UNIVERSITI MALAYSIA SARAWAK
UNIMAS

Faculty of Computer Science and Information Technology

Facial Expression Synthesis using Kernel Approach

Marcella Peter

Master of Science
2020

Facial Expression Synthesis using Kernel Approach

Marcella Peter

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

2020

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: Marcella Peter

Matric No.: 16020055

Faculty of Computer Science and Information Technology

Universiti Malaysia Sarawak

Date: 4 February 2020

ACKNOWLEDGEMENT

Praise to the Lord! The dream that I once thought impossible has come true. Completing this thesis has been a wonderful journey. I would not have been here without the support from them.

My family especially to my beloved parents, Peter and Justina, ‘pounsikou au gigina kumaa di’koyu’ for always believing in me from beginning until the end. My sisters, Collete and Melissa for their constant support.

I would also like to extend my gratitude to my supervisor, Dr Jacey-Lynn Minoi, for her endless support and patience throughout my Master studies. Without her constructive comments and constant encouragement, I could never have finished my thesis successfully. My co-supervisor, Dr. Irwandi Hipni Mohamad Hipiny, it is a great opportunity to be working with you. I am grateful to my thesis examiners for evaluating my research work and their positive comments. Special thanks to MyBrain15 Scholarship for sponsoring my tuition and the Centre for Graduate Studies for their guidance in completing this study.

Huge appreciation to Khairunissa, Woon Chee, Faizol, Annes, Anisia and Ima for their constant support. They always share their knowledge and keep me up to date on my surroundings. To Kak Suriani, thank you for answering my never-ending questions, offers me for a ride and ideas when I am lost in ambiguity.

Finally, to all my friends, while their names are not mentioned here, I appreciate your home cooked meals, for bringing me out to eat and socialise just to make sure my life is well-balanced.

ABSTRACT

Recently, facial identity and emotion study has gained some interest from researchers especially in the works of integrating human emotions and machine learning to improve the current lifestyle. Emotions are initially expressed through facial expression and followed by body language to deliver information. By nature, emotions can be easily expressed such as happiness, sadness, and surprised. However, in computer language, it is still a challenging task especially synthesising realistic facial expression. Therefore, various methods have been proposed to synthesise better facial expression systems that include learning-based and statistical-based approaches. Most of these approaches applied linear methods and the most commonly used one is the Principal Component Analysis (PCA). PCA is a linear transformation technique and can be used for reducing high dimensional data, extracting facial features from an input, transforming the extracted features to represent a face via a face model and subsequently extended for face recognition system. However, linear transformations may lead to some information loss along the way. Furthermore, the facial structure of a face in itself is complex to be expressed using a linear method. Therefore, in this study, a kernel-based method is proposed to deal with the linear approach problems on transformation and projection. This study explored the potential of using nonlinear kernel approach for synthesising neutral facial expressions 3D geometric face models to improve the performance and recognition rates. The kernel approach employed in the research is a novel modified kernel-based Active Shape Model whereby it employed mean template-based face model. The results from the modified kernel method is then compared with the linear-based Active Shape Model and the outcome of the face recognition is used to evaluate the resulting synthesised neutral facial expression. Experiment results have recorded the highest recognition rate with 100% of true positive and have also shown that the recognition outperformed the linear Active Shape Model. The qualitative results of the synthesis have also shown the almost (if not) real facial expressions of the subject. In conclusion, the proposed modified kernel-based Active Shape Model using template-based approach can improve the synthesis of facial expression which then would increase the performance of the recognition rates. Future work would include to further investigating the effect of adjusting expression intensity on the shape model of the synthesised facial expression by integrating the nonlinear approach into an automated face recognition system and applying optimisation approach to improve the efficiency of the modified kernel-based Active Shape Model.

Keywords: Neutral facial expression synthesis, 3D face model, kernel active shape model

Sintesis Ekspresi Wajah Menggunakan Pendekatan Kernel

ABSTRAK

Kajian mengenai wajah dan emosi telah menarik minat penyelidik terutamanya dalam kerja-kerja mengintegrasikan emosi manusia dan pembelajaran mesin untuk memperbaiki gaya hidup semasa. Emosi pada mulanya dinyatakan melalui ekspresi wajah dan diikuti oleh bahasa tubuh untuk menyampaikan sesuatu maklumat. Secara semulajadi, kita dapat dengan mudah meluahkan emosi seperti kebahagiaan, kesedihan dan kejutan tetapi dalam bahasa komputer, ia masih menjadi tugas yang mencabar terutamanya mensistesiskan ekspresi wajah yang realistik. Oleh itu, banyak pendekatan telah dicadangkan untuk mensintesis ekspresi muka dengan lebih baik termasuklah pendekatan berasaskan pembelajaran dan pendekatan berasaskan statistik. Kebanyakan pendekatan ini menggunakan kaedah linear dan yang paling biasa digunakan adalah Analisis Komponen Utama (PCA). PCA adalah teknik transformasi linear dan boleh digunakan untuk mengurangkan data dimensi yang tinggi, mengekstrak ciri-ciri wajah dari input, mengubah ciri yang diekstrak untuk mewakili wajah melalui model muka dan seterusnya diperluaskan untuk sistem pengenalan wajah dan aplikasi. Walau bagaimanapun, transformasi linear mungkin akan mengakibatkan kehilangan maklumat di sepanjang proses tersebut. Tambahan lagi, struktur wajah wajah itu sendiri adalah rumit untuk diungkapkan menggunakan kaedah linear. Oleh itu, dalam kajian ini, satu kaedah berasaskan kernel dicadangkan untuk menyelesaikan masalah linear. Kajian ini meneroka potensi pendekatan kernel tak linear melalui sintesis wajah neutral berasaskan model geometri 3D untuk meningkatkan prestasi dan kadar pengecaman wajah. Pendekatan kernel yang digunakan adalah Model Bentuk Aktif berasaskan kernel yang telah diubahsuai di mana ia menggunakan model wajah berasaskan templat. Kaedah kernel yang dicadangkan ini telah *dibandingkan dengan Model Bentuk Aktif berasaskan konvensional linear dan hasil daripada kadar pengenalan wajah digunakan untuk menilai ekspresi muka neutral yang telah dihasilkan. Keputusan eksperimen yang diperolehi telah mencatatkan kadar tertinggi dengan 100% positif benar dan juga menunjukkan bahawa hasil pengenalan wajah tersebut adalah lebih baik berbanding Model Bentuk Aktif linear. Manakala hasil kualitatif sintesis wajah neutral yang diperolehi menunjukkan ekspresi wajah subjek adalah hampir nyata (kalau tidak) nyata. Secara konklusinya, cadangan Model Bentuk Aktif berasaskan kernel yang telah diubahsuai menggunakan pendekatan berasaskan templat boleh meningkatkan sintesis ekspresi wajah yang mana akan meninngkatkan prestasi kadar pengenalan wajah. Kerja pada masa hadapan merangkumi penyiasatan tentang kesan pelarasan intensiti ekspresi pada model bentuk ekspresi wajah yang disintesis dengan mengintegrasikan pendekatan tak linear ke dalam sistem pengenalan wajah secara automatik dan menerapkan pendekatan pengoptimuman untuk meningkatkan kecekapan model bentuk aktif berasaskan kernel yang telah diubahsuai.*

Kata kunci: *Sintesis ekspresi wajah neutral, model wajah 3D, model bentuk aktif kernel*

TABLE OF CONTENTS

	Page
DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
ABSTRAK	v
TABLE OF CONTENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS	xviii
CHAPTER 1: INTRODUCTION	1
1.1 Overview	1
1.2 Research Problem	4
1.3 Research Objectives	6
1.4 Research Scope	6
1.5 Research Flow	6
1.6 Organisation of the Thesis	8
CHAPTER 2: LITERATURE REVIEW	10
2.1 Introduction	10
2.2 Facial Expression Representation	11

2.3 Approaches of 3D Facial Expression Synthesis 14

2.3.1 Interpolation 14

2.3.2 MPEG-4 Facial Animation Parameter (FAP) 15

2.3.3 Morphing 17

2.3.4 Statistical-based 18

2.3.5 Learning-based 19

2.3.6 Discussions 22

2.4 Research Works on Synthesising Neutral Facial Expression 24

2.5 Brief Overview on Statistical Shape Model Construction 31

2.5.1 PCA based ASM Construction 31

2.6 Motivation of using Kernel PCA as Nonlinear Approach 33

2.7 Summary 35

CHAPTER 3: DATASETS AND PRE-PROCESSING 36

3.1 Introduction 36

3.2 Facial Dataset Acquisition 37

3.2.1 VisionRT 3D Dataset 37

3.2.2 BU3DFE Dataset 38

3.3 Pre-processing of Datasets 38

3.3.1 Facial Features 40

3.3.2 Registration Transformation 40

3.3.3 Pre-processing Steps 44

3.4 Summary 51

CHAPTER 4: PROPOSED METHOD 52

4.1 Introduction 52

4.2 Modified Kernel ASM for Neutral Expression Synthesis 52

4.3 Kernel Active Shape Model 57

4.3.1 Kernel PCA 57

4.3.2 Centring Data in Kernel 60

4.4 Implementation of Proposed mKASM Method 61

4.5 The Kernel PCA-based Face Recognition 66

4.6 Summary 69

CHAPTER 5: EXPERIMENT RESULTS AND ANALYSIS 70

5.1 Introduction 70

5.2 Test 1: Kernel PCA versus PCA Face Recognition 71

5.2.1 Test 1: Experiment 1 and 4 73

5.2.2 Test 1: Experiment 5 to 12 using VisionRT dataset 75

5.2.3 Test 1: Experiment 13 to 60 using BU3DFE dataset 77

5.2.4 Discussions 82

5.3 Test 2: Neutral Expression Synthesis using mKASM 83

5.3.1 VisionRT Qualitative Results and Analysis 84

5.3.2 BU3DFE Qualitative Results and Analysis 86

5.3.3 Discussions 93

5.4 Test 3: Evaluation of the Synthesised Neutral Faces using Face Recognition 94

5.4.1 VisionRT Quantitative Results and Analysis 96

5.4.2 BU3DFE Quantitative Results and Analysis 98

5.4.3 Discussions 110

5.5 Additional Experiment 110

5.6 Comparison of Kernel ASM and Linear ASM 111

5.7 Summary 117

CHAPTER 6: CONCLUSION AND FUTURE WORK 118

6.1 Introduction 118

6.2 Research Conclusion 118

6.3 Research Contributions 119

6.4 Research Limitations 120

6.5 Future Work 120

REFERENCES 121

APPENDICES 129

LIST OF TABLES

		Page
Table 2.1	Emotions coded by Action Units	12
Table 2.2	Summary of facial expression synthesis approaches	20
Table 2.3	Comparisons of linear and nonlinear approach	24
Table 2.4	Research works on synthesising neutral facial expression	26
Table 2.5	Kernel PCA in different applications	34
Table 3.1	The 13 manual landmarks placement selected due to their anatomical distinctness	44
Table 5.1	Experiment sets for Test 1	72
Table 5.2	Experiment sets for Test 2 using mKASM	83
Table 5.3	Experiments sets for Test 3	95

LIST OF FIGURES

		Page
Figure 1.1	A person's expression of smiling, angry and neutral face (Papatheodorou, 2006)	1
Figure 1.2	Analysis-by-synthesis to reconstruct a model	3
Figure 1.3	A facial expresion synthesis pipeline	8
Figure 2.1	The 11 muscle of faces responsible for creating facial expressions (Minoi, 2009)	11
Figure 2.2	Happiness (AU6 + AU12) motion FACS of a face	13
Figure 2.3	Example of AUs (Minoi, 2009)	13
Figure 2.4	Synthesised smile face by interpolation technique (Pighin et al., 2006)	15
Figure 2.5	From left: Neutral face with FAPs and AUs, synthesised image and 3D model (Patel & Zaveri, 2013)	16
Figure 2.6	Generation of new synthesised expression (c) by mapping (a) with (b) (Patel & Zaveri, 2013)	16
Figure 2.7	Example of learning-based experimental result. From left: Input neutral face, target expression face, ground truth by 3D scanner and the output facial expression (Liang et al., 2013)	20
Figure 2.8	(a) Target neutral face, (b) Initialised face model, (c) Synthesised output (Abbound, Davoine & Dang, 2004)	25
Figure 2.9	Reconstruction of facial expressions between a neutral and sad face (Minoi, Thomaz & Gillies, 2011)	27

Figure 2.10 One expression per row. From top-down order is angry, disgust, fear, happy, sadness and surprise. (a)and (d) are the input expressional face, (b) and (e) are ground truth neutral, and (c) and (f) are the synthesised neutral results (Pan et al., 2010) 28

Figure 3.1 The pre-processing steps of 3D face dataset (Papatheodorou, 2006) 39

Figure 3.2 The face *A* transformation of point *a* into its corresponding point *b* in face *B* (Minoi, 2009) 41

Figure 3.3 An example of linear transformation (Papatheodorou, 2006) 42

Figure 3.4 An example of non-rigid transformation (Papatheodorou, 2006) 43

Figure 3.5 The 13 anatomical landmarks on face surface (Papatheodorou, 2006) 45

Figure 3.6 Rigid landmark registration (Papatheodorou, 2006) 46

Figure 3.7 Before and after non-rigid landmark registration 47

Figure 3.8 Non-rigid landmark registration (Papatheodorou, 2006) 47

Figure 3.9 Distance colour map after non-rigid registration (Papatheodorou, 2006) 49

Figure 3.10 Distance colour map between two surface faces based on non-rigid landmark registration and non-rigid surface registration (Papatheodorou, 2006) 49

Figure 3.11 The 83 landmarks selected manually in BU3DFE (Yin et al., 2006) 50

Figure 3.12 A subject from pre-processed VisionRT dataset showing samples of three expressions. From left: Neutral, Frowning and Smiling 50

Figure 3.13 A sample of seven expressions from pre-processed BU3DFE dataset. From left: Neutral, Angry, Fear, Disgust, Happy, Sad and Surprised 50

Figure 4.1 Formulation of Modified Kernel Active Shape Model 53

Figure 4.2 Proposed mKASM algorithm for synthesising neutral expression in 3D faces 54
Figure 4.3 The overview of the implementation 62
Figure 4.4 Kernel ASM module 63
Figure 4.5 Synthesis module 64
Figure 4.6 Face Recognition module 65
Figure 4.7 Show 3D Face module 66
Figure 4.8 Kernel-based PCA face recognition flowchart 68
Figure 5.1 Generic face recogniton framework 71
Figure 5.2 Testing Neutral 1 to Neutral 1 as training Sample in VisionRT dataset 74
Figure 5.3 Testing Neutral to Neutral as training Sample in BU3DFE dataset 74
Figure 5.4 Face recognition rate of VisionRT dataset 76
Figure 5.5 Recognition rate of testing Angry expression to Neutral expression as training sample 77
Figure 5.6 Recognition rate of testing Disgust expression to Neutral expression as training sample 78
Figure 5.7 Recognition rate of testing Fear expression to Neutral expression as training sample 79
Figure 5.8 Recognition rate of testing Happy expression to Neutral expression as training sample 79
Figure 5.9 Recognition rate of testing Sad expression to Neutral expression as training sample 80
Figure 5.10 Recognition rate of testing Surprise expression to Neutral expression as training sample 81

Figure 5.11 A sample of incomplete facial structure from Surprised face data 81
Figure 5.12 Three subjects from VisionRT frown expression synthesised to neutral face 85
Figure 5.13 Three subjects from VisionRT smile expression synthesised to neutral face 86
Figure 5.14 BU3DFE's subject synthesised neutral face from angry expression 87
Figure 5.15 BU3DFE's subject synthesised neutral face from disgust expression 88
Figure 5.16 BU3DFE's subject synthesised neutral face from fear expression 89
Figure 5.17 BU3DFE's subject synthesised neutral face from happy expression 90
Figure 5.18 Average happy face level 3 90
Figure 5.19 Average neutral face from the whole dataset 90
Figure 5.20 BU3DFE's subject synthesised neutral face from sad expression 91
Figure 5.21 BU3DFE's subject synthesised neutral face from surprise expression 92
Figure 5.22 Rate of recognition and RMSE results based on VisionRT's smile as target sample expression 97
Figure 5.23 Rate of recognition and RMSE result based on VisionRT's frown as target sample expression. 97
Figure 5.24 BU3DFE Angry sample recognition results for original angry faces from level 1 until level 4 expression intensities and their corresponding synthesised neutral expression 99
Figure 5.25 BU3DFE Angry sample RMSE for original angry faces from level 1 (A01) until level 4 (A04) expression intensities and their corresponding synthesised neutral expression 99

Figure 5.26 BU3DFE fear sample recognition results for original fear faces from level 1 until level 4 expression intensities and their corresponding synthesised neutral expression 101
Figure 5.27 BU3DFE fear sample RMSE for original fear faces from level 1 (F01) until level 4 (F04) expression intensities and their corresponding synthesised neutral expression 101
Figure 5.28 BU3DFE disgust sample recognition results for original disgust faces from level 1 until level 4 expression intensities and their corresponding synthesised neutral expression 103
Figure 5.29 BU3DFE disgust sample RMSE for original disgust faces from level 1 (D01) until level 4 (D04) expression intensities and their corresponding synthesised neutral expression 103
Figure 5.30 BU3DFE happy sample recognition results for original happy faces from level 1 until level 4 expression intensities and their corresponding synthesised neutral expression 105
Figure 5.31 BU3DFE happy sample RMSE for original happy faces from level 1(H01) until level 4 (H04) expression intensities and their corresponding synthesised neutral expression 105
Figure 5.32 BU3DFE sad sample recognition results for original sad faces from level 1 until level 4 expression intensities and their corresponding synthesised neutral expression 107
Figure 5.33 BU3DFE sad sample RMSE for original sad faces from level 1(SA01) until level 4 (SA04) expression intensities and their corresponding synthesised neutral expression 107

Figure 5.34 BU3DFE surprise sample recognition results for original surprise faces from level 1 until level 4, and their corresponding synthesised neutral expression 109

Figure 5.35 Improvement made on an incomplete surprise face sample from BU3DFE. (a) The original surprise expression. (b) The synthesised neutral from surprise. (c) The ground truth of neutral expression 109

Figure 5.36 BU3DFE surprise sample RMSE for original surprise faces from level 1 (SU01) until level 4 (SU04) expression intensities and their corresponding synthesised neutral expression 109

Figure 5.37 Shape effects on varying b parameters 111

Figure 5.38 Comparison synthesised neutral face of kernel ASM and ASM based approach using VisionRT 113

Figure 5.39 Comparison synthesised neutral face of kernel ASM and ASM based approach using BU3DFE 113

Figure 5.40 Comparison of synthesised neutral face based on kernel ASM approach and ASM approach using VisionRT dataset 114

Figure 5.41 Comparison of mean faces of smile, neutral and frown from VisionRT dataset 115

Figure 5.42 Comparison of synthesised neutral face based on kernel ASM approach and ASM approach using angry and disgust samples in BU3DFE dataset 116

Figure 5.43 Comparison of mean faces of disgust, neutral and happy from BU3DFE dataset 117

LIST OF ABBREVIATIONS

2D	Two dimensional
3D	Three dimensional
ASM	Active Shape Model
AU	Action Unit
FACS	Facial Action Coding System
FAP	Facial Animation Parameter
Kernel PCA	Kernel Principal Component Analysis
KNN	K-Nearest Neighbour
mKASM	Modified kernel-based Active Shape Model
MPEG-4	Moving Picture Experts Group-4
PCA	Principal Component Analysis
RMSE	Root Mean Square Error

CHAPTER 1

INTRODUCTION

1.1 Overview

Facial expression is an important source of social communication to express emotions, such as happy, fear, sad, disgust, surprise and angry. An example is as shown in Figure 1.1, a person is expressing three different facial expression. From an expression, one could tell the emotional state of a person, and this would help to regulate a better social interaction and spoken conversation with others.

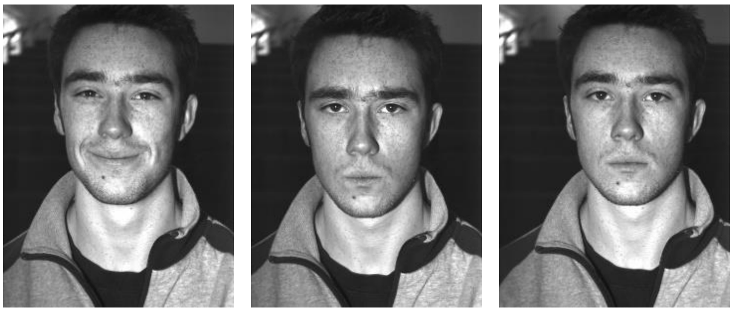


Figure 1.1: A person’s expression of smiling, angry and neutral face (Papatheodorou, 2006)

Researchers took this fact to implement facial expression for the enhancement of the current technologies, especially in areas of facial synthesis and analysis to synthesise a photorealistic image of human faces for medical or human-computer interaction purposes. According to Wang et al. (2011), the technique of facial synthesis can also be used for face recognition. Humans can recognise facial expressions immediately without much effort as compared to computer. However, there exists the presence of face variants in data, and they are due to facial expressions, lightings, poses and facial hairs. These face variants could lead to low performance of recognition. Therefore, there is a need to understand these variations in order to develop a reliable and robust face recognition. A robust face recognition system is able to deal with multiple variations that would possibly exist within a range of different images or data of the same face (Kouzani, 1999).

This research project will deal with the nonlinear approach using kernel method in dealing with neutral expression synthesis invariant of facial expression. Cheng, Zheng and Wang (2007) stated that the analysis of facial expressions remains a challenge for computer because facial expressions change the whole location of facial points. In case of illumination and gestures, which both are rigid transformations, they rely on a fixed facial point such as the nose tip. The topic of facial expression is also chosen due to the fact that facial expression variant will always exist in both two dimensional (2D) and three dimensional (3D) geometric model (Minoi, Thomaz & Gillies, 2011). In two-dimensional (2D) area, synthesising is difficult because of illumination and gestures interferences. But 3D shape data is independent on these physical attributes whereby its identity and expressions are accounted. By incorporating facial expression through the facial synthesis in face recognition, the system performance in terms of face identification can be improved. Existing successful facial recognition systems (Ueda & Okajima, 2019; Agianpuye & Minoi, 2014; Pan et al., 2010) utilised expressionless or neutral facial expression as a training database to achieve a higher rate of recognition. Their efforts also have shown that facial and facial expression synthesis is still ongoing research in computer vision and image processing domain. Furthermore, capturing the exact expressionless face as recorded in the authentication system on the spot is impossible due to the natural and random way of a person’s behaviour and emotional state. Therefore, the system is unable to recognise a person since the query face is not available in the database.

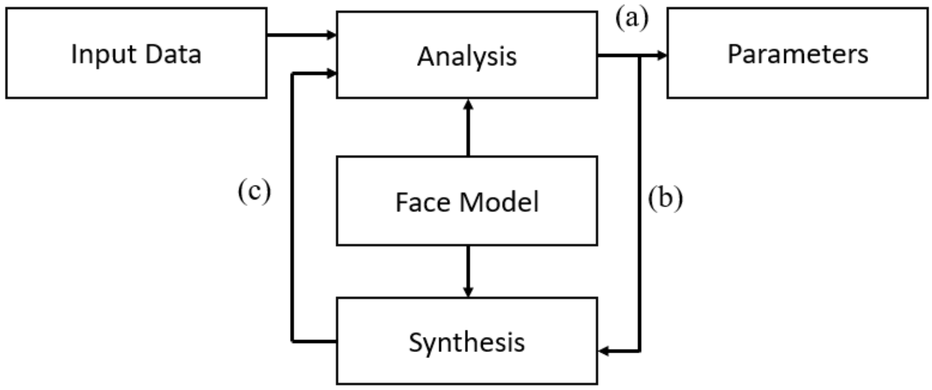


Figure 1.2: Analysis-by-synthesis to reconstruct a model

A model reconstruction approach called analysis-by-synthesis has been employed in computer vision. This paradigm involves recovering shape by comparing a target image with the image obtained using computer graphics-based rendering (Amin & Gillies, 2007). There are three main processes involved as shown in Figure 1.2. The first step as shown in Figure 1.2 (a), an input data extract coefficient such as shape, texture and expression for a model. The second step Figure 1.2 (b), a model is rendered based on the derived coefficient. Lastly, as presented in Figure 1.2 (c), the synthetic model undergoes another analysis to estimate the changes in the model parameters that hold of the shape attributes between the synthetic models with the current data. This approach is highly employed to recover incomplete or missing data. Researchers working in improving face recognition have also taken an interest in this approach. The synthesis of facial expression provides a sequence of remodelled expressional images that a face could have. The synthetic expression can be used in the facial analysis to achieve a better performance of the face recognition system in terms of identification and classification of face images.

There are many approaches that have been proposed to synthesise better facial expression that includes learning-based and statistical-based approaches. Most of these approaches incorporate linear methods - the Principal Component Analysis (PCA) or nonlinear methods, such as Kernel PCA. Jollife (1990) introduced PCA to emphasise variation and bring out strong patterns in a dataset by calculating the main axes of variations. PCA is used in various purposes to reduce the dimension of the dataset by extracting facial features and transforming the extracted data into a most discriminant space that would represent a face model in a lesser dimension. This space is usually useful for many analysis and recognition applications. However, some issues arise in linear method whereby the linear transformations and projections may lead to some information to be lost along the way and the human face structure in itself is a complex model to be simplified with just using a linear approach. Hence, kernel PCA allows a generalisation of the linear method to extract patterns in a nonlinear way. Schölkopf et al. (1998) proposed the kernel PCA, which is widely used for nonlinear feature extraction in face recognition system and building nonlinear shape model of faces. The method is similar to the Support Vector Machine (SVM) and it has proven to be useful for various applications for denoising.

1.2 Research Problem

In many face recognition systems, facial variants are contributors to low recognition rates. In 3D data sets, facial expression variant is distinctly a problem. Facial expression consists of emotional expressions, non-expressional expression and conversation expressions. Synthesising facial expression is a challenging task. Based on current literature, no one method would robustly work to synthesise all facial expressions. Although various synthesising methods have been developed, such as the deep architecture Generative Adversarial Network (Song et al., 2018), Expression Proportion Distribution (Agianpuye & Minoi, 2014) and Morphable model (Blanz, 2006), the study of synthesising faces while retaining individual traits is still a great challenge especially on those extreme higher level of emotional expressions, fear, surprise and disgust that moves most facial muscles and the mandible. Synthesising facial expression can be divided into two types;

i) Synthesising different facial expressions

Synthesising different facial expression refers to adding facial expression into a neutral face. Synthesising a realistic face from different facial expression is complex because facial expressions are generated in nonlinear way. Human face is a complex form composed by skeletons, muscles, and skin which makes it so elusive to analyse (Liu et al., 2011). But due to the nature of human face that is nonlinear, synthesising a realistic face is rather difficult. The facial muscles that generate facial expressions, behave in nonlinear way. Usually, a face data is represented by a linear model in a cartesian space, which is 2D. However, this is insufficient in case of modelling a 3D face because of the shape is rather nonlinear. Due to this, only few researchers have discussed on the topic of nonlinear in facial expression synthesis.

ii) Synthesising neutral facial expression

On the other hand, Blanz (1999) stated that synthesising neutral expression from expressional faces is more challenging than vice versa. Removing facial expression or neutralising is much more complex than adding facial expression into neutral face because it requires one's neutral face and at the same time, still need to preserve one's individual traits. Furthermore, only few literatures found that actually work on synthesising neutral expressions.

Most linear approach are commonly used in facial expression synthesis. Therefore, this research only looked at the possibility of using nonlinear approach for improving face recognition rate by synthesising neutral faces given faces and facial expression are generated in nonlinear way.

1.3 Research Objectives

The following are the research objectives:

i. To synthesise neutral faces using kernel approach.
ii. To evaluate the synthesised neutral faces.

1.4 Research Scope

The research will focus on synthesising face surfaces using nonlinear discriminant method. Existing basis face datasets to be used in the study is a 3D geometric facial data containing facial expressions captured using the VisionRT (Papatheodorou, 2006) and BU3DFE (Yin et al., 2006). Both datasets have been pre-processed using (Papatheodorou, 2006) methods. Template-based Active Shape Model (ASM) and kernel ASM as nonlinear methods will be employed in the synthesis.

1.5 Research Flow

In general, the facial expression synthesis pipeline consists of three main steps. First, the particular model of a subject is obtained and fitted into a predefined prototype mesh. Next, the constructed face model is deformed to produce facial expression. The final stage, wrinkles and vascular effects are considered for added realism. In this research, the facial expression synthesis pipeline is extended and organised into four phases. The pipeline consists of data acquisition and pre-processing, shape model construction, synthesis of new expression, and evaluation which are shown in Figure 1.3.

Step 1 - Data Acquisition and Pre-processing

This step explains the acquisition of datasets that will be used for evaluation of the proposed method. This phase also describes the type of datasets and how raw facial data is pre-processed. The raw data undergoes pre-processing to make it easier to use in subsequent stages. Chapter 3 will expound in detail of this stage.

Step 2 - Shape Model Construction

Selection of method to represent a facial expression model is done based on the reviews on existing facial expression synthesis approaches in Chapter 2. Then, based on the selected model, a proposed method is designed to synthesise neutral facial expression. This phase also includes the construction of a facial shape model for the synthesis process. A detailed explanation of the shape model construction and the proposed synthesis method can be referred in Chapter 4.

Step 3 - Synthesis of New Expression

In this phase, the proposed method is implemented using MATLAB programming. The details of the experiment will be discussed in Chapter 4.

Step 4 – Evaluation

In this phase, the result from the implementation of neutral facial expression synthesis is presented and analysed. The performance of newly synthesised face is measured through visual inspection and fed into face recognition system to test the accuracy. The result will be compared with the benchmark. This will be further detailed in Chapter 5.

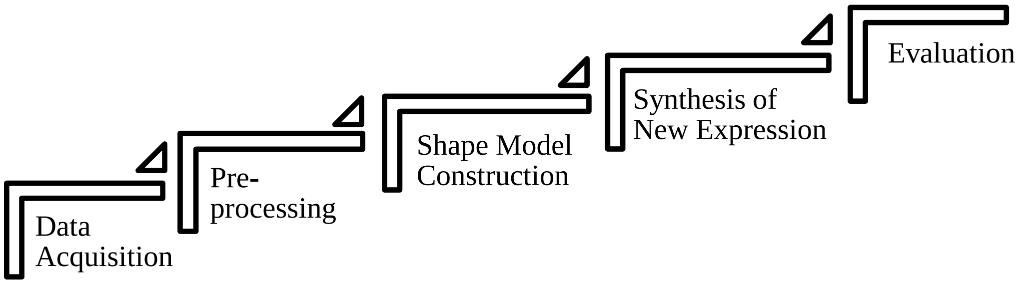


Figure 1.3: A facial expresion synthesis pipeline

1.6 Organisation of the Thesis

This thesis has seven important chapters organised as following:

Chapter 1: Introduction

In this chapter, a preview of facial expression works in face recognition field is presented. Research problem that leads to the idea of using kernel approach is also discussed. In addition, research objectives, research scopes and research flow are also defined in this chapter.

Chapter 2: Literature Review

This chapter gives brief explanation of facial expression, synthesis approaches and some recent works of analysis by synthesis. Also, a discussion in light of the recent works related to facial expression synthesis and chosen kernel approach is presented in this chapter.

Chapter 3: Datasets and Pre-processing

This chapter explains how the 3D face database is acquired and how the facial data scans are pre-processed before proceeding with the synthesis process.

Chapter 4: Proposed Method

This chapter discusses the method implemented in this research. The proposed algorithm, modified kernel Active Shape Model (mKASM) is presented with a diagram. The implementation of the proposed mKASM is discussed in this chapter.

Chapter 5: Experiment Results and Analysis

Several experiments are conducted using different types of facial expressions to evaluate the proposed method. In this chapter, the overall results for the synthesis are further discussed with quantitative and qualitative analysis.

Chapter 6: Conclusion and Future Work

This final chapter concludes the whole thesis. This includes the summary of research contributions based on the analysis and discussions from previous chapter. The limitations of the proposed method are presented along with some improvement which can be made in the future.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Facial analysis and synthesis are crucial in research involving human-centred applications. Face analysis works on the extraction of information, such as head, pose and facial feature motion especially around the eyes and the mouth. Face synthesis refers to the generation of a shape model from a set of parameters that controls pose, gaze and facial expressions. Combination of facial analysis and synthesis would be useful in research and approaches in the area of psychological, 3D animations (Thies et al., 2019), plastic surgery rendering (Bottino et al., 2012), human-centred surveillance (Chen et al., 2018), and authentication systems or notably also known as face recognition. Recently, the implementation of face recognition system has been launched at the airport as additional biometrics to count and identify the on boarding passengers (Nair, 2019). Other than that, face synthesis also used for facial aging (Riaz et al., 2019). The target face is synthesised to a specific age parameter before performing face recognition. Generally, facial synthesis works are closely related to face analysis due to their purpose towards better face recognition.

Facial expression synthesis is a process of generating new face shape from a given face without affecting facial characteristics of the initial face (Agianpuye & Minoi, 2014). Synthesising facial expression is a challenging task and requires high computational in order to generate a realistic facial expression. In this chapter, existing facial expression synthesis approach are reviewed and mainly focusing on model-based approach. The approach incorporates shape variation of a subject face in terms of skin surfaces and muscles which has been widely acknowledged for its successful applications in face analysis domain. Since synthesising expression will involve changing of shapes across the face, it is required to have a face model to interpret the generated face in an image or 3D form. This chapter aims to discuss several approaches to synthesise facial expression by identifying the strengths and limitations of each approach. Synthesising facial expression methods can be considered into two major categories: the 3D geometric manipulation and 2D image manipulation. This research will focus on several 3D geometric manipulation approach that has been widely used for synthesising realistic facial expressions on 3D face, such as interpolation, MPEG Facial Animation Parameter, morphing, statistical-based and learning-based.

2.2 Facial Expression Representation

This section discusses an overview about the how facial expression is described by using a system. Psychologists acknowledged the six basic emotion based facial expressions: angry, fear, disgust, happy, sadness and surprise. In 1978, psychologist Paul Ekman and Wallace V. Friesen studied the relationship between facial expression and deception (Ekman & Friesen, 1978). The facial appearance is controlled by single or combination of facial muscles as shown in Figure 2.1.

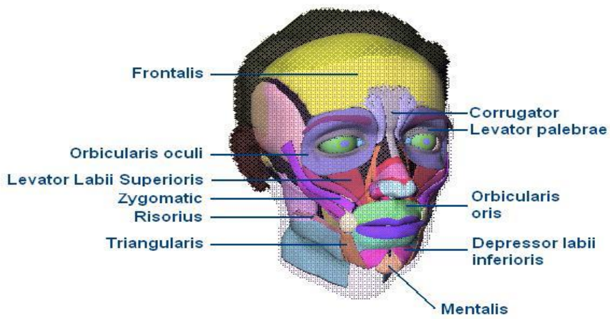


Figure 2.1: The 11 muscle of faces responsible for creating facial expressions (Minoi, 2009)

Their study lead to a manual called the Facial Action Coding System (FACS) to describe the behavior of facial muscles that lead to creation of facial expression (Ekman & Friesen, 1978). According to the FACS scheme, any facial expression results from the mixture of facial muscles. The FACS describes the contraction and relaxation of the facial muscle in terms of Action Unit (AU). The component of AUs can be expanded to expression intensity, facial asymmetry and time duration. In order to generate a facial expression, FACS is manually coded by arranging AUs components and temporal muscles segments as shown in Table 2.1. For an example, based on three level of intensity, a smiling face can be expressed using coded AUs: (AU6+AU12), (AU6+AU12+AU25) and (AU6+AU12+AU25+AU26). The Figure 2.2 presents an example of a face portraying happiness motion based on FACS. There are about sixty AUs to describe a multiple number of facial expressions. Some of the AUs examples can be found in the Figure 2.3. Note that neutral facial expression is defined with 0 AU because there is no facial muscle contraction on the face. There are some facial expression methods which can be found in the next section utilising FACS as a reference for manipulation over facial muscles.

Table 2.1: Emotions coded by Action Units

Emotion	Action Units (AUs)
Neutral	0
Happiness	6,12,25,26
Surprise	1,2,5,26,27
Disgust	9,10,16,17,25,26
Fear	1,2,4,5,20,25,26,27
Sadness	1,4,6,7,11,15,17,25,26
Anger	2,4,5,7,10,17,22,23,24,25,26,27

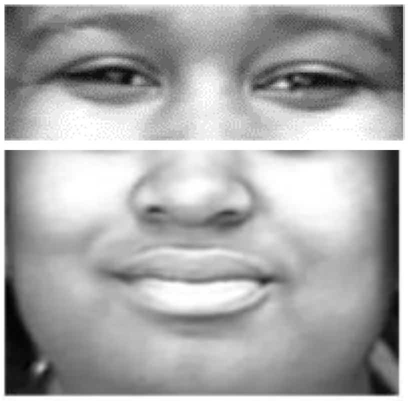


Figure 2.2: Happiness (AU6 + AU12) motion FACS of a face

Figure 2.3: Example of AUs (Minoi, 2009)

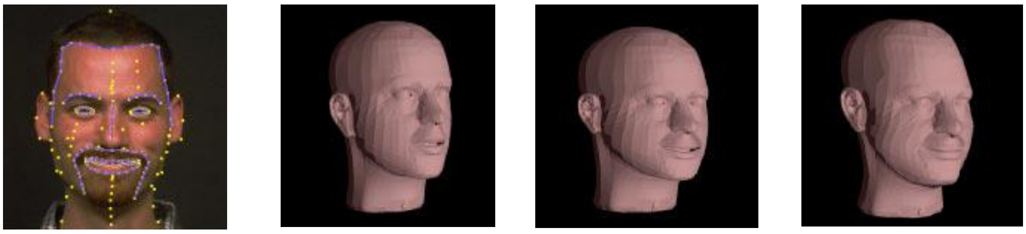
2.3 Approaches of 3D Facial Expression Synthesis

Analysis and synthesis of facial expression research integrate the knowledge of human anatomy in order to reliably imitate and decode a realistic emotion. The following section discusses the capability of facial expressions synthesis approach.

2.3.1 Interpolation

The interpolation method is used for a smooth animation transition between two keyframes over a normalised time interval. The advantage of the interpolation method is its simplicity because a basic facial expression is easily created within a short duration. However, the resulting synthesised expression face turned out to be unrealistic and the expression range is limited to the facial data sample available. As mentioned by Agianpuye (2015), a wide range of facial expression changes can be generated by a combination of simultaneous image interpolation.

An estimated 3D expressive face is synthesised using simple linear interpolation technique by deforming the generic face mesh to fit into a geometry of a subject's face (Pighin et al., 2006). The modelling process of their method is shown in Figure 2.4. Multiple views of images of a subject showing an expression are taken, digitised and are manually annotated with corresponding feature points. The 3D coordinates of the annotated corresponding features are calculated to form a generic model. Then, to mimic the subject's face, the generic model is deformed into 13 corresponding facial points. The new estimated 3D face model portraying a specific facial expression is produced by blending additional correspondence between facial points and texture coordinates as shown in Figure 2.4 (d). The estimated face displayed look more expressive and convincing. However, limitation of this approach is the face model may require more than one head pose from one subject for model fitting process and additional blending tools to make the face to look more realistic.



(a) Facial features marked with set of curves

(b) Generic 3D face model

(c) Adaptation of 13 feature points after pose estimation

(d) Interpolated smile face after 99 additional correspondence

Figure 2.4: Synthesised smile face by interpolation technique (Pighin et al., 2006)

2.3.2 MPEG-4 Facial Animation Parameter (FAP)

The MPEG-4 video coding standard presents a face model for animation based on elementary deformations of facial feature points that are also associated to FAP. This approach generates new facial expression by manipulation of facial parameter from a facial image data (Arya et al., 2014). The manipulation of facial expression involves warping technique using bilinear transformation on a face mesh. According to Malatesta et al. (2009), visemes and expressions are two high-level parameters in facial animation parameter. The MPEG-4 FAP blends shapes of the AUs from FACS to describe expressions. Yet, the facial expression generated using this approach is less realistic.

Patel and Zaveri (2013) generated the six basic facial expressions using this approach on 3D and 2D face data. The experiment is conducted using BU3DFE face database (Yin et al., 2006). First, the distinct features of face such as the mouth, face contours, eyes and eyebrows are identified. Then, the generic 3D model is mapped to a face specific 3D model through geometric transformation. The morphing between two expression models produce a new facial expression. Figure 2.5 illustrates the implementation of MPEG-4 FAP approach. Meanwhile Figure 2.6 shows generation of new facial expression.

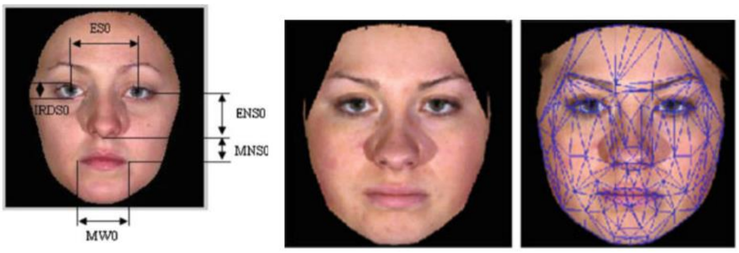


Figure 2.5: From left: Neutral face with FAPs and AUs, synthesised image and 3D model (Patel & Zaveri, 2013)

(a) Sad Expression (b) Surprise Expression (c) New expression

Figure 2.6: Generation of new synthesised expression (c) by mapping (a) with (b) (Patel & Zaveri, 2013)

2.3.3 Morphing

Morphing method divides the whole facial surface like eyes, nose and mouth, into several regions. Each individual model is constructed for each region before blending all individual models back together as predefined model, similar to shape blending in FAP. The synthesis is similar to statistical-based model which consists of learning the face's variations starting from training examples and deriving linear models of shape and texture, but the difference is that the face examples need high resolution laser scans to accommodate colour and 3D texture. Blanz and Vetter (1999) generated new expressive face on a 3D morphable face. Their framework derives a morphable face from a manually labelled example set of 3D face model using linear transformation of model coefficients into vector space representation. The method renders 3D expressional face using model coefficients that consist of facial shape and texture parameters of the examples that are constricted within the vector spanned by the face database (Blanz & Vetter, 1999). New expression is synthesised using the performance-based technique that Blanz and Vetter (1999) proposed by recording two different expressional face scans of the same subject, and then adding the expressional residue differences of model parameters to another subject with neutral expression. Other than facial expression, the method also able to synthesise other facial attributes such as gender, fullness of face, double chin and nose shapes. However, an extended work to test the 3D morphable model by Blanz (2006), mentioned that the face recognition performance is fully independent to the accuracy of the synthesised face. This could be due to the effort of optimising the facial shape and texture parameters simultaneously may lead to higher complexity to compute and may result in inaccurate shapes and textures.

2.3.4 Statistical-based

Statistical-based shape models are derived based on real data. The advantage of this method is the synthesis can be performed in real-time and cost of computing realistic expressive face is relatively low (Agianpuye, 2015). Active Shape Model (ASM) and Active Appearance Model (AAM) proposed by Cootes et al. (1995) are widely implemented in facial analysis and synthesis. ASM is used as a statistical shape model for objects by actively deforming an example object to fit into new targeted object shape. Within a set of training examples, mean shape and statistical model are derived and represented by the Point Distribution Model (PDM), used for controlling the shape model. It can be deduced that statistical techniques could be used to find the coefficient range for neutral facial expression through the mean points and distribution model.

The statistical model based on ASM can be extended with statistical appearance model based on AAM. Wang et al. (2008) demonstrate 3D dynamic expressive face synthesis using least square conformal maps and additional feature correspondence using AAM on four types of facial expressions, smile, sadness, surprise and anger. The difference between AAM and ASM to model the 3D shape model is the AAM made up of 3D geometrical points with 2D texture points while ASM only uses 3D geometrical points as shape component. There are still many works to be explored in this area especially towards nonlinear approach to improve the quality of facial expression analysis and synthesis for the purpose of improving facial recognition.

2.3.5 Learning-based

This approach drawn most attention in recent works on modelling 3D face. The learning-based approach often assume that a pair of target expressive face source is available for transferring. Learning-based approach recovers the 3D shapes of faces according to the information shared by the 3D shape and 2D image subspace (Liang et al., 2013, 2015). A coupled dictionary learning proposed by Liang et al. (2015) generated a transitional 3D expressive face through combination of linear coefficients. In order to derive the intermediate 3D landmark, the generic representation of 2D and 3D landmark is attained through the coupled dictionary. The local deformation affects the vertices that are located close to the landmarks. Their method produced insufficient details around eyes and eyebrows as shown in Figure 2.7. Based on the figure, from left displays the input neutral face, the target expressive face, the ground truth of the expressive face from the 3D scanner and the last column shows the result of their proposed method. Based on the two faces showing neutral and target disgust, modelling without texture will lead to confusion of what emotion is expressed. Moreover, their method only limited to 2D front view images as input.

Furthermore, a nonlinear approach where a 3D shape is constructed through nonlinear manifold embedding and alignment proposed by Wang and Yang (2010). The 3D model is learned by using nonlinear dimensionality reduction technique to establish the local model for each facial patch. The local surface model is then generated for each 3D shape patch (Wang & Yang, 2010).

(a) (b) (c) (d)

Figure 2.7: Example of learning-based experimental result. From left: Input neutral face, target expression face, ground truth by 3D scanner and the output facial expression (Liang et al., 2013)

Deep learning and manifold learning frameworks are also used for various facial features in face modelling process (Lv et al., 2019). This framework could achieve better facial recognition compared to statistical based methods because the method gives dramatic improvements in expression reconstruction, but the frameworks are too complicated to implement. This is because of the manifold or the network model is difficult to interpret and could not be traced easily. Other than that, the training requires a large number of databases in order to achieve good result. Table 2.2 summarises the strength and weaknesses on the existing works of synthesising facial expression in 3D.

Table 2.2: Summary of facial expression synthesis approaches

Approach	Authors	Method	Advantage	Disadvantage
Interpolation	Pighin et al. (2006)	Linear interpolation, Blend shape model	Its simplicity	Limited expression range and require more than one pose to model face and need additional blending tools to make the face more realistic.

Table 2.2 continued

MPEG-4 FAP	Patel & Zaveri (2013)	Warping trough bilinear transformation	Efficient	The FAP does not represent realistic movements directly and impossible to create an expression that does not exist in dataset.
Morphing	Blanz & Vetter (1999)	Morphable model	Very robust 3D scanner captures accurate 3D face	Computationally expensive and the example set of 3D face model is manually labelled, difficult to implement.
Statistical-based	Wang et al. (2008)	AAM	Widely explored by research community Shape and appearance are jointly encoded by model parameters	Dependent on initialisation and rely on PCA for modelling variation
Learning-based	Liang et al. (2015)	RBF Network and Coupled dictionary learning	Efficient	Cannot handle non-frontal facial image, and does not include high expression intensity, Parameter learning algorithm is slow.
	Lv et al. (2019)	Manifold learning and alignment Poisson Based deformation	Better recognition compared to statistical model due to good reconstruction of expression Realistic face expression model	Network model is difficult to learn and implement, require large number of datasets to achieve good result. Does not include high expression intensity.

2.3.6 Discussions

This section introduces a few of facial expression synthesis approaches. Based on the literature discussed, several synthesising approaches applied on a 3D face data are outlined and some interesting trends in the research are found. The first is the type of database used in evaluating the approaches. The face databases are acquired from computer generated blend face shapes, conversion from single or multi view 2D face image and using 3D camera scanner to capture human subjects. Secondly, the face data undergoes pre-processing before proceeding to the analysis and synthesis process. Another thing to note is that the preferred method of synthesising 3D faces method varies depending on the purpose since some might only work better at specific tasks. For example, a statistical-based method is mostly applied in face recognition while interpolation and morphing approach usually implemented in visual animation and computer games.

In general, facial expression synthesis approaches are categorised into parameter based approach, example-based, and learning-based. Parameter-based parameterise 3D expressive faces and are dependent on their shape and manipulation of the parameter set. Approaches in this category add textures to enhance the appearance of realistic 3D faces but the facial shape that masked with textures are difficult to analyse. Interpolation and MPEG-4 Facial Animation Parameter are examples of the parameter-based approach. Example-based or template-based synthesis learns face variations from training examples and requires an example face model to synthesise new facial expression as demonstrated by the morphable model and statistical shape model. Both models used linear transformations to derive coefficient in order to represent a face model. This transformation is also known as feature extraction. However, morphable model derives shape and texture coefficients while statistical shape model derives coefficients via statistical tools, specifically the PDM and mean. Learning-based approach also learns shape information through training faces. Coefficients calculated from the training are used to generate new facial expression. The most commonly used approach is statistical-based, and learning based. While learning-based approach generates more realistic faces as presented in Liang et al. (2015) and Pan et al. (2010), it is complex to compute due to many convolutional layers involved in training and a large number of databases is needed to achieve good result. Statistical-based approach are much more flexible to deform the shape model and requires smaller database to train. For that reason, this research will look more on the statistical-based modelling that fall under example-based approach.

Basically, feature extraction is a core component in face recognition pipeline. Useful shape or facial features are extracted to define the objects in an image clearly. The facial features that made up the template face or example face allows a priori knowledge of the expected shape to be used in the matching process. Feature extraction also a process of dimensionality reduction whereby an initial set of raw data is reduced to more manageable group of data. All facial features in its representation need to be considered to get more accurate facial feature as the accuracy or robustness of the example model is depending the feature extraction method. Facial feature extraction can be divided into linear and nonlinear approach. Table 2.3 shows the comparison of linear and nonlinear approach in facial expression synthesis domain. The linear approach extracts distinct features such as nose tip which does not change much to generate rigid model. Whereas nonlinear approach extracts distinct and hidden features to generate non-rigid face model. Non-rigid model is more personalised compared to rigid model because not only the generic facial features like eyes, nose and mouth are extracted, but non-rigid shapes such as wrinkles and facial expressions are also considered. The similarities of both approaches are both able to extract distinct facial features and synthesise expression by adding or removing expressional residue. The difference is that nonlinear approach able to extract more features that will give more details to the shape model compared to linear approach. The most basic nonlinear approach that this study will look into is kernel method.

Table 2.3: Comparisons of linear and nonlinear approach

Criteria	**Linear**	**Nonlinear**
Strength	Extract distinct features	Extract distinct and discover hidden deformations
Model behaviour	Rigid	Non-rigid
Focus	General template	Person-centric template
Model parameters	Linear model of shapes	Non-linear model of shapes
Synthesis	Adding or removing expressional residual differences of model parameters	

2.4 Research Works on Synthesising Neutral Facial Expression

There are many works found on synthesising different facial expression which some examples can be found in the previous section. According to Chen et al. (2014), neutral facial expression synthesis enhances the performance of expressive face recognition with a single neutral sample in gallery per subject. Despite that, at present, only few works are found on synthesising neutral expressions. Table 2.4 presents few literatures found on neutralising facial expression.

Abbound, Davoine and Dang (2004) proposed a modified Active Appearance model to synthesise different facial expressions. Linear regression is used to construct two facial expressional models, namely the appearance parameters and the facial expression intensity parameter. The idea is to directly combine the normalised training shapes and textures vectors and performing a single PCA on the resulting combined vectors (Abbound, Davoine

& Dang, 2004). The experiment results as shown in Figure 2.8 show a satisfactory neutralised expression. Figure 2.8(a) shows the target neutral expression, followed by an initialised face model as first estimation consists of mean shape and texture. While Figure 2.8(c) is the synthesised output that converges to the target neutral face.

Figure 2.8: (a) Target neutral face, (b) Initialised face model, (c) Synthesised output (Abbound, Davoine & Dang, 2004)

Chang, Zheng and Wang (2007) proposed a Bilinear Identity-Expression Model to synthesise different facial expression to neutral face and vice versa on four types of facial expressions, neutral, surprise, smiling and angry. The researchers trained bilinear model from a 3D face data directly using least square to extract facial features and recover missing data using ASM to synthesise facial expressions. The experiment results have shown that the neutralised face had great similarity with the original neutral face image and helped solving problems in facial expressions related applications.

Table 2.4: Research works on synthesising neutral facial expression

Paper	Dataset	Methods	Results
Abbound, Davoine & Dang (2004)	2D Face CMU. IMM	AAM (ASM & texture) Linear regression model	• Synthesised different expressions to neutral & vice versa in 2D • The expressions are neutral, anger, disgust, fear, joy, surprise and sadness
Chang, Zheng & Wang (2007)	2D Face	Bilinear Identity-Expression Model ASM	• Synthesised different expressions to neutral & vice versa in 2D & 3D model • The expressions are neutral, surprise, smiling, & anger
Pan et al. (2010)	3D Face BU3DFE	Learning based Nonlinear Regression model Poisson-based Deformation	• Neutralising from angry, disgust, fear, happy, sad, & surprise • Does not consider expression intensity • Synthesised neutral faces are visually realistic
Minoi, Thomaz & Gillies (2011)	3D Face BU3DFE	ASM Tensor-based multivariate Statistical Discriminant Model	• Neutralising from angry, disgust, fear, happy, sad & surprise • Novel in synthesising expressions at different intensities
Chen, Bai & Hua (2014)	2D Face CMU	Bilinear kernel rank regression Piece-wise Affine Warp Poisson based Learning	• Neutralising face from smile, surprise, squint, disgust & scream • Nearest Neighbour classifier 45.49% (Average improvement)
Agianpuye (2015)	3D Face BU3DFE, VisionRT 3D	ASM Expression Proportion Distribution (EPD)	• Neutralising from angry, disgust, fear, happy, sad & surprise • Considers expression intensity • Highest recognition rate 92.6% (Angry expression)
Barbosa, Dahia & Segundo (2019)	3D Face Bosphorus	Learning based Convolutional Neural Network Model	• Main purpose is removing facial expression to improve face recognition accuracy • Does not consider realistic face & expression intensity • Lowest Equal Error Rate: 0.061903 (Local Binary Pattern Histogram)

Minoi, Thomaz and Gillies (2011) proposed a tensor based multivariate statistical discriminant method for modelling facial shapes and for recognition. The approach extracts facial expression using statistical discriminant method and reconstruct facial shapes that include emotional expressions and neutral faces via analysis-by-synthesis. The tensor model provides the facial shape to gradually change according to desired facial expression by moving along the most discriminant direction. Figure 2.9 is presenting one of the reconstruction results. The resulting faces are showing smooth transition and realistic expressions. Besides, the neutralised faces contributed to an improvement in recognition rate when tested in face recognition.

Figure 2.9: Reconstruction of facial expressions between a neutral and sad face (Minoi, Thomaz & Gillies, 2011)

Pan et al. (2010) is believed to be the first to apply a nonlinear approach by using a learning-based method to synthesise neutral facial expressions. Pan et al. (2010) applied a learning-based approach called radial basis function regression model to remove facial expression and constructed the 3D face model based on Poisson-based deformation algorithm. Neutral expression is reconstructed by removing facial expression residue from samples and use expression residue from the input expressive face model (Pan et al. 2010). They used BU3DFE dataset to carry out the experiment and have shown quite realistic results as presented in Figure 2.10. However, their approach does not consider the effect of expression with high intensity that could influence the quality of resulting face.

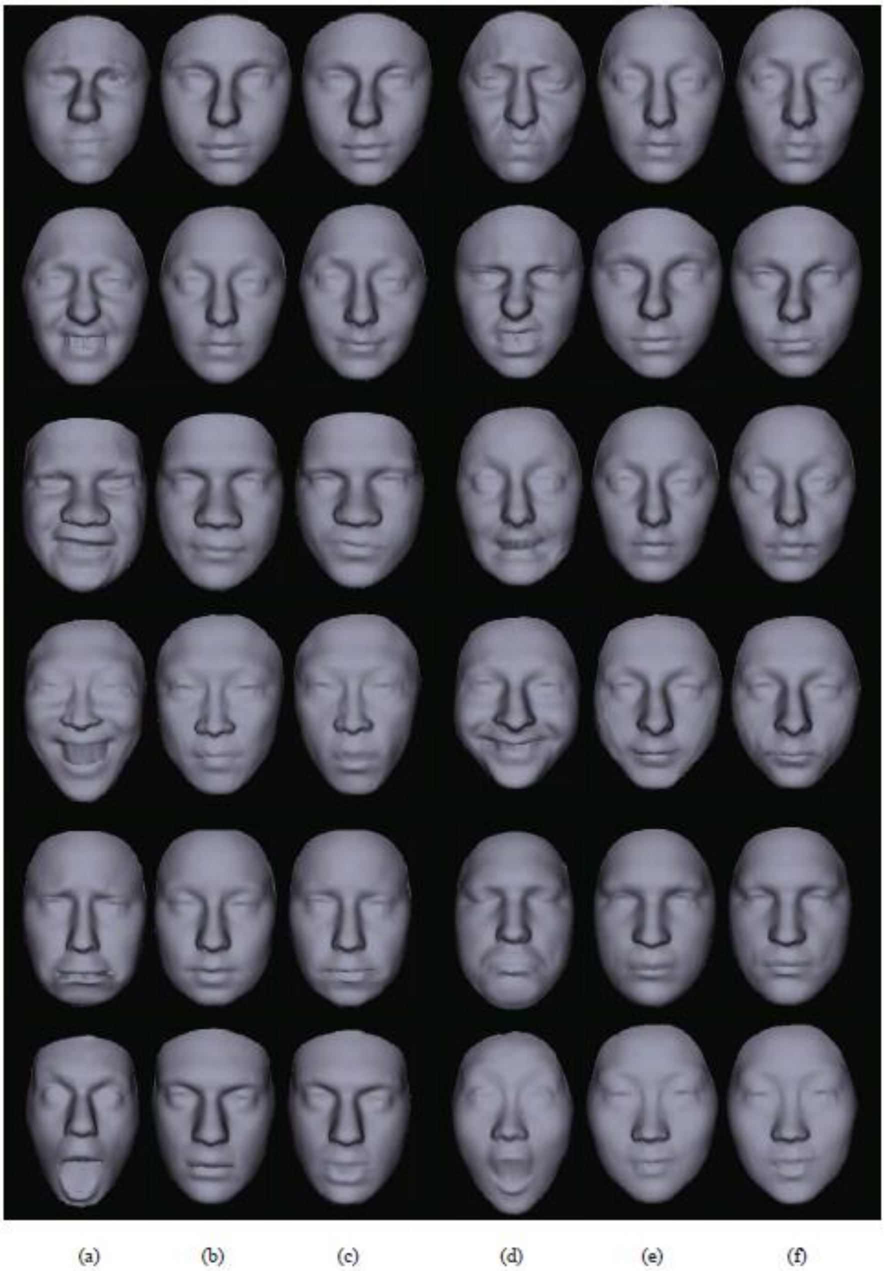


Figure 2.10: One expression per row. From top-down order is angry, disgust, fear, happy, sadness and surprise. (a)and (d) are the input expressional face, (b) and (e) are ground truth neutral, and (c) and (f) are the synthesised neutral results (Pan et al., 2010)

Chen, Bai and Hua (2014) applied a nonlinear approach to neutralise facial expression by proposing the Virtual Subspace Bilinear Kernel Rank Reduced Regression (VSBKRRR) algorithm to improve expression-invariant face recognition. First, both neutral and expressional images are warped to a neutral template using piecewise affine warping.

Then, an expression mask is established by the residual of the two images via learning-based Poisson fusion. Virtual neutralised face image with person-specific attributes preserved is generated by learning-based Poisson fusion from the VSBKRRR neutral image as source, piecewise affine warping image as target face and expression mask as clone area. The experiment results have shown that resulting synthesised neutral faces are almost identical to the ground truth neutral images. When the synthesised neutral face is tested using the Nearest Neighbour classifier, the proposed method outperformed the benchmark, which referred to face recognition without the facial expression neutralisation, by margins of 45.49% above average recognition rate.

The existing work that is also parallel to the direction of this research is proposed by Agianpuye (2015). The proposed method is a modified statistical approach called Expression Proportion Distributions method designed to synthesise a neutral expression on a given non-neutral face. The face model is constructed using ASM. By adding mean neutral to the expressional residual from the training sample with its mean, a new neutral expression is generated. The proportion of neutral expression is enhanced by combining neutral and expressional face in a training sample. Their results showed a promising synthesised expressive face where the synthesised neutral face is nearly identical to the ground truth especially the eyebrows and eyes. This research will adopt this method particularly the addition of mean neutral into the training expression residues to synthesise neutral expression.

Barbosa, Dahia and Segundo (2019) proposed an expression removal algorithm for the purpose of improving face recognition and disregarded realistic results. A 2D orthogonal projection of the 3D input face is passed through a convolutional neural network to regress into a 2D orthogonal projection of the same face but without facial expressions. The neutralised faces are evaluated using four different face recognition and has shown an improvement in the accuracy of face recognition when the expressions are removed.

Based on the Table 2.4, most of the methods applied in existing works is linear approach especially ASM. From the literatures, nonlinear approach has the capability to extract more complex features than linear. For that complexity, only few works found that use nonlinear methods to build face model and subsequently used for synthesising neutral expression in 3D. Besides, based on literature so far, none of them exploring nonlinear ASM, specifically kernel ASM for neutralising 3D face. Most synthesis established by the expressional residue from coefficients or parameter set. Therefore, this study will be looking not only removing the expression from face using kernel but also retaining the identity of the face. This approach falls under nonlinear feature extraction in face recognition pipeline.

This research proposed a two-step method for synthesising neutral expression, shape model construction and synthesis. The reconstruction of neutral expression from expressional face part is similar to performance-based technique in morphing. However, instead of manually measuring the facial feature points for model parameters, the statistical measure via kernel Active Shape Model (kernel ASM) is adopted to derive the nonlinear shape model parameters. The nonlinear shape model is used to construct a pre-model for synthesising new facial expression. The next section will explain the construction of a statistical shape model.

2.5 Brief Overview on Statistical Shape Model Construction

Statistical model synthesises expressions based on a collection of face databases. A realistic facial expression can be generated in real time with low computational cost. One of the examples of statistical-based method is the Active Shape Model (ASM). ASM is one of the well-known methods used in computer vision for discovering deformation pattern of an object. Cootes et al. (1995) first introduced ASM, by iteratively deform to fit an example of object into a new image. A set of points controlled by the shape model is used to represent the object so that the variation can be seen in a training set of labelled examples.

The ASM uses Principal Component Analysis (PCA) to capture the statistics of the training shapes. PCA as a statistical tool, is used to emphasise variation and extract strong patterns in a dataset by calculating the main axes of variation and manipulate the parameters of the axes to produce new shapes.

2.5.1 PCA based ASM Construction

To construct an ASM, a training set is created where landmarks outline the key facial shape properties. The chosen landmarks must be consistent throughout the training set. Alignment of the training set is then applied to align the labelled examples into a common geometric coordinate by translations, rotations and scaling. Then, a correspondence is established between each point of each example in the training set to form a multidimensional space.

Each geometrical shape or example in the training set is represented as a vector denoted as $X = [x_1 y_1 z_1, \dots, x_n y_n z_n]$, where $x_i y_i z_i$ represent the 3D geometric points of an example data and n is the number points.

The PCA algorithm is described as the following:

i. The mean expression, μ is calculated from the training sample, where N is the number of samples,

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad \text{Equation 2.1}$$

ii. Covariance matrix, C is then computed:

$$C = \frac{1}{N}\sum_{i=1}^{N}(x_{i-}\ \mu)(x_{i-}\mu)^T \qquad \text{Equation 2.2}$$

iii. Eigen decomposition from C is then constructed where the *k* principal components are the orthonormal eigenvectors, defined as *w*, correspond to the *k* largest eigenvalues of λ, and S denoted as the training set vector.

$$Sw_k = \lambda_k w_k \qquad \text{Equation 2.3}$$

$$w_k{}^T w_k = 1 \qquad \text{Equation 2.4}$$

In the original ASM algorithm, shape vector denoted as $s,$ is constrained into a plausible shape, represented as $X,$ by projecting s onto the eigenvectors, *W*, where *W* = $(w_1, \dots, w_k)$. These points are approximated by a linear shape model formula also called as the Point Distribution Model as shown in Equation 2.5 where *b* is defined in Equation 2.6.

$$\mathrm{X} \approx \mu + \mathrm{Wb} \qquad \text{Equation 2.5}$$

$$b_N = W^T(\mathrm{s} - \mu) \qquad \text{Equation 2.6}$$

As mentioned by Eguizabal et al. (2018), the *w*'s could be interpreted as the modes of the shape while λ's measure the strength of the modes. Point distribution model allows regularisation in case when *b* parameter is large in magnitude (Cootes et al., 1995). Cootes et al. (1995) stated that a new shape is created with similar attribute of the original training shape sample by regularising the *b* parameter with the limit of $\pm\sqrt[3]{\lambda_k}$.

2.6 Motivation of using Kernel PCA as Nonlinear Approach

This section discusses the use of kernel PCA in several applications and provide justification of why kernel PCA is chosen. PCA is also known as linear dimensionality reduction technique that are commonly used for extracting features. Technically, in facial analysis and synthesis, PCA is a linear method which is effectively used to reduce the dimension of a face image and retain the essential information of the original image (Li et al., 2016). Yet, PCA can only extract linear information during dimensional reduction and this process will cause some information loss as the original data that are relatively nonlinear is not well preserved. This will eventually lead to a lower rate of facial recognition (Li et al., 2016). One of the alternatives to extract features while preserving both linear and nonlinear information is through an extension to the PCA called Kernel Principal Component (kernel PCA), proposed by Schölkopf et al. (1998). Kernel PCA allows generalisation of the traditional PCA for reducing dimensional feature while retaining both linear and nonlinear information. This could be realised with the use of dot product from an input data into higher dimension space and perform linear PCA in that space to extract nonlinear features in a lower dimensional. However, the dot product may lead to expensive operations since high dimensional space is involved. One good thing of using kernel compared to other nonlinear methods such as Local Linear Embedding and Isomap, is the use of 'kernel trick' where substituting the dot product by any kernel function as nonlinear mapping and work on linear PCA without directly working in the higher dimension space. Therefore, kernel PCA is valued for its simplicity and easy to implement (García-González et al., 2020). Due to this, kernel PCA is appreciated as nonlinear dimensionality reduction method in diverse application as shown in Table 2.5. There is also potential that more hidden information can be extracted by using kernel PCA instead of PCA into the ASM.

Furthermore, by nature, human face is non-rigid. The facial structure appeared as nonlinear attributes due to the presence of lighting, pose, wrinkles, physical scars and expressions. The linear method is unable to express all these variations adequately since some of the distinctive values within a face are ignored during the transformation. The ability of the kernel approach to extract the nonlinear data is supported by Cui et al. (2012) where they experimented with a similar approach in X-ray chest image retrieval and implied kernel PCA is effective in extracting feature space and enhanced the precision of retrieving the image. Wang et al. (2015) also combine several kernel functions and tested them using face recognition. Their work performed favourably compared to conventional approach in terms of accuracy and speed.

Table 2.5: Kernel PCA in different applications

Authors	**Domain**	**Remarks**	**Dataset**
García-González et al. (2020)	Image processing	Kernel PCA is used for pre-image reconstruction.	Video frames of open hand that closes and opens on
Elkhadir et al. (2016)	Network security	Kernel PCA overcomes PCA in detecting denial of service (DOS) and probing attacks, especially once the KNN Classifier is used.	KDDcup99 network traffic data
Cui et al. (2012)	Image processing	Kernel PCA is effective in extracting feature space and enhanced the precision of retrieving the image	X-ray images
Kirschner et al. (2011)	Medical image segmentation	Kernel PCA is applied with ASM to determine the correct vertebra class during segmentation	Thoracic vertebrate image

Table 2.5 continued

Romdhani et al. (1999)	Face synthesis	Construct nonlinear ASM using kernel PCA of faces with pose invariant of 2D faces	2D face datasets

2.7 Summary

This chapter presented the facial expression parameters to describe the six archetypal facial expressions. The existing works of synthesis of facial expression also pointed out some of their strength and drawbacks. There are many potential applications using the synthesised facial expressions as noted in introduction section. Based on the review, only few works found that implement nonlinear approach to neutralise expression. A brief introduction of statistical shape model construction and how the linear ASM works is explained. The last section includes motivations of using kernel and some of their applications in different domains. In the next chapter, facial databases and the pre-processing stage will be explained.

CHAPTER 3

DATASETS AND PRE-PROCESSING

3.1 Introduction

This chapter discusses an overview of the datasets that are used for neutral expression synthesis. The datasets have already been pre-processed by Papatheodorou (2006) and Yin et al. (2006). Pre-processing plays a vital role in determining the performance of face recognition and this might also affect the result of the experiment in this research. Pre-processing step is crucial because raw 3D surfaces captured during data acquisition tend to have holes, spikes or large number of vertices, topologies and alignments which generally also called as 'noise' that usually made the data harder to be used (Amin & Gillies, 2007). Hence, a separate chapter is made to explain the pre-processing steps done by Papatheodorou (2006) and Yin et al. (2006) to clean the raw face scans before using the cleaned facial expression datasets for further processes.

Face is a complex three-dimensional (3D) surface with fine details such as wrinkle while skin adds an extra texture layer to the three-dimensional facial surfaces (Amin & Gillies, 2007). The ability to manipulate and reconstruct a two 3D shapes depends strongly on the choice of shape representations. Volumetric representation and surface representation are commonly used to represent a face data. The former one is usually used in medical domain while surface representation is more popular in facial animations and recognition domain. This is due to the surface face data allows shape deformation and realistic rendering (Amin & Gillies, 2007).

In this research, the data used are 3D face surface scans. 3D face is chosen instead of 2D due to its availability and ability to explore the surface information in 3D space which 2D image is limited to do so.

3.2 Facial Dataset Acquisition

The datasets are VisionRT 3D (Papatheodorou, 2006) from Imperial College London and BU3DFE (Yin et al., 2006) from Binghamton University. 3D face data is generated from 2D images and a sensor to capture 3D structure such as optical, laser scanning and acoustic sensors. According to Papatheodorou (2006) and Yin et al. (2006), both facial datasets are acquired from images captured using the stereo 3D camera scanner and merged to produce the 3D face meshes.

3.2.1 VisionRT 3D Dataset

The face dataset consists of 60 (males and females) subjects posing smile, frown and neutral facial expression. There are two levels of neutral facial expression resulting 120 face models while frown and smile each has 60 face models. There is only one level of intensity available on this facial expression database. The raw data is captured in two forms, the original 3D shape data and grey-scale textures. In total after pre-processed, there are 240 surface geometry models and 240 surface textures. The subjects are from 8 South Asians, 6 East Asians, 1 Afro-Caribbean and 45 Caucasians. Most of the subjects are students from 18 to 35 years old. The 3D face models are acquired using a VisionRT 3D stereo camera system (Papatheodorou, 2006). 3D shape is extracted using two cameras meanwhile another camera captured the grey-scale texture. The grey-scale texture faces are stored in bitmap formats.

3.2.2 BU3DFE Dataset

The raw database originally has 100 subjects (males and females) which ranging from age 18 to 70 years old and majority are undergraduate students, graduate students and faculty staffs. The subjects are from a variety of ethnicities including White, Black, East-Asian, Middle East Asian and Hispanic Latino. However, at present, only 54 pre-processed 3D surface face data are available because one need to purchase in order to have access to the whole database. The database consists of two factors, emotion and their intensity. Seven universal emotion pre-sets are acquired from the subjects by performing facial expressions in front of the 3D face scanner. The emotions pre-sets are angry, disgust, fear, happy, sad, surprise and neutral. Except neutral expression, multiple intensities from low, middle, high to highest that reflect different spontaneity level of other expressions are requested from each subject. The 3D scanner fired four instant shots to capture the four intensities of each expression. The neutral facial expression has 54 faces with respect to 54 subjects. Meanwhile other remaining expressions has four expression intensity levels, with a total of 24 poses per subject. Thus, in total after pre-processed, there are 1350 surface geometry models in this dataset.

3.3 Pre-processing of Datasets

Considering both 3D face datasets used in this research have already been pre-processed, this section explains the important steps involved in the work. The pre-processing ensures the 3D face surface is normalised with the same number of vertices and uniform orientation represented by all the face models. As illustrated by Amin and Gillies (2007), Figure 3.1 shows the complete pre-processing steps of 3D surface face dataset as stated by Papatheodorou (2006).

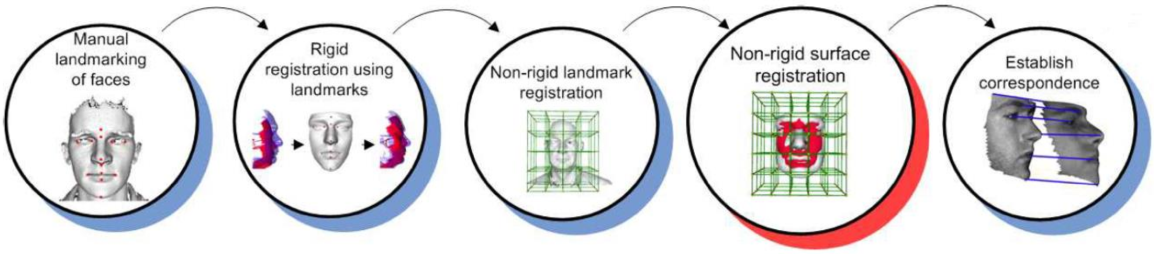


Figure 3.1: The pre-processing steps of 3D face dataset (Papatheodorou, 2006)

Figure 3.1 summarises the overall pre-processing steps. The step-in red shows an additional step proposed by Papatheodorou where non-rigid surface registration is performed for smoother meshes before the correspondence is established. Once the corresponding points has been established, the location of surface points are copied over the template mesh.

Subjects with different facial expressions in the VisionRT and BU3DFE datasets established the 3D correspondence points or aligned using registration algorithm that consists of rigid and non-rigid registration. The registration generates definite linear correspondence between 3D face data points that have differences in geometric configuration and expressions. Based on Minoi (2009), rigid registration corrects pose variation without necessary achieving perfect alignment while non-rigid registration aligns corresponding areas exactly with each other. Initially, the whole surface points from a 3D face data is used as the basis for registration.

In VisionRT face datasets, Papatheodorou (2006) applied iterative closest point algorithm which is an example of rigid registration, thin-plate spline warping and B-spline warping as non-rigid registration to establish the 3D correspondence points.

3.3.1 Facial Features

Facial features are used to represent the prominent attributes of a face such as the eyes, eyebrows, nose and mouth. The integration of this features assists the recognition of face as a whole (Pal et al., 2020). Facial features are also important in many applications that related to faces for instance face alignment and feature extraction for facial expression synthesis. The most commonly used fiduciary facial features are the four eye corners, the tip of the nose and the two mouth corners. Vezzetti and Marcolin (2012) stated that three dimensional (3D) landmarks are extracted from face with the help of an algorithm or manually placed on the face before geometrical features are extracted to perform further process such as face recognition. The VisionRT and BU3DFE datasets are 3D static face data.

3.3.2 Registration Transformation

Registration is used to find the transformation *T* that will map a surface point on a data set to its corresponding point in another dataset. If *a* is a surface point on face *A*, and *b* is the surface point on face *B*, then the relationship is written as:

$$b = T(a) \qquad \text{Equation 3.1}$$

where $a = (a_x, a_y, a_z)$ and $b = \left(b_x, b_y, b_z\right)$ are two corresponding points on two different 3D faces. Figure 3.2 presents the transformation in 2D space. The 3D face is naturally a geometrical structure. During the registration process, geometric distortions are attained through the application of transformation model. The domain of transformation is local when transformation is applied to a certain part of the data and global when applied to the whole data. There are three types of transformation: Rigid, Affine and Non-Rigid transformation.

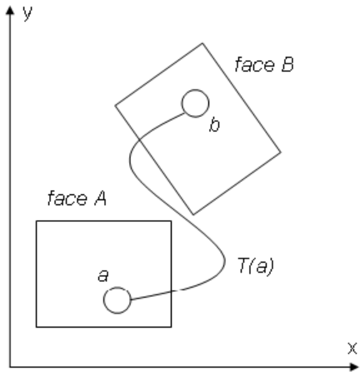


Figure 3.2: The face *A* transformation of point *a* into its corresponding point *b* in face *B* (Minoi, 2009)

a. Rigid Transformation

A rigid transformation is a geometric transformation that preserves all object distances and internal angles. It expressed as a combination of rotation **R** and translation *t* using:

$$b = \mathbf{R}(a) + t \qquad \text{Equation 3.2}$$

The matrix vector is expanded to:

$$\mathbf{T}_{rigid}(a_x, a_y, a_z) = \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \qquad \text{Equation 3.3}$$

where the 3×3 matrix **R** is an orthogonal matrix and t= $[t_x, t_y, t_z]^T$ is the displacement vector describing the translation component.

b. Affine Transformation

Affine transformation preserves parallel lines but not length and angles. The equation is expressed as:

$$b = \mathbf{A}(a) + t \qquad \text{Equation 3.4}$$

where matrix **A** is a matrix represented as:

$$\mathbf{T}_{affine}(a_x, a_y, a_z) = \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \qquad \text{Equation 3.5}$$

The 3×3 matrix **A** is describing the scale, shear and rotation components of the transformation. Figure 3.3 illustrates the rigid and affine shape transformation.

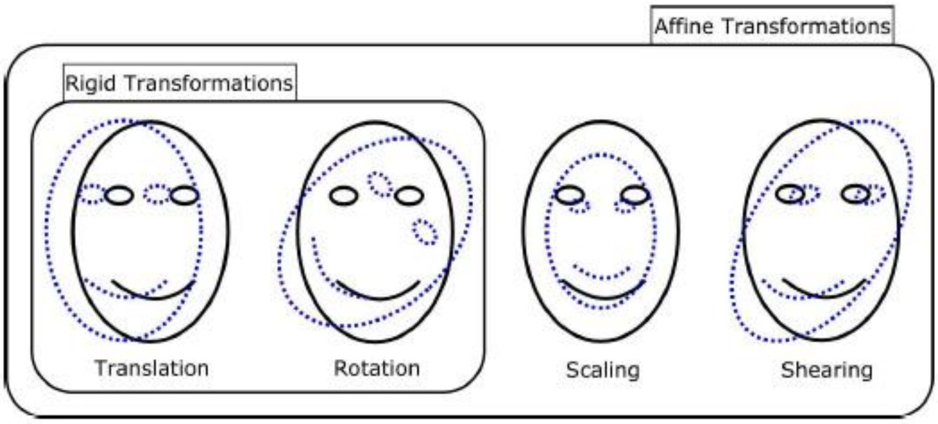


Figure 3.3: An example of linear transformation (Papatheodorou, 2006)

c. Non-Rigid Transformation

A non-rigid or free-form geometric transformation models the local shape deformation efficiently. Free form is applied when one surface must be freely deformed to fit another. In the face cases, free-form transformation is required where affine transformation is unable to map a neutral face on a smiling face. Figure 3.4 illustrates the use of the non-rigid transformation where neutral face is deformed by smiling face.

A simple form of non-rigid transformation is the spline-based transformation. In both datasets, the B-spline transformation model based on Lee et al. (1997) is adopted and extended for 3D surfaces by Papatheodorou (2006). Further details of this method could be found in his work. Practically, the object space is defined by control points or landmarks and the deformation is smoothly interpolated at intermediate points. The movement of the grid points causes regular displacement of each object point in the space. The interpolation condition is expressed as:

$$\mathrm{T}_{interpolation}(a_i) = b_i \qquad i = 1, \dots, n \qquad \text{Equation 3.6}$$

where a_i represent the location of control point in the target surface and b_i is the location of the corresponding control point in the source surface.

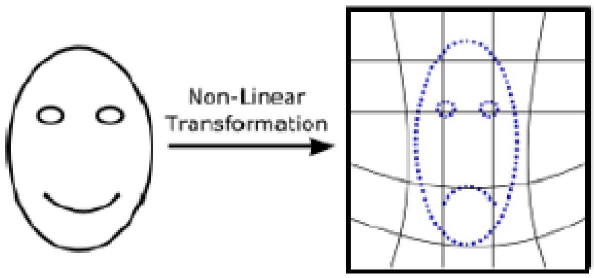


Figure 3.4: An example of non-rigid transformation (Papatheodorou, 2006)

3.3.3 Pre-processing Steps

The facial data scans are visualised in 3D using an open-source visualisation library (www.vtk.org). The selection of face landmark is done manually. A mouse input is used to select the points on the 3D surface. With the visualisation library, the face can be rotated and this made the selection of the correct landmarks easier. However, the cheeks are difficult to put landmark since there is no distinguishable anatomical points across all faces. The landmarks were placed on anatomically distinct points of the face to ensure proper correspondence (Papatheodorou, 2006). Some of the landmark placement include local feature information such as the size of mouth and nose; and the global feature information such as the location of the eyebrows. Papatheodorou selected 13 landmarks and stated that it is enough to build a model since previous works has shown that there are only small differences between utilizing 11 and 59 landmarks. Table 3.1 shows landmarks chosen while Figure 3.5 shows an example face with manual landmarked. The centre of mass of all surfaces is moved to the origin of the coordinate system and rigid registration is done afterwards. According to Papatheodorou (2006), the centring compensates for large differences in the distance between subjects as it increases the chances of correct pairing between the two face surface points.

Table 3.1: The 13 manual landmarks placement selected due to their anatomical distinctness

Anatomical Points Landmarked	
Points	**Landmark Description**
Glabella (1 point)	Area in the centre of the forehead between the eyebrows, above the nose which slightly protruding.
Eyes (4 points)	Both the inner and outer corners of the eyelids are landmarked.

Table 3.1 continued

Nasion (1 point)	The intersection of the frontal and two nasal bones of the human skull where there is a clearly depressed area directly between the eyes above the bridge of the nose.
Nose Tip (1 point)	The most protruding part of the nose.
Subnasal (1 point)	The middle point at the base of the nose.
Lips (4 points)	Both left and right corners of the lips, the top point of the upper lip and the lowest point of lower lip are landmarked.
Gnathion (1 point)	The lowest and most protruding point on the chin.

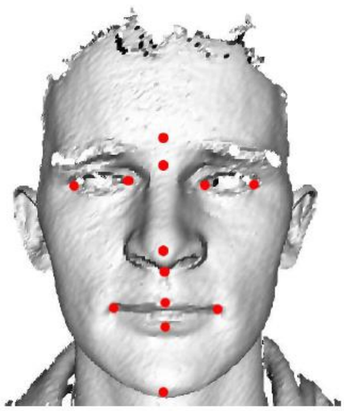


Figure 3.5: The 13 anatomical landmarks on face surface (Papatheodorou, 2006)

a. Rigid Landmark Registration

First, mean landmarks of all anatomical landmarks on face surfaces are calculated to find the mean position for each landmark. Then the rigid transformation on each landmark is computed with the mean to register all face surfaces to one another. As a result, a rough correspondence between all the face surfaces and mean surface is established. Figure 3.6 on top row shows two faces which are represented by the blue and red colour areas respectively, are aligned to the mean landmarks, meanwhile at bottom row shows a frontal 2D projection of outer landmarks of the same faces before and after rigid landmark registration.

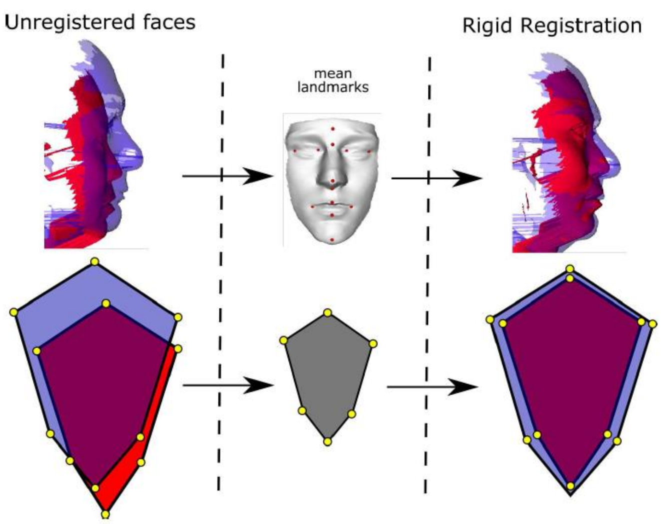


Figure 3.6: Rigid landmark registration (Papatheodorou, 2006)

b. Non-rigid Landmark Registration

The non-rigid registration is used to maximise the correspondence between the faces after rigid registration. In rigid landmark registration case, the point to point correspondences are roughly aligned resulting many surface points are not correctly paired. Thus, the face surfaces still vary in size and shape. Thus, non-rigid landmark registration is required to pair the correct corresponding points. Figure 3.7 shows the corresponding points before and after non-rigid landmark registration. The triangle points denote the manually placed landmarks that need to be aligned while the circle points represent points around the landmarks. Papatheodorou (2006) applied free-form deformation using a multi resolution B-spline approach to model the local non-rigid deformation. The deformation grid control points are gradually adjusted to reduce the distance between aligned landmarks. Figure 3.8 shows two face surfaces represented by the blue and red colour area respectively, aligning themselves to the mean landmark while the bottom row shows the outline polygon of outer landmarks of the same faces before and after non-rigid registration.

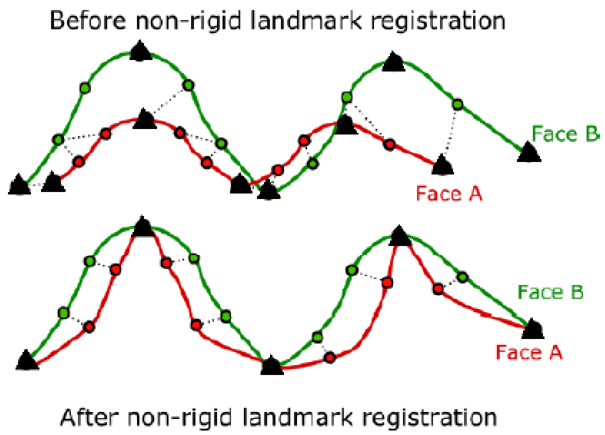


Figure 3.7: Before and after non-rigid landmark registration

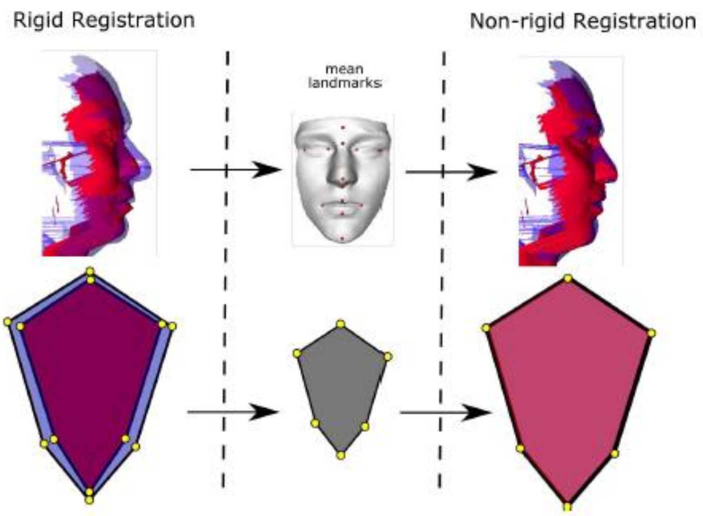


Figure 3.8: Non-rigid landmark registration (Papatheodorou, 2006)

c. Establishing Correspondences

The landmarks obtained through rigid and non-rigid registration in previous steps will be established using a face template. The template is free from artefacts and is not part of the population (Minoi, 2009). The surface point distribution is normalised manually and then wrapped using B-spline transformation. Given a face A, the result of carrying out rigid and non-rigid registration could be represented using the template face A'. Both A and A' are closely aligned around the landmark region as illustrated in Figure 3.8 after non-rigid registration which resulted to Figure 3.9. Figure 3.9 shows a face where each point is colour-coded to show distance to the closest point on the template face. Based on the map, the blue coloured coded signifies the larger distance while the red colour signifies the smallest distance between the two faces. After completing the registration, the template face mesh is used to resample the surface of A'. This subsequently regularised the number of 3D points in A' and connected the points around the edge to cover holes and removes spikes on the template mesh. Finally, a new cleaned face surface having the topology of the template mesh and geometry of A is created by projecting the template to the inverse of the registration transformation.

Figure 3.10 shows distance colour map of a face to compare non-rigid landmark registration and non-rigid surface registration. The distance between a face surface A and a referred surface template A' are displayed through colour code. From the result observed, the mean distance between the face and base mesh has reduced globally after non-rigid surface registration. Figure 3.10 (a) shows the closest points noted as red stripes between two surfaces are distributed around the landmarks as compared to face in Figure 3.10 (b) where the two surfaces are distributed globally.

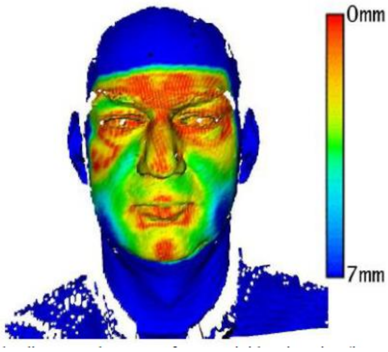


Figure 3.9: Distance colour map after non-rigid registration (Papatheodorou, 2006)

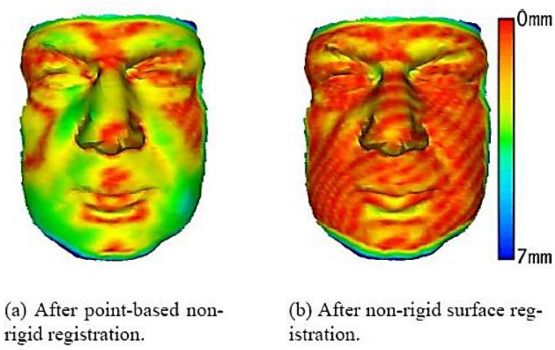


Figure 3.10: Distance colour map between two surface faces based on non-rigid landmark registration and non-rigid surface registration (Papatheodorou, 2006)

The same pre-processing steps are applied in BU3DFE. The only difference is Yin et al. (2006) selected 83 landmarks to establish the 3D correspondences while Papatheodorou (2006) use only 13 landmarks. Figure 3.11 shows the 83 landmarks selected in BU3DFE dataset. Yin et al. (2006) provide model orientation using a normal vector with respect to the frontal projection plane. Then, a triangular plane is formed by giving three vertices picked from two eye corners and a nose as the centre.

Figure 3.11: The 83 landmarks selected manually in BU3DFE (Yin et al., 2006)

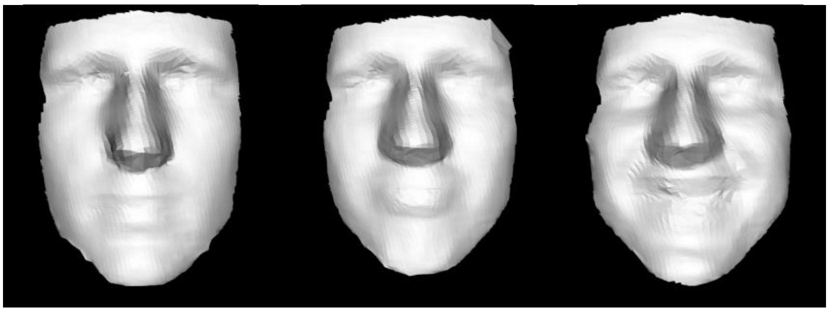


Figure 3.12: A subject from pre-processed VisionRT dataset showing samples of three expressions. From left: Neutral, Frowning and Smiling

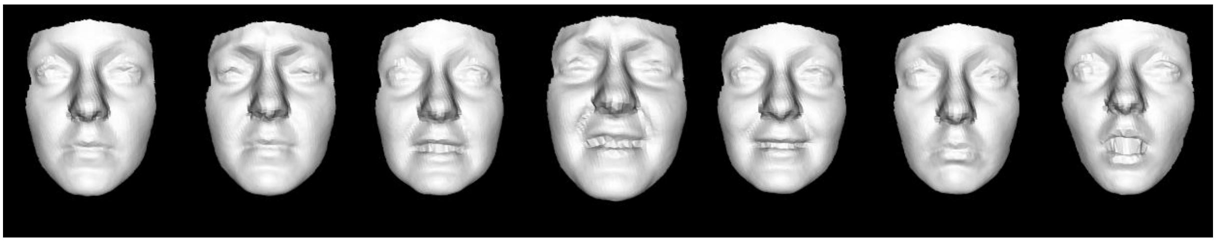


Figure 3.13: A sample of seven expressions from pre-processed BU3DFE dataset. From left: Neutral, Angry, Fear, Disgust, Happy, Sad and Surprised

Figure 3.12 and Figure 3.13 presents a sample of the pre-processed datasets from VisionRT and BU3DFE respectively. As shown in the figure, the surface depth quality is visibly poor for VisionRT compared to BU3DFE where facial features are well-defined especially around the eyes and teeth. This is due to the number of landmarks selected for the registration process, where VisionRT is based on 13 landmarks while BU3DFE is based on 83 landmarks. PCA is used to evaluate the pre-processing techniques by investigating the distribution of the 3D correspondence of each subject in the face space. As recorded in Papatheodorou (2006), the quantitative tests have shown positive results and through human visual inspection, the reconstructed cleaned faces were visually acceptable. The cleaned datasets could be used as ground truth to assist algorithms for 3D model segmentation and 3D feature detection (Yin et al., 2006).

3.4 Summary

The 3D face datasets used in this research are described along with the method applied to pre-process them. The raw face scans which acquired using scanners has several flaws such as the presence of holes and spikes. Besides that, the raw face scans are randomly aligned and there is no known point to point correspondence. Due to this shortcoming, there is a need for the raw face data to be processed before they could be used for shape analysis. In the next chapter, our proposed method to synthesised neutral facial expression will be introduced.

CHAPTER 4

PROPOSED METHOD

4.1 Introduction

This chapter discusses the proposed modified kernel Active Shape Model (mKASM) used in synthesising neutral expression on realistic 3D face dataset. The main goal of the research is to remove any form of facial expression from given expression and transforming it into a neutral expression while keeping the identity of each person using a kernel approach. The proposed method adopts the statistical shape model approach that falls under the example-based domain to build nonlinear shape model. In this approach, a neutral facial expression is synthesised on 3D faces by moving the surface points that represent the face along the appropriate neutral expressive direction in the training set space.

4.2 Modified Kernel ASM for Neutral Expression Synthesis

Inspired with the statistical shape model, specifically, the ASM by Cootes et al. (1995) which could be found in Chapter 2.5, Equation 2.5 and Expression Proportion Distribution (EPD) algorithm by Agianpuye (2015), a modified kernel Active Shape Model (mKASM) is proposed to synthesise new neutral expression. This research employed non-rigid template based on kernel ASM which is an extension of rigid template based on ASM. After the non-rigid template is generated, the template is modified by adding a mean weight that is adopted from Agianpuye (2015). Agianpuye (2015) proposed that the synthesis of new neutral facial expression from expressive faces is constructed by the addition of the mean of original neutral expression to the projected training face to increase the proportion of neutral expression. Hence, this research also adopted the EPD method by the substitution of the mean expression with the mean neutral from the expressional model. This will increase the neutral expression proportion. In addition to this, another parameter called Identity Features is proposed and to be added into the projected training face to increase the proportion of the subject's identity. Figure 4.1 illustrates how the proposed mKASM is formulated.

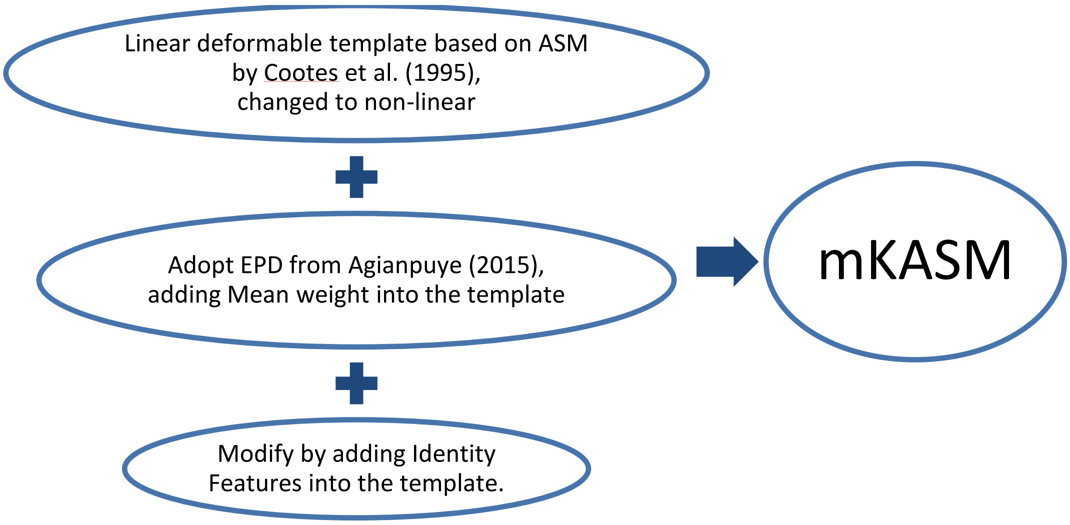


Figure 4.1: Formulation of Modified Kernel Active Shape Model

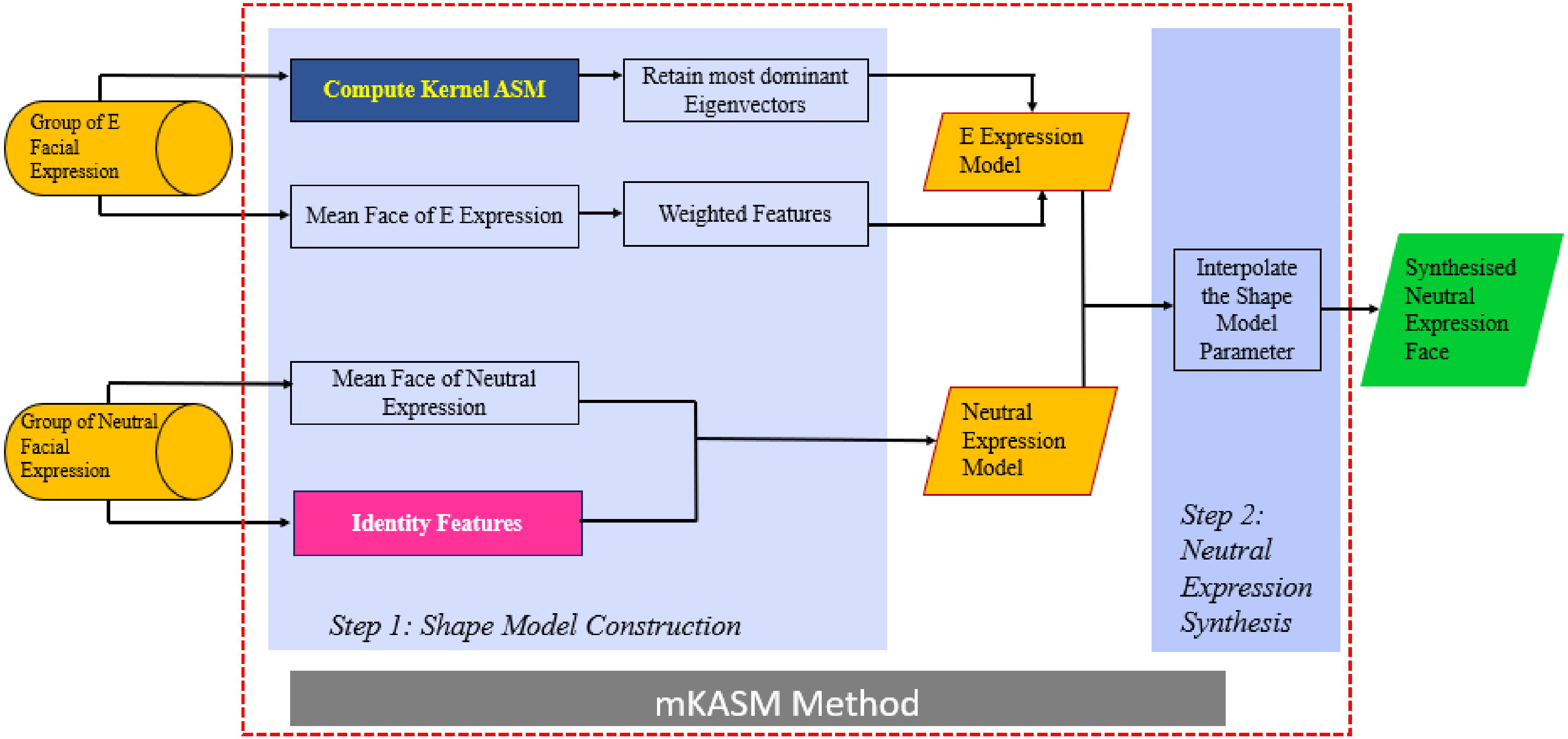


Figure 4.2: Proposed mKASM algorithm for synthesising neutral expression in 3D faces

Figure 4.2 illustrates the framework of the proposed algorithm, mKASM from preparing training sets until the new synthesised neutral face is generated. There are two important steps which could be elaborated as the following:

a. Step 1: Shape Model Construction

In this step, the training sets are prepared from selected datasets, the VisionRT and BU3DFE face datasets. Each group represents one facial expression denoted as **S**. Group E could represents any facial expressions except neutral expression from the dataset. Each training set labelled as Smile, Frown, Angry, Disgust, Fear, Happy, Sad and Surprised. A group of E expression is represented as a vector, $\mathrm{S}_E = [x_1, \ldots, x_n]$, where each x represents a 3D face data and ***n*** is the number of faces in each training set.

The mean face, μ_E is calculated and subtracted with the vector S_E, to compute the Weighted Features, defined as b_E of the test faces, that holds the test's expression features:

$$\mu_E = \frac{1}{n}\sum_{i=1}^{n} S_i \qquad \text{Equation 4.1}$$

$$\mathrm{b}_E = (\mathrm{S}_E - \mu_E) \qquad \text{Equation 4.2}$$

Then kernel ASM is executed for each set of group E to construct the basis of shape model called E Expression Model. Further details of this method will be explained in the next Section 4.3. The E Expression Model consists of eigenvalues and eigenvectors that are sorted in order of the highest eigenvalues. As stated by Yambor et al. (2002) the highest eigenvalues carry the most informative characteristics of the trained set. Since eigenvectors represent the variation in the training set, the lower eigenvectors show less variation and more noise. Therefore, only the first eigenvector is selected here because it represents the most dominant features of the faces in the training set.

Based on the Weighted Features, b_E and eigenvectors, w^{Φ} from kernel ASM the test faces, X is represented by E Expression Model,

$$X \approx \mu_E + w^{\Phi} b \qquad \text{Equation 4.3}$$

$$b_E = (S_E - \mu_E) \qquad \text{Equation 4.4}$$

A group of neutral facial expression from the selected dataset, either VisionRT or BU3DFE, is grouped into another training set denoted as N.mean, $\mu_{\boldsymbol{N}}$ is computed from the group N to get the mean face of the original neutral expression as shown in Equation 4.5. The n represents the number of faces in N set while vector $S = [x_1, \ldots, x_n]$ represents the points of each face in N.

$$\mu_{\boldsymbol{N}} = \frac{1}{n} \sum_{i=1}^{n} S_i \qquad \text{Equation 4.5}$$

The Identity Features is computed by subtracting the group **N** with the mean, $\mu_{\boldsymbol{N}}$ to increase the proportion distribution of neutral expression and to retain the subject's identity.

$$I_{\boldsymbol{N}} = (S_{\boldsymbol{N}} - \mu_{\boldsymbol{N}}) \qquad \text{Equation 4.6}$$

b. Step 2: Neutral Expression Synthesis

Mathematically, the proposed mKASM method is expressed in Equation 4.7, where Z depicts the synthesised neutral face, hereby called as Expression Shape Model. In order to synthesise neutral expression, both parameters from E Expression Model and Neutral Model computed earlier are fed into the Expression Shape Model, as shown in Equation 4.7. The synthesised neutral expression denoted as Z is generated when a test face S with E expression, is passed through the mKASM algorithm. The I_N parameter carries the identity traits of a test face S. Meanwhile, μ_N represents a neutral mean face that carries the original neutral expression features. The weight of the test face from Equation 4.2 is represented by b_E and the first eigenvector is represented by w^{Φ}. After synthesis, the synthesised neutral faces will be stored into a new group to be evaluated.

$$Z \approx I_N + \mu_N + w^{\Phi}\sqrt{b_E} \qquad \text{Equation 4.7}$$

As previously pointed in Chapter 2, facial expression has a different level of intensities. Considering some of the face portrays an extreme expression such as the opening of mouth and widening of cheeks from surprise and fear expressions, the mKASM's ability to synthesise a perfect neutral face might be affected by this. The quality of the newly synthesised neutral expression will be investigated and explained in Chapter 5.

4.3 Kernel Active Shape Model

Corresponding to linear ASM introduced in Chapter 2, a new facial shape is generated by manipulation of the shape model in Equation 4.3. The only difference between ASM and kernel ASM is the linear and nonlinear principal components that are used to construct the shape model. This section explains how nonlinear components, hereby called eigenvectors, are computed using Kernel-based Principal Component Analysis (Kernel PCA). Kernel PCA is an extension of PCA.

4.3.1 Kernel PCA

Kernel PCA is a method for performing a nonlinear form of PCA pioneered by Schölkopf et al. (1997). The comparison of Kernel PCA and PCA algorithm is included in Appendix A and Appendix B. Kernel PCA projects the input space to a feature space through nonlinear mapping and then computes PCA in the feature space to generate nonlinear components. This nonlinear mapping is realised by a kernel function. The kernel functions which have been commonly used by researchers are Linear, Polynomial and Gaussian kernels. Each choice of kernel function may solve different nonlinear problems.

This algorithm of kernel PCA is adopted from Schölkopf et al. (1998). First, for simplicity, assume a set of n centred or with zero mean samples. A sample vector, x is fit into the input space, R^n.

$$x_n = [x_1, \dots, x_n]^T \in R^n, \qquad \text{Equation 4.8}$$

The purpose of PCA is to find the directions of projection that get the most out of the variance C, which is corresponding to finding the eigenvalues, λ and eigenvectors, w from the covariance matrix which is defined as C:

$$C = \frac{1}{n} \sum_{i=1}^{n} x_i \, x_i^T \qquad \text{Equation 4.9}$$

Which can be diagonalized with non-negative eigenvalues satisfying Equation 4.10, where, $\lambda \geq 0$ and, $w \in R^n$.

$$Cw = \lambda w \qquad \text{Equation 4.10}$$

In kernel PCA, to transform data for extracting nonlinear features, each vector x is projected from the input space, R^n to a higher dimensional feature space, R^f by a nonlinear mapping function, Φ, where the dimensional size of feature space denoted as f is larger than the input space, n.

$$\Phi: R^n \rightarrow R^f, f \gg n \qquad \text{Equation 4.11}$$

It should be pointed out that the dimensionality of the feature space could be huge. In R^f, the corresponding eigenvalue problem is,

$$C^{\Phi} w^{\Phi} = \lambda w^{\Phi} \qquad \text{Equation 4.12}$$

where, $C^{\Phi}$is a covariance matrix:

$$C^{\Phi} = \frac{1}{n}\sum_{\mathrm{i}=1}^{n} \Phi(x_i)\, \Phi(x_i)^T \qquad \text{Equation 4.13}$$

All solution of nonlinear w^{Φ} with $\lambda \neq 0$ lies in the span of $\Phi(x_1), \dots, \Phi(x_n)$, there exist coefficients a_i such that,

$$w^{\Phi} = \sum_{i=1}^{n} a_i \Phi(x_i) \qquad \text{Equation 4.14}$$

The Equation 4.13 does not need to be computed explicitly, a kernel function is used instead. To express the relationship in terms of inner-product kernel, a $n \times n$ matrix K is defined by,

$$K_{ij} = K(\Phi_i, \Phi_j) = \Phi(x_i).\Phi(x_j) \qquad \text{Equation 4.15}$$

which can be diagonalized with non-negative eigenvalues satisfying Equation 4.16. The eigenvalue problem becomes,

$$n\lambda K a = K^2 a \qquad \text{Equation 4.16}$$

$$n\lambda a = K a \qquad \text{Equation 4.17}$$

where a denotes a column vector with entries $a_1, \dots, a_n$. The eigenvalue problem shown in Equation 4.17 is solved to find solutions to Equation 4.16. The above derivation assumes that all the projected samples, $\Phi(\mathrm{x})$ are centred in R^f . Centring data will be further discussed in the next section.

Final step is for the purpose of principal component extraction, after Equation 4.14, the vectors in R^f can now be projected to a lower-dimensional space spanned by the eigenvectors w^{Φ}. Let x be a test sample whose projection onto w^{Φ} is $\Phi(\mathrm{x})$, the corresponding nonlinear principal components is expressed as

$$\Phi: w^{\Phi}.\Phi(\mathrm{x}) = \sum_{i=1}^{n} a_i\big(\Phi(x_i)\,.\,\Phi(x_i)\big) = \sum_{i=1}^{m} a_i\,(x_i) \qquad \text{Equation 4.18}$$

where the first q $(1 \leq q \leq n)$ nonlinear principal components or the eigenvectors w^{Φ} are extracted using the kernel function without expensive operation that explicitly projects samples to high dimensional space R^f. In this research, Gaussian Kernel function is used by substituting it into Equation 4.15. The Equation 4.19 expressed the Gaussian Kernel, where, X and Y are referring the inner product from the K_{ij} matrix of Equation 4.15, whereas parameter σ controls the width of the Gaussian kernel.

$$K(x, y) = \exp\left(-\frac{\|X-Y\|^2}{2\sigma^2}\right) \qquad \text{Equation 4.19}$$

To summarise all, there are four necessary steps to perform kernel PCA. First, compute kernel matrix Equation 4.15 using Equation 4.19, then centre the kernel using K Gram matrix in Equation 4.22, solve the eigenvalue problem in Equation 4.17 and finally compute projections using eigenvector corresponding to the map as presented in Equation 4.18. The choice of number for kernel width parameter σ in Equation 4.19 is usually depends on the user. In this study, using $\sigma = 100$ gives the highest recognition rate. Hence, 100 is chosen.

4.3.2 Centring Data in Kernel

This section explains the algorithm to centralise the vectors $\Phi(\mathrm{x})$ in R^f space, which can also be found in Schölkopf et al. (1998). In the beginning of Equation 4.8, assumed the mapped data is centred in R^f,

$$\sum_{i=1}^{n} \Phi(x_i) = 0 \qquad \text{Equation 4.20}$$

Now, to centre the mapped data, note that given a map denoted as Φ and any set of observations $x_1, \dots, x_n$, the points will be centred,

$$\ddot{\phi}(x_i) := \Phi(x_i) - \frac{1}{n}\sum_{i=1}^{n} \Phi(x_i) = 0 \qquad \text{Equation 4.21}$$

Thus, after the assumption in Equation 4.7, then the covariance matrix Equation 4.13 is defined and calculate the dot product matrix Equation 4.15 in R^f. As the centred data not available, the matrix K cannot be computed directly in Equation 4.17. Instead, to substitute the kernel matrix K, the equation is expressed in terms of its non-centred counterpart which is called as the K Gram matrix as shown in Equation 4.22,

$$\widetilde{K}_{ij} = K - 1_n K - K 1_n + 1_n K 1_n \qquad \text{Equation 4.22}$$

where 1_n is the $n \times n$ matrix with all elements equal to $\frac{1}{n}$ (Bishop, 2006). Now, $\widetilde{K}$ could be computed from K and then solve the eigen decomposition problem in Equation 4.14.

4.4 Implementation of Proposed mKASM Method

This section presents the implementation of the proposed mKASM. The performance of the proposed mKASM method is evaluated by using two different datasets. As mentioned in Chapter 3, the face datasets used in this experiment have already been pre-processed, hence this step is omitted. The implementation for synthesising neutral expression and face recognition for validation of the newly synthesised neutral face is carried out using MATLAB. MATLAB is chosen as a platform because it provides simple debugging and clear understanding of each vector and matrix computations. Meanwhile, the synthesised face is rendered in 3D environment for visual inspection using a customised application in C++ programming language via the Visual Studio development tools. The pre-processed face datasets were originally in VTK file formats. Therefore, the open source VTK library (Schroeder et al., 2006) is imported which supports computer graphics and image processing to visualise the 3D face.

The implementation is divided into four modules, three of the modules that were developed in MATLAB are Kernel ASM module, Synthesis module and Face Recognition Module. The remaining module, Show 3D Face, is developed in C++ environment for 3D visualisation of the face. Figure 4.3 shows the proposed framework of the whole implementation of neutral facial expressions synthesis based on mKASM method. As shown in the figure, each module is labelled accordingly where the orange colour box indicates the proposed mKASM method.

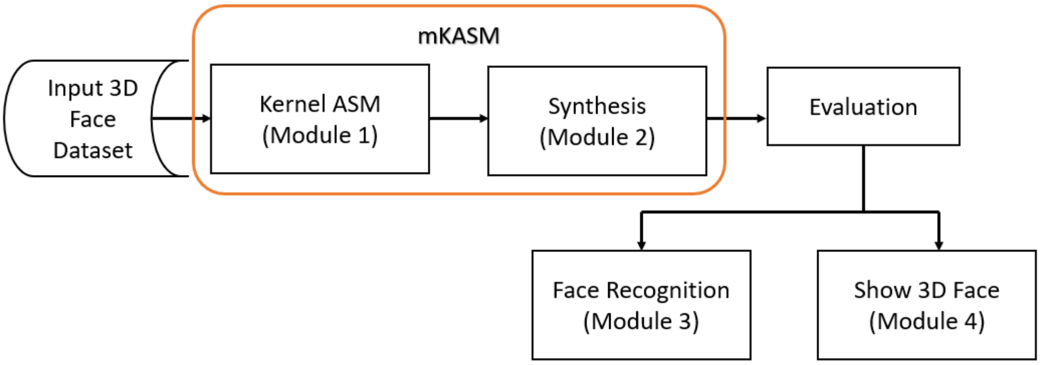


Figure 4.3: The overview of the implementation

Figure 4.4 illustrates the first module, Kernel ASM. This module generates a group of E Expression Model and Neutral Expression Model. The group E could be substituted with any of the group expression except neutral expression. The mean face is calculated to get the average face. Then, the group E face shape data aligned their position to one another by centring the position using translation and remove the rotation pose. The shape model is

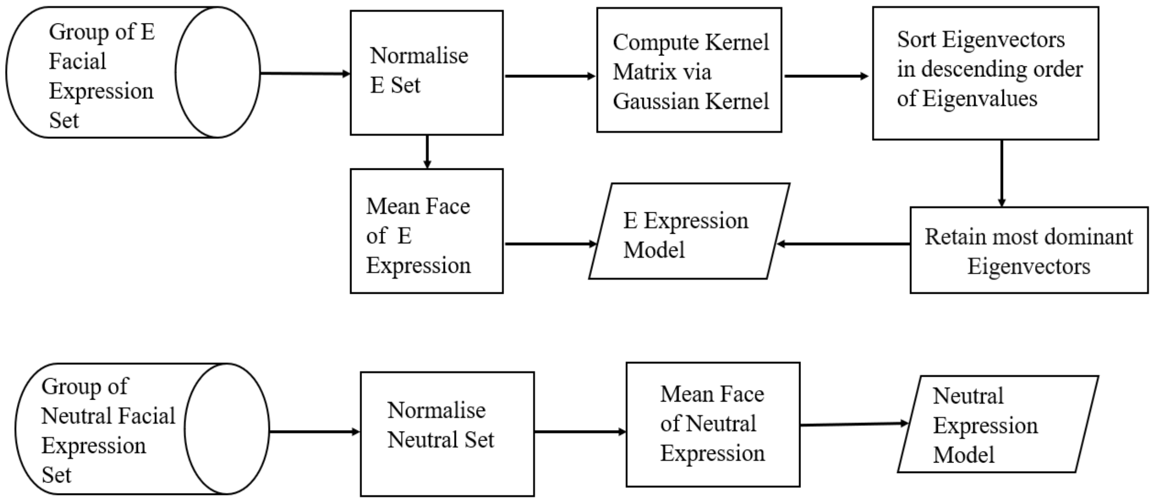


Figure 4.4: Kernel ASM module

built by computing kernel PCA on the aligned group E face data. The results are saved as E Expression Model. As shown in the Figure 4.4, the neutral expression group sample is also aligned, and the mean face is calculated to generate neutral expression model.

Figure 4.5 presents the process of mKASM method in Synthesis module. The E and Neutral expression model parameters are extracted and computed the Expression Shape Model as shown in Equation 4.7 previously. The synthesised neutral expression faces are stored into a new dataset. The newly synthesised neutral expression face will be evaluated through face recognition and visual inspection using rendering scheme.

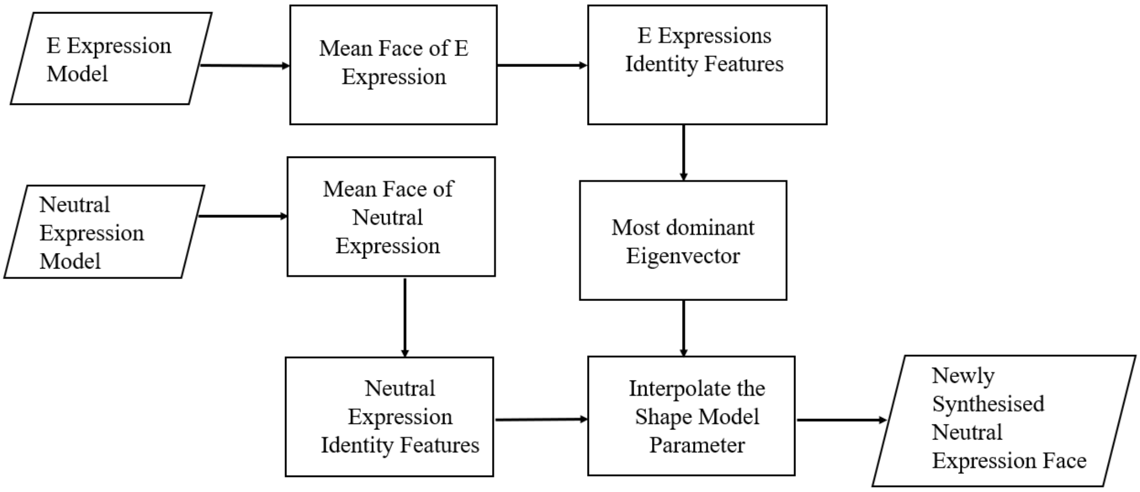


Figure 4.5: Synthesis module

How much of the newly synthesised neutral face is able to retain the respective subject's identity, are determined through eigenvectors. This is because the most dominant eigenvectors correspond to the highest eigenvalues also hold the identity information of the subjects in the training set. Here, the kernel-based PCA face recognition is employed to carry out this investigation. The face recognition is also served as a quantitative evaluation of the proposed mKASM's effectiveness. The original neutral expression is used as training set, while the synthesised neutral expressions is set as the test samples. The Face Recognition module as shown in Figure 4.6 is implemented to test the synthesised neutral face with the trained original neutral expression to determine the newly reconstructed neutral expression is correctly recognised. This experiment will be explained in detail in Section 4.5. The rate of recognition achieved will tell how much the proposed neutral expression synthesis method retains the identity of the test subject after the synthesis process.

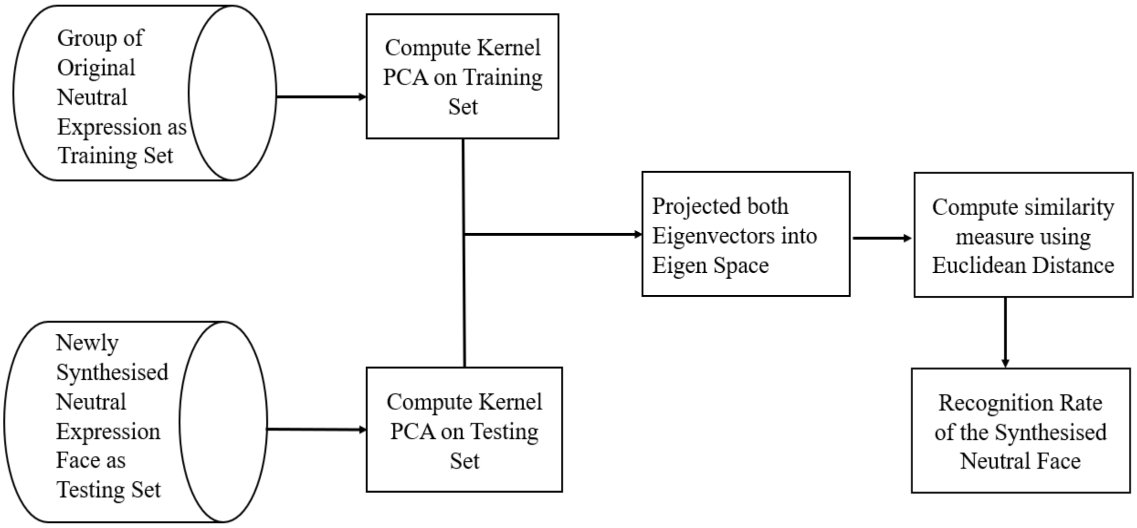


Figure 4.6: Face Recognition module

Next, the effectiveness of the mKASM method also qualitatively evaluated by visually observing the similarities in the facial attributes between the original test face, the newly synthesised neutral face and the ground truth of the neutral face. Figure 4.7 illustrates the process of rendering the 3D face by using the Show 3D Face module. First, the synthesised 3D face points are converted into polygonal data points in VTK file format. Then, the system will read the input VTK file and defined as an actor. The example of rendered 3D face output is shown at Figure 4.7.

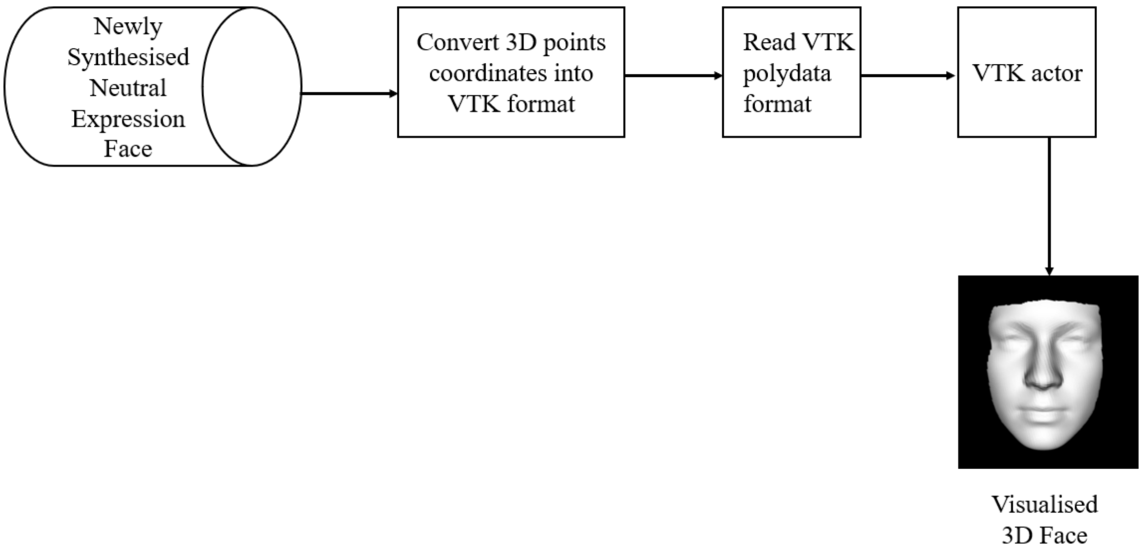


Figure 4.7: Show 3D Face module

4.5 The Kernel PCA-based Face Recognition

This section describes the Face Recognition module in detail. The newly synthesised neutral face could be validated through a simple face recognition system based on kernel PCA. The recognition rate is served as quantitative evaluation of the effectiveness of the mKASM method in producing the synthesised neutral face. The analysis of the result will also give an insight on how much the identity is retained after the synthetisation is held.

The facial recognition system is developed to identify a face by comparing and matching the face over a dataset of known faces. The rate of recognition of the synthesised neutral expression will be compared with the original expressional face. Prior to the synthesis, the first experiment is testing the original expressional face with the original neutral expression. Then, the synthesised neutral expression is tested with the original neutral expression. The recognition rate achieved from the former test will be compared with the later test.

The newly synthesised neutral faces are saved into a new 3D dataset and will be used as testing sample. Meanwhile original neutral expression will be used as training sample. The nonlinear principal components of both training and testing sample are extracted respectively using kernel PCA method, which could be referred at Section 4.3.1. The training samples are projected into the kernel PCA subspace. Recognition is performed by projecting the testing sample into the kernel PCA subspace and defined as query face. To classify the query face, the similarity is measured through Euclidean distance and classical classifier, k Nearest Neighbour (KNN). The minimum distance recorded will determine the subject class of the query face. The value of k for KNN in this experiment is 1 which means the query face is assigned to the subject class of the single nearest neighbour. Below shows the definition of Euclidean Distance, where φ_A is defined as the projected query face, while φ_B is defined as training face and M is the total number of samples.

$$Euclidean\ distance(\varphi_A, \varphi_B) = \|\varphi_A - \varphi_B\| = \sqrt{\sum_i^M (\varphi_A - \varphi_B)^2} \qquad \text{Equation 4.23}$$

The similarity between the projected query face *A* and training face *B* is measured by the Euclidean distance measure as shown in Equation 4.23. The predicted face class is assigned with the minimum distance which is also the nearest neighbour. The minimum distance indicates the highest possibility of similar faces between the projected training face and query face. The recognition rate is then calculated by the total number of correctly recognised predicted face. Figure 4.8 presents the flowchart on how the kernel PCA face recognition is carried out.

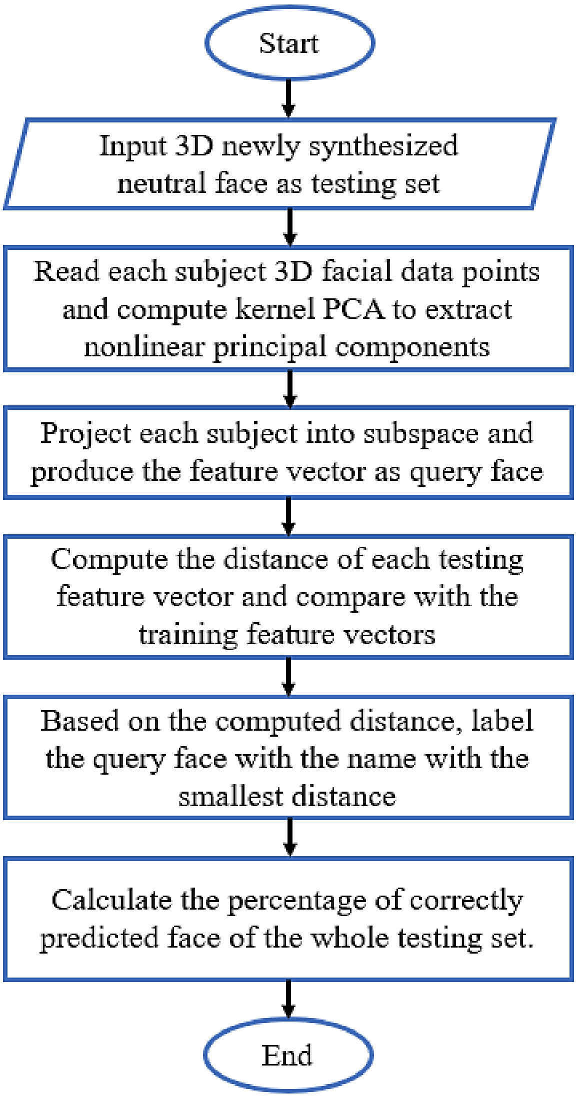


Figure 4.8: Kernel-based PCA face recognition flowchart

4.6 Summary

In this chapter, the proposed mKASM is introduced and presented with a diagram to explain how neutral facial expression is synthesised from a group of expressional faces. The details on the works of kernel ASM is also explained. Other than that, this chapter explains the implementation of proposed mKASM method. The four modules involved in this implementation, Kernel ASM, Synthesis, Face Recognition and Show 3D face, are also presented. This research is more interested on the nonlinear perspective of the synthesis of 3D facial expression. In the next chapter, the mKASM will be evaluated and discussed in detail.

CHAPTER 5

EXPERIMENT RESULTS AND ANALYSIS

5.1 Introduction

This chapter present and analyse the results of implementing the mKASM on VisionRT and BU3DFE datasets. The experiments are categorised into three main tests. In Test 1, which is a preliminary study, the goals are to look for comparison between kernel PCA and PCA-based face recognition and to test the quality of existing datasets. Test 2 is the synthesis experiments, where the proposed mKASM is implemented. The screen capture results of synthesised 3D faces will be shown later. The results are qualitatively analysed by visual inspection. Test 3 is performed to evaluate the synthesised neutral face from Test 2. The synthesised neutral faces are stored and used as testing sample before fed into the face recognition system. In this experimentation, Test 1 is compared with Test 3 as before and after synthesis. The recognition results are presented in graphs to be analysed quantitatively.

Figure 5.1 illustrates the generic framework of face recognition to show the whole process of this research. The framework has a feature extractor and a matcher. The green highlighted box denoted pre-processing which is performed before feature extraction, though in this research, the face datasets have been pre-processed, hence the input data would be the pre-processed 3D face data. After input data, facial features are extracted by the feature extractor, classify the faces using simple classifier, the k-Nearest Neighbour and finally passed through the matcher as query for comparison against the reference or trained face from the controlled database. The recognition rates are calculated at the end of each experiment. This research is focused more on the feature extraction part as pointed in the red box, where the mKASM is done.

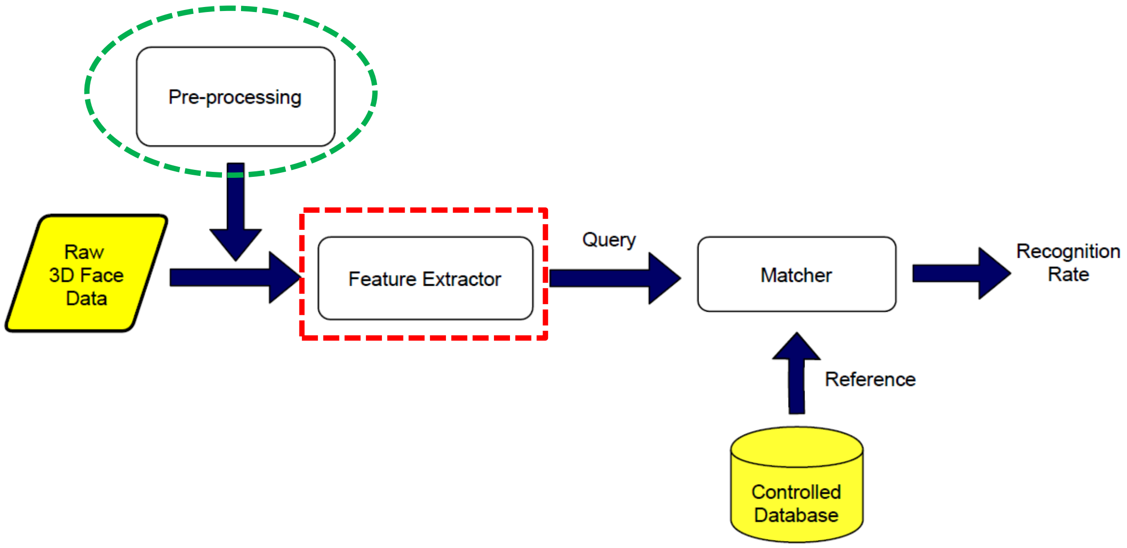


Figure 5.1: Generic face recogniton framework

5.2 Test 1: Kernel PCA versus PCA Face Recognition

This test was conducted on both datasets through kernel-based face recognition to evaluate the quality of face datasets. Additionally, to investigate the kernel potential by using kernel PCA as nonlinear feature extraction method directly on 3D expressional faces which are fed into the face recognition system. The face recognition is developed based on kernel PCA and comparison of performance is made with the PCA face recognition. Face recognition algorithm conducted in this test is similar to the algorithm explained in Section 4.5. However, the only difference is the use of kernel PCA or linear PCA method to extract the nonlinear or linear principal components. The face recognition system is developed under MATLAB platform.

The experiment is arranged based on the concept of bootstrapping. Table 5.1 shows the experiments arrangements. The test set is referring to the input test data while the train set is the controlled database. For an example, first experiment set using VisionRT dataset, the input test face would be Neutral 1, extract facial features using PCA method, and perform template matching with trained Neutral 1. After all test faces are tested, the recognition rate is calculated.

Table 5.1: Experiment sets for Test 1

Datasets	Experiment No.	Test	Train	Method
VisionRT	1	Neutral 1	Neutral 1	PCA
VisionRT	2	Neutral 1	Neutral 1	Kernel PCA
BU3DFE	3	Neutral	Neutral	PCA
BU3DFE	4	Neutral	Neutral	Kernel PCA
VisionRT	5	Frown	Neutral 1	PCA
	6	Frown	Neutral 2	
	7	Smile	Neutral 1	
	8	Smile	Neutral 2	
	9	Frown	Neutral 1	Kernel PCA
	10	Frown	Neutral 2	
	11	Smile	Neutral 1	
	12	Smile	Neutral 2	
BU3DFE	13	Angry 01	Neutral	PCA
	14	Angry 02	Neutral	
	15	Angry 03	Neutral	
	16	Angry 04	Neutral	
	17	Angry 01	Neutral	Kernel PCA
	18	Angry 02	Neutral	
	19	Angry 03	Neutral	
	20	Angry 04	Neutral	
	21	Disgust 01	Neutral	PCA
	22	Disgust 02	Neutral	
	23	Disgust 03	Neutral	
	24	Disgust 04	Neutral	
	25	Disgust 01	Neutral	Kernel PCA
	26	Disgust 02	Neutral	
	27	Disgust 03	Neutral	
	28	Disgust 04	Neutral	
	29	Fear 01	Neutral	PCA
	30	Fear 02	Neutral	
	31	Fear 03	Neutral	
	32	Fear 04	Neutral	
	33	Fear 01	Neutral	Kernel PCA
	34	Fear 02	Neutral	
	35	Fear 03	Neutral	
	36	Fear 04	Neutral	

Table 5.1 continued

BU3DFE	37	Happy 01	Neutral	PCA
	38	Happy 02	Neutral	
	39	Happy 03	Neutral	
	40	Happy 04	Neutral	
	41	Happy 01	Neutral	Kernel PCA
	42	Happy 02	Neutral	
	43	Happy 03	Neutral	
	44	Happy 04	Neutral	
	45	Sad 01	Neutral	PCA
	46	Sad 02	Neutral	
	47	Sad 03	Neutral	
	48	Sad 04	Neutral	
	49	Sad 01	Neutral	Kernel PCA
	50	Sad 02	Neutral	
	51	Sad 03	Neutral	
	52	Sad 04	Neutral	
	53	Surprise 01	Neutral	PCA
	54	Surprise 02	Neutral	
	55	Surprise 03	Neutral	
	56	Surprise 04	Neutral	
	57	Surprise 01	Neutral	Kernel PCA
	58	Surprise 02	Neutral	
	59	Surprise 03	Neutral	
	60	Surprise 04	Neutral	

5.2.1 Test 1: Experiment 1 and 4

This experiment is done to ensure the face recognition system runs properly. Therefore, the same Neutral expression faces is used for training and testing samples. The principal components used are usually based on the number of samples, where VisionRT dataset has 60 subjects and BU3DFE dataset has 54 subjects. In Figure 5.2 for VisionRT dataset, the results starting to record a consistent 100.00% recognition rate from 10 principal components onwards until 60 principal components. Meanwhile for BU3DFE dataset presented in Figure 5.3, the recognition achieved 100% starting from 10 principal components onwards until 54 principal components. The results have shown that the face recognition is working well and can be used for next tests using the face recognition system to calculate the recognition rate of other facial expressions. As mentioned in Wang and Tanaka (2019), the number of nonlinear principal component chosen is equal to the number of samples. Based on the results, both datasets achieved 100% when the principal component use equals to the total number of samples for each expression, which in case of VisionRT, the total is 60, and BU3DFE, the total is 54. Thus, 60 and 54 are used for the following experiment.

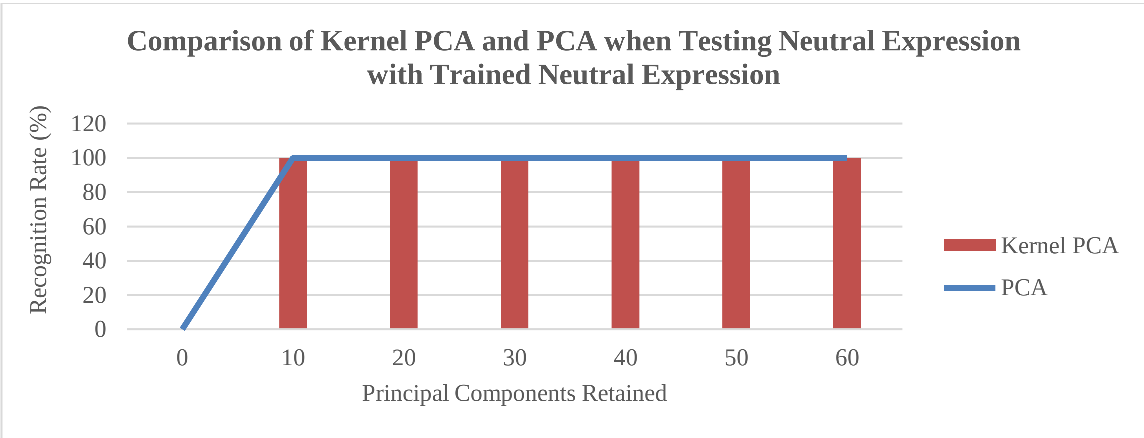


Figure 5.2: Testing Neutral 1 to Neutral 1 as training Sample in VisionRT dataset

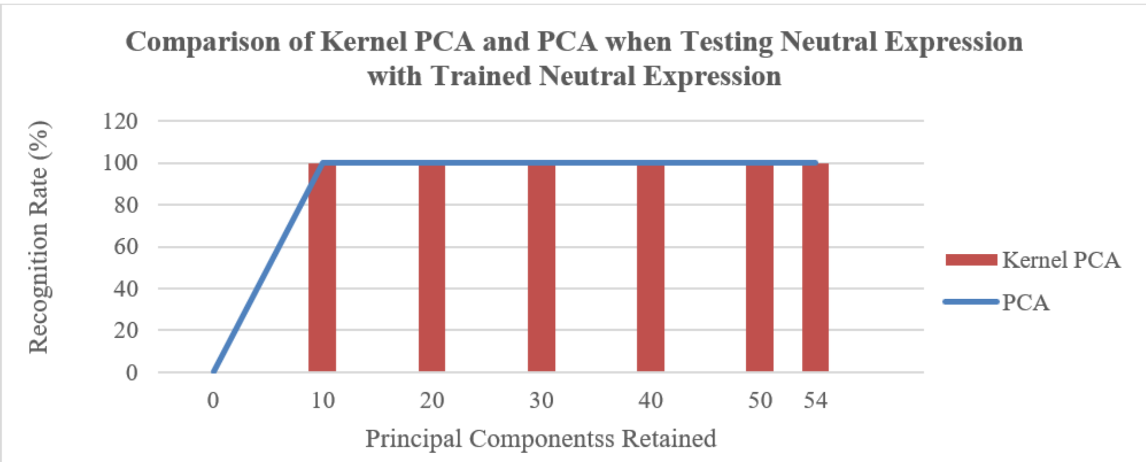


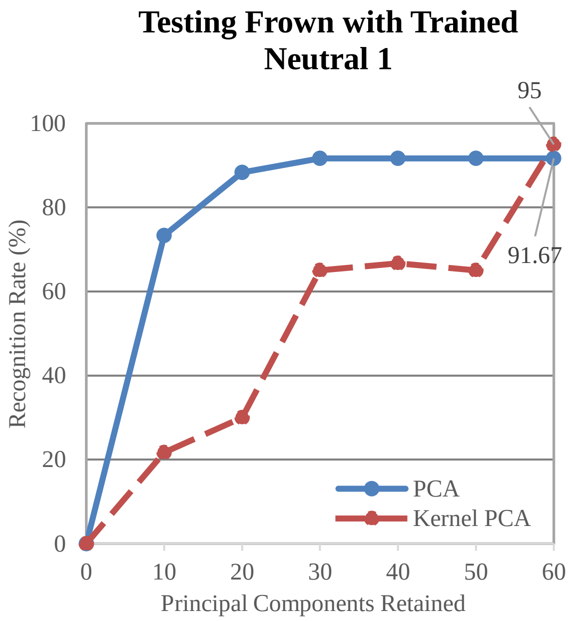
Figure 5.3: Testing Neutral to Neutral as training Sample in BU3DFE dataset

5.2.2 Test 1: Experiment 5 to 12 using VisionRT dataset

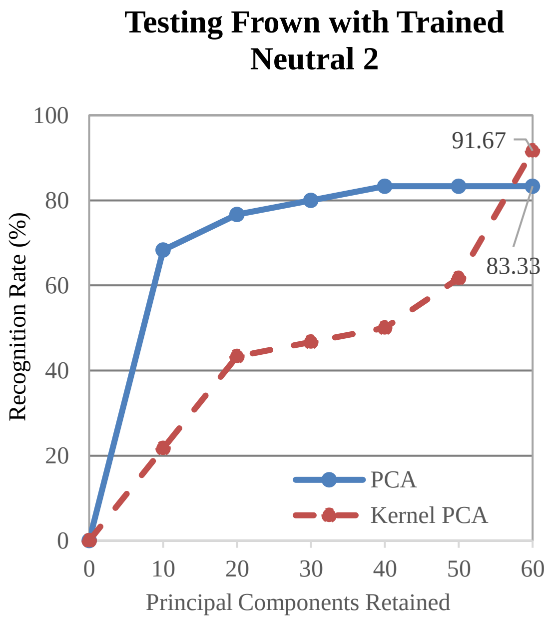
There are four groups of expression used for this experiment, Frown, Smile and two types of neutral expression labelled as Neutral 1 and Neutral 2. In this test, the number of principal components used for both kernel PCA and PCA methods is according to the total number of samples which is 60 subjects. This is because when using the total number of subjects, the maximum recognition rate is achieved for both methods and also the number is chosen to standardise both methods for comparison of the result later.

As shown in Figure 5.4 are the results of four testing expression samples on neutral expression. As observed, the recognition rate gradually increased as the number of principal components retained increased especially using the kernel PCA method. The result of testing Frown with Neutral 1 shown in Figure 5.4(a) achieved 95.00% with kernel PCA, while 91.67% with PCA alone. While the test between Frown and Neutral 2 achieved 91.67% with kernel PCA and 83.33% with PCA. The second-row of the graph presents smile expression as testing sample. From the result, there is an extreme difference in the recognition rate between kernel PCA and PCA. This shows that nonlinear components are better in projecting feature faces in face space that leads to better representation of a subject. In addition, the facial surface structure of a smile also involved more muscles actions than frown. This justifies the approximate 10.00% differences between the highest percentage of recognised frown and smile tests sample.

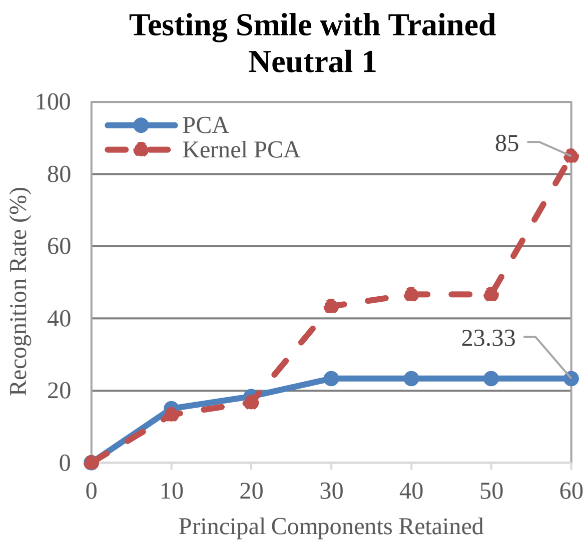
In comparison with two levels of neutral sample used as training set, the Figure 5.4 (a) shows Neutral 1 achieved higher recognition than Neutral 2 in Figure 5.4 (b). This might be due to the pre-processing step which affects the quality of the face data. Other than that, the Neutral 2 might have more shape variation compared to Neutral 1 which contributes the lower rate of recognition. Therefore, in this study, only the Neutral 1 will be used as training sample to test with the synthesised neutral facial expression.



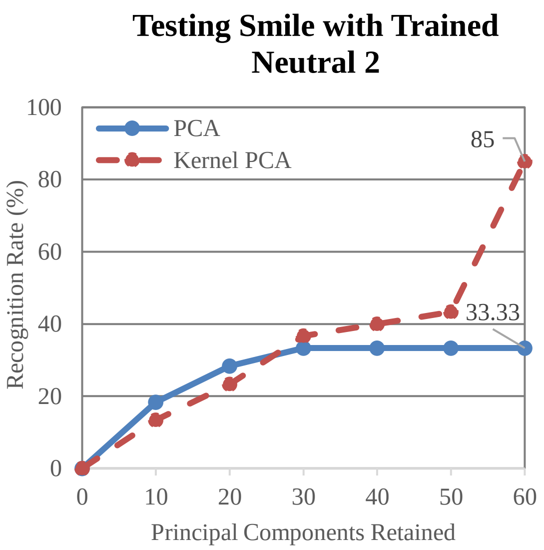
(a) Recognition rate of Neutral 1 as training sample with Frown



(b) Recognition rate of Neutral 2 as training sample with Frown



(c) Recognition rate of Neutral 1 as training sample with Smile



(d) Recognition rate of Neutral 2 as training sample with Smile

Figure 5.4: Face recognition rate of VisionRT dataset

5.2.3 Test 1: Experiment 13 to 60 using BU3DFE dataset

There are six groups of expression in this dataset, angry, fear, disgust, happy, sad and surprise which are organised as testing sample, while neutral expression as training sample. The number of principal components used for both kernel PCA and PCA methods for this dataset is according to the total number of samples, 54 subjects. This is because when using the total number of subjects, both methods achieved the maximum recognition rate and the number is also chosen to standardise both methods for comparison of the result.

Figure 5.5 shows the recognition result with 54 principal components of four intensity levels of angry expression tested with trained neutral expression based on kernel PCA and PCA approaches. The kernel PCA has shown greater percentage of recognition compared to PCA alone over the four expression intensity. The bar graph also shows that level 1 angry achieved the highest recognition compared to level 4 angry. This is due to the muscle extreme of angry expression increases over levels.

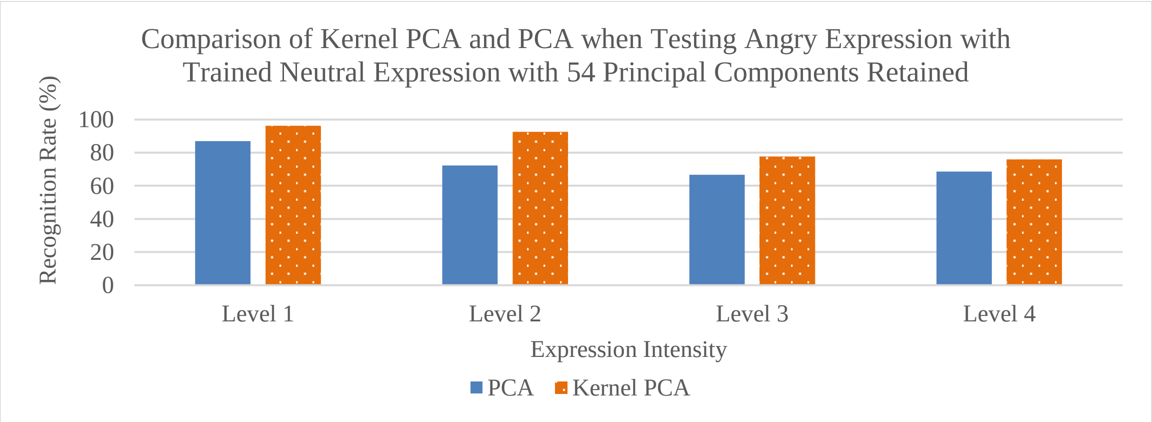


Figure 5.5: Recognition rate of testing Angry expression to Neutral expression as training sample

Meanwhile Figure 5.6 shows the comparison of kernel PCA and PCA approaches when disgust expression is tested on face recognition based on 54 principal components. The results shown in the Figure 4.12 indicate that kernel PCA works better recognition than PCA regardless the four levels of expression intensities. The recognition rate is almost competitive at level 1 disgust where PCA achieved 87.04% and kernel PCA achieved 88.59%. A significant different between the percentage of recognised faces using kernel PCA and PCA over the level 2, level 3 and level 4 disgust expression shows that kernel PCA extracts better features compared to PCA.

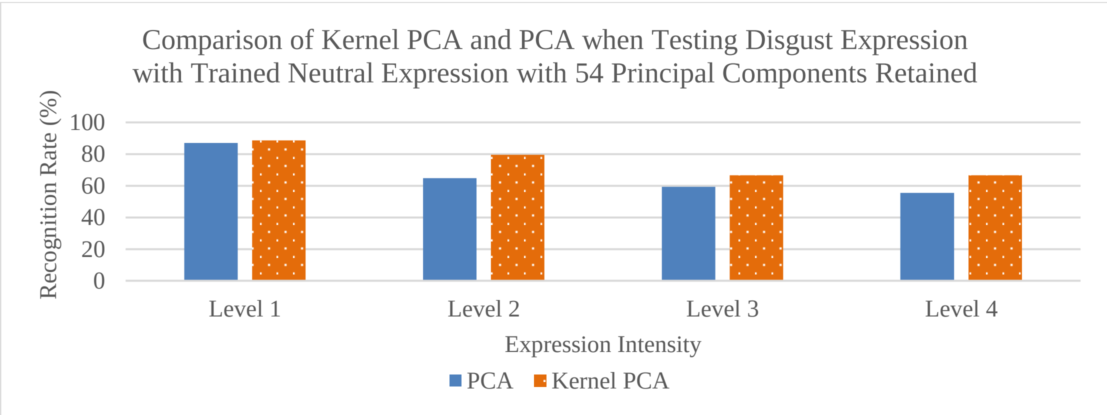


Figure 5.6: Recognition rate of testing Disgust expression to Neutral expression as training sample

Figure 5.7 presents the percentage of recognised faces using fear as testing sample and neutral as training sample based on 54 principal components. Kernel PCA has shown better performance with the highest percentage of 90.74% at level 1 fear expression as compared to PCA with 64.81%. From the graph, it can be seen that the recognition rate for both approaches decreases when performed over the level 2, level 3 and level 4 expression intensity, however, kernel PCA still able to outperform the PCA recognition result.

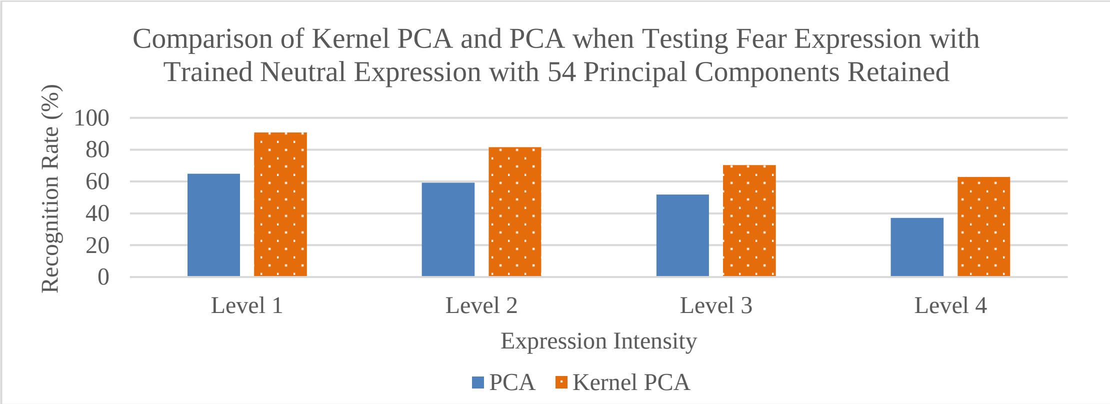


Figure 5.7: Recognition rate of testing Fear expression to Neutral expression as training sample

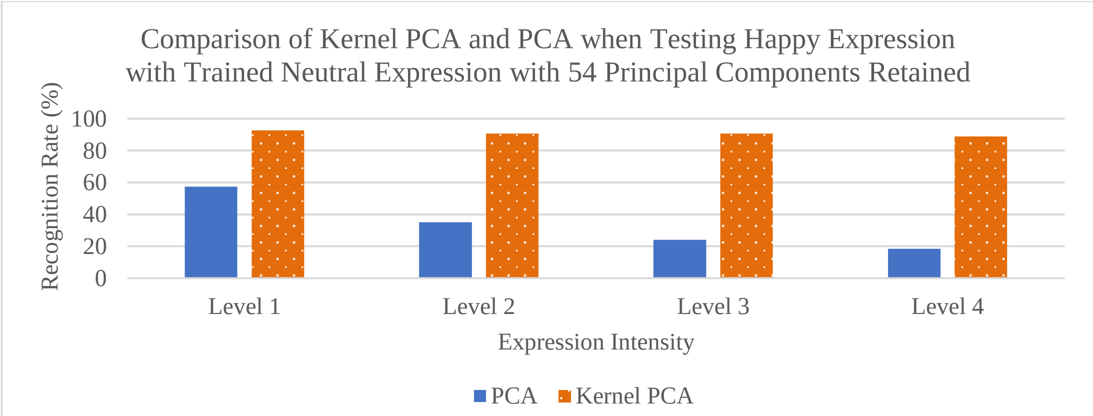


Figure 5.8: Recognition rate of testing Happy expression to Neutral expression as training sample

Figure 5.8 shows the comparison result of kernel PCA and PCA testing happy expression with neutral expression as training sample and 54 principal components are retained. Notable difference in recognition rate can be seen from the results, where kernel PCA achieved a higher percentage with 92.59% consistently for level 1 until level 3 as compared to PCA with 57.41%, 35.19% and 24.07% respectively.

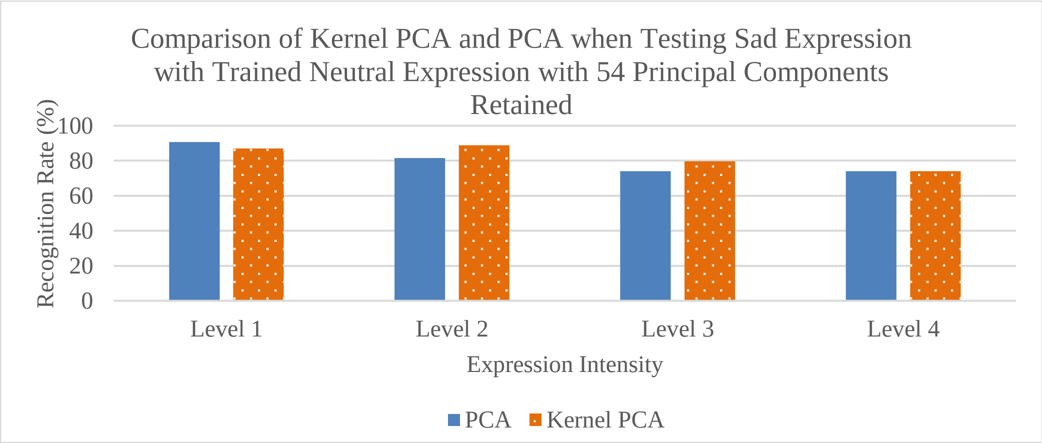


Figure 5.9: Recognition rate of testing Sad expression to Neutral expression as training sample

Figure 5.9 depicts recognition rate based on 54 principal components using sad expression tested on trained neutral expression. The graph shows the comparison of both approaches. The recognition rate for this expression show a comparative result with kernel PCA achieved a slightly higher percentage than the PCA for level 2 and level 3. As shown in the graph, the percentage of recognised faces are the same for both approaches at level 4 with 74.07%. One possible reason to this is the facial structure of sad is almost similar to neutral expression in this dataset.

Figure 5.10 shows the comparison of kernel PCA and PCA approaches in face recognition with 54 principal components retained, using surprise expression as testing sample and neutral expression as training sample. The highest recognition achieved in level 1 by kernel PCA with 94.44% and 55.56% with PCA. Based on the results, kernel PCA has shown that it performed better recognition compared to PCA for all levels except for level 3. This is due to the dataset itself where some of the level 3 surprise pre-processed dataset are found to be incomplete.

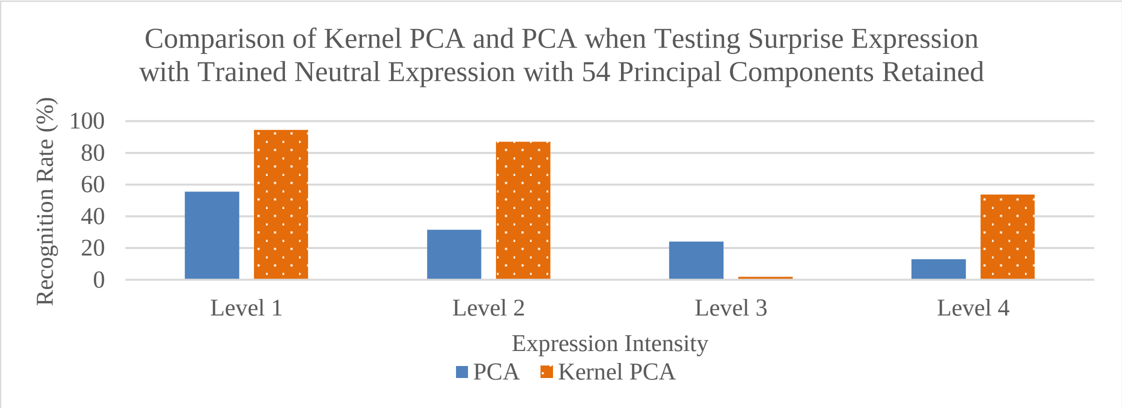


Figure 5.10: Recognition rate of testing Surprise expression to Neutral expression as training sample

The 3D face rendered in Figure 5.11 is showing an 3D geometrical facial structure that barely showing a face. As a result, the overall performance of testing the sample is affected. One possible reason of the incomplete face data could be due to error during pre-processing stage, where the raw data is not cleaned or pre-processed enough. Nevertheless, this error is beyond the scope of this study.

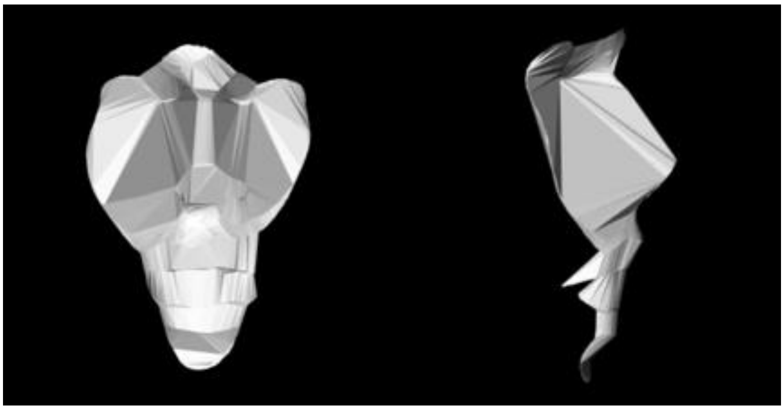


Figure 5.11: A sample of incomplete facial structure from Surprised face data

5.2.4 Discussions

The test 1 results provide a valuable insight towards the quality of both pre-processed 3D datasets. Besides, the kernel potential is also unveiled. Based on the overall results, the recognition rate from using kernel PCA to extract nonlinear facial components is very encouraging. The recognition results indicated that the kernel principal components are more sensitive to 3D faces data as compared to linear principal components. This proves that kernel PCA has the potential to extract better facial features compared to linear PCA for better rate of recognition.

With that, it could be deduced that kernel PCA outperformed PCA in terms of discovering hidden deformation in a face data. This also supports the use of kernel PCA for constructing expression shape model in the proposed mKASM. Furthermore, a problem is found in the Surprised face data sample as shown in Figure 5.11. The incomplete data affected the whole sample and contributed to a very low recognition rate. This issue motivates this research as there is a need to perform facial expression synthesis on the incomplete facial shape in order to recover back the original facial shape to become more identifiable.

Based on this research objectives, using the proposed facial expression synthesis approach, the mKASM, the newly synthesised neutral face will improve the face recognition rate. In order to realise this goal, mKASM is implemented into the same datasets used in this test for easier comparison afterwards. The comparisons can be found in Section 5.4.

5.3 Test 2: Neutral Expression Synthesis using mKASM

Since this study is more interested in removing the expression, the neutral expression synthesis is carried out using the proposed mKASM in this test. Table 5.2 shows the overview of neutral expression synthesis experiments using VisionRT and BU3DFE datasets. There are VisionRT The Train E column is referring to the group E expression as first input and the second input is Train Neutral that refers to the Neutral group. For example, Experiment 1 is synthesising frown to neutral expression, using VisionRT dataset. The input Train E is from frown expression group and input Train Neutral is Neutral 1 group. After all trained samples are synthesised, the newly synthesised neutral faces are stored into a new group and labelled as "Frown to Neutral". The new group will be tested using face recognition in Test 3. Though, before that, this test rendered the synthesised neutral faces in 3D environment to validate the results first by observing the visual changes from the original expression as before synthesis and converted neutral expression as after synthesis whether it still preserves some of the subject's distinct characteristics. In addition to that, the original neutral expression or ground truth from the datasets is also added for comparison.

Table 5.2: Experiment sets for Test 2 using mKASM

Datasets	Exp #	Train E	Train Neutral	Label
VisionRT	1	Frown	Neutral 1	Frown to Neutral
	2	Smile		Smile to Neutral
BU3DFE	3	Angry Level 1	Neutral	(A01) Angry to Neutral
	4	Angry Level 2		(A02) Angry to Neutral
	5	Angry Level 3		(A03) Angry to Neutral
	6	Angry Level 4		(A04) Angry to Neutral
	7	Disgust Level 1		(D01) Disgust to Neutral
	8	Disgust Level 2		(D02) Disgust to Neutral
	9	Disgust Level 3		(D03) Disgust to Neutral
	10	Disgust Level 4		(D04) Disgust to Neutral
	11	Fear Level 1		(F01) Fear to Neutral
	12	Fear Level 2		(F02) Fear to Neutral

Table 5.2 continued

	13	Fear Level 3		(F03) Fear to Neutral
	14	Fear Level 4		(F04) Fear to Neutral
	15	Happy Level 1		(H01) Happy to Neutral
	16	Happy Level 2		(H02) Happy to Neutral
	17	Happy Level 3		(H03) Happy to Neutral
	18	Happy Level 4		(H04) Happy to Neutral
	19	Sad Level 1		(SA01) Sad to Neutral
	20	Sad Level 2		(SA02) Sad to Neutral
	21	Sad Level 3		(SA03) Sad to Neutral
	22	Sad Level 4		(SA04) Sad to Neutral
	23	Surprise Level 1		(SU01) Surprise to Neutral
	24	Surprise Level 2		(SU02) Surprise to Neutral
	25	Surprise Level 3		(SU03) Surprise to Neutral
	26	Surprise Level 4		(SU04) Surprise to Neutral

5.3.1 VisionRT Qualitative Results and Analysis

This section dedicates to Experiment 1 and 2 results. The VisionRT dataset has only one intensity level for frown and smile expression. Figure 5.12 shows faces of three subjects showing comparison between frown as the original expression and the corresponding synthesised neutral expression with respect to the ground truth. From the results, it can be seen that all subjects managed to retain their identity after performing the mKASM. The synthesised neutral expression looks almost similar to the actual neutral expression especially around the mouth area. There are subtle changes on the eyebrows and cheek after the expression neutralisation. Notice the eyebrow of second subject are more relaxed after the synthesis. From the synthesised neutral face, the third subject’s cheek appearance also similar to the original neutral face.

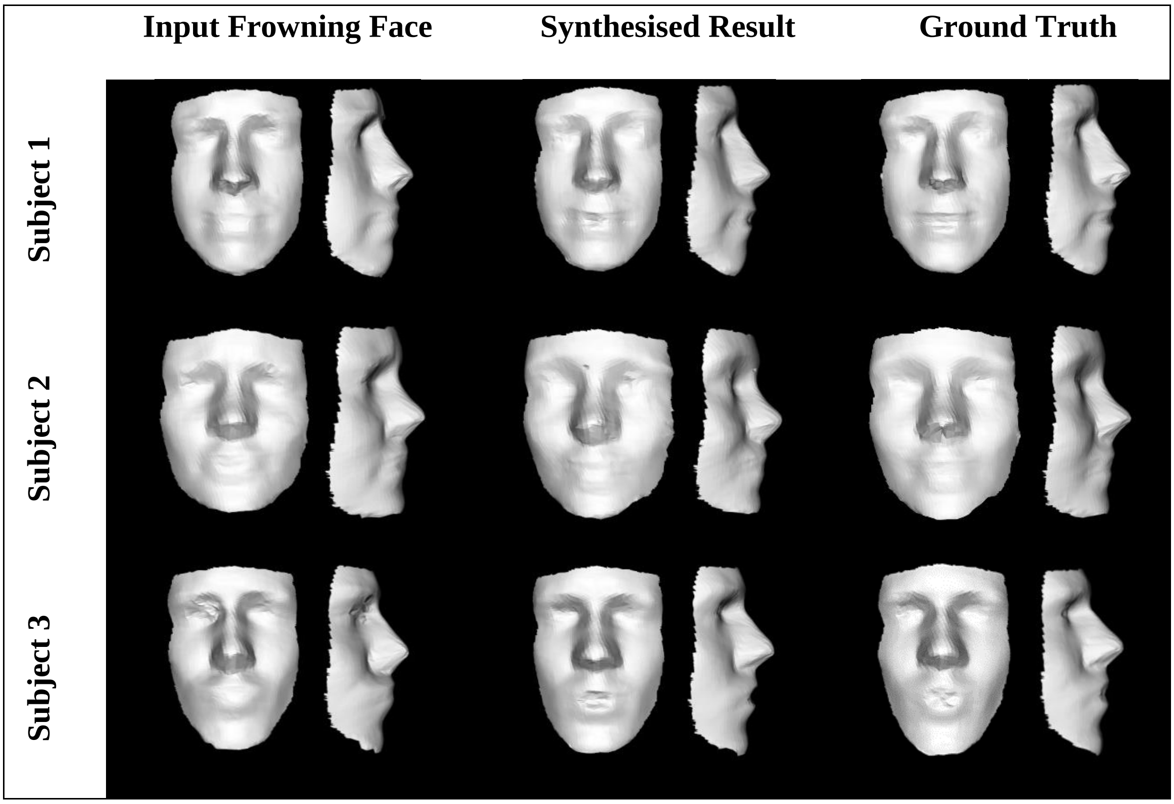


Figure 5.12: Three subjects from VisionRT frown expression synthesised to neutral face

Figure 5.13 shows the comparison between smile expression and the corresponding synthesised neutral expression with the actual neutral expression of three subjects as ground truth. Visible changes after neutral expression synthesis from smile expression can be seen around the cheek muscle. The cheek in synthesised neutral expression is not as relaxed as the ground truth neutral expression. Based on the result, the proposed method has successfully synthesised smile and frown expression into neutral expression from a face while retaining the subject’s identity. The qualitative results of testing the proposed method on VisionRT dataset have shown a positive outcome. In the next section, the proposed method is tested on BU3DFE dataset which has different expression intensities.

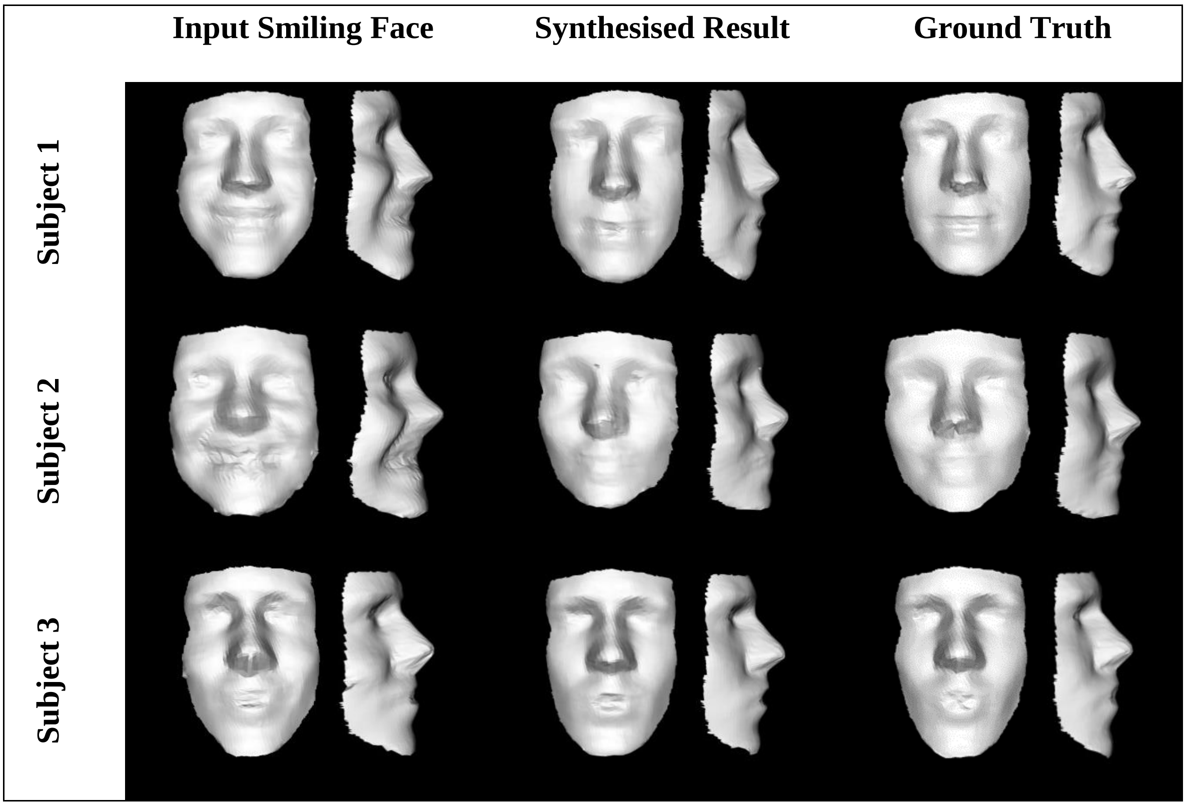


Figure 5.13: Three subjects from VisionRT smile expression synthesised to neutral face

5.3.2 BU3DFE Qualitative Results and Analysis

This section shows the experiment results of synthesising neutral expression from Experiment 3 until 26 using BU3DFE dataset. The dataset has four level of intensity for each of the six expression groups: angry, disgust, fear, happy, sad and surprise. The screen captured image of 3D face of the original expressional face and the ground truth are included as well in the figure as comparison with the synthesised neutral face output.

Figure 5.14 presents a 3D face of a subject's synthesised neutral face from angry expression. The first column of the figure represents original angry expression from the lowest to highest expression intensities, the next column presents the corresponding synthesised neutral expression and last column shows the actual neutral expression from the dataset. Based on the Figure 5.14, after synthesising neutral face from angry expression, there are significant changes for all the expression intensities especially the eyebrows and eyes. The eyes are in original state at synthesised face as compared to the angry face where the eyes are posing hard stare. The stiffen muscles around cheeks and eyebrows also appeared to be more relaxed. The tensed lip which is most noticeable at level 3 and level 4, is seen more relaxed at synthesised neutral face.

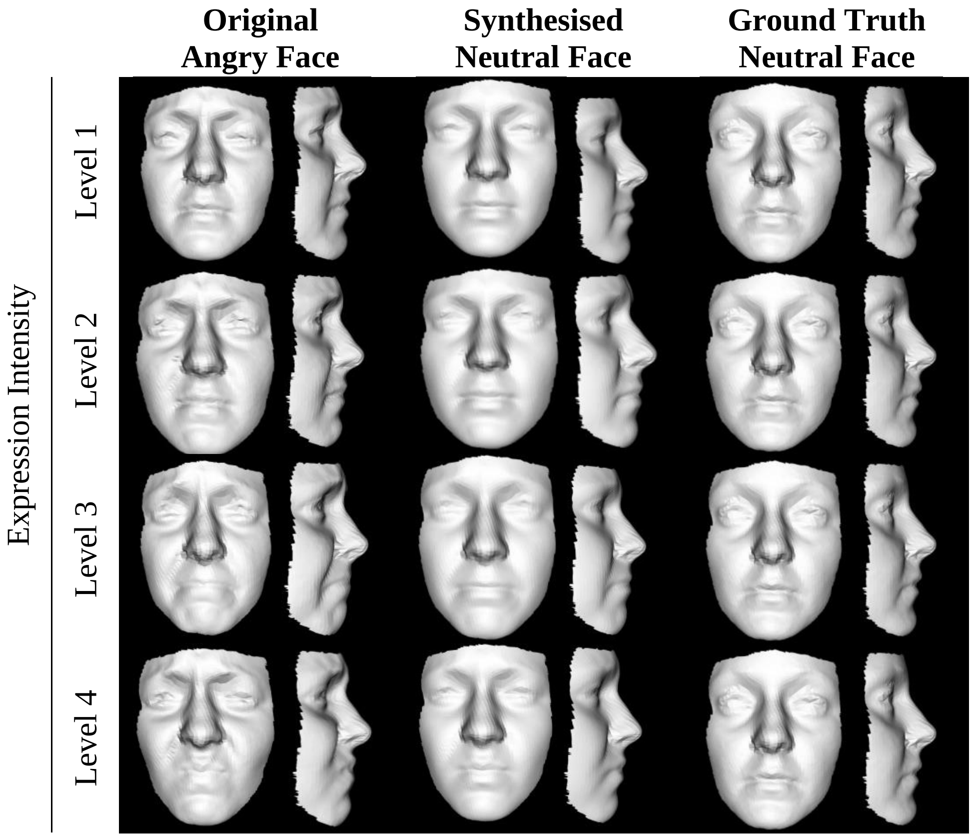


Figure 5.14: BU3DFE's subject synthesised neutral face from angry expression

Figure 5.15 presents a 3D face of a subject's synthesised neutral face from disgust expression. The first column of the figure represents original disgust expression from the lowest to highest expression intensities, the next column presents the corresponding synthesised neutral expression and last column shows the actual neutral expression from the dataset. The original disgust expression face can be seen with exposed teeth. After the synthesis, the mouth is slightly closed and appeared almost similar to the actual neutral face. Besides, the wrinkled nose and harden eyes from disgust face also became more relaxed. Cheeks also appeared to be in original state at synthesised face compared to disgust face where the cheeks are raised.

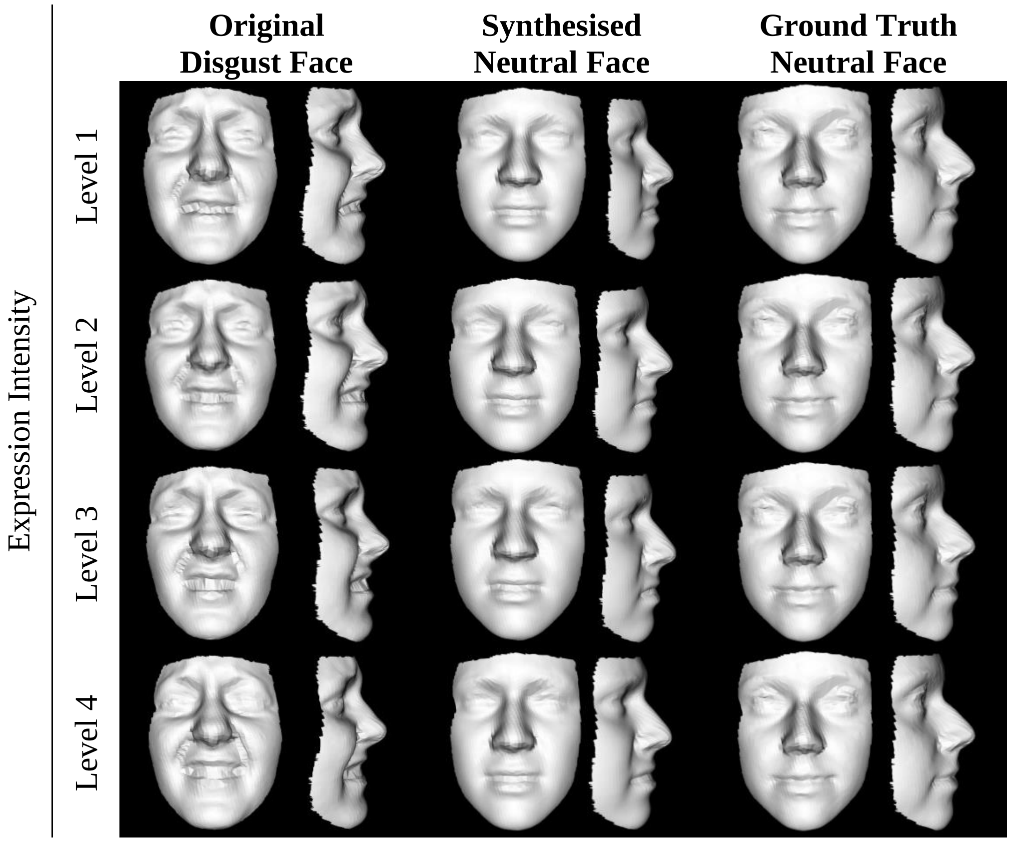


Figure 5.15: BU3DFE's subject synthesised neutral face from disgust expression

Figure 5.16 depicts the result of synthesised neutral expression from fear expression of a subject. The first column of the figure represents original fear expression from the lowest to highest expression intensities, the next column presents the corresponding synthesised neutral expression and last column shows the actual neutral expression from the dataset. The fear expression face is showing raised cheeks and opened mouth. The slightly drawn to the back lips from fear face became more relaxed after the neutral expression synthesis. The cheeks also changed back to its original state. However, the mouth is slightly closed unlike the actual neutral expression which shows the mouth is fully closed. In overall, visually, the identity of the subject is still well-preserved after the mKASM.

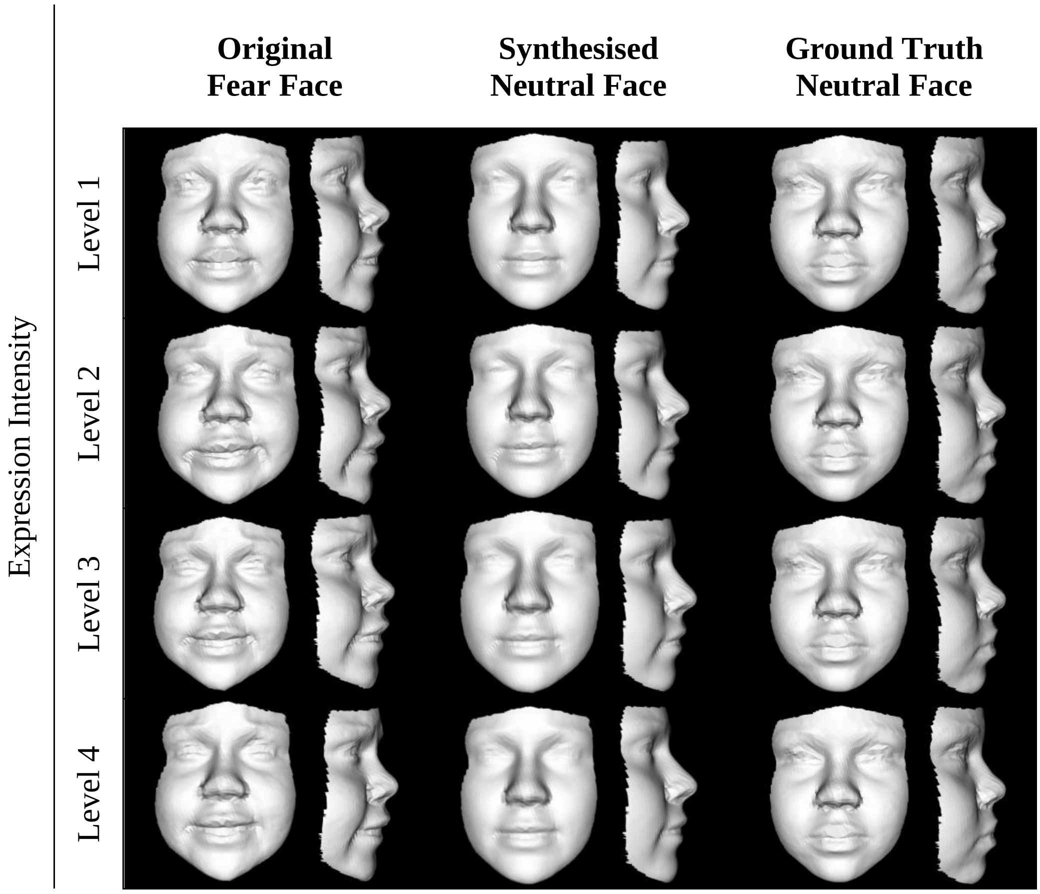


Figure 5.16: BU3DFE's subject synthesised neutral face from fear expression

Figure 5.17 shows the result of synthesised neutral expression from happy expression of a subject. The first column of the figure represents original happy expression from the first until fourth level of expression intensity, the next column presents the synthesised neutral expression from happy and last column shows the ground truth of the neutral expression of the subject in the dataset. After synthesis, it can be seen from the graph that corners of the mouth are drawn back to the original state and are slightly closed. The cheeks are also in a more relaxed state after the synthesis. Based on the result, mKASM is able to transform the happy expression into a reasonable neutral face.

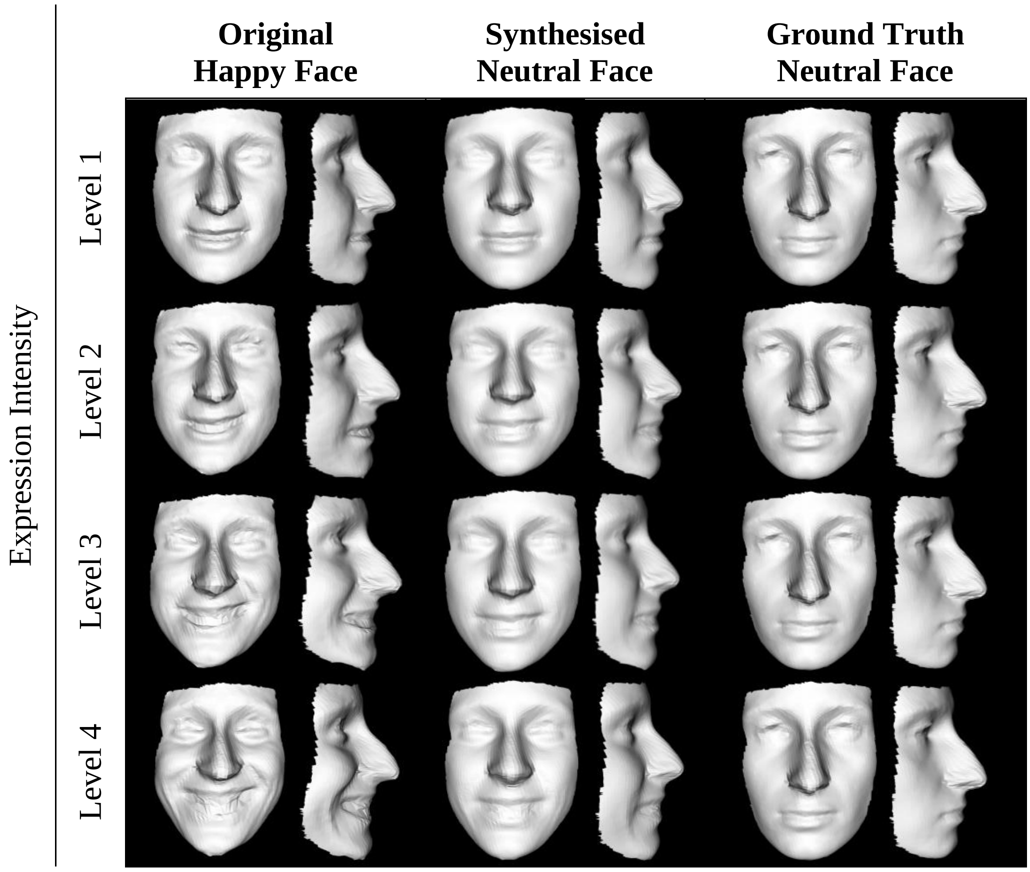


Figure 5.17: BU3DFE's subject synthesised neutral face from happy expression

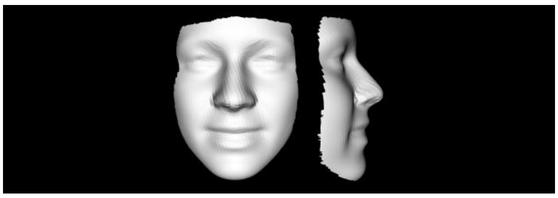


Figure 5.18: Average happy face level 3

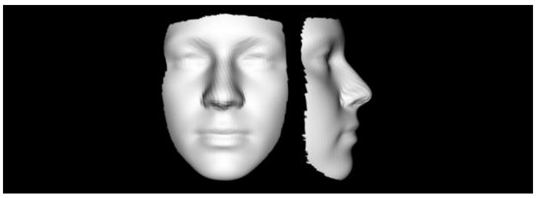


Figure 5.19: Average neutral face from the whole dataset

However, from the Figure 5.17, it can be noticed that the shape of the nose for all four of the synthesised faces is slightly different with the original expressional face. The inconsistency of the shape of the nose might be affected by the geometric shape of average face computed via happy expression sample during the synthesis stage. Figure 5.18 and Figure 5.19 show the differences between the average face for happy expression and neutral expression. Note that both of the shape of the nose is almost similar with the synthesised neutral face.

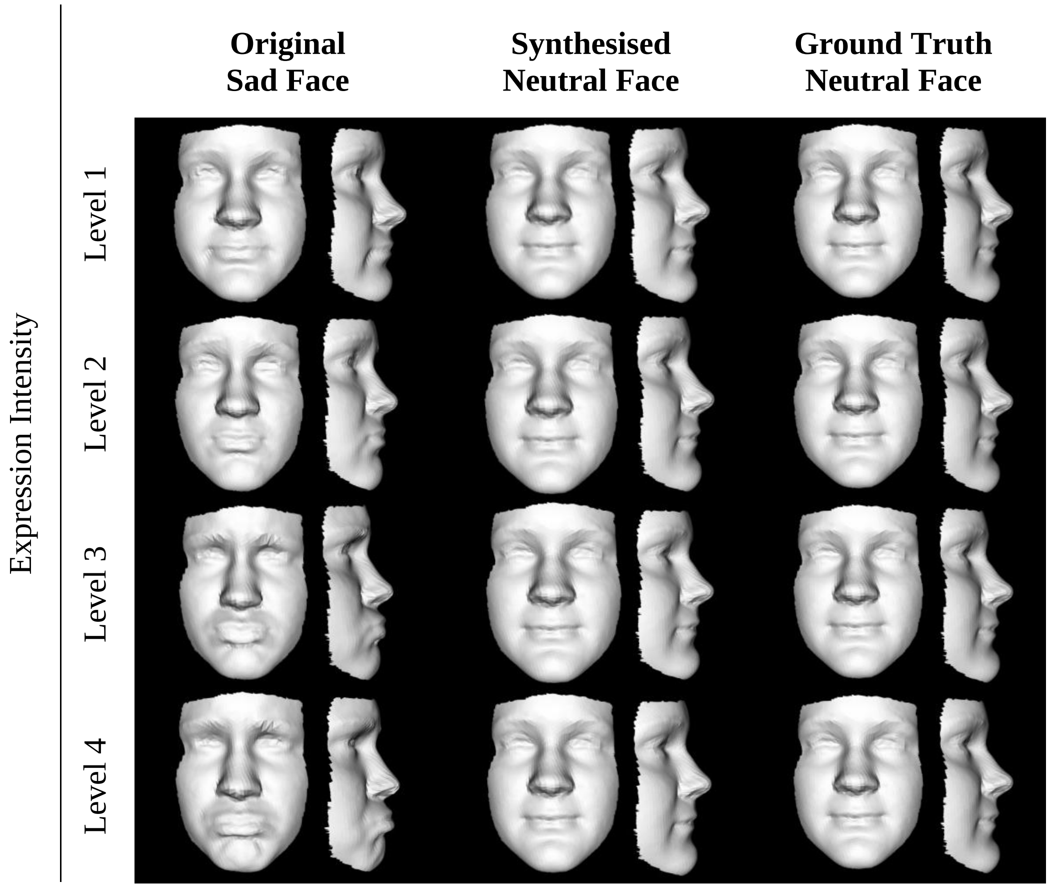


Figure 5.20: BU3DFE’s subject synthesised neutral face from sad expression

Figure 5.20 shows the comparison of synthesised neutral expression from a subject expressing sadness. The first column of the figure represents original sad expression from the first until fourth level of expression intensity, the next column presents the synthesised neutral expression from sad and last column shows the ground truth of the neutral expression of the subject in the dataset. The result shows that the synthesised neutral faces are likely similar to the ground truth neutral face. The mouth from sad expression is successfully synthesised to the subject's original state. As such at level 4 sad expression face, the corner of mouth which is drawn down became more relaxed after synthesised neutral face. The synthesised eyebrow also is not as strained as the original sad expression face.

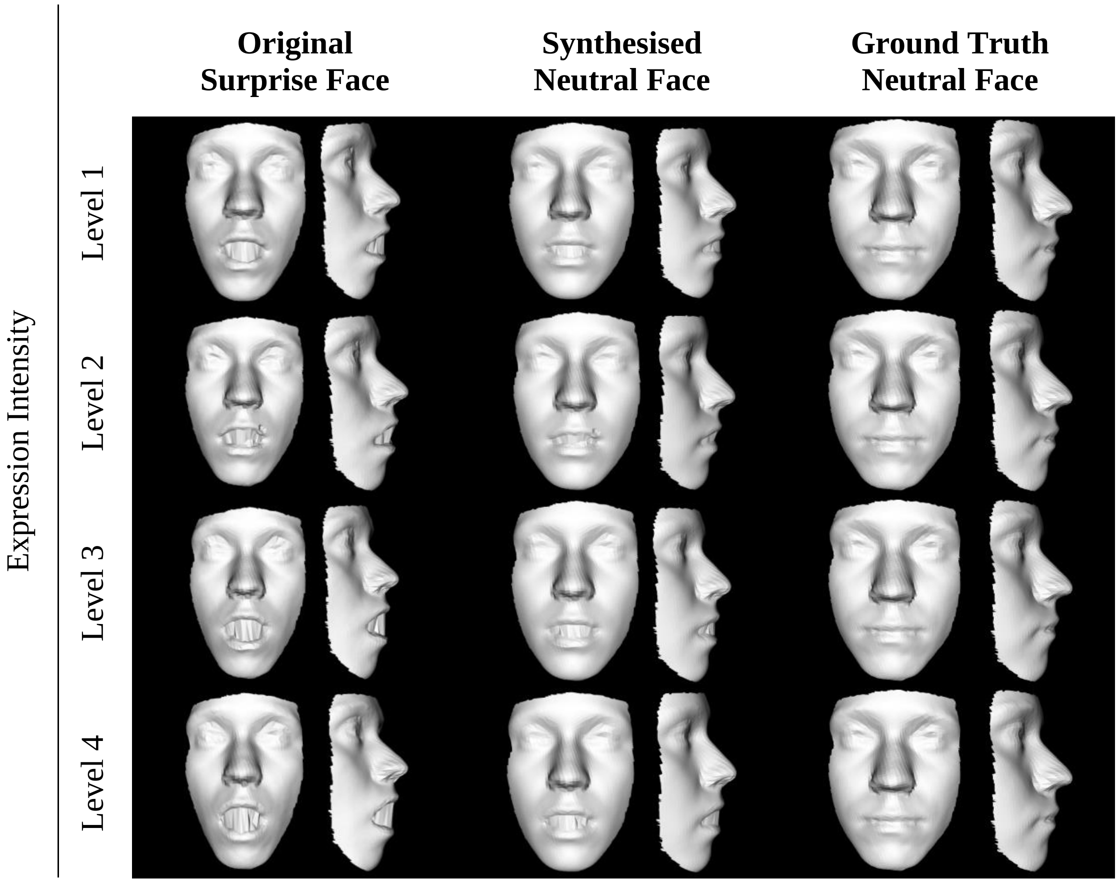


Figure 5.21: BU3DFE's subject synthesised neutral face from surprise expression

Figure 5.21 shows the result of synthesised neutral expression from surprise expression of a subject. The first column in the figure represents original surprise expression from the first until fourth level of expression intensity, the next column presents the synthesised neutral expression from surprise and last column shows the ground truth of the neutral expression of the subject in the dataset. The opened mouth in original surprise expression is slightly closed after synthesised. The raised eyebrows have also changed into its original state as well after performing the mKASM. From the results, it can be deduced that the synthesised neutral faces from surprise expression are visually acceptable and the identity is well preserved for all expression intensities.

5.3.3 Discussions

Overall, this test implements mKASM to neutralise facial expression to both VisionRT and BU3DFE datasets. In terms of neutral expression, the synthesised results for all experiment sets have shown subtle changes on cheek, eyebrow and jaw. Based on the qualitative analysis, the identity of the synthesised neutral face did not fall too far from the original expression of a particular subject. From the figures shown in previous sections, the subject still retains some of the distinct attributes such as the shape structure of face, nose and eyes after the synthesis process. Hence, it can be concluded that all the converted faces achieved an acceptable neutral expression. However, there is a case when the synthesis does not work well where the distinct characteristics such as the shape of the nose should not be changed. This can be observed in Figure 5.17, where the nose shape of synthesised neutral from happy expression is different than the original face. Further improvement could be done in future by combining linear and nonlinear models because face does have both linear and nonlinear movements. Therefore, the first objective to neutralise facial expression from an expressional face without losing the person’s identity is fulfilled.

5.4 Test 3: Evaluation of the Synthesised Neutral Faces using Face Recognition

This section presents the quantitative evaluation of the proposed method. Face recognition is used to verify whether the synthesised neutral face belongs to the same person. If the system able to recognise the synthesised neutral face correctly, then this will prove that the proposed method is effective in neutralising while preserving the identity of the person. The comparison of recognition rates before and after the neutral expression synthesis will ensure the effectiveness of the proposed synthesis method. Hence, in this test, the performance between synthesised neutral faces from Test 2 and original neutral expressions is observed, by feeding them into the face recognition system, which can be found in Chapter 4.5. The results will be compared with recognition results from using kernel PCA-based face recognition in Test 1. The system is using kernel PCA-based method as feature extraction and original neutral expression as training sample. The kernel PCA-based method is chosen because it has shown better recognition results compared to PCA in Test 1. The Root Mean Square Error (RMSE) is also applied as additional quantitative measurement to evaluate the performance of mKASM algorithm. According to Pan et al. (2010), the RMSE between two face model X and Y is defined as:

$$RMSE(X,Y) = \frac{Dist(X,Y)+Dist(Y,X)}{2} \qquad \text{Equation 5.1}$$

$$Dist(X,Y) = \sqrt{\frac{\sum_{t=1}^{n} \|x_t - y_t\|^2}{n}} \qquad \text{Equation 5.2}$$

where n is the number of points on X; x_t is a point on X, y_t is a point of Y which is nearest to x_t. To quantitatively measure the performance for each synthesised neutral face, the RMSE distance between the synthesised neutral face and its neutral face of ground truth is computed. The RMSE distance of Test 1, between the original expressional face and its neutral face is also computed as comparison. The lower the RMSE value, the higher similarity of the synthesised neutral face and the ground truth face.

Table 5.3: Experiments sets for Test 3

Datasets	**Exp #**	**Test Set**	**Train Neutral**
VisionRT	1	Smile to Neutral	Neutral 1
	2	Frown to Neutral	
BU3DFE	3	(A01) Angry to Neutral	Neutral
	4	(A02) Angry to Neutral	
	5	(A03) Angry to Neutral	
	6	(A04) Angry to Neutral	
	7	(D01) Disgust to Neutral	
	8	(D02) Disgust to Neutral	
	9	(D03) Disgust to Neutral	
	10	(D04) Disgust to Neutral	
	11	(F01) Fear to Neutral	
	12	(F02) Fear to Neutral	
	13	(F03) Fear to Neutral	
	14	(F04) Fear to Neutral	
	15	(H01) Happy to Neutral	
	16	(H02) Happy to Neutral	
	17	(H03) Happy to Neutral	
	18	(H04) Happy to Neutral	
	19	(SA01) Sad to Neutral	
	20	(SA02) Sad to Neutral	
	21	(SA03) Sad to Neutral	
	22	(SA04) Sad to Neutral	
	23	(SU01) Surprise to Neutral	
	24	(SU02) Surprise to Neutral	
	25	(SU03) Surprise to Neutral	
	26	(SU04) Surprise to Neutral	

Table 5.3 presents the experiment sets to evaluate the synthesised neutral faces generated from Test 2 which are used as testing set, while the existing neutral expression is fixed as training set. The analysis of the result from this test will give an insight on how much the subject identity is retained after the synthetisation is conducted. For example, Experiment set 1 from VisionRT dataset, using the Smile to Neutral group (synthesised

neutral from smile) as testing input set and Neutral 1 group as training set are passed through the face recognition system. Then, the recognition rate and RMSE are recorded. The number of principal components used is based on the experiment results in Test 1, where maximum recognition result is obtained when the number of principal components used is equal to the total number of samples. Therefore, in Test 3 experiments, 60 principal components used for VisionRT dataset and 54 principal components for BU3DFE dataset.

5.4.1 VisionRT Quantitative Results and Analysis

Standard cross validation method is employed where each test sample is grouped according to smile expression, frown expression and neutral expression. Each group has 60 3D face data corresponds to 60 subjects. The results are presented by two sets of tests. Test 1 is the recognition result from the Test 1 that was previously done in Section 5.2.2 and their corresponding RMSE values. Meanwhile the second set is Test 3 which is referring to the experiment results based on Table 5.3 and their corresponding RMSE values.

Figure 5.22 shows performance results between synthesised neutral expression and the original smile expression. Based on the recognition results, Test 3 achieved higher recognition with 100% which shows that all faces are correctly recognised compared to Test 1 with 85% recognition. The Test 3 achieved a relatively low RMSE value which also confirms that the synthesised face model computed from mKASM is almost similar to the actual neutral expression of the specific subject.

Figure 5.23 shows the recognition rate and RMSE of Test 1 and Test 3 based on frown sample expression. The graph shows an improvement on the rate of recognition from Test 1 to the synthesised neutral expression in Test 3. Test 1 recorded less than 100% recognition rate because some of the subject's identity characteristics were lost while performing the nonlinear dimensionality reduction. While after synthesis, all faces are correctly recognised as shown in Test 3. Besides, the RMSE value of the synthesised face in Test 2 is lower than Test 1 which also supports the recognition results. Based on the tests conducted, it can be stated that the proposed mKASM is effective in synthesising neutral expression on expressional test face without losing much of the subject's identity.

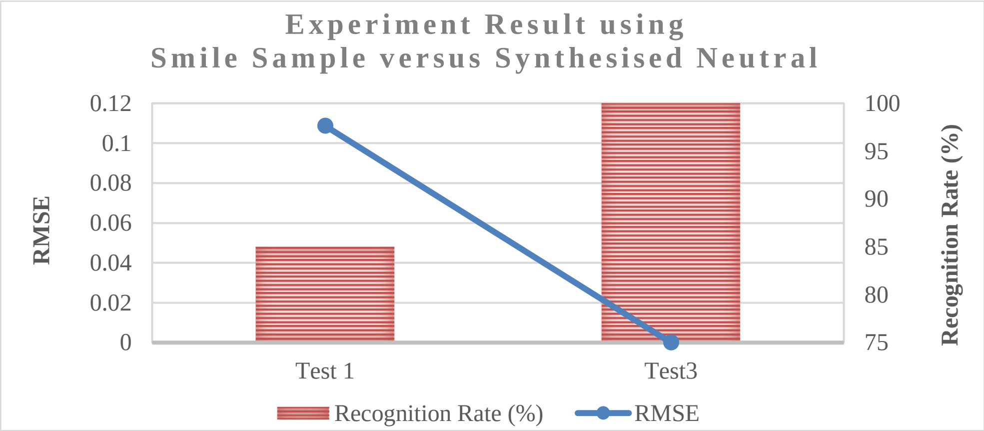


Figure 5.22: Rate of recognition and RMSE results based on VisionRT's smile as target sample expression

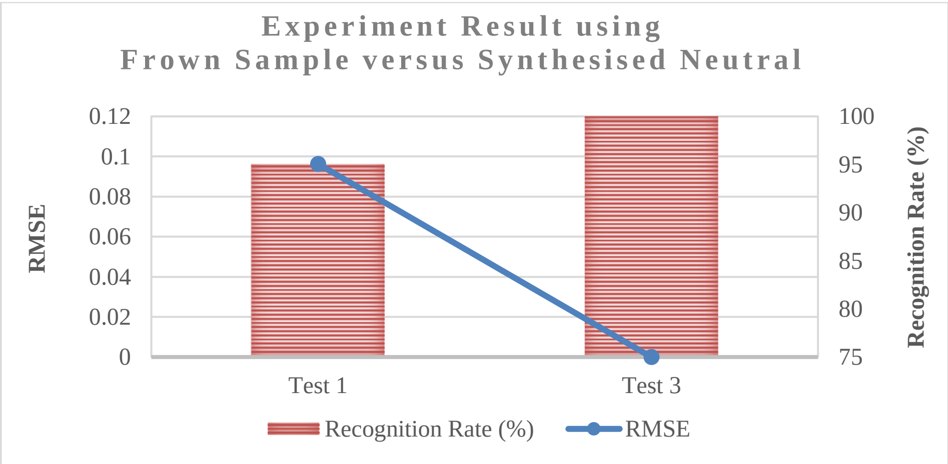


Figure 5.23: Rate of recognition and RMSE result based on VisionRT's frown as target sample expression.

5.4.2 BU3DFE Quantitative Results and Analysis

In order to evaluate the performance of the proposed method, all the 3D faces of six expressions with four levels of intensities and the neutral expression were used. Standard cross validation is employed where each test sample is grouped according to Angry to Neutral (A), Fear to Neutral (F), Disgust to Neutral (D), Happy to Neutral (H), Sad to Neutral (SA) and Surprise to Neutral (SU), whereas Neutral expression as training sample. Each group has 54 face data that corresponds to 54 subjects. Similar to previous datasets, two test sets are presented in this section. Test 1 refers the experiments results taken in Test 1 that was previously done in Section 5.2.3. Test 3 is testing synthesised neutral face with their ground truth neutral expression based on Table 5.3. In contrast with the previous experiment, for this dataset, the results are presented into two parts because this dataset has different level of expression intensities. The first part is the face recognition result between the Test 1 as original facial expressions and Test 3 as synthesised neutral expressions for level 1, level 2, level 3 and level 4 expression intensity. Second part is the comparison of RMSE value between Test 1 and Test 3 for level 1, level 2, level 3 and level 4 expression intensity.

a. Angry Expression Synthesised to Neutral Quantitative Evaluation

Face recognition is executed on the BU3DFE angry sample expression for each level of intensities. From the original angry expression corresponds to Test 1 result displayed in Figure 5.24, a decline can be seen in the recognition rate when the expression intensity increases. This is due to extreme facial deformations that gave a large expression residue between the expressional face and neutral expression face. The maximum recognition achieved by the original angry expression is 96.30% at level 1 while the minimum rate is at the level 4 expression with 75.93%. After the angry expression synthesised to neutral expression as indicated by Test 3, the recognition rate is raised to 100% for all level of expression intensities.

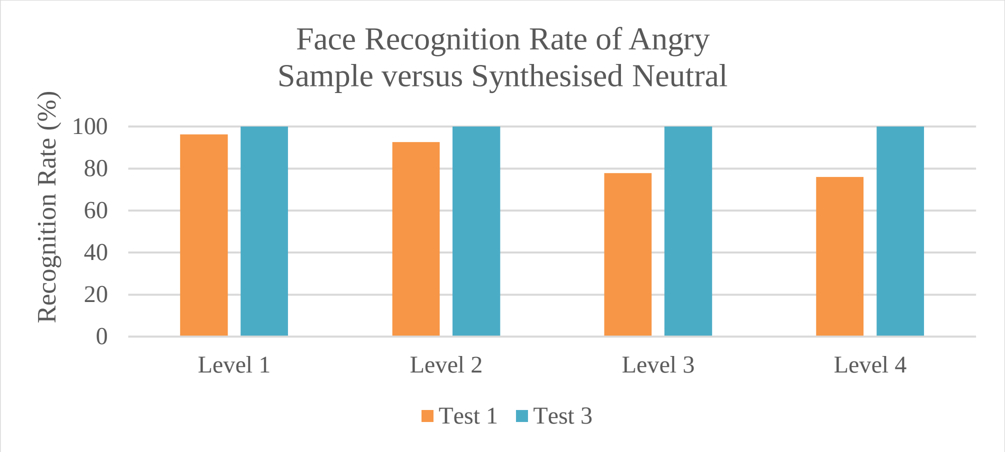


Figure 5.24: BU3DFE Angry sample recognition results for original angry faces from level 1 until level 4 expression intensities and their corresponding synthesised neutral expression

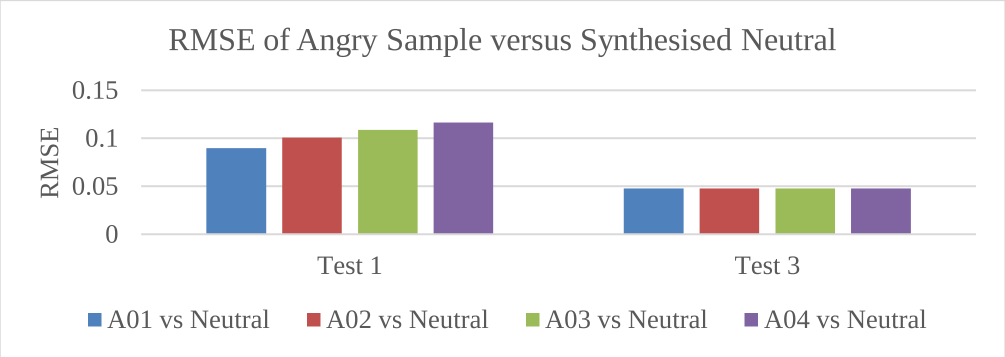


Figure 5.25: BU3DFE Angry sample RMSE for original angry faces from level 1 (A01) until level 4 (A04) expression intensities and their corresponding synthesised neutral expression

Figure 5.25 shows the RMSE result of using BU3DFE angry sample to synthesise neutral expression. The RMSE's pattern is increasing as the level of expression intensifies. This shows that there is a change of expression from one class to another in Test 1.

Meanwhile in Test 3 which is the corresponding synthesised neutral face from each level of expression class in Test 1, have shown that the RMSE values are closer to zero. This proved that the synthesised neutral in Test 3 are similar to the original neutral expression of the specific subject in the angry sample.

b. Fear Expression Synthesised to Neutral Quantitative Evaluation

Figure 5.26 shows experiment result when the synthesised neutral from BU3DFE fear expression sample is tested on trained original neutral expression sample. To compare the effectiveness of the neutralised face, the original fear expression sample is also tested on trained original neutral expression sample. The experiment is carried out for each level of expression intensity. Based on the figure, Test 1 achieved maximum recognition with 90.74% at level 1 while the minimum rate is at the third level expression with 62.96%. After the neutral expression synthesis in Test 3, the 100% recognition for all level of expression intensities indicates that the synthesised faces are correctly recognised.

Figure 5.27 shows the RMSE value of using BU3DFE fear sample versus synthesise neutral expression. The RMSE value in Test 1 is increasing following the level of expression transitions from fear level 1 (F01) until level 4 (F04). Meanwhile in Test 3 which is the corresponding synthesised neutral face for each level of fear expression class in Test 1 have shown that the RMSE values are closer to zero. This confirms that synthesised neutral from angry sample as test faces are similar to the particular subject in the original neutral expression.

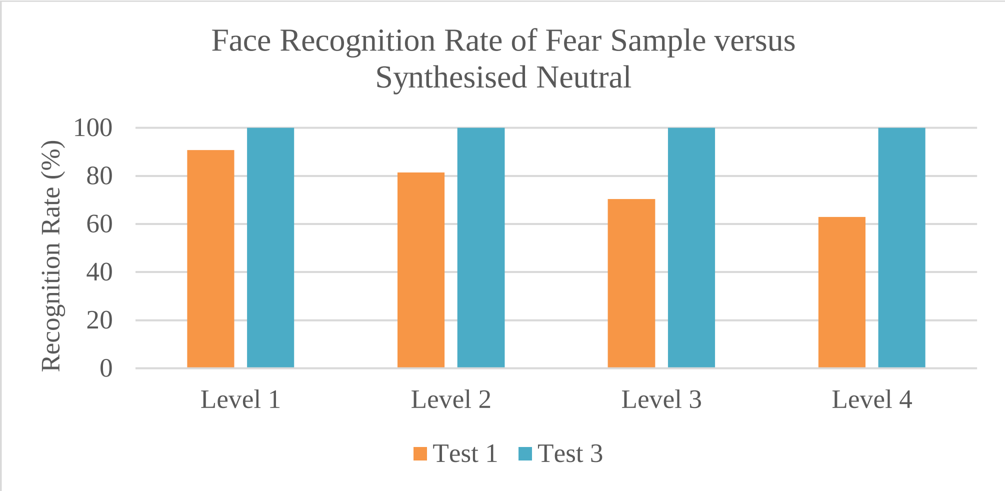


Figure 5.26: BU3DFE fear sample recognition results for original fear faces from level 1 until level 4 expression intensities and their corresponding synthesised neutral expression

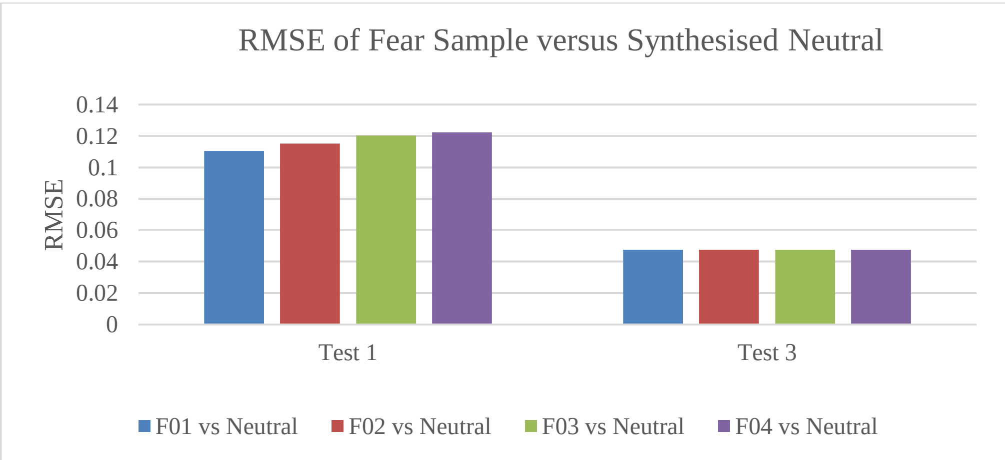


Figure 5.27: BU3DFE fear sample RMSE for original fear faces from level 1 (F01) until level 4 (F04) expression intensities and their corresponding synthesised neutral expression

c. Disgust Expression Synthesised to Neutral Quantitative Evaluation

Figure 5.28 shows the comparison of recognition rate between synthesised neutral from BU3DFE disgust expression indicated as Test 3 versus the original expression from Test 1. The experiment is carried out for each level of expression intensity according to Table 5.3. Based on the result for Test 1, the maximum recognition achieved is 88.59% at level 1. After the neutral expression synthesis, as represented by Test 3 level 1, the recognition rate has increased from 88.59% to 100%. The other three level of disgust expression intensities also recorded 100% recognition which indicates that the synthesised faces are correctly recognised and shown an improvement in face recognition rate compared to using the original disgust expressional face directly as conducted in Test 1.

Figure 5.29 shows the RMSE value of using BU3DFE disgust sample to synthesise neutral expression. Test 3 is the corresponding synthesised neutral face of each level of disgust expression class in Test 1. Noticed that in the figure, Test 3 for all four intensities achieved the RMSE value of approximately to 0.05, which is lower than the RMSE value in Test 1. The result confirms that the synthesised neutral from Test 3 are similar to the original neutral expression of the respective subject in the disgust sample.

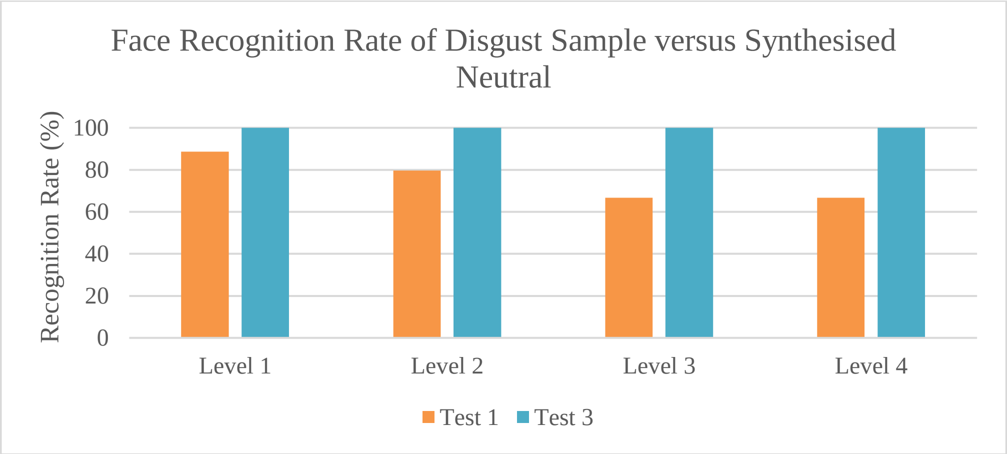


Figure 5.28: BU3DFE disgust sample recognition results for original disgust faces from level 1 until level 4 expression intensities and their corresponding synthesised neutral expression

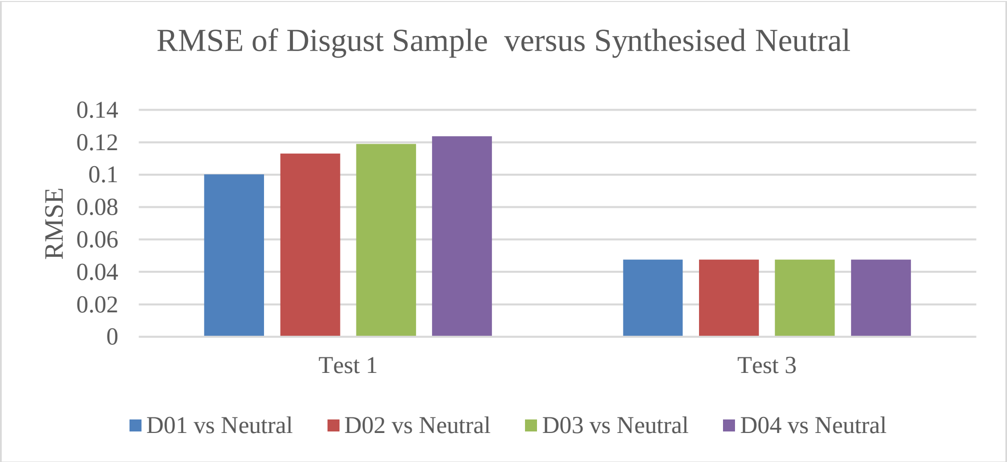


Figure 5.29: BU3DFE disgust sample RMSE for original disgust faces from level 1 (D01) until level 4 (D04) expression intensities and their corresponding synthesised neutral expression

d. Happy Expression Synthesised to Neutral Quantitative Evaluation

Figure 5.30 shows the face recognition result when the synthesised neutral from BU3DFE happy sample is tested on original neutral expression training sample labelled as Test 3. As for comparison with the neutralised face, the original happy expression sample is also tested on trained original neutral expression sample represented by Test 1. The experiment is carried out for each level of intensities. Based on the result, maximum recognition achieved by Test 1 is 92.59% at level 1 and the lowest at level 4 with 88.89% recognition rate. The rate of recognition is decreasing as the expression intensity increases. After the neutral expression synthesis, the 100% recognition rate is recorded in Test 3 for all four level of happy expression intensities. This indicates that all the synthesised faces are correctly recognised.

The newly synthesised neutral expression from happy expression also evaluated using the RMSE to measure the similarities between the synthesised face with the ground truth neutral expression. As shown in Figure 5.31 is the RMSE results of Test 1 and Test 3. The test 1 consists of happy expression according to the sequential intensity level, where the level 4 being the highest intensity. Whereas Test 3 consists of the result after each level of happy expression sample is synthesised to neutral expression. Based on the result, lower RMSE is achieved after the synthesis in Test 3 compared to Test 1. This shows that the synthesised neutral faces have almost similar characteristics with respective subjects' original neutral expression in happy expression sample.

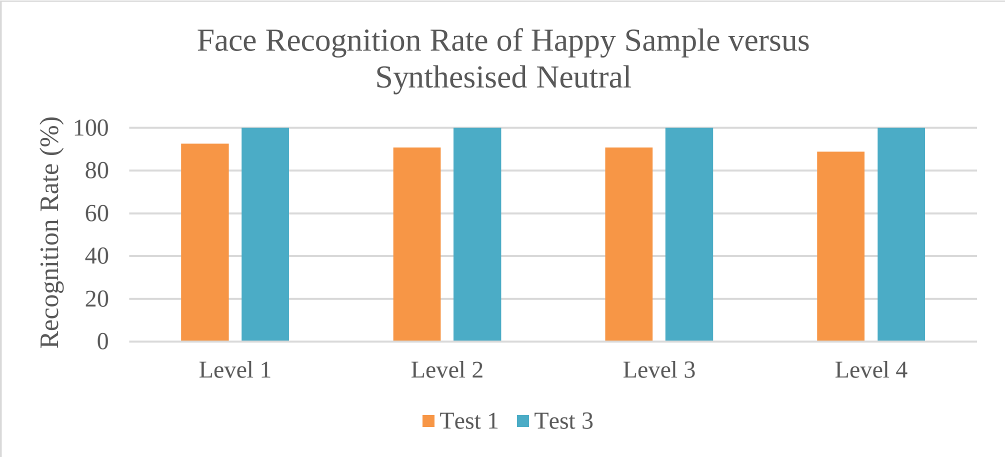


Figure 5.30: BU3DFE happy sample recognition results for original happy faces from level 1 until level 4 expression intensities and their corresponding synthesised neutral expression

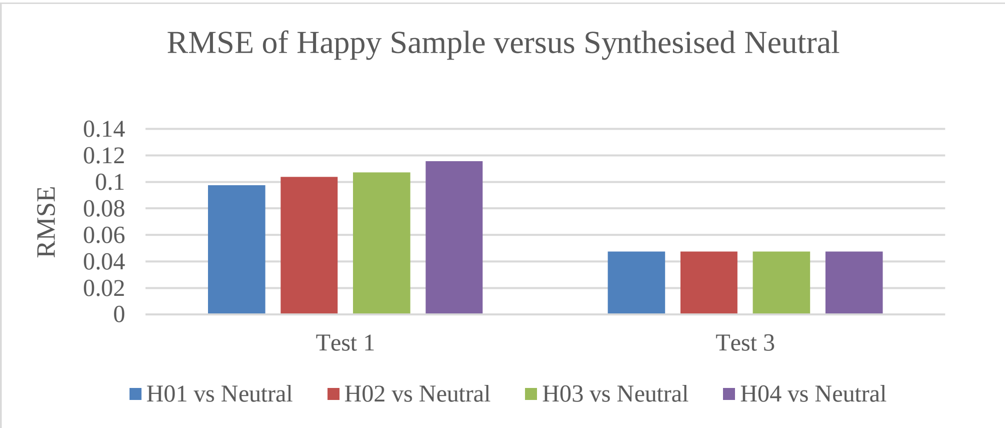


Figure 5.31: BU3DFE happy sample RMSE for original happy faces from level 1(H01) until level 4 (H04) expression intensities and their corresponding synthesised neutral expression

e. Sad Expression Synthesised to Neutral Quantitative Evaluation

Figure 5.32 shows the face recognition result comparison between Test 1 as BU3DFE's sad expression before synthesis and Test 3 as after synthesis. The experiment is carried out for each level of expression intensity. From the result in Test 1, it can be seen that maximum recognition achieved is 88.89% at level 2 while 74.07% at level 4 is the lowest rate. These results are likely due to shape deformations in each of the expressional face in addition to different expression intensities. After the neutral expression synthesis, the recognition rates from Test 1, level 1 until level 4 have increased to 100% which can be found at all levels in Test 3. This shows that the synthesised faces are correctly recognised by the face recognition system.

Figure 5.33 shows the RMSE value of using sad sample to synthesise neutral expression. The RMSE value increases following the expression intensities from sad level 1 (SA01) until level 4 (SA04). Meanwhile for all level in Test 3 that represents the corresponding synthesised neutral face of each level of sad expression class in Test 1, have shown that all RMSE values closer to zero. This confirms that the neutralised face and the original neutral expression of the specific subject in the sad expression sample are likely similar.

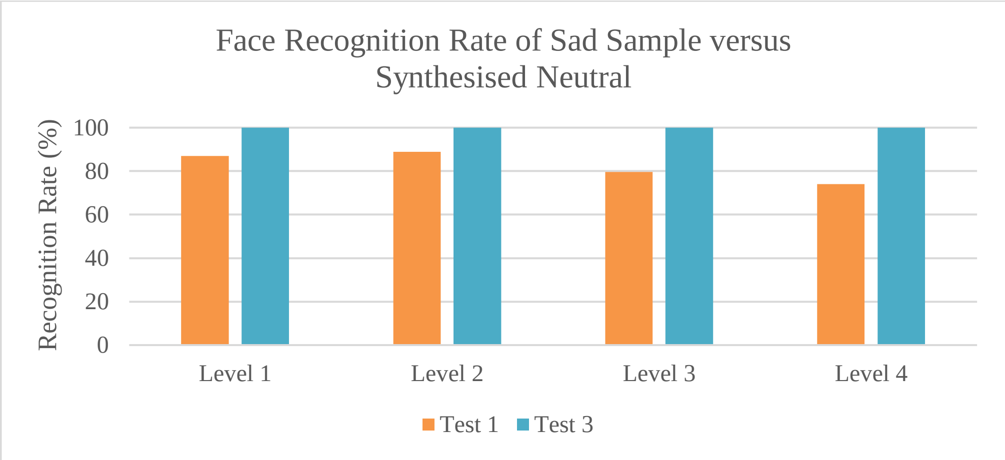


Figure 5.32: BU3DFE sad sample recognition results for original sad faces from level 1 until level 4 expression intensities and their corresponding synthesised neutral expression

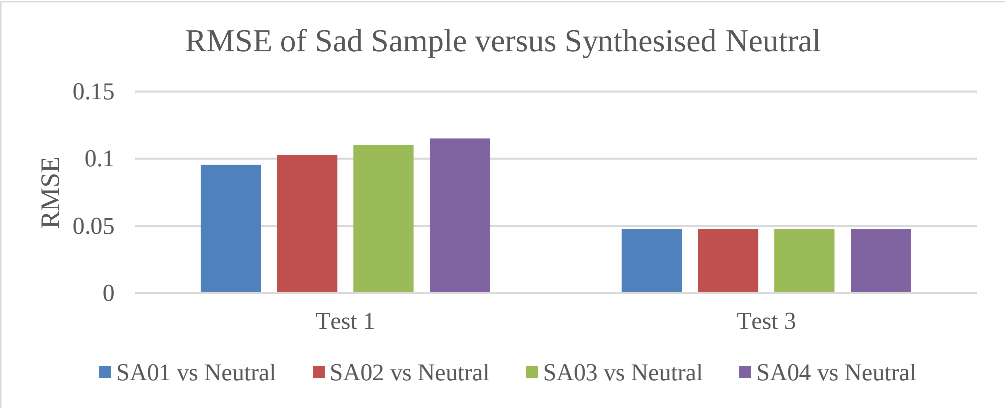


Figure 5.33: BU3DFE sad sample RMSE for original sad faces from level 1(SA01) until level 4 (SA04) expression intensities and their corresponding synthesised neutral expression

f. Surprise Expression Synthesised to Neutral Quantitative Evaluation

Figure 5.34 shows Test 3 experiment result when the synthesised neutral from BU3DFE surprise expression sample is tested on trained original neutral expression sample, compared with Test 1 as before the synthesis. The results in Test 1 indicates that maximum recognition achieved by testing original surprised expression with neutral is 94.44% at level 1 while the minimum rate is at the level 3 expression with 1.85%. This is due to the incomplete face data as shown in Figure 5.35 (a) that affected the whole data sample. However, a neutral expression can be reconstructed from the deformed facial shape using the proposed mKASM. The reconstructed neutral expression can be found in Figure 5.35 (b), while Figure 5.35 (c) shows the ground truth neutral face of that particular subject as comparison. The synthesised face at level 3 in Test 3 achieved 100% recognition rate. This shows that the system is able to correctly recognise the synthesised neutral face for each specific subject. After the neutral expression synthesis as represented by Test 3, the 100% recognition rate achieved for all level of expression intensities indicates that all the synthesised faces are correctly recognised.

Figure 5.36 shows the RMSE value of using BU3DFE surprise sample to synthesise neutral expression. A clear trend can be seen from the result, where the RMSE value is increasing following the level of expression changes from surprise level 1 (SU01) until level 4 (SU04). SU03 in test 1 recorded the highest error value with 0.14 following the poor recognition rate mentioned previously. Meanwhile in Test 3 which represents synthesised neutral faces of each level of surprise expression class in Test 1, have shown that the RMSE values are closer to zero. This proves that the synthesised neutral faces are almost similar to the original neutral expression face of the respective subject in the surprise sample.

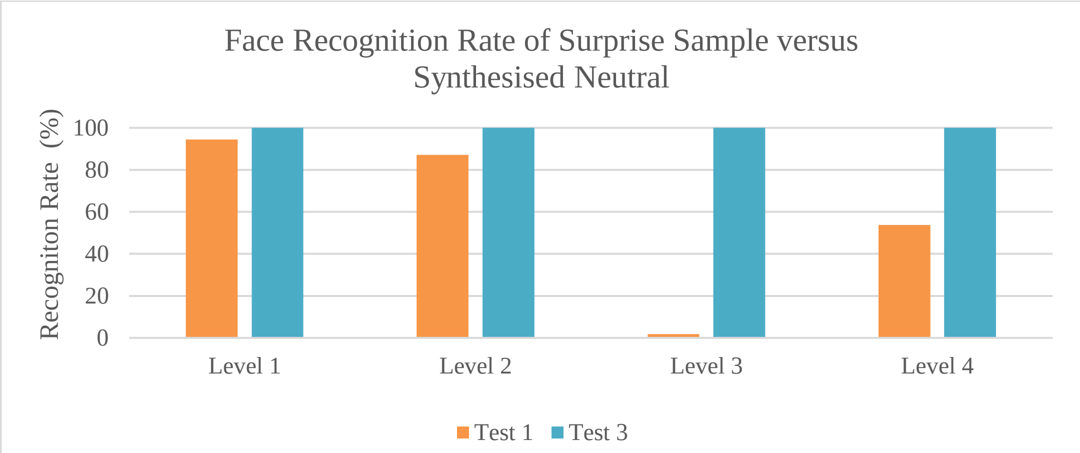
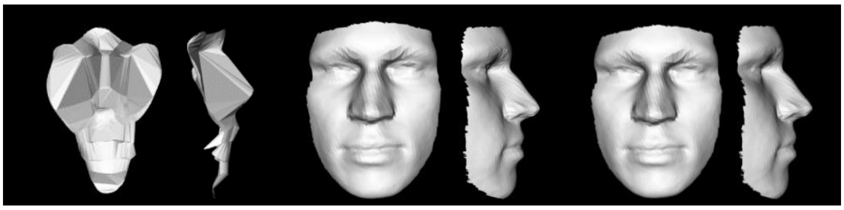


Figure 5.34: BU3DFE surprise sample recognition results for original surprise faces from level 1 until level 4, and their corresponding synthesised neutral expression



(a) (b) (c)

Figure 5.35: Improvement made on an incomplete surprise face sample from BU3DFE. (a) The original surprise expression. (b) The synthesised neutral from surprise. (c) The ground truth of neutral expression

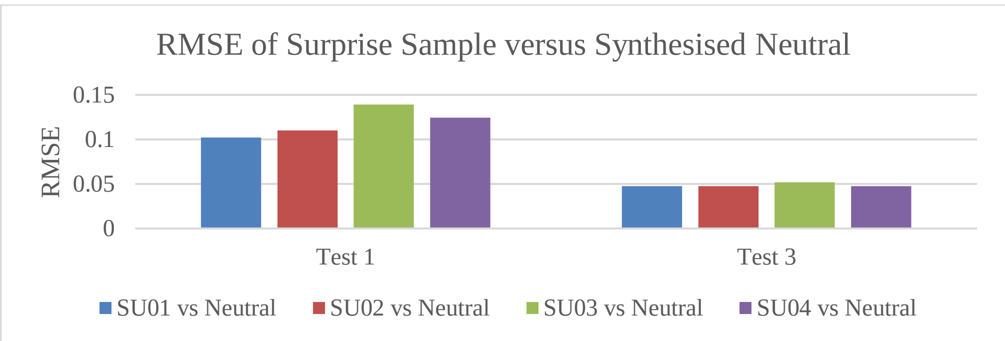


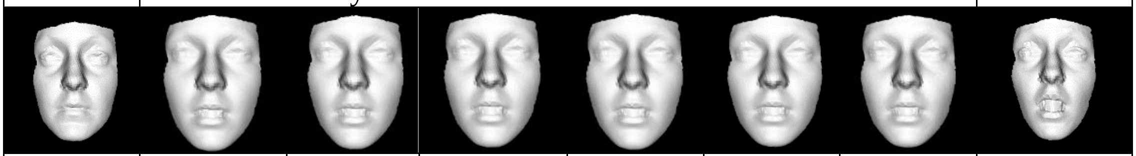
Figure 5.36: BU3DFE surprise sample RMSE for original surprise faces from level 1(SU01) until level 4 (SU04) expression intensities and their corresponding synthesised neutral expression

5.4.3 Discussions

To conclude the overall results in this section, it can be seen that the synthesised neutral faces in Test 3 have always produce a higher recognition rate as compared to the original facial expression in Test 1 for both VisionRT and BU3DFE datasets. This further confirms the effectiveness of the proposed method, where through neutralisation of expressional faces, the recognition rate will improve. The low RMSE value of the synthesised neutral face also confirms this claim. Positive results were achieved especially for higher level of expression intensities from the synthesised neutral faces represented by Test 3 compared to the original expressional face represented by Test 1. The recognition also recorded 100% which shows that the synthesised neutral expression is highly similar to the expected neutral expression of a particular subject. The most significant improvement can be seen at Test 3 synthesised neutral from surprise in Figure 5.34. The poor recognition rate is because of the error in the face data noted previously in Test 1, Figure 5.11. After neutral expression synthesis, the incomplete face is recovered and achieved 100% recognition rate. Therefore, the recognition rate and RMSE positive results have shown that the mKASM can be integrated into the existing face recognition system to enhance the rate of recognition.

5.5 Additional Experiment

Cootes et al. (1995) claimed that more similar shape from the training set can be generated by applying limits to the statistical shape coefficients. As mentioned in Chapter 2.6, experiment on the limit of b shape parameter from the Expression Shape Model can been done to observe the effects of the overall synthesised expression. Thus, an additional experiment has been done but from the results obtained, this approach is not valid. This is because the resulting face as shown in Figure 5.37 produced similar effects for all limits $\pm\sqrt[3]{b}$. Also, the opened mouth is only slightly closed. Therefore, the range of expression changes from face with expression until neutral expression could not be further explored. This issue is also supported by Kirschner et al. (2011) where they stated that the approach of placing limits in the kernel ASM as in the linear ASM is not valid in general. Further investigation is required to determine exactly how the b shape parameter affects the overall shape of the synthesised face.



	Synthesised Neutral from mKASM						
Original Neutral	-3	-2	-1	1	2	3	Original Surprise
	Number of b Parameter, $\pm\sqrt[3]{b}$						

Figure 5.37: Shape effects on varying b parameters

5.6 Comparison of Kernel ASM and Linear ASM

In this section, the performance of kernel ASM with its counter linear approach, ASM is compared. To implement the linear ASM, only the kernel ASM module is substituted with linear ASM module and proceeded to other steps within the same mKASM framework. The linear ASM algorithm can be found in Chapter 2.5.1. The four criteria that are used to measure the comparison are based on complexity, computational time to construct each expressional shape model, face recognition rate and visual comparison.

a. Complexity

Both methods are easy to implement. With either kernel ASM or linear ASM, the generated shape model performed expression synthesis without many small steps involved. However, kernel ASM has two extra steps in the kernel-based PCA for kernel matrix computation and centring the kernel matrix as compared to ASM, which PCA only computes the covariance matrix before performing the eigen decomposition to compute the eigenvalues and eigenvectors.

b. Time Cost

For the second criteria, kernel ASM require approximately 1570 seconds to generate an expressional shape model compared to ASM which only spent approximately 6 seconds. The kernel ASM consumed more time due to higher dimensionality computation involved when computing the kernel PCA.

c. Rate of Recognition

After the neutral expression synthesis, synthesised neutral faces based on linear ASM were tested using the face recognition module. Figure 5.38 and Figure 5.39 presents the rate of recognition comparison between the linear face and nonlinear based synthesised neutral face using VisionRT and BU3DFE dataset respectively. Results in Figure 5.38 indicates that both methods performance are comparative. This could be due to the neutralised faces using both ASM and Kernel ASM have adequately retained significant facial attributes of every face data that lead to the maximum recognition rate. To further investigate the differences, visual comparisons was done in next section to observe the differences. While in the figure 5.39, it can be observed that kernel ASM shows higher and stable recognition rate compared to ASM. This can be especially observed the three expressions, Disgust level 3 and level 4, Fear level 1, and Happy level 3 and level 4.

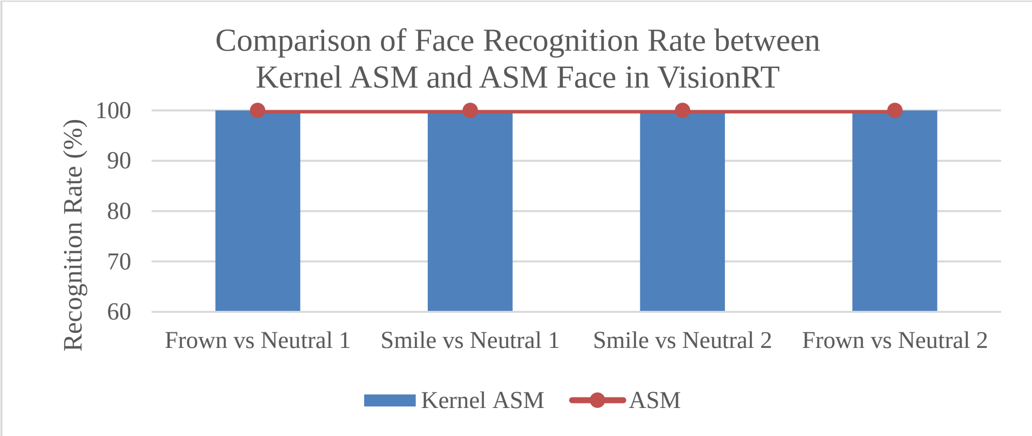


Figure 5.38: Comparison synthesised neutral face of kernel ASM and ASM based approach using VisionRT

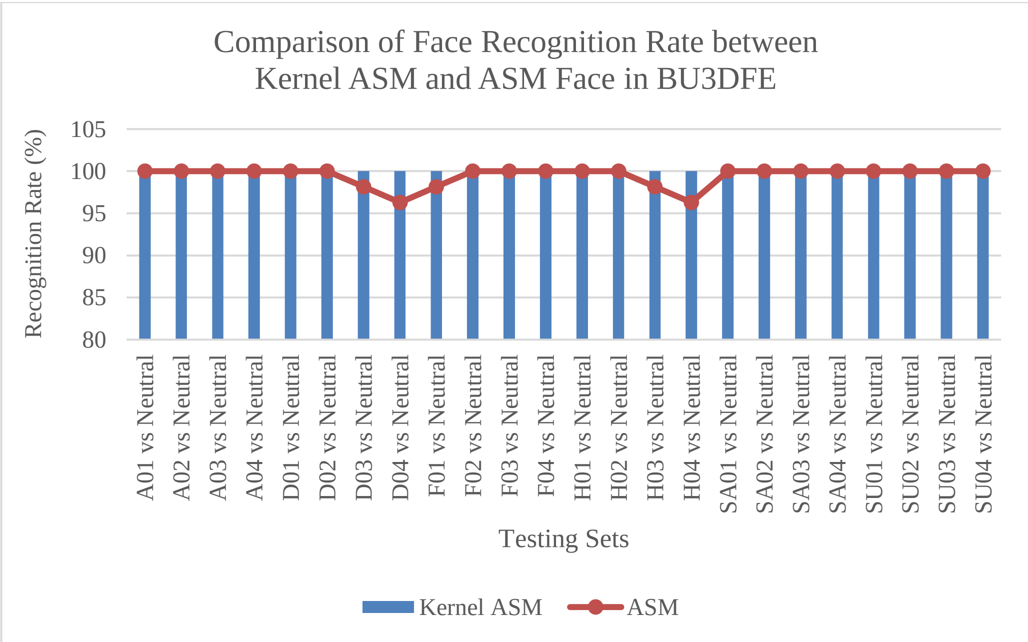


Figure 5.39: Comparison synthesised neutral face of kernel ASM and ASM based approach using BU3DFE

d. Visual Comparison

Figure 5.40 shows sample of three subjects from VisionRT dataset following the competitive results in Figure 5.38. Subtle differences can be observed around the eyes and mouth of the rendered synthesised 3D faces for ASM and Kernel ASM. The eyes and mouth are more defined in kernel ASM compared to ASM. Figure 5.41 shows the average face for the original smile, neutral and frown expression for comparison with the result in Figure 5.40. It can be noticed that the second subject from smile ASM still retained some smile characteristics compared to the kernel ASM. The synthesised face from smile expression based on ASM of the second subject also appeared with slightly different shape than the rest of synthesised face with the same subject. This shows that although quantitatively, the performance of both methods is competitive, subtle differences can be observed visually. The results in Figure 5.40 have shown that synthesis using kernel ASM generates more details compared to the linear ASM.

Synthesised Neutral Expression from Smile		**Synthesised Neutral Expression from Frown**	
Kernel ASM	**ASM**	**Kernel ASM**	**ASM**

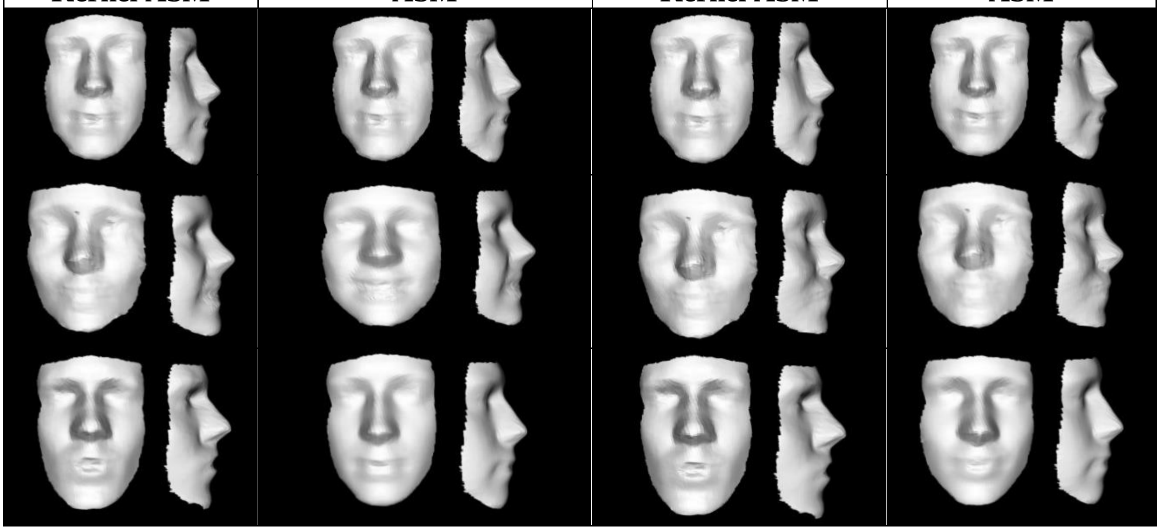


Figure 5.40: Comparison of synthesised neutral face based on kernel ASM approach and ASM approach using VisionRT dataset

Mean Faces		
Smile	**Neutral**	**Frown**

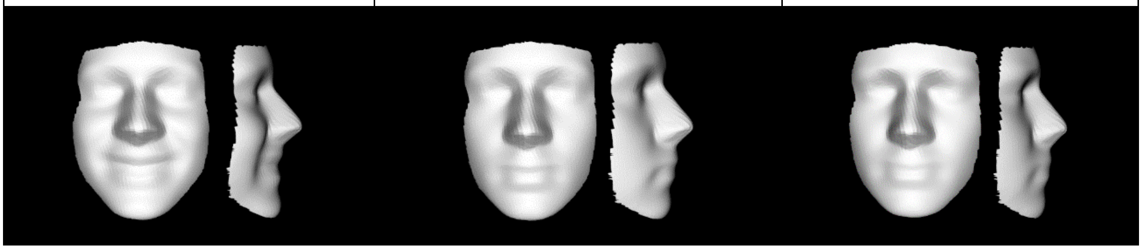


Figure 5.41: Comparison of mean faces of smile, neutral and frown from VisionRT dataset

Figure 5.42 presents a subject sample from BU3DFE dataset portraying synthesised neutral expression from disgust and happy expressions, with four levels of expression intensities. Figure 5.43 shows the average face for the original disgust, neutral and happy expression for comparison with the result in Figure 5.42. From Figure 5.42, it can be observed that there are significant differences found over the eyes and mouth of the rendered synthesised 3D faces for ASM and Kernel ASM. The eyes and mouth are much more relaxed and closer to neutral expression at synthesised faces under kernel ASM compared to ASM. The cheekbone structure of the neutralised faces from kernel PCA also appeared to be more relaxed than PCA. The biggest differences are found at level 3 and level 4 disgust where the synthesised neutral faces using ASM still retains some disgust attributes around the nose and mouth as compared to kernel ASM. Furthermore, the level 4 synthesised neutral from disgust has different facial structure when compared with mean neutral faces in Figure 5.43. The poor results might contribute to lower recognition rate by ASM compared to kernel ASM, which can also be observed in Figure 5.39. Besides, when the synthesised neutral expression from happy in the Figure 5.42 is observed, the transitions from level 1 to level 4 intensity under kernel ASM is much better than ASM in terms of appearance and still able to identify that all four faces are the same subject. Whereas, in ASM level 1 synthesised neutral from happy, the face is appeared to be sadness instead of neutral expression. Despite all mouth of neutralised faces from disgust and happy expressions are still slightly opened, the synthesised neutral face that appears almost the same as the mean neutral as shown in Figure 5.43, which is from level 1 disgust and happy expression under kernel ASM. Hence, it can be deduced that kernel ASM outperformed ASM.

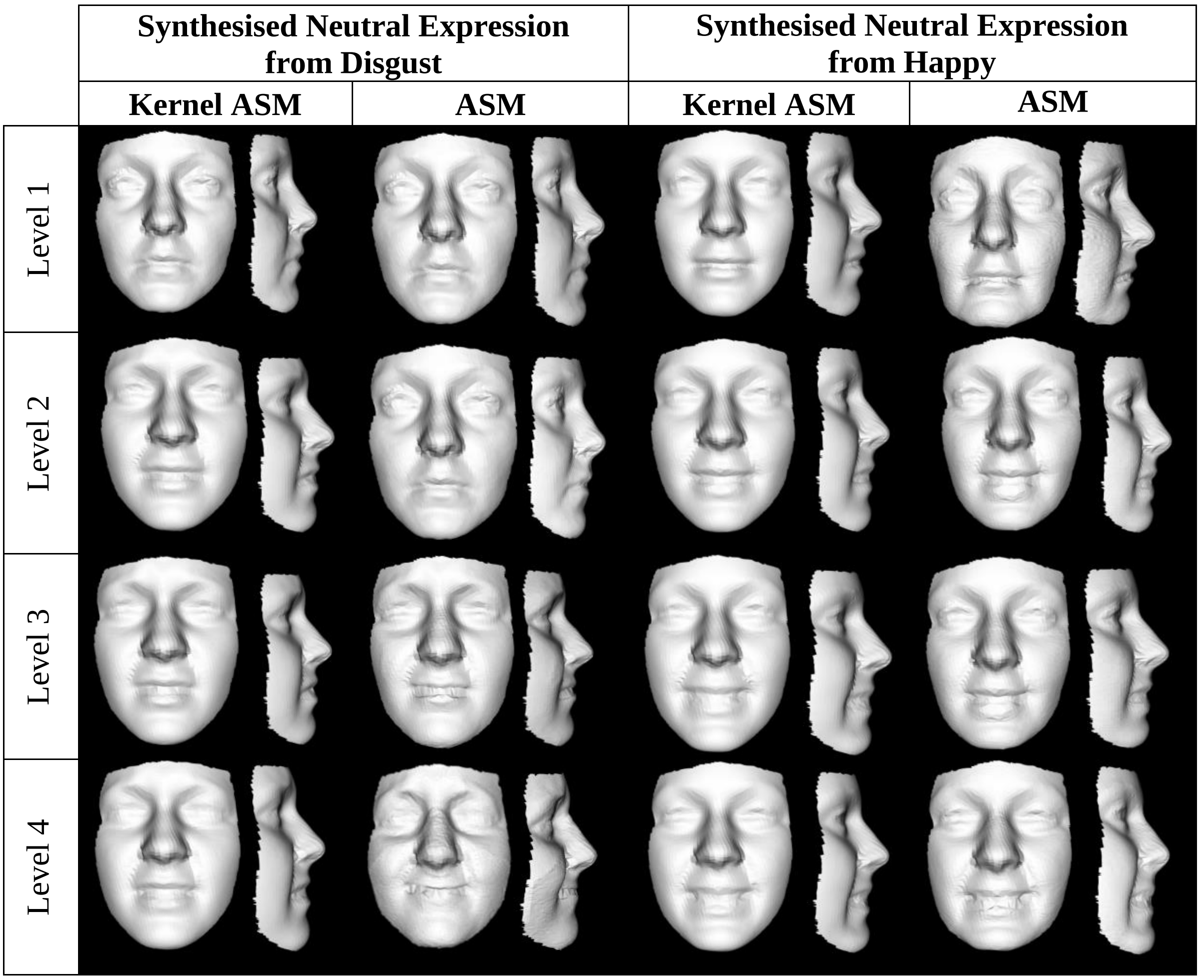


Figure 5.42: Comparison of synthesised neutral face based on kernel ASM approach and ASM approach using angry and disgust samples in BU3DFE dataset

Mean Faces		
Disgust	**Neutral**	**Happy**

Figure 5.43: Comparison of mean faces of disgust, neutral and happy from BU3DFE dataset

5.7 Summary

In this section, three tests are successfully conducted on VisionRT and BU3DFE databases respectively. A lot of time was taken to perform the experiments because it involves more than 100 datasets. Nevertheless, the proposed mKASM method is still able to synthesise neutral expression on a test subject while maintaining some significant attributes that made up the identity of the test subject. The results are presented and analysed through 3D rendering and graph visualisation. Based on the overall results, the mKASM successfully improves the recognition rate. Next section will discuss on the contributions and limitations from this research with several possible future improvements.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Introduction

This chapter sums up the whole implementation of this research. It summarises contributions of this study associated to its body of knowledge. Next, limitations encountered during the study are described. Finally, the last section recommends the possible future work in this research direction.

6.2 Research Conclusion

Synthesising facial expression is challenging especially removing expression or neutralising expression. This research aimed to explore the kernel potential in synthesising neutral facial expression to improve face recognition. The modified Kernel Active Shape Model (mKASM) algorithm is proposed to realise this. In this research, the experiments are categorised into three tests. The first test is comparing performance of kernel PCA versus PCA based face recognition. From the results, it can be inferred that kernel PCA outperformed PCA. The kernel parameter and the number of eigenvectors that will give the best recognition results also obtained from this experiment. The second test is synthesising neutral expressions using the proposed mKASM. The synthesised faces are rendered in 3D environment for visual comparison with the original facial expression as before synthesis and the ground truth. Most of the synthesised faces are successfully neutralised. The resulting synthesised faces are stored to be used for final test. The last test is to evaluate the synthesised neutral faces using kernel PCA based face recognition and the results shown better recognition as compared to feeding the expressional face directly into the system.

In conclusion, this study introduced a nonlinear method to synthesise neutral expression and has shown improvement in face recognition. Kernel is a very complex method compared to PCA and require more processing time. Yet, with this complexity, the proposed approach that is embedded into the face recognition system has shown that it does robust enough to achieve a high recognition rate. In addition, mKASM able reconstruct an incomplete face data into a more recognisable subject portraying neutral expression. Therefore, the proposed method enhances the performance of the existing linear approach. The synthesising approach is different than other method because this research utilised the whole face surface data points as facial features instead of selecting feature points or facial landmark for modelling the shape variations.

Overall, objectives of this research have been fulfilled. The main objective is to improve face recognition by synthesising facial expression using kernel approach. The first objective is realised by using different facial expressions to be synthesised into neutral expression through the proposed kernel approach. The second objective was evaluating the newly synthesised neutral faces. Based on the evaluations, the synthesised neutral faces improve the face recognition. This research may provide a good start for novice researcher who want to explore kernel in facial image processing and computer vision.

6.3 Research Contributions

The contribution of this research is the proposal of nonlinear approach called modified Kernel Active Shape Model (mKASM) to synthesise neutral expression from faces with different kinds of facial expressions while keeping the identity of the person for the purpose of improving recognition rate. Other than that, a kernel-based PCA method for 3D facial recognition has been implemented to recognise faces and it produced better face recognition rate as compared to the linear PCA-based method. In some cases, PCA is found to give higher recognition rate when the expression intensity is lower. However, the recognition rate of a subject with PCA method was a little lower than the kernel PCA. Hence, the kernel-based PCA has shown better performance in extracting more information as compared to linear PCA. The proposed mKASM is particularly useful for reconstruction of facial shape model which is often missing in the training sample. This method can be used as well in different applications that need to retain the identity of the person and where the facial expression variant is the one would be changed.

6.4 Research Limitations

One of the limitations of this research is, there is only one set of face database with expression intensities. The database has help uncover the limitation of the proposed mKASM when extreme expression intensity case is present. Based on the results from additional experiment that has been done earlier as shown in Section 5.5, the shape parameter for varying expression intensity is not valid to be implemented in the proposed method. Hence, the shape varieties of the synthesised neutral faces could not be explored.

6.5 Future Work

The future work outlined are presented with the objective to improve and expand the research. The proposed method will be further investigated with the use of expression intensity adjustment to improve the poor synthesis results like surprise expression. Another possible future work would be integrating the proposed mKASM into an automated face recognition system. Lastly, one can look into optimising the time to compute the high dimensional processing to improve the efficiency of the overall performance.

REFERENCES

Abboud, B., Davoine, F., & Dang, M. (2004). Facial expression recognition and synthesis based on an appearance model. *Signal Processing: Image Communication, 19*(8), 723–740. https://doi.org//10.1016/j.image.2004.05.009

Agianpuye, A. S., & Minoi, J. L. (2014). Synthesizing neutral facial expression on 3D faces using active shape models. In *2014 IEEE Region 10 Symposium* (pp. 600-605). IEEE. https://doi.org/10.1109/tenconspring.2014.6863105

Agianpuye, S. (2015). Synthesizing neutral facial expressions on 3D faces. Master's thesis, Universiti Malaysia Sarawak.

Amin, S. H., & Gillies, D. (2007). Analysis of 3D face reconstruction. In *14th International Conference on Image Analysis and Processing* (pp. 413-418). IEEE. https://doi.org/10.1109/iciap.2007.4362813

Arya, G. J., Kumar, K. A., & Rajasree, R. (2014). Synthesize of emotional facial expressions through manipulating facial parameters. In *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies* (pp. 911–916). IEEE. https://doi.org/10.1109/iccicct.2014.6993088

Barbosa, L. A., Dahia, G., & Segundo, M. P. (2019). Expression removal in 3D faces for recognition purposes. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)* (pp. 753-758). IEEE. https://doi.org/10.1109/BRACIS.2019.00135

Bishop, C. M. (2006). *Pattern recognition and machine learning* (1st ed.). Springer.

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 187-194). ACM. https://dl.acm.org/doi/pdf/10.1145/311535.311556

Blanz, V. (2006). Face recognition based on a 3D morphable model. In *7th International Conference on Automatic Face and Gesture Recognition* (pp. 617-624). IEEE. https://doi.ieeecomputersociety.org/10.1109/FGR.2006.42

Bottino, A., De Simone, M., Laurentini, A., & Sforza, C. (2012). A new 3-D tool for planning plastic surgery. *IEEE transactions on biomedical engineering*, *59*(12), 3439-3449. https://doi.org/10.1109/TBME.2012.2217496

Chang, J., Zheng, Y., & Wang, Z. (2007). Facial expression analysis and synthesis: a bilinear approach. In *2007 International Conference on Information Acquisition* (pp. 457-464). IEEE. https://doi.org/10.1109/ICIA.2007.4295777

Chen, X., Qing, L., He, X., Su, J., & Peng, Y. (2018). From eyes to face synthesis: A new approach for human-centered smart surveillance. *IEEE Access*, *6*, 14567-14575. https://doi.org/10.1109/ACCESS.2018.2803787

Chen, Y., Bai, R., & Hua, C. (2014). Personalized face neutralization based on subspace bilinear regression. *IET Computer Vision*, *8*(4), 329-337. https://doi.org/10.1049/iet-cvi.2013.0212

Cootes, T., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models - their training and application. *Computer Vision and Image Understanding*, *61*(1), 38–59. http://dx.doi.org/10.1006%2Fcviu.1995.1004

Cui, W., Chen, S., Yu, T., & Ren, L. (2012). Feature extraction of x-ray chest image based on kpca. In *Proceedings of 2012 2nd International Conference on Computer Science and Network Technology, ICCSNT 2012* (pp. 1263–1266). https://doi.org/10.1109/iccsnt.2012.6526153

Eguizabal, A., Schreier, P. J., & Ramírez, D. (2018). Model-order selection in statistical shape models. In *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing* (pp. 1-6). IEEE. https://doi.org/10.1109/MLSP.2018.8516941

Ekman, P. & Friesen, W. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press.

Elkhadir, Z., Chougdali, K., & Benattou, M. (2016). Intrusion detection system using pca and kernel pca methods. In *Proceedings of the Mediterranean Conference on Information & Communication Technologies 2015* (pp. 489-497). Springer. https://doi.org/10.1007/978-3-319-30298-0_50

García-González, A., Huerta, A., Zlotnik, S., & Díez, P. (2020). *A kernel Principal Component Analysis (kPCA) digest with a new backward mapping (pre-image reconstruction) strategy*. arXiv preprint. https://arxiv.org/abs/2001.01958v1

Jolliffe, I.T. (1986). Principal component analysis and factor analysis. In *Principal Component Analysis*. Springer Series in Statistics. https://doi.org/10.1007/978-1-4757-1904-8_8

Kirschner, M., Becker, M., & Wesarg, S. (2011). 3D active shape model segmentation with nonlinear shape priors. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 6892, pp. 492–499). Springer. https://doi.org/10.1007/978-3-642-23629-7_60

Kouzani, A. Z. (1999). Facial expression synthesis. In *Proceedings 1999 International Conference on Image Processing, Cat. 99CH36348* (Vol. 1, pp. 643-647). IEEE. https://doi.org/10.1109/ICIP.1999.821713

Lee, S., Wolberg, G., & Shin, S. Y. (1997). Scattered data interpolation with multilevel b-splines. *IEEE Transactions on Visualisation and Computer Graphics*, *3*(3), 228-244. https://doi.org/10.1109/2945.620490

Li, L., Liu, S., Peng, Y., & Sun, Z. (2016). Overview of principal component analysis algorithm. *Optik - International Journal for Light and Electron Optics*, *9*(127), 3935–3944. https://doi.org/10.1016/j.ijleo.2016.01.033

Lv, C., Wu, Z., Wang, X., & Zhou, M. (2019). 3D facial expression modelling based on facial landmarks in single image. *Neurocomputing*, *355*, 155-167. https://doi.org/10.1016/j.neucom.2019.04.050

Liang, H., Liang, R., Song, M. and He, X. (2015). Coupled dictionary learning for the detail-enhanced synthesis of 3D facial expressions. *IEEE Transactions on Cybernetics*, *46*(4), 890-901. https://doi.org/10.1109/TCYB.2015.2417211

Liang, H., Song, M., Xie, L., & Liang, R. (2013). Personalized 3D facial expression synthesis based on landmark constraint. In *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference* (pp. 1-6). IEEE. https://doi.org/10.1109/APSIPA.2013.6694270

Liu, X., Xia, S., Fan, Y., & Wang, Z. (2011). Exploring non-linear relationship of blend shape facial animation. *Computer Graphics Forum*, *30*(6), 1655-1666. Blackwell Publishing Ltd. https://doi.org/10.1111/j.1467-8659.2011.01852.x

Malatesta, L., Raouzaiou, A., Karpouzis, K., & Kollias, S. (2009). MPEG-4 facial expression synthesis. *Personal and Ubiquitous Computing*, *13*(1), 77-83. https://doi.org/10.1007/s00779-007-0164-1

Matthews, I., Xiao, J., & Baker, S. (2006). On the dimensionality of deformable face models. *Technical Report CMU-RI-TR-06-12*.

Minoi, J. L. (2009). Geometric expression invariant 3D face recognition using statistical discriminant models. PhD Thesis, Imperial College London.

Minoi, J. L., Thomaz, C. E., & Gillies, D. (2011). Synthesizing 3D face shapes using tensor-based multivariate statistical discriminant methods. In *International Conference on Informatics Engineering and Information Science* (pp. 413-426). Springer. https://doi.org/10.1007/978-3-642-25483-3_34

Nair, S. (2019, April 26). AirAsia uses facial recognition for self-check-in and boarding at Senai Airport. *The Star Online*. https://www.thestar.com.my/tech/tech-news/2019/04/26/register-your-face-on-airasia-app-for-a-seamless-experience-at-senai-airport

Pal, M., Ghosh S., & Sarkar, R. (2020). Modification of existing face images based on textual description through local geometrical transformation. In J. K. Mandal & D. Bhattacharya (Eds.), *Emerging Technology in Modelling and Graphics* (pp. 159-170). Springer. https://doi.org/10.1007/978-981-13-7403-6_16

Pan, G., Han, S., Wu, Z., & Zhang, Y. (2010). Removal of 3D facial expressions: A learning-based approach. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 2614-2621). IEEE. https://doi.org/10.1109/CVPR.2010.5539974

Papatheodorou, T. (2006). 3D face recognition using rigid and non-rigid registration. PhD Thesis, Imperial College London.

Park, J. M., Choi, H. C., & Oh, S. Y. (2010). Non-rigid 3D face shape reconstruction using a genetic algorithm. In *IEEE Congress on Evolutionary Computation* (pp. 1-6). IEEE. https://doi.org/10.1109/CEC.2010.5586177

Patel, N., & Zaveri, M. (2013). 3D facial model reconstruction, expressions synthesis and animation using single frontal face image. *Signal, Image and Video Processing*, *7*(5), 889-897. https://doi.org/10.1007/s11760-011-0278-9

Pighin, F., Hecker, J., Lischinski, D., Szeliski, R., & Salesin, D. H. (2006). Synthesizing realistic facial expressions from photographs. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference* (pp. 19-28). ACM. https://doi.org/10.1145/1185657.1185859

Riaz, S., Ali, Z., Park, U., Choi, J., Masi, I., & Natarajan, P. (2019). Age-invariant face recognition using gender specific 3D aging modelling. *Multimedia Tools and Applications*, *78*(17), 25163-25183. https://doi.org/10.1007/s11042-019-7694-1

Romdhani, S., Gong, S., & Psarrou, A. (1999). A multi-view nonlinear active shape model using kernel pca. In *Proceedings of the British Machine Vision Conference 1999* (pp. 483-492). BMVA. https://doi.org/10.5244/c.13.48

Schölkopf, B., Smola, A., & Müller, K. R. (1997). Kernel principal component analysis. In *International Conference on Artificial Neural Networks* (pp. 583-588). Springer. https://doi.org/10.1007/BFb0020217

Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, *10*(5), 1299-1319. https://doi.org/10.1162/089976698300017467

Schroeder, W. J., Martin, K. M., & Lorensen, W. E. (2006). *The visualization toolkit* (4th ed.). Kitware.

Song, L., Lu, Z., He, R., Sun, Z., & Tan, T. (2018). Geometry guided adversarial facial expression synthesis. In *Proceedings of the 26th ACM International Conference on Multimedia* (pp. 627-635). https://dl.acm.org/doi/pdf/10.1145/3240508.3240612

Thies, J., Zollhöfer, M., & Nießner, M. (2019). Deferred neural rendering: Image synthesis using neural textures. *ACM Transactions on Graphics, 38*(4), 1-12. https://doi.org/10.1145/3306346.3323035

Ueda, J., & Okajima, K. (2019). Face morphing using average face for subtle expression recognition. In *2019 11th International Symposium on Image and Signal Processing and Analysis* (pp. 187-192). IEEE. https://doi.org/10.1109/ISPA.2019.8868931

Vezzetti, E., & Marcolin, F. (2012). 3D human face description: Landmarks measures and geometrical features. *Image and Vision Computing, 30*(10), 698–712. https://doi.org/10.1016/j.imavis.2012.02.007

Wang, D., Lu, H., & Yang, M. H. (2015). Kernel collaborative face recognition. *Pattern Recognition*, *48*(10), 3025-3037. https://doi.org/10.1016/j.patcog.2015.01.012

Wang, Y., Liu, Z., & Guo, B. (2011). Face synthesis. In *Handbook of face recognition* (pp. 521-547). Springer. https://doi.org/10.1007/978-0-85729-932-1_20

Wang, S., Gu, X. D., & Qin, H. (2008). Automatic non-rigid registration of 3D dynamic data for facial expression synthesis and transfer. In *2008 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-8). IEEE. https://doi.org/10.1109/CVPR.2008.4587791

Wang, D., & Tanaka, T. (2019). Kernel principal component analysis allowing sparse representation and sample selection. *ECTI Transactions on Computer and Information Technology*, *13*(1), 9-20. https://doi.org/10.37936/ecti-cit.2019131.187506

Wang, X., & Yang, R. (2010). Learning 3D shape from a single facial image via non-linear manifold embedding and alignment. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 414-421). IEEE. https://doi.org/10.1109/CVPR.2010.5540185

Yambor, W. S., Draper, B. A., & Beveridge, J. R. (2002). Analyzing pca-based face recognition algorithms: Eigenvector selection and distance measures. *Series in Machine Perception and Artificial Intelligence*, *50*, 39–60. https://doi.org/10.1142/9789812777423_0003

Yin, L., Wei, X., Sun, Y., Wang, J., & Rosato, M. J. (2006). A 3D facial expression database for facial behavior research. In *7th International Conference on Automatic Face and Gesture Recognition* (pp. 211-216). IEEE. https://doi.org/10.1109/FGR.2006.6

Zhao, B., Gao, L., Liao, W., & Zhang, B. (2017). A new kernel method for hyperspectral image feature extraction. *Geo-spatial Information Science*, *20*(4), 309-318. https://doi.org/10.1080/10095020.2017.1403088

APPENDICES

Appendix A: Kernel PCA Algorithm

Input: sample $S = [x_1, \dots, x_n]^T \in R^n$,

Output: top Φ of principal components

(1) Compute kernel matrix K: $K_{ij} = K(\mathrm{x}_i, \mathrm{x}_j)$ using Gaussian kernel, $K(x, y) = \exp\left(-\frac{\|X-Y\|^2}{2\sigma^2}\right)$

(2) Centre K: $\widetilde{K}_{ij} = K - \frac{1}{n}K - K\frac{1}{n} + \frac{1}{n}K\frac{1}{n}$

(3) Solve eigenvalue problem of $\lambda w^{\Phi} = C^{\Phi} w^{\Phi}$: $n\lambda a = \widetilde{K}a$,

where a denotes a column vector with all entries of the projected samples $\Phi(\mathrm{x})$,

which in turn completely determines the eigenvectors, $w^{\Phi} = \sum_{i=1}^{n} a_i \Phi(x_i)$.

(4) w^{Φ} , is the nonlinear principal components.

Appendix B: PCA Algorithm

Given N D-dimensional vectors (Notice usually $N < D$),

Input: sample $X_n = [x_1, \ldots, x_n]^T \in R^n$,

Output: top P of principal components

(1) Compute the mean vector μ (D-by-1 vector): $\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$

(2) Subtract each X_i by mean and get: $\Psi_i = X_i - \mu$

(3) Form the covariance matrix C of all the Ψ_i: $C = \frac{1}{N}\sum_{i=1}^{N} \Psi_i \; \Psi_i^T$

(4) Solve eigenvalue problem: $\lambda w = Cw$

(5) Preserve the P largest eigenvectors, w^P where λ is the P^{th} eigenvalue of the selected eigenvalues of Ψ, $\lambda_P \geq \lambda_{P+1}$, and $w_P w_P^T = 1$.

5) w^P, is the linear principal components.

Appendix C: List of Publications

1. **Peter, M.,** Minoi, J. L., & Hipiny, I. H. M. (2019). 3D face recognition using kernel-based pca approach. In *Computational Science and Technology* (pp. 77-86). Springer. https://doi.org/10.1007/978-981-13-2622-6_8
2. **Peter, M.**, Minoi, J. L., & Rahman, S. A. (2020). Neutral Expression synthesis using kernel active shape model. *Indonesian Journal of Electrical Engineering and Computer Science, 20*(1), 150-157. http://doi.org/10.11591/ijeecs.v20.i1.pp150-157