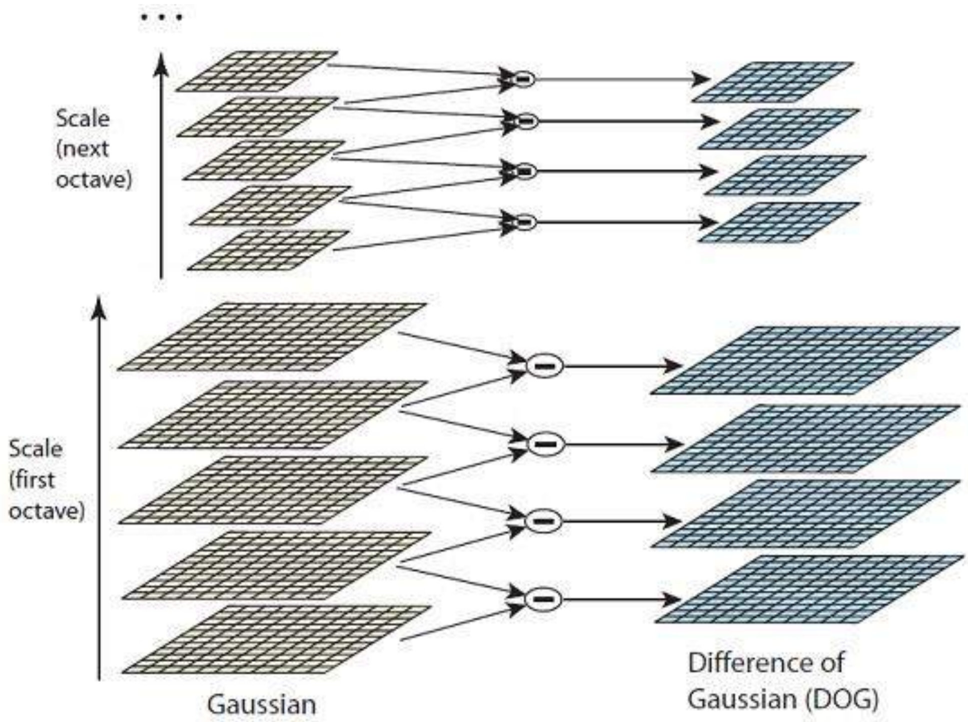


Figure 3.8: Difference of Gaussian Method (Rattani & Tistarelli, 2009)

For each octave of scale space, the initial image is repeatedly convolved with Gaussians to produce the set of scale-space images. The adjacent Gaussian images are then subtracted to produce the Difference of Gaussian images, which help to down-sampled it by factor of two. The process is then repeated until the maximum iteration is reached. This method has mathematically made the Laplacian of Gaussian to be scale invariant (Lowe, 2004). After detected all the key points' locations, low-contrast and edges key points are eliminated. Then, the next step involves orientation assignment to each remaining key points so that it will be invariant to rotation. An orientation histogram with 36 bins is then created where the neighbours of key points are gathered depends on the scale, gradient magnitude and direction of the region. This orientation histogram then produced the key descriptor. 36 bins orientation histogram then formed into a 16x16 block, which then divided into 16 sub-blocks of 4x4 sizes. 8 bins histogram is then constructed from the sub-blocks which representing a vector from the key descriptor. These values are then extracted for training process.

3.4 Classification

Classification is the most crucial part in machine learning process. The optimal result to determine the best classifier are said to be subjective to the problem domain (Cover & Hart, 1967). In the direction of choosing the best classifier, three classifiers were selected for the experiment which are SVM, Decision Tree and KNN. The classifiers are set and trained using *k*-fold cross-validation method in all experiments in this research. At each fold, nine subsets are used to train the classifier while one subset is used for validation.

3.4.1 Support Vector Machine

Support Vector Machine (SVM) is a machine learning algorithm that has a role in separation of classes as a classifier. It applies statistical learning theory by dividing training data using a decisions boundary, called kernel (Cristianini & Shawe-Taylor, 2000). Generally, the boundary exists in the format or hyperplane in multidimensional space. Theoretically, the fundamental in implementing SVM is the mathematical programming techniques and kernel functions (Burbidge & Buxton, 2001). In real-world applications, problems are usually complex. Thus, SVM exists with tuning parameters (regularization parameter, gamma and kernel) in order to achieve good results based on the tested dataset.

While using SVM, normally researcher confronts with the choice of which kernel to use thenceforth how to specify the parameters. As for kernel, the researcher has to choose an order to use linear, polynomial or Gaussian kernel. The suitable type of kernel depends on the problem itself. The regularization parameter will affect the margin of the hyperplane, as the larger its value the smaller the margin and increase the accuracy for the classification. Meanwhile, the gamma parameter indicates the distance between the classification lines with the point to be considered. The higher the gamma value, the range for consideration to the line will be smaller. For this research, the kernel for each experiment is tested using all kernel and the best result are chosen as the representative of SVM classifier.

SVM is considered as a prominent classifier for face analysis experiments (Chu et al., 2017; Muhammad et al., 2017; Valstar et al., 2017). Its performance will be further evaluated against the other classifier in Chapter 4.

3.4.2 Decision Tree

Decision Tree is one of basic classification algorithm in machine learning. Brief idea of decision tree is to divides the plotting area of graph into smaller part by introducing lines repetitively. The process will iterate until either has meet pure classes, in which the part of the group contains only a single class category or until the criteria of the classifier are met. The process is called impurity, where it traces each group for available members to divide on (Quinlan, 1969). Usually for big amount of data, a threshold percentage of impurity is used to stop the iteration for faster performance but may affect the accuracy (Rokach & Maimon, 2008). The impurities are measured by entropy, which are calculated by the Equation 3.1, where H represent the entropy and $p(x)$ is the probability of function x.

$$H = -\sum p(x) \log p(x)$$ Equation 3.1

At every branching stage, the decision will select the best information gain. Information gain is calculated by the Equation 3.2:

$$Information\ Gain\ (n) = Entropy(x) - (weighed\ average * entropy(child))$$
Equation 3.2

When information gain is 0, means that the members are not divided and thus classified. For the experiment, it is best to use the maximum number of splits for the decision tree classifier. The use of Decision Tree for face analysis had said to be supportive when it merges the decision with other classifier (Kumar et al., 2017), but not much to be highlighted when act alone (Zhang et al., 2018).

3.4.3 K Nearest Neighbors

Apt from its name, KNN algorithm classify data by refer to majority vote of its neighbours. The numbers of neighbours are defined as k, for example if $k = 1$, the nearest neighbour that available to assign the class of the data is 1 or single nearest neighbour. The higher value of k, the range for voting will be bigger. Cover and Hart (1967) have explored the admissibility of the number of KNN and it occurred that single-NN has strictly lower probability than other KNN, but still not conclude that single-NN is the best approach for face analysis. Further explanation, given a point x ∈ X, the distance of each point of the data are calculated. The majority votes from the closest range of first k will classify the point x.

One of the problems occurred when a tie of voting from the neighbours usually happened when more than 1 k is assigned. Another problem is once the input vector is classified there is no indication of its weight as a member to the particular class (Cover & Hart, 1967). But still, the use of KNN in facial analysis is significant and show promising results (Bo & Li, 2018; KaviPriya & Muthukumar, 2018; Zinia et al., 2018), thus worth to be include in comparison.

3.5 Emotion Mapping

Discrete labels of emotions from the basic 7 facial expression have its limitations in describing the emotions. Thus, the experiment are expended from the discrete output and projecting them into a 2D emotional space, called Whissell Space (Whissell, 1989) as shown in Figure 3.9.

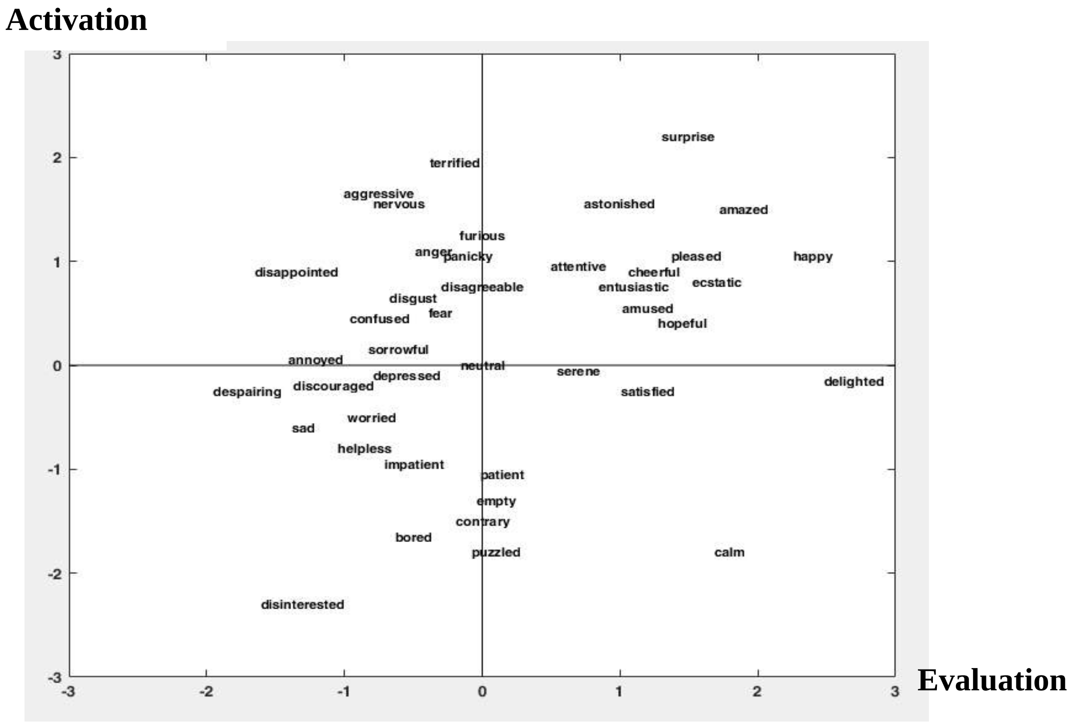


Figure 3.9: The Whissell's Evaluation-Activation Space (Whissell, 1989)

The assignment from the facial expression classification process has the final output of confidence value (CV) corresponding to each expression in percentage. Certain rules are applied to obtain affective weight to expose and dispose emotional incompatibilities. The work from Plutchik (1980) defined the "emotion orientation" values as a series of affect words. The Plutchik's Wheel of Emotions (Plutchik, 1980) showed in Figure 3.10 highlighted that eight primary emotions that are classified into opposite polar, which are "happy" and "sadness", "acceptance" and "disgust", "fear" and "anger", "surprise" and "anticipation".

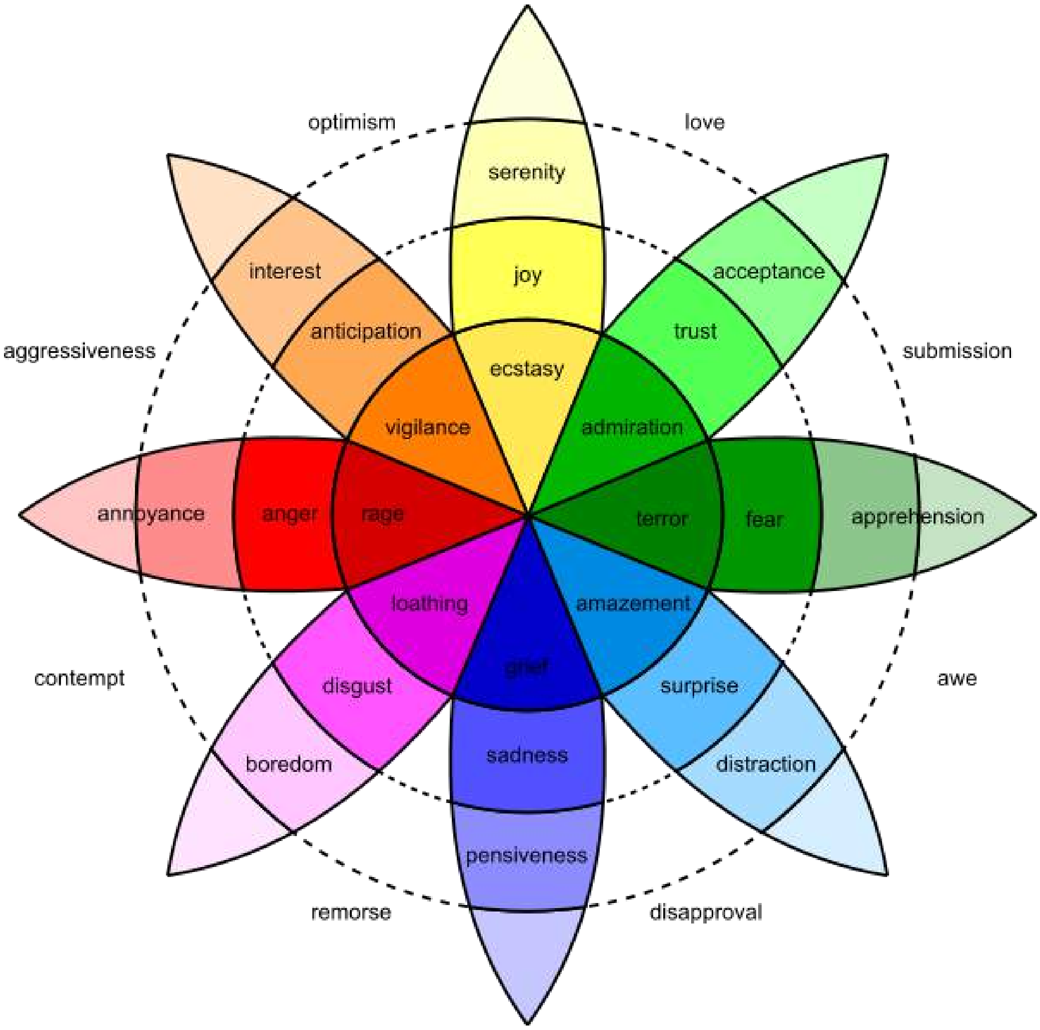


Figure 3.10: Plutchik's Wheel of Emotions (Plutchik, 1980)

One of the rules applied that if emotional incompatibilities are detected, the one that has a closer emotional orientation to the detected emotion is chosen. Then, the coordinate for the affective values are calculated with the Equation 3.3, where x_i and y_i are the initial coordinate of the 7 basic expressions ("angry", "disgust", "fear", "happy", "neutral", "sad" and "surprise") whereas $CV(E_i)$ represent the confidence values for each basic expressions appeared in a given spontaneous expression.

$$\boldsymbol{x} = \frac{\sum_{i=1}^{7} x_i CV(E_i)}{\sum_{i=1}^{7} CV(E_i)} \text{ and } \boldsymbol{y} = \frac{\sum_{i=1}^{7} y_i CV(E_i)}{\sum_{i=1}^{7} CV(E_i)} \qquad \text{Equation 3.3}$$

The affective values are then chosen by applying the shortest distance formula from obtained coordinate (x, y).

3.6 Summary

In this research, the first Malaysian ethnicity facial database that consists of 4 major ethnics in Malaysia, which are Chinese, Iban, Indian and Malay were developed. The database consists of datasets that are categorised by subjects, expressions and ethnicities. The facial analysis conducted in this research involved three major methods which are facial feature extraction, classification and emotion mapping. The feature extractions are using HOG, LBP and SIFT descriptors, whereas the classification processes are using SVM, Decision Tree and KNN classifiers. Lastly, the spontaneous expressions via emotional mapping process by using Whissel Space is discussed.

CHAPTER 4

RESULTS AND ANALYSIS

In this chapter, the results from the testing experiments explained in previous chapter are discussed. The accuracy is used to evaluate the performance of the feature and classifiers and also analysed the confusion matrix from each experiment. The level of accuracy is calculated from the percentage of successful classification group of data in each experiment (Equation 4.1), success rate is calculated using Equation 4.2, whereas the confusion matrix is defined in Table 4.1.

$$\boldsymbol{Accuracy = \frac{True\ Positive(TP) + True\ Negative(TN)}{Condition\ Positive\ (P) + Condition\ Negative(N)}} \quad \text{Equation 4.1}$$

$$Success\ Rate = \frac{True\ Positive(TP)}{Condition\ Positive\ (P)} \quad \text{Equation 4.2}$$

Table 4.1: Confusion Matrix Table

		Actual Class	
		X	**Y**
Predicted Class	**X**	$\frac{TP}{P} \times 100\%$	$\frac{False\ Positive(FP)}{N} \times 100\%$
	Y	$\frac{False\ Negative(FN)}{P} \times 100\%$	$\frac{TP}{P} \times 100\%$

These results will determine which feature and classifier that execute the best results for each experiment. Data confusion also evaluated from the confusion matrix. All images from the database are in grayscale and scaled to 100 x 100 pixels at the beginning of the experiments.

4.1 Ethnicity Classification

This testing experiment is conducted to calculate the performance in classifying the data into four ethnic groups, which are Chinese, Malay, Iban and Indian. There are total of 1491 images from 167 subjects to be classified. First, the performance of the facial features are tested, which are HOG, LBP and SIFT feature and classify it using the same classifier, in which the SVM classifier with 10-fold cross-validation test are chosen. The results of accuracy along with the success rate for each ethnicity from each different feature are shown in Table 4.2.

Table 4.2: Results of Accuracy and Success Rate for Each Ethnicity from Different Features (SVM Classifier)

Feature	Accuracy (%)	Success Rate (%)			
		Chinese	Iban	Indian	Malay
HOG	94.60	95.44	94.18	96.13	92.63
SIFT	93.83	93.31	91.78	96.77	92.96
LBP	75.03	70.08	62.96	84.23	78.57

As can be observed, HOG features had the best accuracy results, which are 94.60%, when classified using SVM classifier compared to the others. SIFT features show competitive results which is 93.83% and quite reliable in detecting Indian and Malay ethnicity. While LBP performance is low compared to the other two features with 75.03% accuracy. HOG is expected to get better result than SIFT in large-scale detection because HOG signifies the overall image compared to SIFT which is more decisive in describing a specific point in the images. Whereas, LBP may have some difficulties in dealing with diverge illumination from the images as it describes based on the intensity patterns of the images.

Next the performances of different classifiers (SVM, Decision Tree and KNN classifiers) are compared using the same facial features, which in this case the HOG feature are selected. Similarly, the results of accuracy and the success rate for each classifier all using 10-fold cross-validation test are compared and the results are presented in Table 4.3.

Table 4.3: Results of Accuracy and Success Rate for Each Ethnicity from Different Classifiers (HOG Feature)

Classifier	**Accuracy (%)**	**Success Rate (%)**			
		Chinese	**Iban**	**Indian**	**Malay**
SVM	94.60	95.44	94.18	96.13	92.63
Decision Tree	54.00	50.42	46.08	61.07	55.77
KNN	96.90	96.41	92.21	99.25	98.77

The results show the performance of the best setting for each classifier. Above all, KNN classifier shows the best accuracy which is 96.90%. It leads the other classifiers in determining Chinese, Indian and Malay ethnicity with the percentage of 96.41%, 99.25% and 98.77% respectively. The best setting for the KNN classifier is using $k = 1$ neighbour in this experiment, as the results are compared using 50 and 100 number of neighbour. The lower number of neighbour minimized the error rate because the prediction is always the closest to the training data itself. For SVM classifier, the best accuracy (compared to Linear) shown when the Kernel function are set to Cubic, which are 94.60%. It shows a reliable result and the highest for determining Iban ethnicity. And lastly, the Decision Tree which are set to Complex Tree (with maximum 100 of splits), which showed the best result compared to Simple Tree and maximum 50 splits as a higher value will lead to more specific results. However, it still failed to perform well and show competitive result with 54.00% accuracy.

Thus, from this it can be concluded that the best pair so far to determine the ethnicity for this experiment are the HOG features and KNN classifier. Table 4.4 shows a confusion matrix for the combination experiment.

Table 4.4: Confusion Matrix of 10-fold Cross-Validation Test using HOG Feature and KNN Classifier (Ethnicity Data Group)

Actual Class	**Predicted Class (%)**			
	Chinese	**Iban**	**Indian**	**Malay**
Chinese	98.31	1.25	0.25	0.25
Iban	1.93	97.69	0	0
Indian	1.10	2.18	99.25	0.98
Malay	0.06	4.36	0.50	98.77

From the table, Indian is the most distinctive ethnic with 99.25% success rate, followed by Malay, Chinese and Iban with 98.77%, 98.31% and 92.21% respectively. It can be said that Indians have outstanding facial features compared to other. As for this experiment, it also may due to the skin intensity of the subjects' images in this groups that highlight the contrast. Meanwhile, Iban have the biggest confusion for this experiment, especially to the Chinese group, but yet still low in percentage which are 1.93% of confusion. From the subject group, Iban also have the highest number from mix marriage parents which may lead to the confusion for Chinese, Indian and Malay ethnicity.

4.2 Basic Facial Expression of Emotion Classification

Similarly, as previous experiment, a testing experiment is conducted to calculate the performance in classifying the data into 7 distinct emotions, which are “angry”, “disgust”, “fear”, “happy”, “neutral”, “sad” and “surprise”. There are total of 1166 images from 167 subjects to be classified. Likewise, the performances of the facial features are tested, which are HOG, LBP and SIFT feature and classify it using the same classifier, in which the SVM classifier with 10-fold cross-validation test are chosen. The comparison results of accuracy, along with the success rate for each ethnicity from each different feature are shown in Table 4.5.

Table 4.5: Results of Accuracy and Success Rate for Each Distinct Emotion from Different Features (SVM Classifier)

Feature	Accuracy (%)	Success Rate (%)						
		Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
HOG	59.10	60.35	67.27	6.92	94.01	65.43	41.92	79.52
SIFT	58.30	60.95	63.69	35.93	82.63	65.43	37.72	62.05
LBP	40.82	55.03	27.38	20.96	65.27	40.12	26.95	50.00

From the table above, HOG features present the best accuracy results, which are 59.10%, when classified using SVM classifiers compared to the others. However, HOG shows the lowest result for “fear” while SIFT shows the highest with 35.93%. Again as HOG feature signify the overall image compared to SIFT which is more decisive in describing specific point of image. Therefore, SIFT can detect the small difference occurred in “fear” as “fear” are easily confused with other expressions, especially to “surprise” and “sad”. SIFT features shows competitive results with 58.30% accuracy and present the best performance to indicate “angry”, “fear” and the same result with HOG for “neutral” expression. LBP features again appear to be the lowest with 40.82% accuracy.

Next, same with the previous, the performance of different classifiers (SVM, Decision Tree and KNN classifiers) are compared using the same facial feature, which in this case the HOG feature is selected. Similarly, the results of accuracy and the success rate for each classifier using 10-fold cross-validation are compared and the results are presented in Table 4.6.

Table 4.6: Results of Accuracy and Success Rate for Each Distinct Emotions from Different Classifiers (HOG Features)

Feature	**Accuracy (%)**	**Success Rate (%)**						
		Angry	**Disgust**	**Fear**	**Happy**	**Neutral**	**Sad**	**Surprise**
SVM	59.10	60.35	67.27	6.92	94.01	65.43	41.92	79.52
Decision Tree	31.30	28.78	27.32	22.65	53.84	29.49	27.35	32.45
KNN	43.80	45.30	61.14	18.57	63.62	29.10	58.99	69.54

The results showed the performance of each classifier. SVM classifier showed the best accuracy which are 59.10% which also ahead of other classifiers in determining “angry”, “disgust”, “happy”, “neutral” and “surprise” with the percentage 60.35%, 67.27%, 94.01%, 65.43% and 79.52% respectively. For KNN classifier, the best performance is when using 100 number of neighbours compared to 1 and 50 number of neighbours which lead to 43.80% accuracy. KKN showed highest results for determining “sad” emotions with 58.99% success rate. And lastly, the Decision Tree which its best results are when set to maximum 50 numbers of splits after compared to Simple Tree and maximum 100 numbers of splits, as the higher number of splits may lead to under fitting. This classifier failed to show competitive results with 31.30% accuracy compared to others. However, it appears to have highest success rate for "fear", even though still low in percentage, which are 22.65%.

Thus, it is concluded that the best pair so far to determine the emotion facial expression from this experiment is using the HOG feature extraction and SVM classifier. Table 4.7 shows a confusion matrix for the combination experiment.

Table 4.7: Confusion Matrix of 10-fold Cross-Validation Test using HOG Feature Extraction and Classify with SVM Classifier (Emotion Data Group)

Actual Class	**Predicted Class (%)**						
	Angry	**Disgust**	**Fear**	**Happy**	**Neutral**	**Sad**	**Surprise**
Angry	60.36	7.10	3.6	3.6	9.47	13.61	2.37
Disgust	5.95	67.27	4.17	10.71	1.79	7.74	2.38
Fear	26.15	22.31	6.92	9.23	11.54	13.08	39.23
Happy	0	2.40	0	94.01	0.60	1.20	1.80
Neutral	6.17	0.62	4.32	6.79	65.43	8.64	5.56
Sad	20.36	8.98	5.99	4.79	14.37	41.92	3.59
Surprise	2.41	1.81	3.01	7.83	4.22	1.20	79.52

From the table above "happy" is the most distinctive emotion with 94.01% success rate, followed by "surprise", "disgust", "neutral" and "angry" with 79.52%, 67.27%, 65.43% and 60.36% respectively among the high success rate. As expected, "happy" and "surprise" appear to be among the highest in accuracy as the two expressions are the most frequently used in communication and therefore quickly identified compared to others (Du & Martinez, 2011). Whereby, "angry", "disgust", "sad" and "fear" expression appear to be difficult to classify as all four are categorized as negative emotions which are often difficult to be classified (Kulkarni et al., 2009). These emotions are also observed to occasionally appear in combinations, in which one of them shows higher intensity (Kulkarni et al., 2009). “Sad” and “fear” appear to be among the lowest success rate with 41.92% and 6.92% respectively. As for “fear”, the biggest confusion for this experiment appears to occur with “surprise”, which can be explained by Roy-Charland et al. (2014) where the difficulty in recognising “fear” may account for the similar visual configuration with “surprise”. As for “sad”, it often appears as micro-expressions, which occurs very brief especially for people who done it spontaneously or without experience in acting the emotions like the subjects for this experiment. This also means that usually hard for people to express their sadness and causing difficulty in detecting sadness. Therefore, the result for “sad” is less adequately recognized and show strong asymmetric confusion with the others, especially “angry” and “neutral”. However, “neutral” and “anger” expression are rarely confused for “sad”.

In other hand, we may observe the results in correspond to the AUs for each emotion. As mentioned in Chapter 2 before, Calvo et al. (2018) concluded that “surprise” usually includes the activation of AU 1 + 2 + 5 + 25 + 26. Note that “fear” activates AU 1 + 5 + 25, which are subsets of those of “surprise”. Therefore, “fear” is confused as “surprise” and not the other way round. Same occurred in “sad” which includes AU 1 + 4 + 15. If the AU 15 is not notable, which in Calvo et al. (2018) only 0.98 AU evidence scores of significance for “sad”, it can be confused with “angry” and “disgust” for the existence of AU 4 (with AU evidence score of 1.55 for sad). It also may be confused with “fear” and “surprise” for the presence of AU 1 (with AU evidence score of 1.38 for “sad”). Unlike “happy”, even though it only includes AU 6 + 12 + 25, each AU have 2.88, 4.06 and 2.07 evidence score respectively, which makes it easier to be distinguished compared to other emotions. As for “anger” in other hand, with AU 4 + 7, have risk to be confused with “disgust” (AU 4 + 6 +

7 + 9 + 10) as it is subset for the expression, and also “sad” if the AU 7 (with AU evidence score of 0.89 for “anger”) is not observable.

4.3 Classify Basic Facial Expression of Emotion of Same Ethnicity

The testing experiment was conducted to compare the performance in detecting the 7 distinct emotions for each specific ethnicity. The emotions are grouped based on their ethnicity and the distribution of the data is as stated in Table 4.8.

Table 4.8: Data Distribution in Detecting Basic Facial Expression of Emotion of Same Ethnicity

Ethnicity	No. of Images	No. of Subjects
Chinese	278	40
Iban	237	34
Indian	320	47
Malay	331	48

The performance is tested using the best combination of HOG feature and SVM classifier with 10-fold cross-validation test. The comparison results of accuracy, along with the success rate for emotions for each ethnicity are shown in Table 4.9, Table 4.10, Table 4.11 and Table 4.12.

Table 4.9: Confusion Matrix of 10-fold Cross-Validation Test using HOG Feature and SVM Classifier (Emotion Data Group-Chinese)

Actual Class	**Predicted Class (%)**						
	Angry	**Disgust**	**Fear**	**Happy**	**Neutral**	**Sad**	**Surprise**
Angry	37.50	2.50	7.50	2.50	20.00	25.00	5.00
Disgust	12.50	47.50	12.50	20.00	0	0	7.50
Fear	22.50	12.50	0	10.00	17.50	2.50	35.00
Happy	0	2.56	0	94.87	0	2.56	0
Neutral	17.50	0	7.50	5.00	55.00	10.00	5.00
Sad	28.21	7.69	0	12.82	15.38	30.77	5.13
Surprise	0	0	5.00	15.00	2.50	2.50	75.00

Table 4.10: Confusion Matrix of 10-fold Cross-Validation Test using HOG Feature and SVM Classifier (Emotion Data Group-Iban)

Actual Class	**Predicted Class (%)**						
	Angry	**Disgust**	**Fear**	**Happy**	**Neutral**	**Sad**	**Surprise**
Angry	44.12	20.59	14.71	2.94	2.94	14.71	0
Disgust	2.94	58.82	5.88	17.65	0	14.71	0
Fear	32.35	14.71	2.94	11.76	8.82	8.82	20.59
Happy	0	0	0	100.00	0	0	0
Neutral	15.15	3.03	9.09	15.15	45.45	6.06	6.06
Sad	17.14	5.71	0	2.86	8.57	62.86	2.86
Surprise	3.03	0	6.06	3.03	9.09	0	78.79

Table 4.11: Confusion Matrix of 10-fold Cross-Validation Test using HOG Feature and SVM Classifier (Emotion Data Group-India)

Actual Class	**Predicted Class (%)**						
	Angry	**Disgust**	**Fear**	**Happy**	**Neutral**	**Sad**	**Surprise**
Angry	74.47	2.13	4.26	2.13	8.51	8.51	0
Disgust	4.35	52.17	2.17	26.09	6.52	2.17	6.52
Fear	20.00	13.33	2.22	6.67	11.11	17.78	28.89
Happy	0	4.26	0	93.62	2.13	0	0
Neutral	9.09	2.27	2.27	9.09	54.55	4.55	18.18
Sad	28.89	8.89	15.56	4.44	11.11	28.89	2.22
Surprise	2.17	0	0	6.52	13.04	4.35	73.91

Table 4.12: Confusion Matrix of 10-fold Cross-Validation Test using HOG Feature and SVM Classifier (Emotion Data Group-Malay)

Actual Class	**Predicted Class (%)**						
	Angry	**Disgust**	**Fear**	**Happy**	**Neutral**	**Sad**	**Surprise**
Angry	41.67	12.50	8.33	6.25	14.58	14.58	2.08
Disgust	14.58	37.50	8.33	18.75	6.25	6.25	8.33
Fear	6.25	18.75	10.42	10.42	8.33	10.42	35.42
Happy	0	6.38	426	82.98	0	2.13	4.26
Neutral	13.33	0	2.22	13.33	51.11	13.33	6.67
Sad	16.67	12.50	10.42	6.25	18.75	29.17	6.25
Surprise	4.26	6.38	12.77	17.02	2.13	2.13	55.32

Overall, it can be concluded that the pattern for the results are almost as in Table 4.7, with lower percentages. This might be due to the decrease number of overall samples in each experiment for each ethnicity. “Happy” still achieved the highest success rate for all ethnicity, and the most performed is Iban ethnic with 100% prediction score. “Fear” again appears to achieved the lowest success rate for all ethnicity, the worst for Chinese ethnic with 0% prediction score. It can be said that Chinese hardly express “fear” and tend to be confused with “surprise” (35.00%), which similarly explained in previous discussion (Experiment 4.2). Fear expression involves activation of AU1, AU5 and AU25, which involving inner brow, upper lid and also lips. This features movements are less noticeable for Chinese ethnic which caused the low percentage. Other than that, “fear” also is significantly confused with “angry” especially for Iban ethnic, which scores the highest among other emotions with 32.35%. From this, the expression of “fear” for certain ethnicity especially Iban might also involve the activation of AU4 or AU7 which caused this confusion.

The next emotion that appears to have low success rate is “sad” for all ethnic where Iban with highest score which is 62.86%. The percentage is interestingly higher compared to other ethnic which can only score around 25% to 30% and only 17.14% confusion with “angry”. It can be said that Iban have their distinctive way in portraying sad expression. It is also notable that for all ethnicity, “sad” are likely to be confused with “angry”, especially for Indian, which share the same success rate of 28.89%. This may be due to the AU 1 and AU 15 being less expressed or there exist the activation of AU7 that caused this confusion. As for “angry”, Indian appear to be the most outstanding in percentage with 74.47% compared to others which are mostly lower than 50%.

4.4 Classification of Ethnicity from Basic Facial Expressions

Testing in determining the ethnicity for each facial expression of emotion image is also conducted. The relevance for this experiment is to observe any differences in term of accuracy in detecting ethnicity when given a facial expression of emotion image. The distribution of the data used for the experiment is as shown in Table 4.13.

Table 4.13: Data Distribution in Detecting Ethnicity from Basic Facial Expressions of Emotion

Facial Expression of Emotion	**No. of Images**
Angry	169
Disgust	168
Fear	167
Happy	167
Neutral	162
Sad	167
Surprise	166

Again, the performances of this experiment are tested using HOG feature and classify using SVM classifier with 10-fold cross-validation test. The comparison results of accuracy, along with the success rate for ethnicity for each facial expression are shown in following tables.

Table 4.14: Confusion Matrix of 10-fold Cross-Validation Test using HOG Features and SVM Classifier (Angry Data Group)

Actual Class	Predicted Class (%)			
	Chinese	**Iban**	**Indian**	**Malay**
Chinese	52.63	35.29	8.16	8.33
Iban	26.32	50.00	2.04	12.50
Indian	10.53	5.88	75.51	8.33
Malay	10.53	8.82	14.29	70.83

Table 4.15: Confusion Matrix of 10-fold Cross-Validation Test using HOG Features and SVM Classifier (Disgust Data Group)

Actual Class	Predicted Class (%)			
	Chinese	**Iban**	**Indian**	**Malay**
Chinese	40.82	37.50	0.00	16.67
Iban	28.57	37.50	2.56	14.58
Indian	14.29	6.25	84.62	8.33
Malay	16.33	18.75	12.82	60.42

Table 4.16: Confusion Matrix of 10-fold Cross-Validation Test using HOG Features and SVM Classifier (Fear Data Group)

Actual Class	Predicted Class (%)			
	Chinese	**Iban**	**Indian**	**Malay**
Chinese	54.05	29.73	2.50	15.09
Iban	29.73	32.43	0.00	20.75
Indian	8.11	13.51	85.00	5.66
Malay	8.11	24.32	12.50	58.49

Table 4.17: Confusion Matrix of 10-fold Cross-Validation Test using HOG Features and SVM Classifier (Happy Data Group)

Actual Class	Predicted Class (%)			
	Chinese	**Iban**	**Indian**	**Malay**
Chinese	52.63	33.33	2.44	8.70
Iban	34.21	35.71	2.44	10.87
Indian	5.26	14.29	82.93	10.87
Malay	7.89	16.67	12.20	69.57

Table 4.18: Confusion Matrix of 10-fold Cross-Validation Test using HOG Features and SVM Classifier (Neutral Data Group)

Actual Class	Predicted Class (%)			
	Chinese	**Iban**	**Indian**	**Malay**
Chinese	61.76	32.43	2.70	11.11
Iban	23.53	48.65	2.70	11.11
Indian	8.82	5.41	91.89	9.26
Malay	5.88	13.51	2.70	68.52

Table 4.19: Confusion Matrix of 10-fold Cross-Validation Test using HOG Features and SVM Classifier (Sad Data Group)

Actual Class	Predicted Class (%)			
	Chinese	**Iban**	**Indian**	**Malay**
Chinese	64.71	27.27	8.33	7.69
Iban	26.47	57.58	0.00	13.46
Indian	0.00	3.03	75.00	15.38
Malay	8.82	12.12	16.67	63.46

Table 4.20: Confusion Matrix of 10-fold Cross-Validation Test using HOG Features and SVM Classifier (Surprise Data Group)

Actual Class	**Predicted Class (%)**			
	Chinese	**Iban**	**Indian**	**Malay**
Chinese	53.85	27.03	4.55	15.22
Iban	23.08	43.24	6.82	10.87
Indian	12.82	16.22	75.00	4.35
Malay	10.26	13.51	13.64	69.57

It can be said that Indian has the most accuracy for every expression, followed by Malay, then Chinese and Iban. The patterns are similar to previous experiment 4.2, except the obvious decrement in percentage. It might be due to the decreasing in number of overall samples for each facial expression. The highest success rate for Indian is from "neutral" data with 91.89%. This may because for neutral expression, their distinctive facial expression is highlighted and not to be confused with other expression. Iban appeared to have the lowest success rate for all expression, especially for "fear" with 32.43%. Amongst all expression, "fear" appears to caused biggest confusion due to high similarities to others. As discussed before in Chapter 4.1, Iban also have the highest number from mix marriage parents, which is 24% of the data, that may lead for the confusion for Chinese, Indian and Malay ethnicity. From this, it can be observed that different expression affects the accuracy in classification the ethnicity.

4.5 Emotion Mapping of Spontaneous Facial Expression

In this emotion mapping experiment, the theory from Chapter 3.5 is performed, in which the end result determines the emotions described from the Whissell Space using subject samples from the spontaneous dataset. Figure 4.1 listed 10 random samples from the dataset.

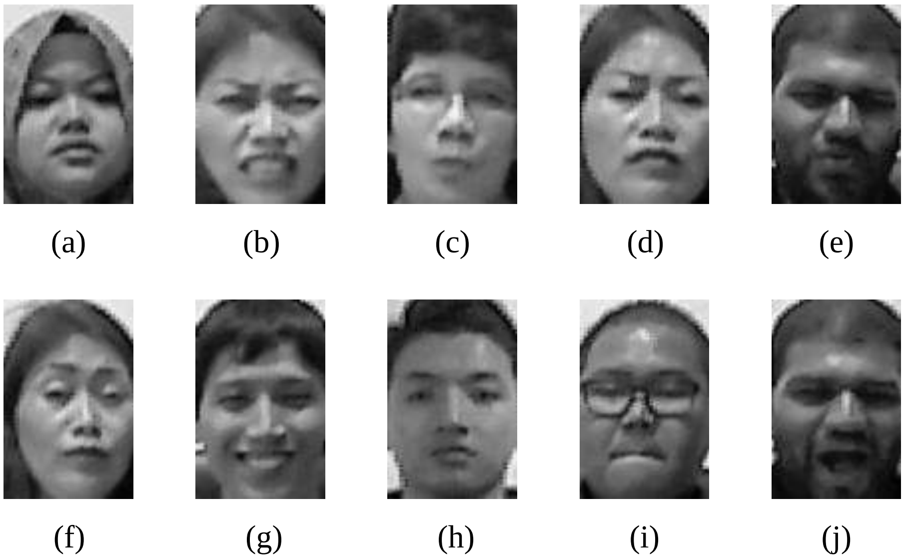


Figure 4.1: 10 Random Samples from Spontaneous Dataset

In this experiment, the classifier models from previous experiment on emotions (Chapter 4.2) are used to obtain the confidence values for each emotion of the subjects from the spontaneous dataset. The closest emotional orientation is chosen according to the Plutchik's Wheel of Emotions and the result from that, the coordinate is obtained using Equation 3.3. The closest affective values are labelled by calculating the shortest distance from the coordinate point. The emotion mapping for subject (a) from Figure 4.1 is shown in Figure 4.2.

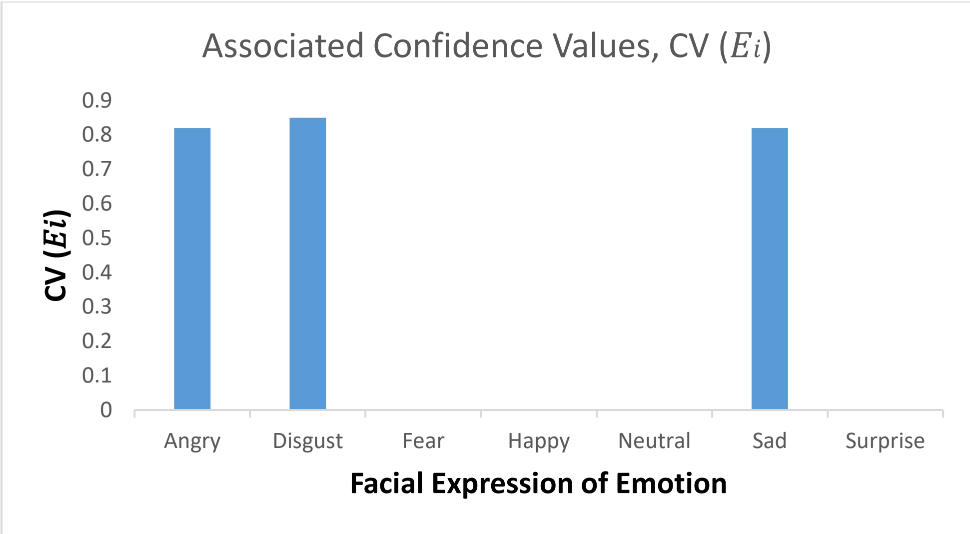


Figure 4.2: Associated Confidence Values of Emotions for Subject (a) from Figure 4.1

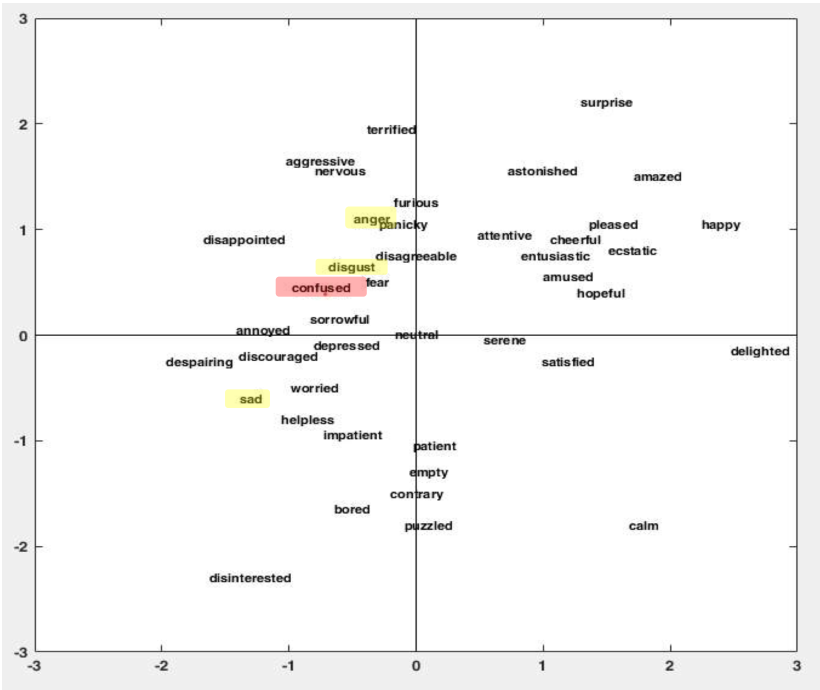


Figure 4.3: Example of Emotional Mapping for Subject (a) from Figure 4.1

The image discrete emotion class assigned confidence value of 82% to "anger", 85% to "disgust" and 82% to "sad", and the rest are assigned zero for subject (a). The coordinates calculated happened to be closest to "confused" at (-0.75, 0.45), thus resulting the predicted affective value. The results for the rest of the subjects are in the Table 4.21.

Table 4.21: Results of Predicted Affective Value for Samples from Spontaneous Dataset

Subject	Confidence Values, CV (E_i)							Predicted Affective Value	Ground Truth Based on 7 Basic Expressions
	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise		
(a)	0.82	0.85	0	0	0	0.82	0	Confused	Angry
(b)	0.81	0.89	0	0	0	0	0.78	Furious	Angry/Disgust
(c)	0.81	0	0	0	0	0	0.91	Astonished	Angry/Fear
(d)	0	0.87	0	0	0	0.82	0	Sorrowful	Angry/Sad
(e)	0.81	0.83	0	0	0	0	0.88	Furious	Disgust
(f)	0	0	0.83	0	0.84	0	0.84	Attentive	Fear
(g)	0	0	0	0.94	0	0	0	Happy	Happy
(h)	0.83	0.87	0	0	0.82	0	0	Fear	Neutral
(i)	0	0.83	0	0	0	0.8	0	Discouraged	Sad
(j)	0	0.82	0	0	0	0	0.77	Furious	Surprise

From this experiment, the affective values of spontaneous expressions from the confidence values of 7 basic expressions are predicted. Compared to the 7 basic expressions, the affective values explained more on the behavioural and emotional values of the spontaneous reactions.

4.6 Summary

The combination of HOG features and KNN classifier is the best recommendation so far for ethnic recognition, whereas HOG features and SVM classifier is best for emotion recognition. Indian ethnicity appears to be the most distinct among others. Meanwhile, "happy", is the most recognizable emotion and "fear" is the hardest to be distinguished. Also from the research, the emotions are better described by using the Whissell Space. It also can be observed that different expression affects the accuracy result in ethnic classification. Iban occurs to have lowest success rate in predicting ethnicity for all expressions.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Introduction

This chapter concluded the overall achievements and findings that have been obtained in developing the research. The observation and results from testing experiments are further reviewed. Limitation and suggestion for potential future works related to the research are also covered in this chapter.

5.2 Conclusion

This research mainly contributes on establishing and analysing MUA: Malaysian Ethnic Group Face Database, which is also the first Malaysian ethnicity facial database and thus a leading study in analysing facial expressions between Chinese, Iban, Indian and Malay ethnicities. The database includes continuous emotion facial expression for “angry”, “disgust”, “fear”, “happy”, “neutral”, “surprise” and “sad” with low to high intensity of the facial expressions. It also consists of spontaneous facial expression with the ground truth of the subjects’ emotion recorded from the reaction of simulating videos. The recordings are allowed to includes obstacles like subjects wearing scarf, hair, cap and glasses. These obstacles are important in order to assure the reliability of the study to be applied into real-world applications.

In order to achieved higher accuracy in face recognition, the selection of facial feature and classifier are crucial. Therefore, for this research, the performance between HOG, LBP and SIFT are calculated to acquire the best facial feature extraction method. The comparison between SVM, KNN and Decision Tree classifier are also executed for both ethnicity and facial expression recognition. From the experiments, it can be observed that HOG feature is compatible and achieving highest accuracy for both emotion and ethnicity recognition. The achievement for the classifiers' performances are competitive for KNN and SVM. The best pair for ethnic classification is using HOG feature and KNN classifier, whereas for emotion facial expression is using HOG feature and SVM classifier.

Between the 4 ethnicities, the ethnic recognition for Indian appears to be the most outstanding with highest accuracy among others. Meantime, "happy" appeared to be the most recognizable emotion facial expression with highest accuracy and "surprise" to be the second. "Angry", "disgust", "sad" and "fear" shows to be less recognizable and "fear" turn out to be the least with lowest success rate among all expressions. Different facial expressions occurred to have influence in the accuracy for ethnic classification.

Researches nowadays moving towards detecting the emotions spontaneously and unintentionally. Thus, better way to describe emotion is compelling to be executed. Thus, emotion mappings are done to describe affective values from the spontaneous facial expressions. Other than that, the emotional mapping also helps in describing the facial expressions from the spontaneous dataset better using the Whissell Space.

The first objective is achieved by discussed results in classifying different facial expression of emotion from each ethnicity. Different ethnicity proved to have different sensitivity to express certain emotions. From the experiment, Iban's "happy" expressions are easily recognisable, whereas Chinese appear to have most confusion in expressing "fear". In this research, intensive analysis in classifying basic facial expression between each ethnicity are discussed. Hence, these partially agreed to the hypothesis, as the ethnicity categorization may improve the facial expression classification using FACS.

The second objective is achieved by classifying emotion mapping of spontaneous facial expression using Whissel Space. The predicted affective values are compared with the ground truth recorded during the data collection. This research showed that the method proved to be successfully describe on the behavioural and emotional values of the spontaneous facial expression compared to the basic facial expression.

The experiment to detect ethnicity from each basic facial expression are also conducted in order to investigate the success rate of ethnicity classification from the specific facial expressions. However, the results show that the success rate in classifying ethnicity from each facial expression of emotion decreased compared to the ethnicity classification from the whole database. Thus, this deny earlier hypothesis that the deformation of facial expression using FACS can improved ethnicity classification. However, the results might due to limited number of data after specific allocation of basic facial expression from the database. Intensive analysis discussed to investigate the cause of confusion from the facial deformation for facial classification.

5.3 Future Work

There are still other approaches in improving the result for facial expression classification of ethnicity. One of the improvements for this project is to increase the number of image samples, because the current number of images are still low which are less than 50 images for each ethnicity. In order to have higher accuracy and precision result, a greater number of samples is suggested. The study can be further explored for other ethnicities to analyse the cross-cultural relation of emotion in multicultural country of Malaysia.

For future work, the database can be improved by includes 3D face dataset into the database. 3D data may contain more information and details from the face compared to 2D data and images. Some research nowadays proven to enhance the accuracy result for facial expression recognition. This 3D dataset can be further analysed as the visual information studies nowadays are actively researching into the 3D data in order to create various application in future.

More suggestions of facial feature and classifier can be further analysed in which might result in improved accuracy. From the thesis, it is proven that different facial feature and classifier affecting the results of ethnic and emotion facial classification. It is also recommended for further research to use combination of classifiers in order to obtain better results and lower the percentage of confusion.

From this, research on the specific AUs appear for each expression for each ethnicity can be conducted, which realistically may improve the recognition accuracy. The experiments from the thesis shows that specific combination of AUs appeared with different intensities appeared in different emotion facial expression. Different ethnicities also portray particular AUs in interpret basic facial expression of emotion. Therefore, specific identification of AUs may lead to further research that lead to increase the result of emotion facial expression recognition. It also helps to minimize false positive and false negative during the classification process.

The idea from this research can be further expanded into software development that can be used to collect data and perform face recognition, ethnic recognition and emotion recognition especially for Malaysian. A lot of cognitive and psychological field may benefit to the development of software that are usable to the end-user. It also an advantage if it can be developed into live video feed recognition. The ability to recognize emotion facial expression from live video may be useful in commercial areas. However, it is expected to encounter various obstacles in order to analyse and perform recognition it in a short time.

REFERENCES

Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 28*(12), 2037-2041.

Bartlett, M. S., Movellan, J. R., Littlewort, G., Braathen, B., Frank, M. G., & Sejnowski, T. J. (2005). Towards automatic recognition of spontaneous facial actions. In What the Face Reveals (Paul Ekman Erika L. Rosenberg, 2nd ed.), pp. 393-412. Oxford University Press.

Beghdadi, A., Asim, M., Almaadeed, N., & Qureshi, M. A. (2018). Towards the design of smart video-surveillance system. *2018 NASA/ESA Conference on Adaptive Hardware and Systems (AHS),* Edinburgh, UK, 6-9 August 2018. IEEE.

Bentin, S., & Deouell, L. Y. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. *Cognitive Neuropsychology, 17*, 35-54.

Berardesca, E., Leveque, J. L., & Maibach, H. I. (2006). *Ethnic Skin and Hair.* CRC Press.

Berretti, S., Del Bimbo, A., & Pala, P. (2006). Description and retrieval of 3D face models using iso-geodesic stripes. In: *Proceedings of the 8th ACM international workshop on Multimedia information retrieval.* ACM.

Berretti, S., Del Bimbo, A., & Pala, P. (2012). Distinguishing facial features for ethnicity-based 3D face recognition. *ACM Transactions on Intelligent Systems and Technology (TIST), 3*(3), 1-20.

Bo, Z. H. A. I., & Li, Z. M. (2018). Face Recognition Based on Fusion Feature of LBP and PCA with KNN. In: *3rd International Conference on Computational Modeling*

Simulation and Applied Mathematics, Wuhan, China, 27-28 September 2018. DEStech Publications.

Bowyer, K. W., Chang, K., & Flynn, P. (2006). A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *Computer Vision Image Understanding, 101*(1), 1-15.

Braun, A., Queiroz, R., Lee, W., Feijo, B., & Musse, S. R. (2018). Persona: A Method for Facial Analysis in Video and Application in Entertainment. *Computers in Entertainment (CIE), 16*(3), 4.

Braun, L., Wolfgang, M., & Dickersin, K. (2013). Defining race/ethnicity and explaining difference in research studies on lung function. *European Respiratory Journal, 41*(6), 1362-1370.

Briggs, B. D., Clevenger, L. A., Clevenger, L. A. H., Penny, C. J., Rizzolo, M., & Sipolins, A. G. (2018). Emotional analysis and depiction in virtual reality. U.S. Patent Application 15/445,335.

Buang, I. N., & Ujir, H. (2019). MUA: Malaysian Ethnicity Recognition. *International Journal of Advanced Trends in Computer Science and Engineering, 8*(1.6), 543 - 548.

Burbidge, R., & Buxton, B. (2001). An introduction to support vector machines for data mining. *Keynote papers, Young OR12*, 3-15.

Calder, A., Rhodes, G., Johnson, M., & Haxby, J. (2011). *Oxford Handbook of Face Perception.* London, U.K.: Oxford University Press.

Calvo, M. G., Fernández-Martín, A., Recio, G., & Lundqvist, D. (2018). Human observers and automated assessment of dynamic emotional facial expressions: KDEF-dyn database validation. *Frontiers in Psychology, 9*, 2052.

Chang, J. C. (2004). *Facial surveillance system and method.* U.S.: Patent Application 10/429,720.

Chang, K. I., Bowyer, K. W., & Flynn, P. J. (2005). An evaluation of multimodal 2D+3D face biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(4), 619-624.

Chen, B., & Nahrstedt, K. (2019). FIS: Facial Information Segmentation for Video Redaction. *2019 IEEE Conference on Multimedia Information Processing and Retrieval*, 360-365.

Chu, W. S., De la Torre, F., & Cohn, J. F. (2017). Selective transfer machine for personalized facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(3), 529-545.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13*(1), 21-27.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In: *2005 IEEE Computer Society Conference on Camera Vision and Pattern Recognition,* San Diego, 20 - 25 June 2005. IEEE.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animal.* London: J. Murray.

Dhall, A., Goecke, R., Lucey, S., & Gedeon, T. (2011). Acted facial expressions in the wild database. *Australian National University, Canberra, Australia, Technical Report TR-CS-11, 2,* 1.

Du, S., & Martinez, A. M. (2011). The resolution of facial expressions of emotion. *Journal of Vision, 11*(13), 24-36.

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences, 111*(15), E1454-E1462.

Eberhardt, J. L., Dasgupta, N., & Banaszynski, T. L. (2003). Believing is seeing: The effects of racial labels and implicit beliefs on face perception. *Personality and Social Psychology Bulletin, 29*(3), 360-412.

Ekman, P. (1973). Cross-cultural studies of facial expression. Darwin and facial expression: A century of research in review. In *Darwin and Facial Expression: A Century of Research in Review* (Ekman, P., Ed.), pp. 169-222. Academic Press, Inc.

Ekman, P. (2001). *Telling Lies: Clues to deceit in the marketplace, politics, and marriage,* 3rd ed., New York: W. W. Norton.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124-9.

Ekman, P., & Friesen, W. V. (1976). *Pictures of Facial Effect.* Palo Alto, CA: Consulting Psychologist Press.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: Investigator's guide.* Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system: The manual on CD-ROM. Instructor’s Guide.* Salt Lake City: Network Information Research C.

Ekman, P., Dalgleish, T., & Power, M. (2001). *Handbook of Cognition and Emotion*, Chihester, UK: Wiley, 1999.

Ekman, R. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS).* USA: Oxford University Press.

El Meguid, M. K. A., & Levine, M. D. (2014). Fully automated recognition of spontaneous facial expressions in videos using random forest classifiers. *IEEE Transactions on Affective Computing, 5*(2), 141-154.

Farkas, L. G. (1994). *Anthropometry of the Head and Face.* New York: NY: Raven Press.

Freeman, J. B., Ambady, N., & Holcomb, P. J. (2010). The face-sensitive n170 encodes social category information. *NeuroReport, 21*(1), 24-28.

Fu, S., He, H., & Hou, Z. G. (2014). Learning race from face: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(12), 2483-2509.

Gazizullina, A., & Mazzara, M. (2019). Spontaneous Emotion Recognition in Response to Videos. In: Mazzara, M., Bruel, J., Meyer, B. and Petrenko, A., *International Conference on Objects, Components, Models and Patterns*, 15 - 17 Oktober 2019. Springer, Cham.

Girard, J. M., Chu, W. S., Jeni, L. A., & Cohn, J. F. (2017). Sayette group formation task (GFT) spontaneous facial expression database. In: *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition*, Washington, DC, 30 May - 3 June 2017, IEEE.

Happy, S. L., Patnaik, P., Routray, A., & Guha, R. (2015). The Indian spontaneous expression database for emotion recognition. *IEEE Transactions on Affective Computing, 8*(1), 131-142.

Hupont, I., Cerezo, E., & Baldassarri, S. (2010). Sensing facial emotions in a continuous 2D affective space. In: *2010 IEEE International Conference on Systems, Man and Cybernetics*, Istanbul, Turkey, 10 - 13 October 2010. IEEE.

Izard, C. E. (2009). Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annual Review of Psychology, 60*, 1-25.

Jilani, S. K., Ugail, H., Bukar, A. M., Logan, A., & Munshi, T. (2017). A Machine Learning Approach for Ethnic Classification: The British Pakistani Face. *International Conference on Cyberworlds*, 170-173.

KaviPriya, A., & Muthukumar, A. (2018). Evaluation of Human Age with FKP Using K-NN. In: *2018 International Conference on Soft-computing and Network Security,* Coimbatore, India, 14 - 16 February 2018. IEEE.

Kulkarni, S. S., Reddy, N. P., & Hariharan, S. I. (2009). Facial expression (mood) recognition from facial images using committee neural networks. *Biomedical Engineering Online*, *8*(1), 16.

Kumar, T., Jangra, S., & Bhushan, S. (2017). Face Recognition with Decision Tree using SVM and SURF. *International Journal of Control Theory and Applications, 10*(15), 173-180.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. In: Verma, I., *Proceedings of the National Academy of Sciences of the United States of America*, USA, 18 December 2001, United States Nation Academy of Sciences.

Lalji, Z., & Good, J. (2008). Designing new technologies for illiterate populations: a study in mobile phone interface design. *Interacting with Computers, 20*(6), 574-586.

Levin, D. T., & Angelone, B. L. (2002). Categorical perception of race. *Perception, 31*(5), 567-578.

Levine, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General, 129*(4), 559-574.

Levine, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology, 22*(6), 1364-1382.

Li, Y., Mavadati, S. M., Mahoor, M. H., Zhao, Y., & Ji, Q. (2015). Measuring the intensity of spontaneous facial action units with dynamic Bayesian network. *Pattern Recognition, 48*(11), 3417-3427.

Liu, P., & Yin, L. (2017). Spontaneous thermal facial expression analysis based on trajectory-pooled fisher vector descriptor. In: *2017 IEEE International Conference of Multimedia and Expo,* Hong Kong, 10 - 14 July 2017. IEEE.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 2*(60), 91-110.

Lu, X., & Jain, A. K. (2004). Ethnicity identification from face images. In: *Proceedings of International Society for Optics and Photonics.* 12 - 16 April 2004.

Lu, X., Chen, H., & Jain, A. K. (2006). Multimodal facial gender and ethnicity identification. In: Zhang, D. Y. and Jain, A. K. (Eds.) *2006 International Conference on Biometrics.* Hong Kong, China, 5 - 7 January 2006. Springer.

Lucey, P., Cohn, J. F., Prkachin, K. M., Solomon, P. E., & Matthews, I. . (2011). Painful data: The UNBC-McMaster shoulder pain expression archive database. In: 9th *IEEE International Conference on Automatic Face and Gesture Recognition Workshops,* Santa Barbara, CA, USA, 21 - 23 March 2011, IEEE.

Ma, J., Yang, B., Luo, R., & Ding, X. (2019). Development of a facial-expression database of Chinese Han, Hui and Tibetan people. *International Journal of Psychology, 3*(55), 456-464.

Martinez, A., & Du, S. (2012). A model of the Perception of Facial Expressions of Emotion by Humans. *Journal of Machine Learning Research, 13*(2012), 1589-1608.

Masood, S., Gupta, S., Wajid, A., Gupta, S., & Ahmed, M. (2018). Prediction of human ethnicity from facial images using neural networks. In *Data Engineering and Intelligent Computing* (Satapathy, S. C., Bhateja, V., Raju, K. S. and Janakiramaiah, B.), pp. 217-226. Singapore: Springer.

Mavadati, M., Sanger, P., & Mahoor, M. H. (2016). Extended disfa dataset: Investigating posed and spontaneous facial expressions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Las Vegas, USA, 26 June - 1 July 2016. IEEE.

Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., & Cohn, J. F. (2013). DISFA: A Spontaneous Facial Action Intensity Database. *IEEE Transactions on Affective Computing, 4*(2), 151-160.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2001). Emotional intelligence as a standard intelligence. *Emotion, 3*, 232-242.

Mays, V. M., Ponce, N. A., Washington, D. L., & Cochran, S. D. (2003). Classification of race and ethnicity: implications for public health. *Annual Review of Public Health, 24*(1), 83-110.

McLain, D., & Kefallonitis, E. (2019). Advances and Distinctions in the Use of Biometric Methods Versus Traditional Methods for Studying the Customer Experience. In *Strategic Innovative Marketing and Tourism*, (Kavoura, A., Kefallonitis, E., and Giovanis, A.) pp. 517-522. Athens Riviera: Springer, Cham,.

Mead, M. (1975). Review of: Darwin and Facial Expression (P. Ekman, ed.). *Journal of Communication, 25*(1), 209-213.

Mohammadi, M. R., Fatemizadeh, E., & Mahoor, M. H. (2015). Intensity estimation of spontaneous facial action units based on their sparsity properties. *IEEE Transactions on Cybernetics, 46*(3), 817-826.

Muhammad, G., Alsulaiman, M., Amin, S. U., Ghoneim, A., & Alhamid, M. F. . (2017). A Facial-Expression Monitoring System for Improved Healthcare in Smart Cities. *IEEE Access, 5*, 10871-10881.

O'Toole, A. J., Harms, J., Snow, S. L., Hurst, D. R., Pappas, M. R., Ayyad, J. H., & Abdi, H. (2005). A video database of moving faces and people. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(5), 812-816.

Palestra, G. (2018). Cognitive Emotions Recognition in e-Learning: Exploring the Role of Age Differences and Personality Traits. In *Methodologies and Intelligent Systems for Technology Enhanced Learning, 9th International Conference*, (Popescu, E., Gil, A. B., Lancia, L., Simona Sica, L., Mavroudi, A.), p. 97. Cham, Switzerland: Springer.

Phillips, P. J., Scruggs, W. T., O'Toole, A. J., Flynn, P. J., Bowyer, K. W., Schott, C. L., & Sharpe, M. (2010). FRVT 2006 and ICE 2006 large-scale experimental results. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(5), 831-846.

Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., & Worek, W., (2005). Overview of the face recognition grand challenge. In: *2005 IEEE Computer Society Conference on Camera Vision and Pattern Recognition*, San Diego, 20 - 25 June 2005. IEEE.

Pittsburgh, I. (2011, Nov 4). *Pattern Recognition. (2011, Nov. 4) PittPatt SDK v5.2.2 Documentation.* Retrieved Jul 2019, 19, from http://www.pittpatt. com/documentation/

Plutchik, R. (1980). *Emotion: A Psychoevolutionary Synthesis.* New York: Harper & Row.

Qu, F., Wang, S. J., Yan, W. J., Li, H., Wu, S., & Fu, X. (2017). CAS (ME)2: A Database for Spontaneous Macro-Expression and Micro-Expression Spotting and Recognition. *IEEE Transactions on Affective Computing, 9*(4), 424-436.

Quinlan, J. R. (1969). Induction of decision trees. *Machine Learning, 1*(1), 81-106.

Rattani, A., & Tistarelli, M. (2009). Robust multi-modal and multi-unit feature level fusion of face and iris biometrics. In *International Conference of Biometrics,* Alghero, Italy, 2 - 5 June 2009. Springer.

Reisenzein, R., Studtmann, M., & Horstmann, G. (2013). Coherence between Emotion and Facial Expression: Evidence from Laboratory Experiments. *Emotion Review, 5*(1), 16-23.

Rokach, L., & Maimon, O. Z. (2008). *Data mining with decision trees: theory and applications*, Vol. 69, Singapore: World Scientific.

Rossion, B., & Jacques, C. (2002). N170: Understanding the time course of face perception in the human brain. In *The Oxford Handbook of Event-related Potential Components* (Luck, S. J. and Kappenman, E. S., Eds.), pp. 115-141. New York: Oxford University Press.

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science, 290*(5500), 2323-2326.

Roy-Charland, A., Perron, M., Beaudry, O., & Eady, K. (2014). Confusion of fear and surprise: A test of the perceptual-attentional limitation hypothesis with eye movement monitoring. *Congnition and Emotion, 7*(28), 1214-1222.

Saeed, A., & Al-Hamadi, A. (2015). Boosted human head pose estimation using kinect camera. In Image Processing (ICIP). In: *2015 IEEE International Conference on Image Processing,* Quebec City, Canada, 27 - 30 September 2015. IEEE.

Saha, P., Bhattacharjee, D., De, B. K., & Nasipuri, M. (2019). A Survey on Image Acquisition Protocols for Non-posed Facial Expression Recognition Systems. *Multimedia Tools and Applications, 78*(16), 23329-23368.

Saragih, J. M., Lucey, S., & Cohn, J. F. (2009). Face alignment through subspace constrained mean-shifts. In: *2009 International Conference of Computer Vision*, Kyoto, Japan, 29 September - 2 October 2009. IEEE.

Saul, L. K., & Roweis, S. T. (2000). An introduction to locally linear embedding. *unpublished. Available at: http://www. cs. toronto. edu/~ roweis/lle/publications. html.* Retrieved from http://www. cs. toronto. edu/~ roweis/lle/publications. html

Sebe, N., Sun, Y., Bakker, E., Lew, M. S., Cohen, I., & Huang, T. S. (2004). Towards authentic emotion recognition. In: *IEEE International Conference on Systems, Man and Cybernetics*, The Hague, Netherlands, 10 - 13 October 2004. IEEE.

Sharma, M., Khera, S. N., & Sharma, P. B. (2019). The troika of artificial intelligence, emotional intelligence and customer intelligence. *International Journal of Social Computing and Cyber-Physical Systems, 2*(2), 119-131.

Shiyuan, H. (2002). Early meanings and application of the terms “ethnos” and “ethnic group”. *Ethno-National Studies,* (4), 1-10.

Souto, T., Silva, H., Leite, A., Baptista, A., Queirós, C., & Marques, A. (2019). *Facial Emotion Recognition: Virtual Reality Program for Facial Emotion Recognition—A Trial Program Targeted at Individuals With Schizophrenia.* Rehabilitation Counseling Bulletin, p.0034355219847284.

Shu, C., Ding, X., & Fang, C. (2011). Histogram of the oriented gradient for face recognition. *Tsinghua Science and Technology, 16*(2), 216-224.

Szeliski, R. (2010). *Computer vision: algorithms and applications.* Springer Science & Business Media.

Tao, X., Liu, Y., Jiang, C., Wang, Z., & Qin, X. (2018). QoE-Oriented Multimedia Assessment: A Facial Expression Recognition Approach. *IEEE MultiMedia, 26*(1), 41-50.

Tarnowski, P., Kołodziej, M., Majkowski, A., & Rak, R. J. (2017). Emotion recognition using facial expressions. *Procedia Computer Science, 108*, 1175-1184.

Tenenbaum, J. B., De Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science 290*, *5500*, 2319-2323.

Tian, Y. I., Kanade, T., & Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(2), 97-115.

Tin, H. H. K., & Sein, M. M. (2011). Race identification for face images. *ACEEE International Journal on Information Technology, 1*(02), 35-37.

Universitet, L. (2012). *Candice 3 Model*. Retrieved 8 2018, from http://www.icg.isy.liu.se/candide/

Valstar, M.F., Sánchez-Lozano, E., Cohn, J.F., Jeni, L.A., Girard, J.M., Zhang, Z., Yin, L. & Pantic, M. (2017). FERA 2017- Addressing Head Pose in the Third Facial Expression Recognition and Analysis Challenge. In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition*, Washington, DC, 30 May - 3 June 2017, IEEE.

Wallhoff, F. (2006). *Facial expressions and emotion database.* Munich, Germany: Technische Universitat Munchen.

Wan, S., & Aggarwal, J. K. (2014). Spontaneous facial expression recognition: A robust metric learning approach. *Pattern Recognition, 47*(5), 1859-1868.

Wang, C., Zhang, Q., Liu, W., Liu, Y., & Miao, L. (2019). Facial feature discovery for ethnicity recognition. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9*(1), 1278.

Wang, L., Xue, Q., Hao, J., & Yu, H. (2019). Research of Human-Machine Interface Evaluation Based on Cogtool. In: *International Conference on Human-Computer Interaction.* Paphos, Cyprus, 2 - 6 September 2019, Springer.

Wang, S., Liu, Z., Lv, S., Lv, Y., Wu, G., Peng, P., Chen, F. & Wang, X., (2010). A natural visible and infrared facial expression database for expression recognition and emotion inference. *IEEE Transactions on Multimedia*, *12*(7), 682-691.

Whissell, C. M. (1989). The Dictionary of Affect in Language. In *Emotion: Theory, Research and Experience* (Plutchik, R. and Kellerman, H., Eds., Vol. 4) pp. 113-131. Academic Press.

Woodworth, R. S. (1938). *Experimental psychology.* New York: Experimental psychology.

Wouters, N., Kelly, R., Velloso, E., Wolf, K., Ferdous, H.S., Newn, J., Joukhadar, Z. & Vetere, F. (2019). Biometric Mirror: Exploring Ethical Opinions towards Facial Analysis and Automated Decision-Making. In: *Proceedings of the 2019 on Designing Interactive Systems Conference,* San Diego, USA, 23 - 28 June 2019, ACM.

Yan, W. J., Li, X., Wang, S. J., Zhao, G., Liu, Y. J., Chen, Y. H., & Fu, X. (2014). CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PloS one*, *9*(1), e86041.

Yin, L., Sun, X. C. Y., Worm, T., & Reale, M. (2008). A high-resolution 3D dynamic facial expression database. In: *8th IEEE International Conference on Automatic Face Gesture Recognition*, Amsterdam, Netherlands, 17 - 19 September 2008. IEEE.

Yoo, S. H., Matsumoto, D., & LeRoux, J. A. (2006). The influence of emotion recognition and emotion regulation on intercultural adjustment. *International Journal of Intercultural Relations*, *30*(3), 345-363.

Yu, K. T. (2018). *Cross-Cultural Examination of the Cognitive Theory of Depression Among Individuals of Chinese and Canadian Descent.* Doctoral dissertation, Faculty of Graduate Studies and Research, University of Regina.

Zhalehpour, S., Onder, O., Akhtar, Z., & Erdem, C. E. (2016). BAUM-1: A spontaneous audio-visual face database of affective and mental states. *IEEE Transactions on Affective Computing*, *8*(3), 300-313.

Zhang, J., Shen, W., Liu, L., & Wu, Z. (2018). Face recognition model based on privacy protection and random forest algorithm. In *The* 27th *Wireless and Optical Communication Conference*, Hualien, Taiwan, 30 April - 1 May 2018, IEEE.

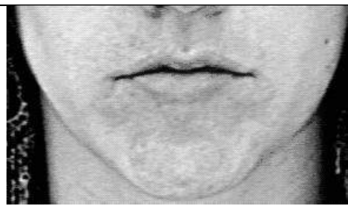
Zhong, C., Sun, Z., & Tan, T. (2009). Fuzzy 3D face ethnicity categorization. In: Tistarelli, M. and Nixon, M. S., *2009 International Conference on Biometrics*, Alghero, Italy, 2 - 5 June 2009, Springer.

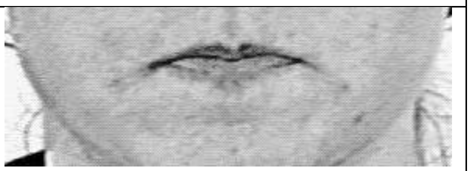
Zinia, S., Azmol, A. I., & Islam, S. (2018). *Performance analysis of different machine learning approaches for single modal facial expression detection.* BRAC University: Doctoral dissertation, BRAC University.

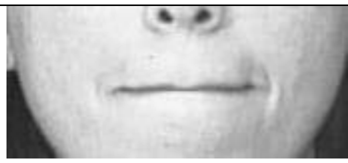
Zuo, H., Wang, L., & Qin, J. (2017). XJU1: A Chinese Ethnic Minorities Face Database. *2017 International Conference on Machine Vision and Information Technology*, 7-11.

APPENDIX

AU	Discription	Facial Muscle Involved	Example Image
1	Inner Brow Raiser	*Frontalis, pars medialis*	
2	Outer Brow Raiser	*Frontalis, pars lateralis*	
4	Brow Lowerer	*Corrugator supercilii, Depressor supercilii*	
5	Upper Lid Raiser	*Levator palpebrae superioris*	
6	Cheek Raiser	*Orbicularis oculi, pars orbitalis*	
7	Lid Tightener	*Orbicularis oculi, pars palpebralis*	
9	Nose Wrinkler	*Levator labii superioris alaquae nasi*	
10	Upper Lip Raiser	*Levator labii superioris*	

11	Nasolabial Deepener	*Zygomaticus minor*	
12	Lip Corner Puller	*Zygomaticus major*	
13	Cheek Puffer	*Levator anguli oris (a.k.a. Caninus)*	
14	Dimpler	*Buccinator*	
15	Lip Corner Depressor	*Depressor anguli oris (a.k.a. Triangularis)*	
16	Lower Lip Depressor	*Depressor labii inferioris*	
17	Chin Raiser	*Mentalis*	

18	Lip Puckerer	*Incisivii labii superioris and Incisivii labii inferioris*	
20	Lip stretcher	*Risorius w/ platysma*	
22	Lip Funneler	*Orbicularis oris*	
23	Lip Tightener	*Orbicularis oris*	
24	Lip Pressor	*Orbicularis oris*	
25	Lips part**	*Depressor labii inferioris or relaxation of Mentalis, or Orbicularis oris*	
26	Jaw Drop	*Masseter, relaxed Temporalis and internal Pterygoid*	

27	Mouth Stretch	*Pterygoids, Digastric*	
28	Lip Suck	*Orbicularis oris*	
41	Lid droop**	*Relaxation of Levator palpebrae superioris*	
42	Slit	*Orbicularis oculi*	
43	Eyes Closed	*Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis*	
44	Squint	*Orbicularis oculi, pars palpebralis*