3D FACIAL EXPRESSION CLASSIFICATION USING A STATISTICAL MODEL OF SURFACE NORMALS AND A MODULAR APPROACH

HAMIMAH UJIR

A thesis submitted to
University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Electronic, Electrical & Computer Engineering
University of Birmingham
August 2012

ABSTRACT

Following the success in 3D face recognition, the face processing community is now trying to establish good 3D facial expression recognition. Facial expressions provide the cues of communication in which we can interpret the mood, meaning and emotions at the same time. With current advanced 3D scanners technology, direct anthropometric measurements (i.e. the comparative study of sizes and proportions of the human body) are easily obtainable and it offers 3D geometrical data suitable for 3D face processing studies. Instead of using the raw 3D facial points, we extracted its derivative which gives us 3D facial surface normals. We constructed a statistical model for variations in facial shape due to changes in six basic expressions using 3D facial surface normals as the feature vectors. In particular, we are interested in how such facial expression variations manifest themselves in terms of changes in the field of 3D facial surface normals. We employed a modular approach where a module contains the facial features of a distinct facial region. The decomposition of a face into several modules promotes the learning of a facial local structure and therefore the most discriminative variation of the facial features in each module is emphasised. We decomposed a face into six modules and the expression classification for each module is carried out independently. We constructed a Weighted Voting Scheme (WVS) to infer the emotion underlying a collection of modules using a weight that is determined using the AdaBoost learning algorithm. Using our approach, using 3D facial surface normal as the feature vector of WVS yields a better performance than 3D facial points and 3D distance measurements in facial expression classification using both WVS and Majority Voting Scheme (MVS). The attained results suggest surface normals do indeed produce a comparable result particularly for six basic facial expressions with no intensity information.

DEDICATION

To my dear father

Ujir Sadi

1953 – 2012

ACKNOWLEDGEMENT

I would like to express my gratitude to my supervisor, Dr Michael Spann whose guidance and support was paramount to this work. He patiently provided the constructive advice and unflagging encouragement for me. I would also like to acknowledge Malaysian Ministry of Higher Education and Universiti Malaysia Sarawak (UNIMAS) for the financial support provided throughout this journey.

My biggest thank you to my husband, Irwandi Hipni and my son,Yahya Jerrel Hipni. Their love and patience are my driving force. To my parents and mother – in – law, Ujir Sadi, Rohana Abd Majid and Julia Dolhassan, thank you for being there when I needed you most. For my siblings, in-laws, nephews and nieces thank you for the unconditional support and devotion.

My friends in Malaysia, United Kingdom, Japan, Australia and other parts of the world were the sources of laughter, joy and support. This thesis would never have been completed without the encouragement of my best friend, Affiza Abd Rahman.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

3DMM	3D Morphable Model
AdaBoost	Adaptive Boosting
AdaBoost.MH	Adaptive Boosting with Multi-Class Humming Lost
AU	Action Unit
AFEA	Automatic Facial Expression Analysis
AAM	Active Appearance Model
ASM	Active Shape Model
BFSC	Basic Facial Shape Component
DCT	Discrete Cosine Transform
DVS	Democratic Voting Scheme
ESC	Expressional Shape Component
FAPs	Facial Animation Parameters
FAPUs	Facial Animation Parameter Units
FER	Facial Expression Recognition
FP	False Positive
HGM	Hierarchical Graph Matching
ICP	Iterative Closest Point
LDA	Linear Discriminant Analysis
LCS	Longest Common Subsequence
MDS	Multi-Dimensional Scaling
MVS	Majority Voting Scheme
OVA	One-versus-All
OVO	One-versus-One
PCA	Principal Component Analysis
PGA	Principal Geodesic Analysis
PGSFS	Principal Geodesic Shape-From-Shading
ROC	Receiver Operating Characteristics

SFAM	Statistical Facial Feature Model
SIFT	Scale-Invariant Feature Transform
SVD	Singular Value Decomposition
SVM	Support Vector Machine
WVS	Weighted Voting Scheme

LIST OF PUBLICATIONS

Conference Paper

- Ujir, H, and Spann, M. 2011.Facial Expression Recognition using MPEG-4 FAP-based 3D MMM. *In the Proceeding of the Third ECCOMAS Thematic Conference on Computational Vision and Medical Image* Processing (VIPIMAGE), Algarve, Portugal.

Book Chapter

- Ujir, H, and Spann, M. 2013. Facial Expression Recognition using MPEG-4 FAP-based 3D MMM. In: Natal, R and Tavares, J.M., ed. *Lecture Notes in Computational Vision and Biomechanics: Topics in Medical Image Processing and Computational Vision, Springer-Verlag*, pp. 33-47.

Journal

- Ujir, H, and Spann, M. 3D Facial Expression Classification using a Statistical Model of Surface Normals and a Modular Approach. *Submitted to Journal of Pattern Analysis and Applications (Springer-Verlag) in December 2012.*

CHAPTER 1

INTRODUCTION

Face processing studies have been carried out over the past few decades for different purposes which began with face recognition and now facial expression classification studies are also emerging. Classification of facial expressions is a challenging problem as the face is capable of complex motions and the range of possible expressions is wide (Sandbach[1] al., 2012). The difference between posed and spontaneous examples along with the wide variations seen between subjects when expressing emotions are the main obstacles in this area. Posed expressions are the expressions a person will produce when she/he was asked to do so while spontaneous examples are acquired spontaneously (Bettadapura, 2012).

Facial expression recognition is an emerging research area spanning several disciplines such as pattern recognition, computer vision and image processing. It brings benefits in human centred multimodal human-computer interaction (HCI) whereas the user’s affective states motivate human action and enrich the meaning of human communication. In HCI, affective computing employs human emotion to build more flexible and natural multimodal (Jaimes et al., 2007). The automatic human affect recognition system will change the ways we interact with computer systems. For example in intelligent automobile system with a fatigue detector, the vigilance of the driver could be monitor and apply appropriate action to avoid the accidents (Ji et al., 2006). With an efficient automated face expression classification, perhaps it will be an aid to the affect-related research community to carry out clinical psychology, psychiatry, and neurosciences research. Such systems could improve the quality of the affect-related research by improving the reliability of measurements and speeding up the currently tedious task of processing data on human affective behaviour (Ekman et al., 2005). In addition, facial expressions are the key component in machine understanding of sign language. Facial expressions change the meaning of adjectives or convey adverbial information as facial expressions are timed to occur with hand movements for signs during specific parts of a sentence (Huenerfauth et al., 2011). Furthermore, American Sign Language signers also use facial expressions to convey emotional subtext. Therefore, an automatic facial expression classification is crucial part in machine understanding of sign language.

Human interactions consist of speech and gestures and humans are more aware of the facial expression of the people they are interacting with, rather than any other non-verbal type of communications such body gestures, postures and eye contacts (DataFace, 2003; Rose-Hulman, 2010). Facial expressions provide the cues of communication in which we can interpret the mood, meaning and emotions at the same time. Therefore, it is important to have accurate and robust expression classification to harness the information available in human expression.

Many approaches for 2D facial expression classification have been proposed in the literature. Unfortunately, most of them suffer from limitations of 2D image acquisition such as illumination changes and head pose variations as well as changes in facial appearance like make – up, glasses etc. The illumination problem is basically the variability of an object's appearance from one image to the next with slight changes in lighting conditions and viewpoints (Vishwakarma et al., 2007). Due to the illumination limitation, it is difficult to handle subtle facial behaviour in 2D modality. In most cases, when 2D facial expression images are employed, a consistent facial pose is used to ensure a good classification performance is achieved.

Facial expression studies have evolved from 2D to 3D modality which has much more to offer. With the advances in 3D scanners, the acquisition of 3D facial structure and motion is now a feasible task. 3D facial expression data remove the problems of illumination and pose that are inherent to 2D modality. Moreover, the expression dynamics which offer out-of-plane movement that cannot be captured with 2D are available in 3D facial data. 3D scanners also generate the 3D point clouds. Therefore, direct facial anthropometric measurements can be carried out and produce 3D facial landmarks as the output. Facial anthropometric refers to the comparative study of sizes and proportions of the human face which include the discriminatory structural characteristics of the human face (Gupta et al., 2010). These facial landmarks are the soft-tissue landmarks which lie on the skin and can be identified on the 3D point clouds. Moreover, these facial landmarks are the points where all faces join and that have a particular biological meaning (Vezzetti et al., 2012). For instance, the facial landmarks with a particular biological meaning such as the nose tip, inner-corner eyes and etc.

Following the success in 3D face recognition, the face processing community is now trying to establish good 3D facial expression classification. 3D geometry contains ample information about human facial expression (Tang et al., 2008). With this in mind, 3D facial expression classification is believed to be the next promising technology.

There are several studies using 3D data in the face processing area. For instance, according to Savran et al., (2012), 3D data can maintain a high performance for lower face action unit (AU) compared to 2D data and it also offers 3D facial surface data. They compared 3D modality *vis-a-vis* 2D modality for AU classification where the 3D data is converted to 2D images of surface curvature to ensure a fair ground of comparison is carried out. For 2D modality, 2D image intensity is extracted.

1.1 Motivation and Contribution

Ceolin (2012) aims to fit the statistical models of shape to 2D facial images and recover the information concerning 3D shape from these images. Ceolin used a 2.5D shape representation based on facial surface normals which is acquired from 2D intensity images using Shape from Shading (SFS). SFS is known to recover surface shape from variations in brightness and it is more natural as it captures features of human vision system. The 2.5D surface normals (or known as facial needle maps) are then used to classify facial expression and gender.

Facial action units (AUs) represent the muscular activity that produces facial appearance changes (Ekman and Friesen, 1978). Sandbach[3] et al., (2012) proposed a new feature descriptor called local normal binary patterns (LNBPs) which is exploited for detection of facial action units (AUs).LNBPs employ the normals of the triangular polygons that form the 3D mesh to encode the shape of the mesh at each point. Surface normal feature is equivalent to encoding the gradient of a 2D intensity image, thus it provides a richer source of information about the shape of the facial mesh than the depth alone. Initially, a circular neighbourhood around each point, specified by a radius r and P points regularly spaced around the circle.The unit normal $\mathbf{n}_p$ at each point v_p in the neighbourhood is found, along with that at the central point $\mathbf{n}_c$, through $x - y$ interpolation of the given points in the mesh. From here, two descriptors are formed: (1) $LNBP_{OA}$, which calculates the scalar of two normals and (2) $LNBP_{TA}$, which calculates the difference of two angles of the normals, the azimuth and the elevation. Feature vectors are then formed for each of the descriptors through the use of histogram. These histograms are concatenated into one large feature vectors.

This work is motivated by the geometrical information such as 3D facial points is easily provided by the 3D scanners. We extracted 3D facial surface normal from the 3D facial points. Surface normals are considered to be more accurate in describing facial surface changes compared to using facial points due to the fact that a surface normal depends on a facial point as well as its neighbouring facial points. Therefore, the face deformations which happen when facial expression occurs can be observed closely. In particular, we are interested in how such facial expression variations manifest themselves in terms of changes in the field of 3D facial surface normals.

The difference between our work and Ceolin's (2012) is that the surface normals is acquired from 2D intensity images using SFS. In our work, the surface normals is calculated using 3D facial points. Our approach differs from Sandbach et al., (2012) in the sense of calculating the normals method. In their work, the unit normal $\mathbf{n}_p$ at each point $v_p$which is regularly spaced at a r radius and P points around the circle is calculated. While in our work, the surface normal of a point is calculated by taking into account the surface normal of the points that are connected to that particular point. This means that no exact amount of points or the size of area is considered. Furthermore, Sandbach used the histograms of the surface normals to form the feature vector, whereas we used the surface normals directly as the descriptor in a statistical model.

Figure1.1 Levels of intensity for Happy expression taken from Frowd et al. (2009)

Each of the basic facial expressions has levels of intensity which depend on the level of intensity of each facial feature. Intensity level of a facial expression is important as it will lead to a false impression of people's emotion if misinterpreted. For example, the smiling face with low intensity can be easily misinterpreted as a neutral facial expression (Beszédeš et al., 2007). Figure 1.1 shows intensity level for Happy expression where we can see each facial feature deforms rather distinctively at each intensity level. The decomposition of a face into several modules promotes the learning of a facial local structure and therefore the most discriminative variation of the facial features in each module is emphasised. In particular, we would like to see how the modular approach improves the classification of 3D facial expression.

Our contribution in this thesis comes in a package of using the established 3D database to classify six basic facial expressions with no intensity information together with a modular approach. Initially, a face is decomposed into several modules. The 3D facial surface normal for each module are computed using a very basic computation which involves 3D facial points. These surface normals are then used in a statistical model to capture the variation of the shape due to facial expression changes in each module. The statistical model generates the shape parameters which are used as the feature vector to classify the facial expression in two different classifiers. The expression classification for each module is carried out independently and therefore each module is expected to have a different classification result from the other modules. In order to infer the emotion underlying a collection of modules, a Weighted Voting Scheme (WVS) is constructed. In WVS, each module carries its own weight which indicates the importance of that particular module to classify the facial expressions. Facial expression with the highest accumulated weight is considered as the facial expression shown by the 3D probe using WVS. The weight is determined using the Adaptive Boosting (AdaBoost) learning algorithm.

1.2 Thesis Outline

The remainder of this thesis is organized into the following chapters.

Chapter 2 provides a thorough review of the literature. It starts with 3D face databases that are publicly available. The next section discusses3D facial expression in general which covers two approaches, one for emotion classification and the other one for AU detection. For both approaches, we discuss the 3D facial features used in the subsequent section. Only 3D facial static data is described as that is used in our study. Then, the statistical modelling used in face processing studies is discussed followed by modular-based works.

In chapter 3, we present the pre-processing and statistical modelling to be used in this work. First we explain the pre-processing steps in our approach, which begins with data extraction and 3D facial points alignment. Principal Component Analysis (PCA) as the statistical shape modelling is mathematically described in the next section.

In chapter 4 we explain the extraction of 3D facial surface normals which is computed straightforwardly from 3D facial points. This chapter also includes an explanation of the different classification approaches namely nearest neighbour classifier and Support Vector Machines (SVM) which are used in this work. The results of 3D facial expression using 3D facial surface normals as the feature vectors using both classifiers are discussed next. For the purpose of evaluation, 3D facial points and 3D distance measurements are also used as the feature vectors in the experiments.

In chapter 5 we discuss the decomposition of a face into several modules. Each module has a collection of facial features associated with Facial Animation Parameters which is the muscular action relevant to AUs. We explain the priority rank of each module and how the weight of each module is computed using AdaBoost. The integration of modules is dealt with using the Weighted Voting Scheme (WVS) approach which is also described. The results of modular 3D facial surface normals and WVS are discussed.

In chapter 6, based on the experimental results found in chapter 4 and 5, several key tables are produced, analysed and discussed.

Finally, Chapter 7 offers some concluding remarks where a summary of the contributions, the weaknesses of our approach as well as future is presented.

CHAPTER 2

LITERATURE REVIEW

In this thesis, we use statistical modelling of 3D surface normals and use a modular approach where a face is decomposed into several modules. Expression classification is performed on each modular independently and the results of the modules are then pooled to infer the underlying expression. Works focusing on finding the best features to represent the facial deformation in facial expression classification are not as extensive as in the face classification area (Vezzetti et al., 2012). This research involves three main themes: (i) 3D facial features, (ii) Statistical model and (iii) Modular-based work. In this chapter, we provide a thorough review of the literature relevant to these topics.

The remainder of this chapter is organized as follows: In section 2.1 we describe 3D face databases that are publicly available. In section 2.2, the basic framework of facial expression classification and two significant goals in 3D facial expression, basic emotion classification and AUs detection are discussed. The 3D facial features used in this area are reviewed in section 2.3 in which we deliberately arrange according to the facial expression goals. Statistical classification is discussed in section 2.4 and modular-based approaches are discussed in section 2.5. The next section discussed about the classifiers used in this area of study. Concluding remarks can be found in section 2.7.

2.1 3D Face Expression Databases

A number of databases have been created in the past two decades for the purpose of face modelling and recognition. This section only discusses about existing static 3D face databases.

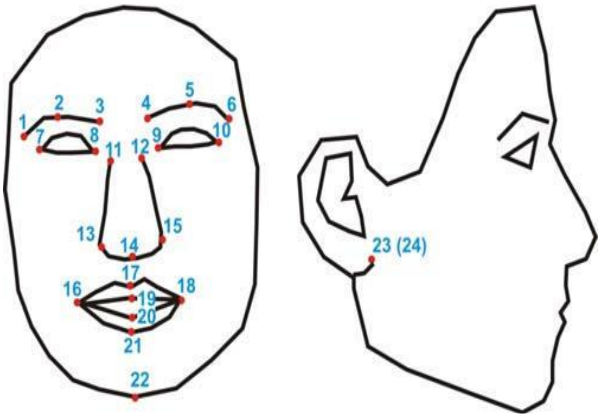
Databases such as GavabDB(Moreno et al., 2004), Benedikt et al. (2010), Blanz et al., (1999), ND-2006 (Faltemier et al., 2007), CASIA (Zhong et al., 2007), York 3D (Heseltine et al., 2008), Texas (Gupta et al., 2010) and the extension version of FRGC dataset, known as FRGC v2 (Philips et al., 2005)are rarely used in facial expression classification studies due to incomplete basic expression set and irregular distribution of the expression variations (Fang et al., 2011).Though databases such as ICT-3DRFE (Stratou, et al., 2011) and Tsalakanidou (Tsalakanidou et al., 2010) offer six basic expressions, the facial landmarks are not provided by the developer. Details of all existing static 3D face databases are summarised in Table 2.1.

Table 2.1 3D face databases containing 3D static expressions.

Name/Database	**Size**	**Content**	**Landmarks**	**Publicly Available?**
BU-3DFE	100 adults	6 basic expressions	83 facial landmarks	Y
Bosphorus	105 adults	6 basic expressions	24 facial landmarks	Y
ICT-4DRFE (Stratou et al., 2011)	23 adults	6 basic expressions, 2 neutral, 2 eyebrows, 4 eye gaze and 1 scrunched face	N/A	Y
Tsalakanidou et al. (2010)	52 adults	6 basic expressions and 11Aus	N/A	N
Benedikt et al. (2010)	94 adults	Smiles and word utterance	N/A	N
Blanz et al. (1999)	200 adults	Neutral faces	N/A	Y
ND-2006 (Faltemier et al., 2007)	888 adults	Neutral and 5 expressions: Happy, Disgust, Sad, Surprise, Random	N/A	Y
CASIA (Zhong et al., 2007)	123 adults	Neutral and 5 expressions: Smile, Laugh, Anger, Surprise, Eyes closed	N/A	Y
GavabDB (Moreno et al., 2004),	61 adults	3 expressions: Open/Closed Smiling and Random	N/A	Y
York 3D (Heseltine et al., 2008)	350 adults	Neutral and 4 expressions: Happy, Anger, Eyes Closed and Eyebrows raised	N/A	Y
Texas (Gupta et al., 2010)	105 adults	Neutral and smiling or talking with open/closed eyes	25 facial points	Y

Based on table 2.1, only two 3D facial expression static databases that are available publicly and provide at least six basic facial expressions and complete with 3D facial landmarks which is Bosphorus Database and BU-3DFE Database. The following sub-sections discuss these two databases.

2.1.1 Bosphorus Database



1. Outer left eye brow
2. Middle of the left eye brow
3. Inner left eye brow
4. Inner right eye brow
5. Middle of the right eye brow
6. Outer right eye brow
7. Outer left eye corner
8. Inner left eye corner
9. Inner right eye corner
10. Outer right eye corner
11. Nose saddle left
12. Nose saddle right
13. Left nose peak
14. Nose tip
15. Right nose peak
16. Left mouth corner
17. Upper lip outer middle
18. Right mouth corner
19. Upper lip inner middle
20. Lower lip inner middle
21. Lower lip outer middle
22. Chin middle
23. Left ear lobe
24. Right ear lobe

Figure 2.1 24 facial landmarks provided by the Bosphorus Database

A multi–attribute database developed by researchers from Bogazici University, Turkey called the Bosphorus database (Savran et al., 2008) was acquired using the Inspeck Mega Capturor II 3D which is a commercial structured-light based 3D digitizer device. With this device, the 3D face is captured by projecting one or more encoded light patterns onto the scene and the deformation on the objects' surfaces is measured to obtain the shape information (Sandbach et al^2., 2012). The weakness of this device is it only allows limited amount of movement due to restriction of the area that is simultaneously covered by the structured pattern and visible by the pattern. For the same reason, the acquired range images may also contain holes. In addition, the sources of lights are visible to the subjects and therefore, the spontaneous expression type of data is not available to be capture using this device.

The Bosphorus database is complete with 24 facial landmark points; provided that they are visible in the given scan (i.e., the right and left ear lobe cannot be seen from the frontal pose), refer to figure 2.1. These facial landmark points are manually labelled with its specific anatomic denotation by the developer, for instance landmark no 14 is denote as the nose tip. Each segmented 3D face consists of approximately 35, 000 points. On the other hand, the texture images are also provided with the resolution of 1600 x 1200 pixels.

There are 105 subjects (60 men and 45 women) with 53 different face scans per subject. The database provides a rich set of expressions, systematic variation of poses and different types of realistic occlusions. Each scan is intended to cover one pose and/or one expression type. Thirty-four facial expressions are composed of a wisely chosen subset of facial Action Units (AUs) of the Facial Action Coding System (FACS), as well as the six basic emotions, as shown in Table 2.2.

Table 2.2 Facial Expression with the Corresponding Facial Action Units (FAUs)

1) Lower FAUs	
Nose Wrinkler – AU9	Lip Puckerer – AU18
Upper Lip Raiser – AU10	Lip Stretcher – AU20
Lip Corner Puller – AU12	Lip Funneler – AU22
Left Lip Corner Puller – AU12L	Lip Tightener – AU23
Right Lip Corner Puller – AU12R	Lip Presser – AU24
Low Intensity Lip Corner Puller – AU12LW	Lips Part – AU25
Dimpler – AU14	Jaw Drop – AU26
Lip Corner Depressor – AU15	Mouth Stretch – AU27
Lower Lip Depressor –AU16	Lip Suck – AU28
Chin Raiser – AU17	Cheek Puff – AU34

2) Upper FAUs	**3) Some FAUs Combinations**
Inner Brow Raiser – AU1	Jaw Drop (AU26) + Low Intensity Lip Corner Puller (AU12LW)
Outer Brow Raiser – AU2	Lip Funneler (AU22) + Lips Part (AU25)
Brow Lowerer – AU4	Lip Corner Puller (AU12) + Lip Corner Depressor (AU15)
Eyes Closed – AU43	
Squint – AU44	

FACS provides descriptive power necessary to describe the details of facial expression (Tian et al., 2001). Action units (AUs) represent the muscular activity that produces facial appearance changes (Ekman et al., 1978). In general, there are 44 AUs but in the Bosphorus database, only a subset of these AUs are collected which consists of those

AUs that are easier to enact. The selected AUs were grouped into 20 lower face AUs, 5 upper face AUs and 3 AUs combinations.

2.1.2 BU-3DFE Database

Yin et al. (2006) from Binghamton University developed theBU-3DFE database and this 3D database contains 100 subjects (56% female, 44% male), ranging in age from 18 years to 70 years old, with a variety of ethnic/racial ancestries, including White, Black, East-Asian, Middle-east Asian, Indian, and Hispanic Latino. The data were captured using the 3DMD dynamic 3D stereo system which is a multi-view stereo type of acquisition. This family of acquisitions employ multiple cameras placed at various known viewpoints from the subjects.

This device records the same simultaneously with constant light sources and does not require flashing lights. Therefore, more natural expressions from the subjects can be recorded. The approach of using multiple cameras in data collection is the constraint as it increases the cost.

The models were created with the resolution in the range of 20,000 to 35,000 polygons which depends on the size of the subject’s face. Each of the six prototypical expressions includes four levels of intensity. A neutral expression is also available however without levels of intensity. There are 25 instant 3D expression models for each subject, resulting in a total of 2500 3D facial expression models in the database.

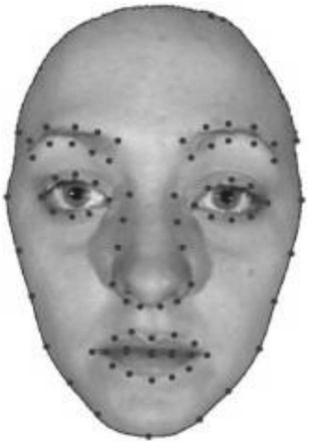


Figure 2.2 The 83 facial landmarks given in BU-3DFE database (Sandbach^2 et al., 2012)

Most of the existing works were evaluated on the BU-3DFE database mainly because of two reasons: (i) it was the first database that is publicly available and (ii) the 83 manually annotated dense landmarks provided with the release (Fang et al., 2011). Furthermore, all face models in this database are cropped from the original scans which greatly facilitate research in face processing studies. Table 2.3 below summarize a comparison between two static 3D face databases. The apparent difference between the two databases are the acquisition method, the level of facial expression intensity, the number of annotated facial landmarks and the extra information provided by the developers such as 3D facial action unit (AUs) data (for Bosphorus Database). The different method to acquire the 3D facial data for each database might influence the successfulness of the approach chosen. In this work, we used Bosphorus Database as we have been granted the access to the database but not to BU-3DFE Database.

Table 2.3 A comparison table between two prominent databases.

Bosphorus Database (Savran et al., 2008)	**BU-3DFE Database (Yin et al., 2006)**
Inspeck Mega Capturor II 3D *(Structured-light based)*	3DMD Dynamic 3D stereo system *(Multi-view Stereo based)*
6 basic expressions	6 basic expressions with four levels of intensity
1 neutral expression	1 neutral expression
24 manually annotated 3D facial landmarks	83 manually annotated 3D facial landmarks
34 Facial Action Units expressions	No Facial Action Units expressions

2.2 3D Facial Expression

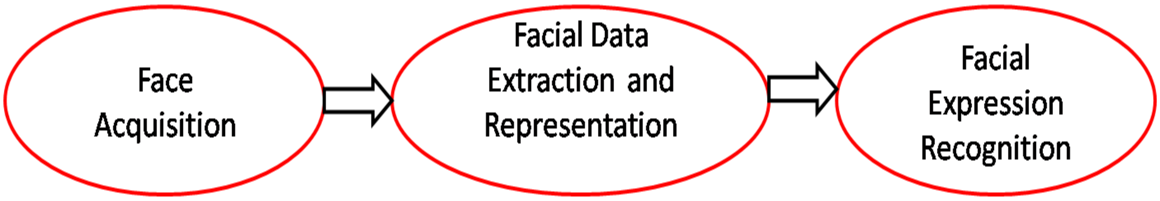

Face Acquisition
Facial Data Extraction and Representation
Facial Expression Recognition

Figure 2.3 Generic Automatic Facial Expression Analysis (AFEA) Framework (Fasel et al., 2003; Tian et al., 2005).

Analysing facial expression includes both a measurement of facial motion and the classification of the expression. General approaches to Automatic Facial Expression Analysis (AFEA) consist of three steps as described in figure 2.3. It begins with face acquisition which involves the detection of the face region. We omitted the first stage because it is not essential in our work.

Based on figure 2.1, the next stage is to extract and represent the facial deformation caused by facial expressions. In the 2D modality, there are two approaches for facial feature extraction in facial expression classification: geometric–based and appearance–based. Apparently, geometric–based methods present the shape and location of facial features such as nose, mouth, eyes and eyebrows. With appearance–based methods, image filters are applied to extract a feature vector. In addition to the geometry and appearance approaches, there are existing methods which used different approaches, for instance (i) using properties of 3D facial landmarks such as 3D facial landmarks distances, ratio distances and 3D curvature features (Soyel et al., 2008; Tang et al., 2008); (ii) morphable models (Mpiperis et al., 2008); (iii) combinations of 3D geometry and 2D texture (Zhao et al., 2010) (iv) mapping from 3D to 2D (Savran et al., 2008).

The last stage of AFEA systems is facial expression classification that includes two goals, basic emotions and facial action units. Both goals are discussed in the following sub-section.

2.2.1 Basic Emotions

From Ekman (1994), facial expressions are claimed to be constant across cultures and universal. However, in their cross culture studies, only six basic facial expressions are considered which are Anger, Disgust, Fear, Happy, Sad and Surprise, as shown in figure 2.2. This explains why most of the work in this area of study attempts to classify only the six basic expressions.

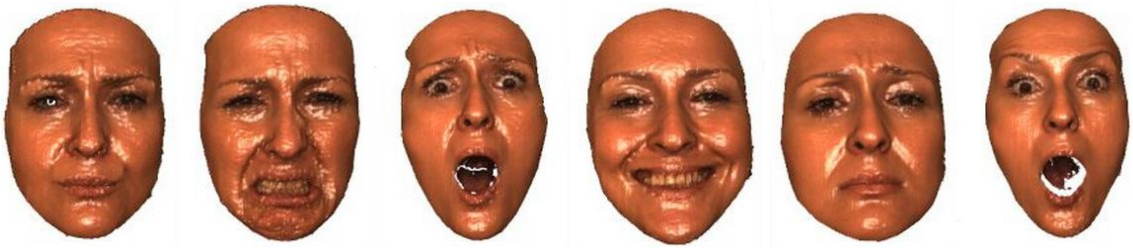


Figure 2.4 Emotion-specified facial expressions. From left: Anger. Disgust, Fear; Happy, Sad, Surprise (Savran et al., 2008)

2.2.2 Action Units (AUs)

The Facial Action Coding System (FACS) was introduced by Ekman and Friesen (1978). FACS is a method of measuring facial activity in terms of facial muscle movements. It consists of over 45 distinct AUs corresponding to a distinct muscle which are essentially facial phonemes that can be assembled to form facial expressions (Kapoor et al., 2003). Facial phonemes here are referring to a collection of facial features that are associated with the construction of any facial expression.

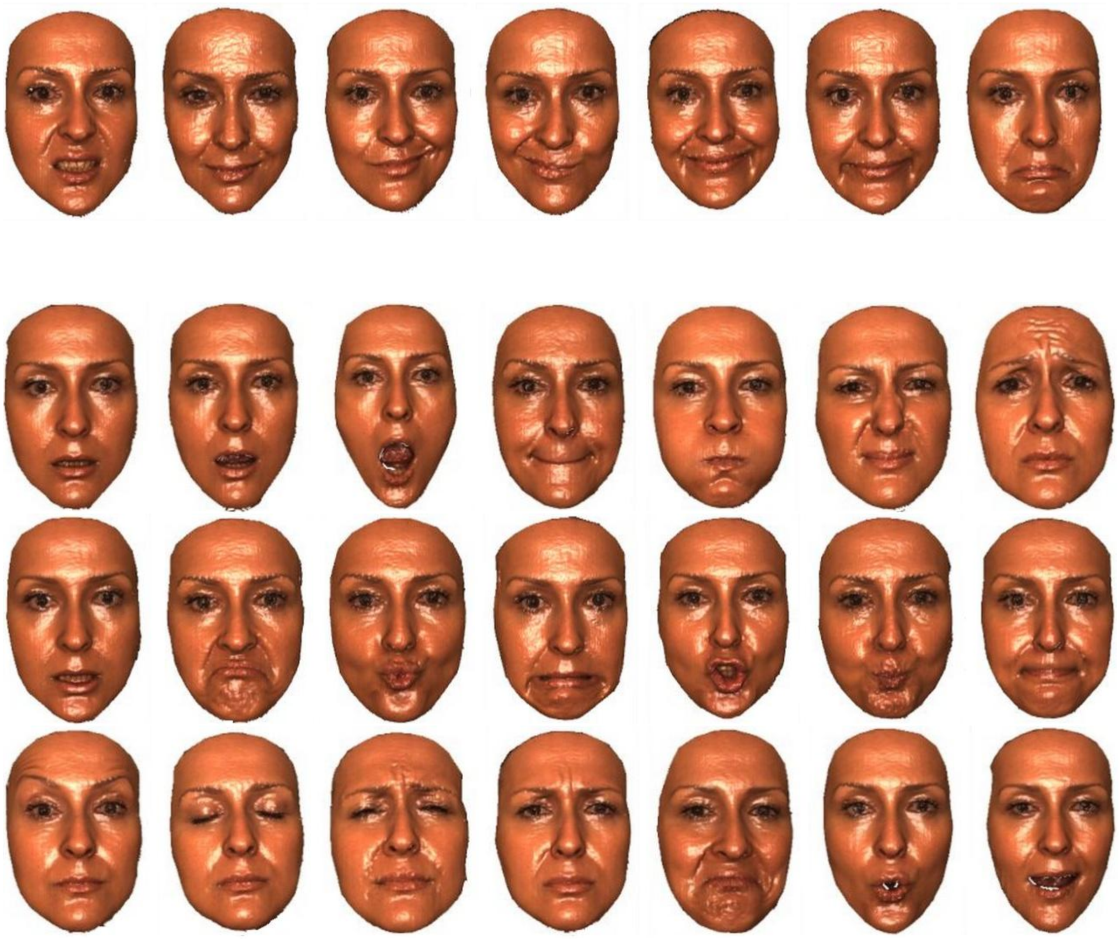


Figure 2.5 Example of AUs taken from Savran et al. (2008)

AUs detection is denoted as another solution to classifying various human emotions instead of only six basic expressions. Velusamy et al. (2011) stated that detecting AUs prior to emotion makes a classification system more suited to a culture independent interpretation which is in contradiction with Ekman’s claim where expressions are constant across cultures. In addition, these basic expressions occur relatively infrequently and emotions are displayed by more subtle changes in one or few discrete facial features such as raising the eyebrows in Surprise (Russell and Fernandez-Dols, 1997). AUs are more flexible in that thousands of anatomically possible facial expressions can be described by a small number of AUs and AU descriptors (Fang et al., 2011). Moreover, according to Velusamy et al. (2011), there are 7000 emotions in practice. However, Parrott (2001) claimed that there are 139 facial expressions that humans are capable of displaying. With this amount of facial expressions to be programmed, fully automated facial expression classification is indeed a long way from being perfected.

2.3 3D Facial Features

Facial features can be classified as being permanent or transient (Bettadapura, 2012). The permanent appearance of the face is formed by the shapes and placement of the bones of the skull, the cartilage,and the soft tissues, including the muscles, fat, and skin, of the face, which also might include a person's genetic background (e.g., race, ethnicity, and family membership), genetic diseases (e.g., Down's syndrome), and more fuzzy concepts such as personality, character, and temperament (DataFace, 2003).The consistent facial feature underlies our attribution of identity to a person and its characteristics also contribute to the relatively static expression of the face. Eyes, lips, eyebrows and cheeks are the permanent features. Facial lines, brow wrinkles and deepened furrows are the example of transient features that appear with changes in expression and disappear on a neutral face.

In the past, the researchers were trying to find the best 3D facial features to represent the salient features used in face recognition and to quantify the facial deformation caused by facial expressions. We focus on 3D facial features using 3D facial static data in this review and 3D facial features normally are the permanent features.

3D facial landmarks were extracted from the face by various researchers in many different ways. Facial landmarks on a face are the real feature while the geometric features that are extractable from the real feature are the measure features. Distances or angles used in any area of face processing studies are considered measures, rather than real features (Vezzetti et al., 2012). However, the nature of these reference points may be geometric, for instance, curvature and shape.

Vezzetti et al., (2012) in their studies reviewed several geometrical features to describe 3D human faces. Among the geometrical measures used for the purpose of 3D face recognition are Euclidean distance, geodesic distance, arc – length distance, ratio – of distances, curvature and shape, shape index and spin images, spin images, 3D SIFT features, depth information and texture information. However in face recognition, the computation of the Euclidean or geodesic distances between facial landmarks is a method widely used.

The 3D facial features studied in 3D facial expression classification is conducted by categorizing it according to the goal of the study, for basic emotions or AUs detection.

2.3.1 Basic Emotions

The first study of classifying facial expressions using 3D data from BU-3DFE was carried out by Wang et al., (2006). They extracted and labelled the primitive 3D surface features (i.e.: flat, peak, ridge, ravine, pit, concave hill, convex hill, convex saddle hill, slope hill, concave saddle hill, ridge saddle and ravine saddle) and derived their statistical distributions to represent the distinct prototypical facial expressions. The expression classification is based on the distribution of the above labels over the face. Thus, the same type of facial expression is expected to share a similar primitive label distribution. However, their partitioned regions do not contain mouth and eyes, which are significant regions in determining facial expression. Their algorithm involves manually labelled facial points in order to obtain more accurate region partitions. Furthermore, this technique requires extensive computation of curvature features which is challenging.

Soyel et al., (2007) used six characteristics distances between 3D facial landmarks to form a distance vector. They only used 11 manually labelled facial landmarks to extract the distances by utilizing facial symmetry. The distance vector is derived for every 3D model and is used to compare faces for facial expression classification. To achieve the person-independent requirement, they normalized the distance vector of an expressional face by the width of the face. They extended their work by introducing an automatic feature selection mechanism (Soyel et al., 2010). In their extended work, all 83 facial features available in BU-3DFE are used to find distances between points. The classification rate obtained using their approach was slightly higher compared to Wang et al. (2006), refer to Table 2.4.

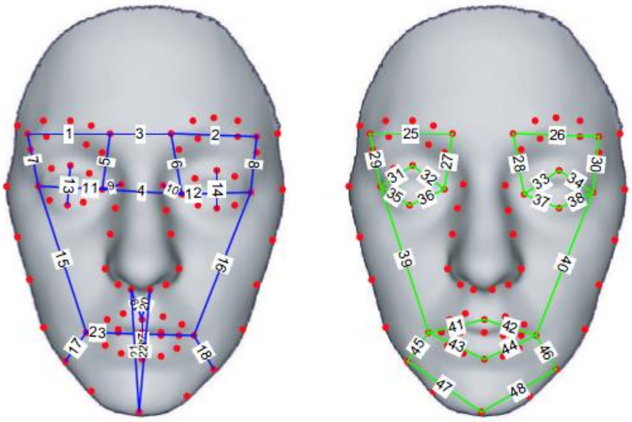


Figure 2.6 The distance features (left) and the slope features (right) (Tang et al., 2008)

Another important result was obtained by Tang et al., (2008) and their work was based on the ratio of distances. A set of 96 features are devised based on properties of the line segments connecting facial feature points on a 3D face model. The features consisted of the normalized distances and slopes of the line segments connecting a subset of the 83 facial feature points (refer to figure 2.6). To ensure the features are person-independent, the distance features are normalized by facial animation parameter units (FAPUs). Using a multi-class Support Vector Machine (SVM) classifier, an 87.1% average classification rate is achieved and the highest classification rate obtained in the experiments is 99.2% for the classification of Surprise.

Mpiperis et al. (2008) used bilinear models for joint 3D identity-invariant facial expression classification and expression-invariant face classification. The bilinear models are developed using the concept of a morphable model which consists of principal components of 3D faces. Each vertex of an adapted deformable model to the face scan is expressed by two independent and weighted sets of coefficients, one for identity and the other for facial expression.

Gong et al., (2009) suggested an automatic facial expression classification approach by exploring shape deformation. The shape of an expressional 3D face is assumed as the sum of two parts, a basic facial shape component (BFSC) and an expressional shape component (ESC). A reference face for each input 3D non-neutral is built by a learning method to separate BFSC and ESC. The BFSC estimation is done using Karhunen – LoeveTransform (KLT) which is closely related to Principal Component Analysis (PCA). The expression descriptors are computed by taking the surface changes between the original expressional face and its BFSC at the eyes and mouth regions. The SVM classifier with RBF kernel is used with expression descriptors as the feature vectors. In their work, the facial landmarks is automatically labelled which is different from Wang et al., (2006), Soyel et al., (2006) and Tang et al., (2008). The expression with the highest classification is Surprise and the lowest is Fear.

Maalej et al., (2010) proposed an approach based on local shape analysis of several relevant regions/patches of a given face scan. Based on the symmetry property of the human face, they optimized their work by using only landmarks laying on half of a face model (left part). The patches centred on a set of landmarks are extracted. Then, a curve-based representation of these patches is applied to capture the deformation between them on different faces under different expressions. The length of the geodesic path is computed and used as the input to the classifiers. A binary type of classification is used where the similarity scores between faces using all patches are computed. AdaBoost and SVM with kernel algorithms are implemented. The average success classification rate is 96.1% and the highest classification is Surprise expression.

Pinto et al., (2011) extracted 2D and 3D descriptors from different scales of wavelet transforms from seven expressions which includes the neutral expression. Then a Sequential Forward Floating Selection algorithm is used to analyse the multi-scale features to select the subset of features that best represents each facial expression.

Berreti et al. (2011) proposed an automatic approach for person-independent facial expression from 3D facial scans. In their work, a set of facial points are detected and SIFT descriptors are computed around the sample facial points of the face are used as a feature vector to represent the face. Before performing classification of the extracted descriptors, a feature selection approach is used to identify a subset of features with minimal redundancy and maximal relevance among the large set of features extracted with SIFT. Finally, the set of selected features are feed to SVM.

Their solution offers three main contributions: (i) to automatically detect facial points located in morphologically salient regions of the face; (ii) a local based description of the face that computes SIFT features on a set of sample points of the face derived starting from 9 facial points; (iii) a solution to feature selection for the identification of the salient SIFT features. Using a multi-class SVM classification on a large set of experiments, an average of 78.43% has been obtained.

As mentioned in the previous section, the BU-3DFE database provides six basic facial expressions with four levels of intensity. Studies conducted by Wang et al., (2006), Soyel et al., (2007), Tang et al., (2008) and Gong et al., (2009) who used this database only used the 2 highest intensities for every kind of expression.

From table 2.1, we can see only two 3D databases offer the complete six basic expressions with facial landmarks. Table 2.4 shows the existing works of 3D facial expression classification focused on six basic facial expressions which started from the release of BU-3DFE database. The comparison attributes are the database, 3D facial features, classifiers, the success percentage for each facial expression as well as the average success rate of the classification. Only one of the existing works used in-house dataset while the rest used 3D facial data from BU-3DFE database. From the average across expressions value, the expression that has the highest success rate is Surprise while the lowest success rate is the Fear expression. If we do not take into account the results from Pinto et al. (2010) which used in-house dataset, this comparison is still indicated as unfair because of the difference in the classifiers used.

Table 2.4 Comparison of 3D Facial Features in Facial Expression Classification Rates.

Author(s)	**Database**	**3D Facial Features**	**Classifier**	**Angry**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**	**Average**
Wang (2006)	BU-3DFE	Surface curvatures	Linear Discriminant Analysis	80.00	80.40	75.00	95.00	80.40	90.80	**83.1%**
Soyel (2007)	BU-3DFE	3D distance- vector	FF Neural Network	85.00	91.70	91.70	95.00	90.70	98.30	**92.07%**
Tang (2008)	BU-3DFE	Ratio of distances	Multiclass SVM (OVO)	86.70	84.20	74.20	95.8	82.50	99.20	**87.1%**
Mpiperis (2008)	BU-3DFE	3D deformable model	Particle Swarm Optimization	75.30	100.00	100.00	100.00	79.10	100.00	**92.3%**
Gong (2009)	BU-3DFE	Surface depth changes	Multiclass SVM	71.41	76.60	62.48	81.21	77.49	88.13	**76.2%**
Maalej (2010)	BU-3DFE	Curve-based	Multiclass SVM (OVO) and AdaBoost	96.50	97.00	94.50	94.67	96.00	97.83	**96.1%**
Pinto (2010)	In-house dataset	2D and 3D wavelet	AdaBoost	90.00	79.00	74.00	90.00	84.00	73.00	**94.8%**
Soyel (2010)	BU-3DFE	3D distance- vector	Tree-PNN	91.70	93.90	90.60	94.10	90.80	98.90	**93.33%**
Berretti (2011)	BU-3DFE	3D SIFT	Multiclass SVM	78.43	77.05	67.50	77.42	78.86	91.31	**78.4%**
Average across expressions				83.89%	86.65%	81.11%	91.47%	84.43%	93.05%	

Due to this difference, Gong et al. (2009) carried a similar experimental setting for 3D facial expression classification using four different 3D facial features (Wang et al., 2006; Soyel et al., 2007; Tang et al., 2008, Gong et al,. 2009). The average classification is computed using SVM classifier in 10-fold cross validation on the two highest intensities expression from BU-3DFE database. Berreti et al. (2011) used the same experimental setting in order to compare the performance of their 3D facial feature. The results reported in table 2.5 are the average classification rates for each of the 3D facial features done by few researchers under the same experimental setting.

Table 2.5 Comparison of several existing works using a similar experimental setting.

	Wang (2006)	**Soyel (2007)**	**Tang (2008)**	**Gong (2009)**	**Berreti(2011)**
Average Classification Rates	61.79%	67.52%	74.51%	76.22%	78.43%

Although this sub-section reviewed 3D facial features in the field of facial expression classification, the work of Ceolin (2012) is worth to be discussed as it used the same concept of facial feature employed in this work which is the surface normals. Ceolin (2012) used a 2.5D facial surface normals (or known as facial needle maps) which is acquired from 2D intensity images using Shape from Shading (SFS), referred to as Principal Geodesic Shape-From-Shading (PGSFS).The PGSFS method relies on a statistical model of facial shape formulated in the needle-map domain using Principal

Geodesic Analysis (PGA). PGA is a generalization of PCA to a non-linear setting of manifolds suitable for use with shape descriptors such as medial representations. The PGSFS method is used to iteratively recover needle-maps that realistically capture facial shape and also satisfy the image irradiance equation as a hard constraint. In other words, the recovered facial needle-maps both encode facial shape information and implicitly capture facial texture information. They demonstrated the visualization of the distances distribution using Multi-Dimensional Scaling (MDS) to embed the faces in a two-dimensional pattern space. They proved that a good separation of different faces under varying expression is plausible using statistical model. However, in their work, facial expression classification results for each six basic expressions are not provided therefore no comparison to other works can be made.

2.3.2 Action Units

The basic emotions occur relatively infrequent (Tian et al., 2001). Human tend to show simple facial motion such as tightening the lips in anger or obliquely lowering the lip corners in sadness (Carroll et al, 1997). To capture the subtlety of human emotion, the research community started to work on the AUs classification. However, due to the lack of FACS-coded databases, the AU-based classification research is not as numerous as in basic emotion type of classification.

Zhao et al. (2010) used their extended Statistical Facial Feature Model (SFAM) to generate feature instances corresponding to AU classes for three different modalities: facial landmark configurations, local texture and local geometry. The SFAM is a partial 3D face morphable model which contains both global variations in landmark configuration (morphology) and local ones in terms of texture and shape around each landmark (Zhao et al., 2009).15 features are extracted from three facial modalities, including multi-scale LBP, shape index, distances between landmarks and landmark displacement. SFAM is learnt by applying PCA to three kinds of training features while preserving 95% of variations for each type of features. Then, the similarity between each feature on a face and its instances are evaluated to obtain a set of similarity scores. Experiments on recognizing 7 AUs and 16 AUs have achieved 94.2% and 85.6% recognition rates respectively.

Savran et al, (2012) compared 3D modality *vis-a-vis* 2D modality for AU classification and they demonstrated that 3D modality is better especially for lower face AUs. They map the 3D data into 2D curvature images with each point in the image representing the curvature of the 3D surface at that point in the 2D plane. The comparison between these two modalities is based on Receiver Operating Characteristic (ROC) curves.

Sandbach[3] et al., (2012) proposed a new feature descriptor; local normal binary patterns (LNBPs), which is exploited for detection of facial action units (AUs).LNBPs employ the normals of the triangular polygons that form the 3D mesh to encode the shape of the mesh at each point. Initially, a circular neighbourhood around each point, specified by a radius r and P points regularly spaced around the circle.The unit normal $\mathbf{n}_p$ at each point v_p in the neighbourhood is found, along with that at the central point$\mathbf{n}_c$, through $x - y$ interpolation of the given points in the mesh. From here, two descriptors are formed: (1) $LNBP_{OA}$, which calculates the scalar of two normals and (2) $LNBP_{TA}$, which calculates the difference of two angles of the normals, the azimuth and the elevation. Feature vectors are then formed for each of the descriptors through the use of histogram. The x-y plane of the mesh is divided into 10x100 equally-sized square blocks and for each of these a histogram is calculated from the calculated binary numbers. These histograms are then concatenated into 1D feature vector suitable for use with the SVMs.

To date, there have been no studies using the BU-3DFE database to classify AUs simply because no AUs data provided by the BU-3DFE database developer. Nevertheless, Sun et al. (2008) manually labelled only 8 AUs in the BU-4DFE database for their partial AU classification which is clearly using 3D dynamic data.

As we mentioned before, there are several works present in the AU detection study; however, AUs mapping to facial expressions is still at a minimum. The recent work in AU-based studies was from Velusamy et al., (2011) where in their work, relationships between AUs and facial expressions are captured as templates strings comprising the most discriminative AUs for each facial expression. The Longest Common Subsequence (LCS) distance is used to calculate the closeness of a test string of AUs with the template string and hence infer the underlying facial expressions. However, Velusamy et al.'s work is based on 2D data.

Table 2.6 shows six basic facial expressions and associated AUs from several studies. Different studies state different Action Units (AUs) which are involved in six basic facial expressions. Ekman et al. (1978) introduced the basic AUs involved and in time, other researchers add/deduct certain AUs to represent the facial expressions. We believed this has to do with the intensity of the facial expressions itself, for instance different studies might focus on a certain degree of intensity. If we look at the Disgust expression, only Zhang et al., (2008) and Savran et al., (2008) agreed that the Disgust expression should have AU9 and AU10.Velusamy et al., (2011) and Lucey et al., (2002) do not even include AU9 and AU10 in the Disgust expression. As a reminder, AU9 is Nose Wrinkler while AU10 is Upper Lip Raiser.

Similarly, in the MPEG-4 standard (Pandzic et al., 2002), the six facial expressions are defined by facial animation parameters (FAPs) which describe how much the facial feature points have to be moved. Raouzaiou et al., (2002) in facial expression modelling provide FAP to AU mapping. However, their mapping has quite a difference with Zhang et al.'s (2008). For example in Raouzaiou et al., (2002), AU6 only consists of two FAPs which are *lift_l_cheek* and *lift_r_cheek* while in Zhang et al., (2008), AU6 also comprised *close_t_l_eyelidandclose_t_r_eyelid*.

Table 2.6 Six basic facial expressions with AUs

	Anger	Disgust	Fear	Happy	Sad	Surprise
Ekman &Friesan (1978)	4+5+7+23	9+15+16	1+2+4+5 +20+26	6+12	1+4+15	1+2+5B+26
Lucey et al., (2002)	4+5+15+17	1+4+15+17	1+4+7+20	6+12+25	1+2+4+15+17	1+2+5+25 +27
Raouzaiou et al., (2002)	2+4+5+7 +17	5+7+10+25	4+5+7+24 +26	26+12+7+6+20	7+5+12	26+5+7+4 +2+15
Deng et al., (2008)	2+4+7+9 +10+20+26	NIL	1+2+4+5 +15+20+26	1+6+12+14	1+4+15+23	1+2+5+15+16+20+26
Zhang et al., (2008)	2+4+7+23 +24	9+10	20+(1+5)+ (5+7)	6+12	1+15+17	5+26+27+ (1+2)
Primary						
Auxiliary	23+7+17+4+2	9+7+4+17 +6	20+4+1+5 +7	12+6+26 +10+23	15+1+4+17+10	27+2+1+5 +26
Savran et al., (2008)	2 +4+7+9+10+20+26	9+10	1+2+4+5+15+16+20 +26	1+6+12+14	1+ 4+ 15+23	1+2+ 5+15+16 +20+26
Velusamy et al., (2011)	17+25+ 26+16	17+25+26	4+5+7+25 +26	16+25+26	4+7+25+26	NIL

2.4 Statistical Approaches

3D face recognition and 3D facial expression classification success is also dependant on mechanism used for recognizing a person or classifying emotions. Among the approaches used in both fields are Principal Component Analysis (PCA) (Turk and Pentland 1991), Linear Discriminant Analysis (LDA) (Belhumeur et al., 1997), Iterative Closest Point (ICP), Active Shape Models (Prabhu et al, 2000), Probabilistic Neural Network (Vinitha et al., 2009), Support Vector Machine (SVM), Statistical Facial Feature Model (SFAM) (Zhao et al., 2009) etc. The most popular ones are PCA and LDA in which they sometimes are coupled with other approaches; for example PCA with the nearest neighbour classifier.

Most methods in face processing studies use dimensionality reduction techniques due to the fact that faces are represented as points in high-dimensional image space. By employing dimensionality reduction, a more meaningful representation is established, therefore, addressing the issue of the ”curse of dimensionality” (Sharath et al., 2011). Dimension reduction is a process of reducing the number of variables under observation. Although face images can be regarded as points in a high-dimensional space, they often lie on a manifold (i.e., subspace) of much lower dimensionality, embedded in the high-dimensional image space. Originally, the main issue in dimensionality reduction is how to properly define and determine a low-dimensional subspace of face appearance in a high-dimensional image space.

The dimensionality reduction approaches are divided into supervised (i.e. Eigenfaces by Turk and Pentland, 1991) and unsupervised (i.e. Fisherfaces by

Belhumeur et al., 1997). In recent years, Eigenfaces and Fisherfaces have attracted much attention especially in the 2D modality of face processing study.

2.4.1 Eigenfaces

Turk and Pentland (1991) introduced the idea of Eigenfaces which became a gold standard in face recognition. In their work, face images are projected into a feature space through PCA which means that a face is represented as a linear combination of a set of basis images.

Pure data driven methods can be applied without knowledge and extract good parameters by using the input data, PCA is an example. PCA is one of the methods in multivariate statistics that encompasses simultaneous observation and analysis of more than one statistical variable. Generally, PCA will reduce the large dimensionality of the data space (observed variables) to the smaller intrinsic dimensionality of the feature space (independent variables) which are needed to describe the data economically (Rady, 2011). In PCA, an orthogonal system is found such that data is best approximated by the minimum number of dimensions and the correlation between different dimensions is minimized.

When applied to face images, PCA yields a set of eigenfaces and these eigenfaces are the eigenvectors that are associated with the largest eigenvalues of the covariance matrix of the observed data. Each face image can be reconstructed based on the weighted average of the principal components of the original training set of face images. PCA projections are optimal for reconstruction from a low dimensional basis; however they may not be optimal from a discrimination standpoint as they do not use class information in the projection (Belhumeur et al., 1997; Turk and Pentland, 1991).

Unsupervised method such as PCA expose statistical properties of the input data to discover relevant features and PCA is widely applied in computer vision applications, particularly face processing studies. The Active Appearance Model (AAM) and 3D Morphable Model are the earliest example of approaches that employed PCA in their framework.

AAM was proposed by Cootes et al., (1998). An AAM contains a statistical model of the shape and grey-level appearance of the object of interest which can generalise to almost any valid example. A year later, the 3D Morphable Model was introduced by Blanz et al. 3D Morphable Model is derived from a data set of 3D face models by automatically establishing correspondence between the examples. It captures the variations observed within a data set of 3D scans of examples and converts their shape and texture into a vector space representation (Blanz et al., 1999).

2.4.2 Fisherfaces

Fisherfaces was introduced by Belhumuer et al. (1997) and the work is based on LDA which is basically an enhancement to PCA. LDA assumes that classes/labels are being used and that features of different classes/labels have a Gaussian distribution with the same covariance matrix but different mean. The approach of the LDA is to project all the data points into new space, normally of lower dimension, which maximises the between-class separability while minimising their within-class variability (Gillies, 2013).

The apparent difference between LDA and PCA is that LDA produce a subspace that maps the sample vectors of the same class to a single spot of the feature representation and therefore the gaps between those of different classes are as clear as possible. Given a number of independent features relative to which the data is described, LDA creates a linear combination of those which yields the largest mean differences between the desired classes.

In PCA, the PCA subspace is determine from the training data. i^{th} image vector containing N pixels in the form of

$$\mathbf{x}^i = \left[\mathbf{x}_1^i, \cdots, \mathbf{x}_N^i\right] \tag{2.1}$$

All p images in the image matrix

$$\mathbf{X} = [\mathbf{x}^{\mathbf{1}}, \cdots, \mathbf{x}^{\boldsymbol{p}}] \tag{2.2}$$

The covariance matrix is computed

$$\Omega = \mathbf{X}\mathbf{X}^{\mathbf{T}} \tag{2.3}$$

The eigenvalues and eigenvectors is solved

$$\Omega\mathbf{V} = \boldsymbol{\Lambda}\mathbf{V}, \tag{2.4}$$

where $\boldsymbol{\Lambda}$ is the vector of eigenvalues of the covariance matrix.

LDA uses PCA subspace as input data, i.e. matrix $\mathbf{V}$ obtained from PCA. The important step in LDA which differentiate it from PCA is to find two scatter matrices referred to as the “between class“ and “within class“ scatter matrices (Mazanec et al., 2008). The within class matrix is defined as follows:

$$\mathbf{S}_{\mathbf{w}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{C}} \mathbf{S}_{\mathbf{i}}, \mathbf{S}_{\mathbf{i}} = \sum_{x \in X_i} (\boldsymbol{x} - \boldsymbol{m}_{\boldsymbol{i}})(\boldsymbol{x} - \boldsymbol{m}_{\boldsymbol{i}})^{\mathbf{T}} \tag{2.5}$$

where $\mathbf{m}_{\boldsymbol{i}}$is the mean of the images in the class and $C$is the number of classes. The between class matrix is defined as:

$$\mathbf{S}_{\mathbf{B}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} n_{\mathrm{i}}(\boldsymbol{m}_{\boldsymbol{i}} - \boldsymbol{m})(\boldsymbol{m}_{\boldsymbol{i}} - \boldsymbol{m})^{\mathbf{T}}, \tag{2.6}$$

where $\mathrm{n}_{\boldsymbol{i}}$is the number of images in the class, $\mathbf{m}_{\boldsymbol{i}}$is the mean of the images in the class and $\boldsymbol{m}$is the mean of all the images. Then generalized eigenvalue problem in LDA is solved using

$$\boldsymbol{S}_{\boldsymbol{B}}\mathbf{V} = \boldsymbol{\Lambda}\mathbf{S}_{\mathbf{w}}\mathbf{V} \tag{2.7}$$

Belhumuer et al. (1997) also carried out a comparison experiment between Eigenfaces and Fisherfaces and they reported that Fisherfaces appears to be the best simultaneously handling variation in lighting and expressions. The most common problem in Fisherfaces is that if the dimension is much larger than the number of training samples per class and as a result, a singular matrix is produced. To overcome this problem, the face image is projected into a face subspace of PCA. Subsequently, the projected PCA vectors are applied to LDA to construct a linear classifier in the subspace. Even though LDA is said to perform better than PCA in classification, LDA requires more computation compared to PCA.

2.5 Modular-Based Work

A pure eigenface system can be fooled by gross variations in the input image (hats, beards, etc). Pentland et al., (1994) introduced the modular eigenspaces (or eigenfeatures) used in face recognition. According to them, the modular description allows for the incorporation of important facial features such eyes, nose and mouth. They showed that eigenfeatures alone were sufficient in achieving a 95% recognition rate in their experiment. By using a combination of eigenfeatures and an eigenface representation, a slight improvement of 98% was obtained. They also showed that a modular representation has the advantage of disambiguating false eigenface matches due to gross variations in the input image.

There are also several studies that employed face decomposition in their work and most of them are based on a linear combination approach. Tena et al., (2011) used a collection of PCA sub-models that are independently trained but share boundaries. Their findings strengthen the hypothesis that a region-based model is better than a holistic approach and the region-based approach increases flexibility for local deformations. Gottumukkal et al., (2003) also showed a significant result especially when there are large variations in facial expression and illumination. The work of Tena et al., (2011) was based on 3D data and 2D data in Gottumukkal et al., (2003). However, there is no facial expression classification results recorded in Tena et al., (2011) as this work is developed for animation purposes while Gottumukkal et al., (2003) was for face recognition.

Gottumukkal et al., (2003) discovered that if the face images are divided into very small regions the global information of the face may be lost and the accuracy of this approach is no longer acceptable. Thus, choosing the size of the modules to represent a face is also vital. Chiang et al. (2009) divided the face into five modules which included the left eye, the right eye, the nose, the mouth, and the bare face with each facial module identified by a facial landmark at the module centre.

2.6 Classifiers

In machine learning and statistics, classification is the problem of identifying to which of a set of categories of a new observation belongs, on the basis of a training set of data containing instances whose category membership is known (Wikipedia[2], 2013). Classification methods are used in many areas like data mining, finance, signal decoding, voice recognition, computer vision, natural language processing or medicine. In this area of study, once the facial features are extracted and selected, the next step is to classify the probe face. Face processing classification algorithms can be roughly divided into two broad families of approaches: (i) learning-based classifiers (sometimes known as parametric classifiers and (ii) non-parametric classifiers.

The learning-based classifiers require intensive learning phase of the classifier parameters. The methods such as Support Vector Machines (SVM) (Tang et al., 2008; Gong et al., 2009; Maalej et al., 2010; Berreti et al., 2011), Boosting, Linear Discriminant Analysis (Wang et al. 2006), Neural Network (Soyel et al., 2007), rule-based (i.e. PSO (Mpiperis et al., 2008), Decision Trees are known as parametric classifiers.

For non-parametric classifiers, the classification is based on the data and therefore, no learning or parameters are required. The basic idea of the nearest-neighbour classifier is to store all labelled instances (i.e., the training set) and compare new unlabelled instances (i.e., the test set) to the stored ones to assign them an appropriate label. Non-parametric classifiers have several important advantages that are not shared by most learning approaches: (i) Can naturally handle a huge number of classes. (ii) Avoid overfitting of parameters and (iii) Require no learning or training phase (Boiman et al., 2008). The most common non-parametric is the nearest-neighbour classifier.

In this work, we are demonstrating the discriminative power of feature set and thus a simple classifier such as nearest neighbour and SVM is sufficient to be used.

2.7 Conclusions

It should be clear from this literature review that research on 3D facial expression classification has not been as extensive as research in 3D face recognition in the last few decades. This is due to the existence of the 3D expression data that has been publicly available only since 2006. We conclude this chapter with a summary of the literature.

Research into 3D facial expression databases, the general framework of facial expression, 3D facial features, statistical modelling methods and classifiers used by the research community were reviewed. In the literature, we can see a gap in the use of the most fundamental feature in 3D which is 3D surface normals to classify 3D facial expressions. In terms of visual appearance, surface normals produce smooth shading across the transition from one triangle to another, making a fundamentally polygonal object look round. Besides that 3D facial surface normals also provide a richer source of information about the shape of the facial mesh than the depth alone. We are interested to investigate the feasibility of surface normals to classify facial expression. Our main objective is to use 3D facial surface normals which are extracted straightforwardly from the provided 3D facial points in 3D facial expression classification.

Most of the studies stressed the regions that are salient to quantify facial expression. These regions contain facial features that reflect the intensity of facial expression shown by the subjects. We aim to find 3D facial surface normals on each region and model it using the statistical approaches. In particular, we aim to see how

3D surface normals are distributed in each region deformation caused by a facial expression.

CHAPTER 3

DATA PRE-PROCESSING AND STATISTICAL MODEL

Statistical models attempt to model the shapes of objects found in images and the model characterizes the variation of shapes within a training set. There are two types of statistical modelling; unsupervised and supervised. One of the well-known statistical models is Principal Component Analysis (PCA) belongs to the unsupervised learning group. Unsupervised learning group is designed to extract common sets of features present in the input data and the examples given to the learner are unlabelled. The only input parameters are the number of dimensions that will be retained in the embedding and the data points. In unsupervised learning the machine obtains neither supervised target outputs, nor rewards from its environment. In a sense, unsupervised learning can be thought of as finding patterns in the data above and beyond what would be considered pure unstructured noise (Ghahramani, 2004).

PCA has been used widely in 2D and 3D to extract facial features, to align facial landmarks as well as to perform face identification. Heseltine et al[1]., (2004) used PCA to reduce the dimensionality of 3D facial surfaces to perform 3D face recognition while Kapoor et al., (2010) used the Mahalanobis distance as the feature vectors in PCA for facial expression classification. Several works using PCA with different feature vectors in the field of face processing, ranging from 2D to 3D, can be found in Gottumukkal and Asari, (2003), Praseeda et al., (2008), Dongcheng, S. and Jieqing (2010) and Tena J.R. et al., (2011). In this work, PCA is used to model the 3D geometrical properties of 3D facial expressions and in this chapter the basic concept of PCA is discussed.

The remainder of this chapter is organized as follows: In sections 3.1 and 3.2, we explain the pre-processing steps in our approach, which begins with data extraction and 3D facial points alignment, respectively. PCA is described in section 3.3. Section 3.4 concludes this chapter.

3.1 3D Face Points Extraction

3D faces with different facial expressions were used in this work and the data was acquired from the Bosphorus database (Savran, et al, 2008). Although the Bosphorus database provides 105 subjects with six basic expressions plus neutral expression, only 65 subjects can be used for training and experiments since these were the only subjects that came with a complete set of six basic facial expressions.

3D faces with labelled expressions with different poses are not available in this database and therefore, all 3D face images of six basic facial expressions are frontal profiles. The Bosphorus database provides 24 manually annotated facial landmarks, provided that they are visible in the scan (see figure 2.1).These facial landmarks are manually labelled with its specific anatomic denotation by the developer, for instance landmark no 14 is denote as the nose tip. However, only 22 of the provided facial landmarks were used as the two facial landmarks (both earlobes) were not visible in the frontal scan and thus the 3D correspondence of both earlobes could not be computed. In this work, the focus is on the Facial Animation Parameters (FAPs) involved in six basic facial expressions. There are six facial landmarks that are visible in the frontal scan and they are associated with the determined FAPS but not provided directly by the Bosphorus database which is the top and bottom of both eyes and centre of both pupils. However, the developer of the Bosphorus database does provide the 3D facial landmarks together with its 2D pixel value correspondence. Therefore, we manually marked the six extra facial landmarks on its 2D image and the 3D correspondences were established manually. The six extra facial landmarks are the centre point of both pupils as well as the lowest and highest points on both eyes (refer to Figure 3.1).

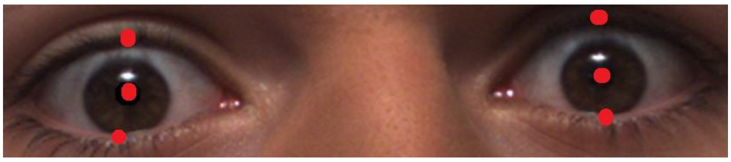


Figure 3.1 Six extra facial landmarks

17 face boundary 3D facial points were also found and added into the system (refer to Figure 3.2).

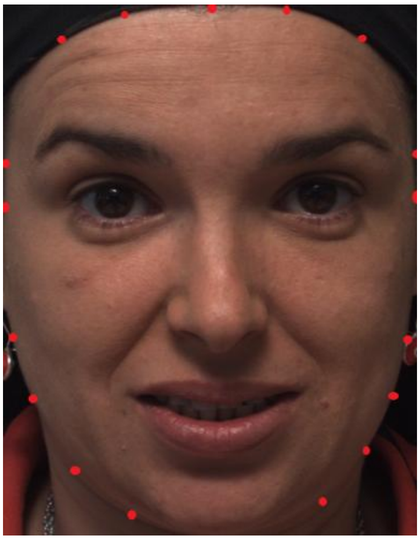


Figure 3.2 17 face boundary 3D facial points

In addition to the 22 provided landmarks, 6 extra landmarks and 17 boundary points, we also added 70 facial points that we computed by finding the average of each triangle's vertices. This is done with the objective of having a rather dense looking 3D face model. Figure 3.3 shows a facial point (point number 4) as an average of three facial points (point number 1, 2 and 3) that form a triangle mesh.

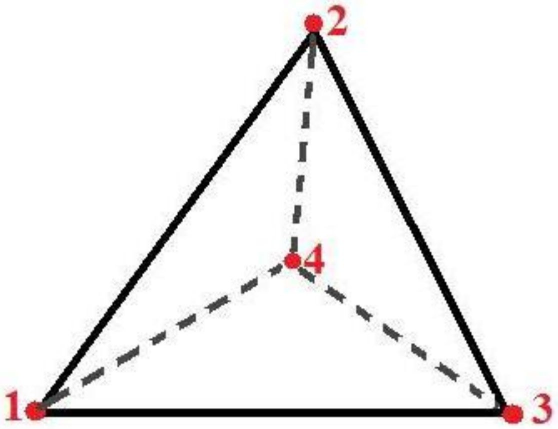


Figure 3.3 Finding an extra facial point

Figure 3.4 shows an example of a 3D face using the complete set of 3D facial points with Delaunay triangulation. Delaunay triangulation is a proximal method that satisfies the requirement that a circle drawn through the three nodes of a triangle will contain no other node (Tchoukanski, 2012).

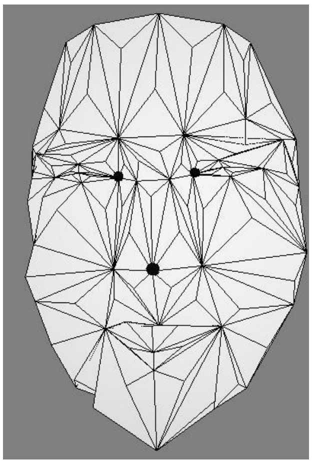


Figure 3.4A 3D face model with happy expression

3.2 3D Facial Points Alignment

3D facial points of each face need to be aligned before any value comparison between the faces takes place. This is to ensure the 3D faces are as closely aligned to each other as possible while keeping the shape unchanged. In this work, a simple affine transformation is employed in the alignment process. The scanned 3D faces normally have hundreds to thousands of 3D facial points; clearly a global transformation for all 3D points requires extensive computation. The alternative solution is to use only three 3D facial landmarks, (l_1, l_2, l_3) in the initial alignment process. Those feature landmarks are specifically the inner left and right eye corner, $l_1$and l_2, as well as the nose tip, l_3; the three black dots in figure 3.4 and 3.5 denote the feature landmarks used. These three landmarks are often used in face alignment because the change of their position in any expression is infrequent (Tian et al., 2001).

A mean shape is used as a reference in 3D facial alignment. The three 3D facial landmarks $(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3)$from neutral expressions from all subjects are used to find the mean shape. As a result, we have three facial points $\left(\bar{\mathbf{l}}_1, \bar{\mathbf{l}}_2, \bar{\mathbf{l}}_3\right)$ that represent the mean shape

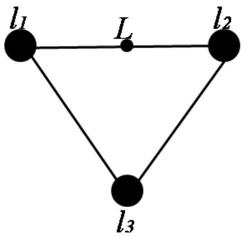


Figure 3.5 Three facial landmarks used in 3D face alignment.

Figures 3.7a and 3.7b show the initial condition before 3D face alignment took place. The red coloured triangle is the mean shape while the uncoloured triangles are the training sets. Each corner of the triangles represents the three feature points in the mean shape as well as in the training sets. M_i is the triangle for each training sets where $i = 1, \dots, k$ and k is the number of training 3D faces. Let

$$\bar{M} = \frac{1}{t} \sum_{i}^{\mathrm{k}} M_i \tag{3.1}$$

where $M_i = [M_1, M_2, \dots, M_{\mathrm{k}}]$ and $M = \{\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3\}$. There are two phases of alignment process. The first phase of the alignment process is the translation to an origin point and only the three significant feature points will undergo the translation process. The origin point refers to l_3 in figure 3.3. The Euclidian distance between $\mathbf{l}_3$ for every M_i with $\bar{\mathbf{l}}_3$ of the mean shape is measured. This means that each M_i has their own set of $\mathbf{distance1}_i$. Let

$$\mathbf{distance1}_i = \left\| \mathbf{l}_3^{\mathbf{i}} - \bar{\mathbf{l}}_3 \right\| \tag{3.2}$$

where $\mathbf{distance1} = (xdistance, ydistance, zdistance)$. Subsequently, all three landmark points of M_i are translated according to the value of the $\boldsymbol{distance1_i}$ obtained.

$$(x, y, z) \rightarrow (x + xdistance, y + ydistance, z + zdistance) \tag{3.3}$$

Figures 3.9(c) and 3.9(d) show the translated points for all $M_i$and we can see that l_3 for all $M_i$are in the same position. Figure 3.6 shows the top view of after the translation process of the mean shape $\bar{M}$ (red rectangle) and M_i(black rectangle). It also shows the l_1 and l_2 of $M_i$and $\bar{M}$. We can see l_1 and l_2 of $M_i$are still not aligned with l_1 and $l_2$of $\bar{M}$.

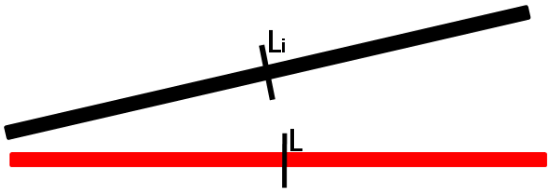


Figure 3.6 An example of after the translation process of the mean shape $\bar{M}$(red rectangle) and M_i(black rectangle) from top view.

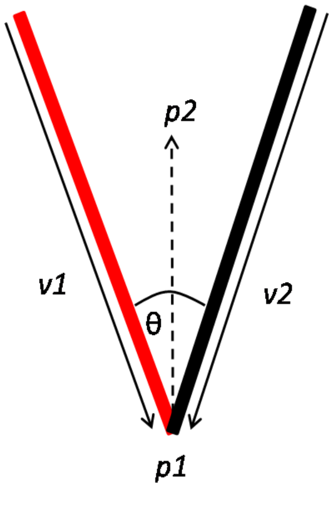


Figure 3.7θ must be determined for the first phase of 3D rotation.from side view

Figure 3.7 shows the initial scenario before the second phase in the 3D alignment phase begins from the side view. The second phase in the alignment process involves rotation of each M_i about an arbitrary line byθ, an angle formed by vectors $\mathbf{v}_1$ and $\mathbf{v}_2$as shown in figure 3.7.This phase is divided into two parts each of which is a 3D rotation. The objective of the first part is to ensure $\mathbf{v}_1$ and $\mathbf{v}_2$is as close as possible and thus the angle θ between $\mathbf{v}_1$ and $\mathbf{v}_2$is kept to a minimum. To do this, we need to determine the θ value and the arbitrary line for each M_i. The midpoint of l_1 and l_2 for both $M_i$and $\overline{M}$, denoted as L_i and L, are computed (see figure 3.4). Then, vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are computed.

$$\mathbf{v_1} = \|L - l_3\| \tag{3.4}$$

$$\mathbf{v}_2 = \|L_i - l_3\| \tag{3.5}$$

Next, the angle θ between $\mathbf{v}_1$ and $\mathbf{v}_2$is determined.

$$\theta = \cos^{-1} \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\|\|\mathbf{v}_2\|} \tag{3.6}$$

The normal to the plane containing $\mathbf{v}_1$and $\mathbf{v}_2$, denoted as $\mathbf{q}$, is calculated.

$$\mathbf{q} = \mathbf{v}_1 \times \mathbf{v}_2 \tag{3.7}$$

The arbitrary line which comprises points $p1$and $p2$, as defined in equations 3.8 and 3.9 is computed. $l_3$in equation 3.8 and 3.9 is clearly l_3 of $\bar{M}$ as l_3 of all M_i have been translated to it in the previous phase.

$$p1 = l_3 \tag{3.8}$$

$$p2 = l_3 + \mathbf{q} \qquad 3.9)$$

With the arbitrary line and θ determined, we need to ensure the direction of the 3D rotation. All three 3D facial landmarks are then rotated using the arbitrary line and two different angles which are θ and$-\theta$. We thus have two sets of newly transformed 3D facial landmarks; one using the θ and the other one is using $-\theta$. Next, the midpoint of l_1 and l_2 for both sets of newly transformed 3D facial landmarks are calculated. The midpoint of each set is denoted as $mp1$ and$mp2$. To decide which angle (θ or $-\theta$) we have to store, the distances between L to $mp1$ and $L$to $mp2$ are computed. The smallest distance of those two tells us the angle that we ought to store. Figure 3.8 described the first part of the 3D rotation result from the top view. From this first part of the rotation, $p1$, $p2$ and θ (or $-\theta$) are stored.

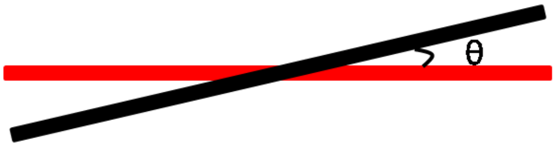


Figure 3.8 An example of the first part of 3D rotation result from top view.

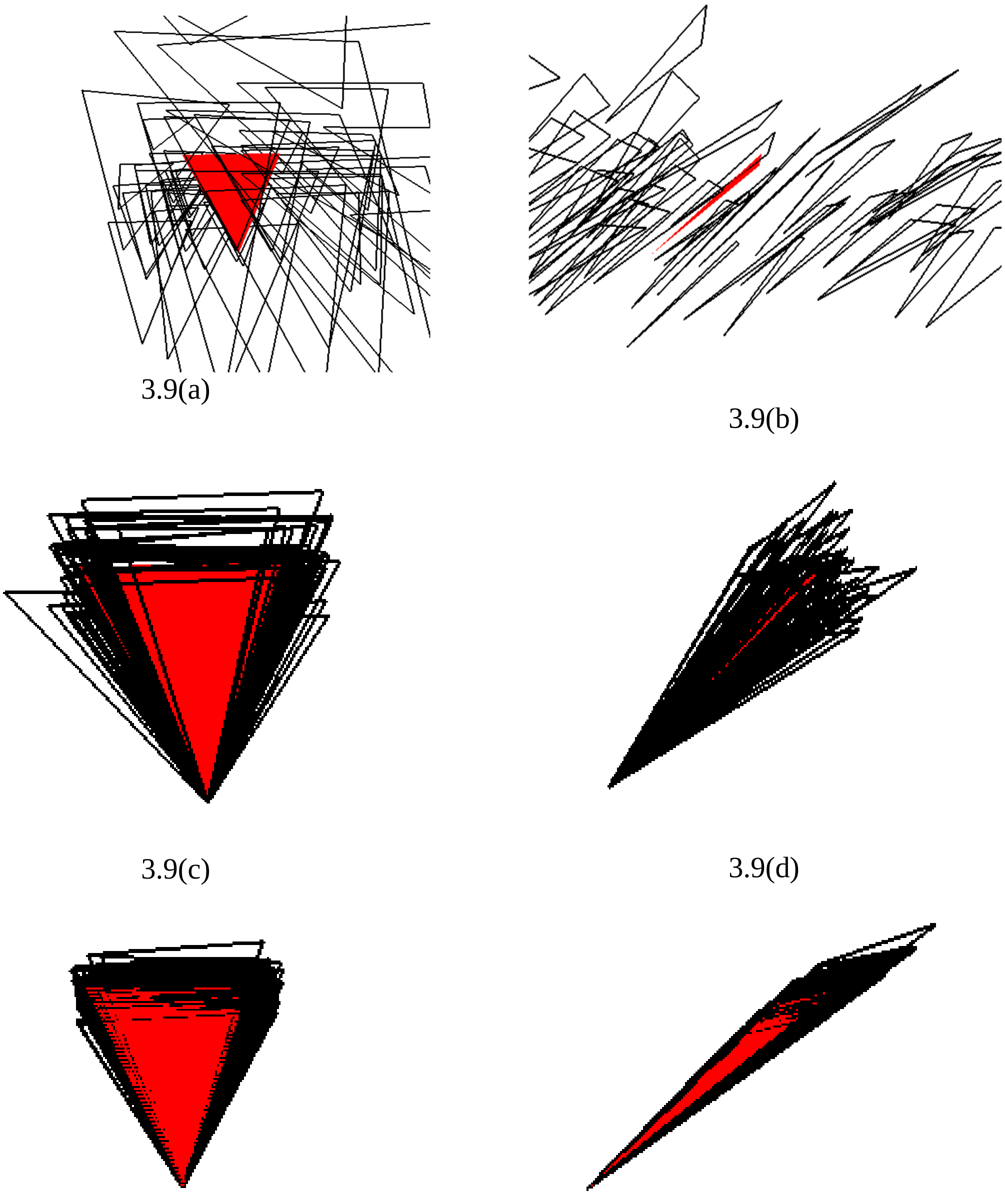
The objective of the second part of 3D rotation is to ensure l_1 and l_2 of M_i and $\bar{M}$ is as close as possible to each other. The second part of the 3D rotation starts with a translation for all 3D landmarks. The Euclidian distance between L_i for every $M_i$with L of the mean shape is measured. Let

$$\mathbf{distance2}_i = ||\mathbf{L} - \mathbf{L}_i|| \tag{3.10}$$

where $\mathbf{distance2}_i = (xdistance, ydistance, zdistance)$. All the landmarks of each M_i are then translated using the$\mathbf{distance2}_i$ value. Then, another vector, $\mathbf{v}_3$ is computed using the new translated 3D landmarks.

$$\mathbf{v}_3 = ||\mathbf{L} - \mathbf{L}_i|| \tag{3.11}$$

$\theta$is set to 0° and a $smallestdistance$ variable is set to 1000. $l_1$is then rotated using the line formed by $\mathbf{v}_3$and l_3 and θ. Next, the new l_1 is stored in an array denoted as $temp$. Then, we change the θ by adding another degree of angle to it.The rotation is repeated again using the same line formed by $\mathbf{v_3}$and l_3 but with the new θ. This iteration ends when $\theta = 360°$. After rotation, the distance between the new rotated $l_1$which has been stored in the array $temp$ and the l_1 of the mean shape $\bar{M}$ is computed. If the currently computed distance is less than the $smallestdistance$ parameter, the θ corresponding to the l_1 is saved. From the second part of the rotation, $\mathbf{v}_3$, l_3 and θ alignment parameters are saved. Figures 3.7e and 3.7f show the final outcome of the 3D face alignment from the frontal and side views.



3.9(a)

3.9(b)

3.9(c)

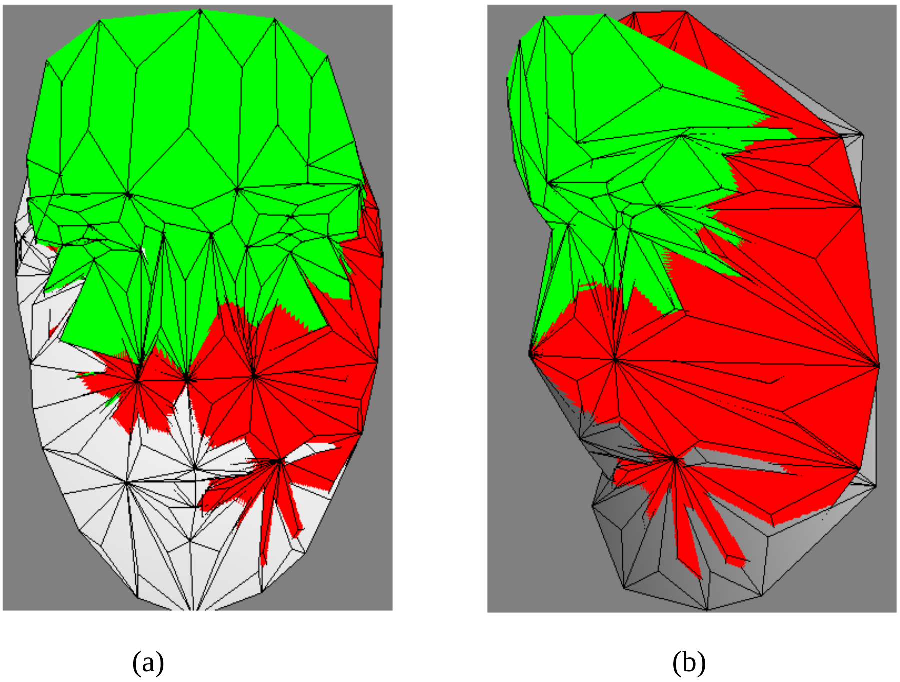
3.9(d)

3.9(e)

3.9(f)

Figure 3.9(a)-3.9(f) The red triangle is the mean shape while the black triangles are the training sets. The left side is the frontal view while the right side is the side view. (3.9(a) and 3.9(b)) Mean shape with the training sets before alignment process. (3.9(c) and 3.9(d)) After the translation transformation to the origin point. (3.9(e) and 3.9(f)) A complete alignment process.

Figures 3.9(a)-3.9(f) show the alignment process for only three landmarks. After all the alignment parameters for each face in the training set are computed, all the 3D landmarks of a face will be aligned according to its assigned facial alignment parameters. Figures 3.10(a) and 3.10(b) show the result of 3D alignment for three faces. The grey face is the mean shape while the green face is a fear expression and the red face is a surprise expression. From the figure we can see the three faces are overlapping onto each other and this means that they were aligned. The difference of each face shape is more apparent from the side view (figure 3.10(b)).



(a) (b)

Figure 3.10 Results of 3D face alignment (a) front biew (b)side view

Table 3.1 Pseudo-Code Illustration for Second Phase in 3D Face Alignment

```
Function Transformation2

**** First part of 3D rotation****
Get mean shape midpoint, $L$ of $l_1$ and $l_2$
Get $M_i$ midpoint, $L_i$ of $l_1$ and $l_2$
$\mathbf{v_1} = |L - l_3|$
$\mathbf{v_2} = |L_i - l_3|$
Get angle, $\theta$ between $\mathbf{v1}$ and $\mathbf{v2}$
Get a normal, $q$ of $\mathbf{v1}$ and $\mathbf{v2}$
$p1 = l_3$
$p2 = l_3 + \mathbf{q}$
For each facial landmark
    Rotate on a line form by $p1$ and $p2$ by $\theta$
    Store output as $array1$
    Rotate on a line form by $p1$ and $p2$ by $-\theta$
    Store output as $array2$
Get $array1$ midpoint of $l_1$ and $l_2$, $mp1$
Get $array2$ midpoint of $l_1$ and $l_2$, $mp2$
Compute distance between $mp1$ and$L$, $dist1$
Compute distance between $mp2$ and $L$, $dist2$
If $dist1 < dist2$ , save $\boldsymbol{\theta}, \boldsymbol{p1}$ and $\boldsymbol{p2}$
Else save $\boldsymbol{-\theta}, \boldsymbol{p1}$ and $\boldsymbol{p2}$

**** Second part of 3D rotation****
$\mathbf{distance2_i} = L - L_i$
Translate all landmarks using $\mathbf{v_3}$ value
$\mathbf{v_3} = L - L_i$
$\theta = 0°$
$smallestdistance = 1000.0$
For each facial landmark
  While $\theta < 360°$
    Rotate on a line form by $l_3$ of face mean shape and $\mathbf{v_3}$ by $\theta$
    Store new $l_1$ in array of $temp$
    $\theta = \theta + 1$
  Get distance for $l_1$ of face mean shape and $l_1$ of $temp$, $dist$
  If $dist < smallestdistance$
    Save $\boldsymbol{\theta}$,$\mathbf{v_3}$ and $\boldsymbol{p1}$
```

Table 3.1 describes the second phase of the 3D face alignment in pseudo-code. All sets of facial landmarks M_i are passed through this function. The 3D translation and rotation parameters of both parts, for each M_i are saved. The emboldened symbols in Table 3.1 refer to the saved parameters in this phase. With these parameters, the remaining coordinates of all 3D facial points of each face are aligned according to the same transformation set.

Table 3.2 Alignment Parameters for M_1

16.34 14.92 -18.88-			→ translation value
-0.09219	2.54 -15.19 40.38	-433.39 -0.60 63.50	→θ (or $-\theta$), $p1$ and $p2$ parameters.
-0.00873	2.54 -15.19 40.38	1.54 14.74 2.55	→θ , $v4$ and l_3 parameters.

Table 3.2 shows an example of the alignment parameters for M_1. The first line is the translation parameters where all 3D facial landmarks of M_1must be translated to 16.34 in the x-direction, 14.92 in the y-direction and -18.88 in the z-direction. The first part of the rotation is in the second line and the first parameter is the angle while the rest are the vectors $\mathbf{v_1}$and $\mathbf{v_2}$. The last line is the second part of the rotation and the order of the parameters is similar to the second line.

In the initial phase of this work, we used a set of faces with six basic expressions to compute a mean shape. However, we are concerned about the different type of expressions from different subject that might influence the face mean shape. We were concerned that the intense expressions (from the six basic expressions) such as surprise and happy give a big impact in the mean shape computation; thus the mean shape will look a lot like the intense expressions. Therefore, we would like to see the dispersion of the aligned training set from the mean shape.

We managed to compute the standard deviation of the aligned training set for both types of the mean shape: mean shape built from six basic expressions and mean shape built from neutral expression only. Table 3.3 shows the result and both standard deviations of the aligned training set are close to each other. From this table, we can conclude that building a mean shape whether using the six basic expression data or neutral faces data gives a similar alignment result.

Table 3.3 Standard deviation of the training set from two types of mean shape

Type of Mean Shape	***Standard Deviation(σ) of the Aligned Training Set***
Mean shape built from six basic expressions	0.680
Mean shape built from neutral expression only	0.675

3.3 Principal Component Analysis

Principal Component Analysis (PCA) is a useful statistical technique that has found application in data compression and it is the simplest eigenvector-based multivariate analysis method. If a multivariate dataset, for example a set of images, is visualised as a set of coordinates in high-dimensional data space, PCA provide will lower the dimensional face image but it still has most informative information.

PCA performs a basis transformation to an orthogonal coordinate system formed by the eigenvectors of the covariance matrices (Hwang et al., 2000). Its operation can be thought as revealing the internal structure of the data in a way which best explains the variance in the data (Wikipedia, 2013). PCA is often used as a method that can get the shape representation of the face by the principal components. In this work, the goal of a PCA is to determine the principal directions of variation of the data within the data cloud.

The PCA computation in this work is based on Turk and Pentland (1994) and Trivedi (2009). Suppose 3D facial points are the feature vectors. Let $\mathbf{\Omega}_i$ be the training 3D facial points of the i^{th} person which has N 3D facial points.

$$\mathbf{\Omega}_i = \begin{bmatrix} s_{x,1}, s_{y,1}, s_{z,1} \\ s_{x,2}, s_{y,2}, s_{z,2} \\ \cdots \\ s_{x,N}, s_{y,N}, s_{z,N} \end{bmatrix}_{\text{Nx3}} \tag{3.12}$$

From equation 3.11, $\mathbf{\Omega}_i$ can be represented as 1-D vector by concatenating each row into a single column vector

$$\mathbf{\Omega}_i = \left[s_{x,1}, s_{y,1}, s_{z,1}, s_{x,2}, s_{y,2}, s_{z,2}, \dots, s_{x,N}, s_{y,N}, s_{z,N}\right]^T_{\text{Nx3x1}} \tag{3.13}$$

The 3D facial points are mean centred by subtracting the mean facial points from each 3D facial points. Let $\bar{\mathbf{m}}$ represent the mean of 3D facial points:

$$\bar{\mathbf{m}} = \frac{1}{k}\sum_{i=1}^{k} \mathbf{\Omega}_i \tag{3.14}$$

where $i = 1, \ldots, k$ and$k$is the number of training 3D faces. Let $\mathbf{d}_i$ be defined as mean centred 3D facial points:

$$\mathbf{d}_i = \mathbf{\Omega}_i - \bar{\mathbf{m}} \tag{3.15}$$

PCA can be derived using the covariance matrix or Singular Value Decomposition (SVD). Here we are going to explain the method that we chose which is using the covariance matrix.

$$\mathbf{\Sigma} = \frac{1}{k} \sum_{i=1}^{k} \mathbf{d}_i \ \mathbf{d}_i^T \tag{3.16}$$

The covariance matrix Σ is a sum of outer vector products and it is a k x$(3N)$ matrix. Essentially the covariance matrixΣexpresses the variation about the mean in each dimension. PCA determines a linear transformation of the data which diagonalizes the covariance matrix of the transformed data. We compute the matrix **V** of eigenvectors which diagonalizes the covariance matrix.

$$\mathbf{V}^{-1}\mathbf{\Sigma}\mathbf{V} = \mathbf{D} \tag{3.17}$$

where **D** is the diagonal matrix of eigenvalues of **Σ.**

Next, we find vectors $\mathbf{u}_j$ and scalars λ_j which are the eigenvectors and eigenvalues of the covariance matrix. Our aim is to seek a set of k orthornormal vectors,$\mathbf{u}_i$ which best describes the distribution of the data. The j^{th} vector, $\mathbf{u}_j$, is chosen such that

$$\lambda_j = \frac{1}{k}\sum_{i=1}^{k}\left(\mathbf{u}_j^T\mathbf{d}_i\right)^2 \tag{3.18}$$

is a maximum, subject to

$$\mathbf{u}_l^T\mathbf{u}_j = \partial_{lj} = \begin{cases} 1 & \text{if} l = j \\ 0 & \text{otherwise} \end{cases} \tag{3.19}$$

To determine the number of principal components to use, we first rank the eigenvalues, λ_j's in decreasing order. We chose the s principal components corresponding to the eigenvalues for which:

$$\sum_{j=1}^{s}\lambda_j > f\sum_{j=1}^{s}\lambda_j \tag{3.20}$$

f is some fraction of the variation in the original dataset that we want to explain in our transformed feature space. It is standard practice in the face processing area to keep 95% to 99% of the total variance.

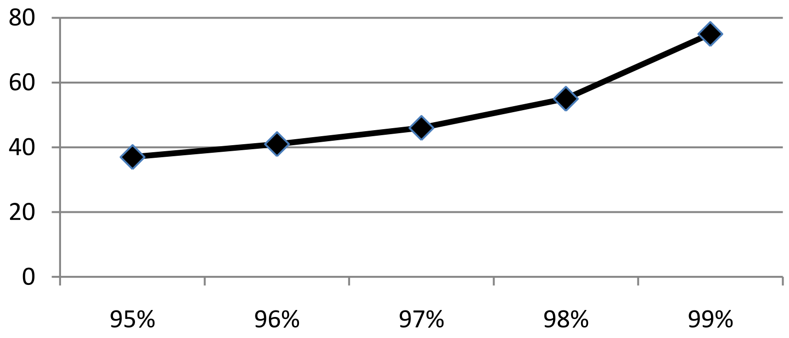


Figure 3.11 Principal components versus the percentage of variance retained.

In this work, we chose to retain 97% of the variance and that is 46 out of 260 principal components in the case of 3D facial surface normals as the baseline feature. The reason we chose 97% of the variance is simply because it is the middle point between 95% and 99%. Figure 3.9 shows number of principal components versus the percentage of variance retained.

At this stage, each $\mathbf{d}_i$ can be represented as a linear combination of the eigenvectors $\mathbf{u}_j$:

$$\mathbf{d}_i = \sum_{j=1}^{k} \omega_{ij} \mathbf{u}_j \tag{3.21}$$

This linearly convex combination is fully controlled by the shape parameters, ω_{ij} , given by:

$$\omega_{ij} = \mathbf{u}_j^T \mathbf{d}_i \tag{3.22}$$

Each set of training 3D facial points is represented on this basis as the vectors which are simply the projection of the data onto the subspace defined by the eigenvectors.

$$\varphi_i = \left[\omega_{i1}, \omega_{i2}, \dots, \omega_{ij}\right]^T \tag{3.23}$$

where$i = 1, \dots, j$.We can create a range of face shapes by varying the shape parameters ω.

$$\widehat{\boldsymbol{\Omega}} = \bar{\mathbf{m}} + \sum_{j=1}^{k} \omega_{ij}\, \mathbf{u}_j \tag{3.24}$$

3.4 Conclusions

In this chapter, we described our pre-processing steps which are common in the 3D face recognition field. The process began with the raw 3D facial points extraction from Bosphorus database and was followed by 3D face alignment. We used 22 manually annotated 3D facial landmarks and another 93 additional 3D facial points. We introduced the PCA computation that has been implemented in this work.

In the next chapter, both pre-processed data and the PCA algorithm will be used to perform 3D facial expression classification.

CHAPTER 4

3D FACIAL SURFACE NORMALS

Combinations of facial features form a human facial expression. Therefore, the deformation of facial features should be a suitable approach in order to determine the facial expressions shown by the subjects. The question is which 3D properties best describe the deformation of facial features so that a higher success rate of facial expression classification can be achieved. These 3D properties should be the significant properties that involves in at least six basic facial expressions and therefore the facial deformation can be easily observe.

The use of 3D facial geometric data and extracted 3D features for facial expression classification has not been widely studied. According to Gökberk et al., (2006), the most frequently used 3D facial features in 3D face classification are 3D point (also called point cloud feature), 3D feature distance (Soyel et al., 2007), curvature-based descriptors (Gökberk et al., 2006) and facial profile curves and 3D shape analysis (Soyel et al., 2007; Wang et al.,2006; Maalej et al., 2010). Generally, the 3D features are extracted and fed into the facial expression/face classification classifiers. In Soyel et al., (2007), distances between 3D facial landmarks were used directly as the input to classifiers in order to classify facial expression.

This chapter is about our approach to facial expression classification using 3D facial surface normals built from 3D facial points as the baseline feature. The preliminary results are presented in this chapter. After the common pre-processing step, 3D facial surface normals are extracted from 3D facial points. Shape weights are then computed from principal components formed by a set of 3D faces. For the purpose of evaluation, facial expression classification using 3D facial points and 3D distance measurements are also carried out. Using shape weights as the input, two chosen classifiers are used in facial expression classification: a simple nearest neighbour classifier and a Support Vector Machine (SVM).In this work, we are demonstrating the discriminative power of feature set and thus a simple classifier such as nearest neighbour and SVM is sufficient to be use.

The remainder of this chapter is organized as follows: In section 4.1 we explain the extraction of 3D facial surface normals from 3D facial points. The classification procedure which includes an explanation of the different classification approaches is described in section 4.2. The results are discussed in section 4.3 followed by a general discussion in section 4.4. Section 4.5 concludes this chapter.

4.1 3D Facial Surface Normals

Verzetti et al., (2012) in their work conclude that results on 3D facial features studies in the facial expression classification area were not as numerous as in face classification studies. Until now, only two significant studies have been frequently referred to in the literature: (1) Euclidean distances for six different facial features (Soyel et al., 2007) and (2) ratio of distances which are based on properties of the line segments connecting a set of particular facial features (Tang et al, 2008).

Our focus is on the advantage of having 3D facial points which are easily provided by the technology 3D scanners on the market. With the availability of raw 3D facial data, extraction of 3D facial surface normals is a straightforward task. For that reason, a model based on 3D facial surface normals is suggested as another 3D geometric measurement feature that could improve facial expression classification rate. A surface normal is a vector that is perpendicular to the tangent plane to a surface at a point .In addition, surface normals are also the features that encode the local directional gradient.

We believe that each expression has a consistent distribution of surface normals which distinguish it from other expressions. When the facial expression changes, the facial points positions also change which will cause its surface normal to change since surface normals is a derivative of facial landmark position. Surface normals are considered to be more accurate in describing facial surface changes compared to using facial points due to the fact that a surface normal is built from a 3D facial point as well as its neighbouring facial points. This has to do with the computation of surface normals which includes every neighbouring facial points of those particular facial features. The similar idea of taking into account the neighbouring facial points into the computation of a facial feature can be found in curvature-based descriptors and surface profiles. However, in this work, we are interested in how such facial expression variations manifest themselves in terms of changes in the field of 3D facial surface normals.

Ceolin (2012) in her work employed a 2.5D representation based on facial surface normals (also known as facial needle map) for gender and facial expression classification. The needle map used is a shape representation that is acquired from 2D intensity images using Shape-from-Shading. In our work, the surface normals are computed using the raw 3D facial points extracted from the Bosphorus database which is different from Ceolin's method. The surface normal of each triangular polygon is calculated using its corner points (i.e.: three 3D facial points).

Let $\mathbf{F_i}$ be a 3D face of the i^{th} subject. $\mathbf{F_i}$is represented by the set of 3D facial points

$$\mathbf{F_i} = \left\{p_1^i, p_2^i, \cdots, p_N^i\right\} \tag{4.1}$$

where the $p^i s$ are the (x, y, z)coordinate of each 3D facial point and N is the number of 3D facial points in the face. At each of the 3D facial points on the facial surface, we encode the facial points using their unit surface normal vectors,

$$\mathbf{\Omega_i} = \left\{s_1^i, s_{2,}^i, \cdots, s_N^i\right\} \tag{4.2}$$

where the$s_k^i s$ are the 3D unit normals $s_k^i = \{s_x, s_y, s_z\}$. Once all of the triangular polygon normals are calculated, the normal for each vertex in the triangulated face data is computed by averaging the normals of the neighbouring polygon surface normals. Figure 4.1 shows an example of the triangular polygon with its vertex normals.

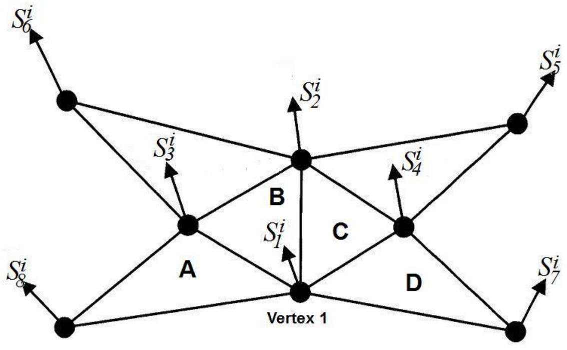


Figure 4.1 Example of triangular polygons with its vertex normals.

For example, *vertex 1* has four neighbouring meshes and in order to get S_1^i, all the normals of four triangular meshesthat include *vertex 1,2, 3, 4, 7* and *8*must be averaged, as shown in figure 4.1. In addition, to calculate surface normals, these vertices have to be in anti-clockwise winding order around the face. To calculate the normal for this mesh, we need to compute the cross product of these vectors followed by normalization to find the unit vector of the normal. Once all surface normals for mesh A, B, C and D are calculated then the average of them is calculated. Besides unweighted average, the other option is to use a weighted average by some importance factor (i.e.: area of a mesh). In this work, the unweighted average of normals is employed is done to avoid any other factor (such as the area of a mesh) to influence the normal computation. Figure 4.2 shows an example of surface normals of a 3D facial surface. The red lines denote the surface normals on a 3D facial surface.

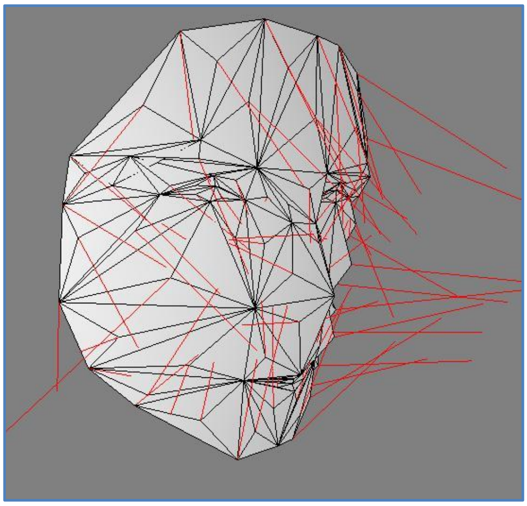


Figure 4.2 Surface normals on a 3D facial surface

4.2 Classification Procedure

After the extraction of 3D surface normals, these features are then used as the input in the statistical modelling stage. In this work, PCA is chosen and it has been described in Section 3.3. A 3D face probe$\boldsymbol{\partial}$, contains N set of 3D surface points. The probe is normalized where $\overline{\mathbf{m}}$ is the mean of 3D surface normals.

$$\boldsymbol{\gamma} = (\boldsymbol{\partial} - \overline{\mathbf{m}}) \tag{4.3}$$

This normalized probe is projected onto the eigenspace (the collection of Eigenvectors) and find out the weightsω_{ij}.

$$\omega_{ij} = \mathbf{u}_j^T \boldsymbol{\gamma} \tag{4.4}$$

for $j = 1, \dots, k$ and k is the number of training 3D faces and $\mathbf{u}_j^T$ are the eigenvectors. The normalized probe $\boldsymbol{\gamma}$ can be represented as:

$$\varphi_i = \left[\omega_{i1}, \omega_{i2}, \dots, \omega_{ij}\right]^T \tag{4.5}$$

In this work, we used two types of classification approaches: Nearest Neighbour Classifier and Support Vector Machine. As mentioned before, we are demonstrating the discriminative power of feature set and thus a simple classifier such as nearest neighbour and SVM is sufficient to be use. The following sections give details on the classification approach.

4.2.1 Nearest Neighbour Classifier

The nearest neighbour is the simplest of all algorithms for classifying a test sample. Given a training set with *n*-classes, a new training sample is classified by calculating the distance to the nearest training class.

L_1distance measurement (or known as Manhattan distance) is the sum of absolute differences between two vectors where L_1 is is achieved by walking ‘around the block’ in order to get from point x to point y. On the other hand, Euclidean distance is a L_2 norm type distance measurement where the length of the straight line between point x to point y is the distance.

Any distance measure can be used, however the most widely used distance metric is the Euclidean distance. In this work, Euclidean distance is used simply because the idea is to measure the shortest distance (i.e. the length of the straight line between two locations) of the projected probe face to the projected training sets. In general, the distance between two points **x** and **y** in a Euclidean space, $\Re^n$ is given by:

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{|x_i - y_i|^2} \tag{4.6}$$

To determine a face class that provides the best description of the probe, the feature vector of the face class which minimizes the Euclidean distance to vector is chosen

$$e_k = \|\varphi - \varphi_k\| \tag{4.7}$$

4.2.2 Support Vector Machine

The Support Vector Machine (SVM) is a supervised learning method that analyses and recognizes patterns. It is inherently a two-class/binary classifier. Given a set of training examples, each data is marked as belonging to one of the two classes and the SVM builds a model that categorizes the new example data to one class or another. SVM map an input sample to a high dimensional feature space and try to find an optimal hyperplane that minimizes the classification error for the training data using the non–linear transformation function (Sebald, et al., 2001). A new example is then predicted to belong to a class based on which side it falls in.

The boundaries between classes are hyperplanes (a line in figure 4.3). The best hyperplane for the SVM means the one with the largest margin (two dashlines in figure 4.3) between the two classes. Margin means the maximal width of the slab parallel to the hyperplanes that has no interior data points (Mathworks, 2012). The vectors near the hyperplanes are the support vectors and the support vectors are the width constrain of the margin. SVM analysis determines the hyperplanes that are oriented so that the margin between the support vectors is maximized.

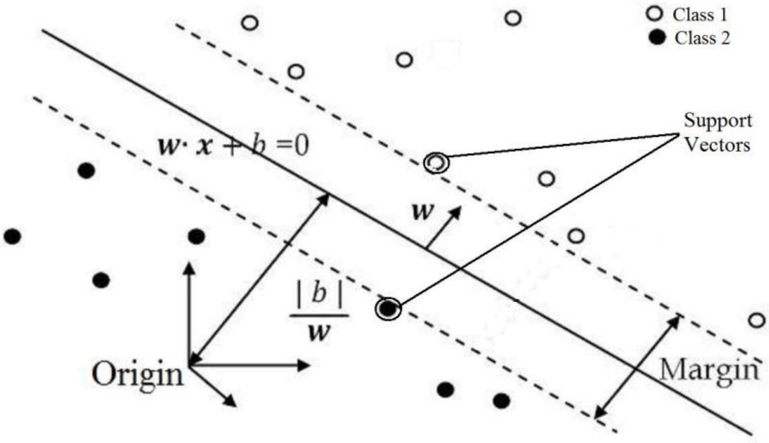


Figure 4.3 SVM concepts of two classes

Facial expression classification is not a binary classification as we have six basic facial expressions involved. Despite being inherently binary, SVM can also solve multiclass problems. Figure 4.4 shows a multiclass SVM problem where we have six classes which are separated by gaps. There are two ways of doing multiclass classification using SVMs: (i) one-versus-all classifiers (OVA) and (ii) one-versus-one classifiers (OVO). The concept of OVA is that a data point is considered as belonging to a class if and only if that particular class accepted it and other classes rejected it. On the other hand, OVO which is also known as "pairwise coupling", solves the multiclass classification by choosing a class that is selected by the most classifiers. In OVO, an SVM classifier is developed for each pair of classes resulting in $N(N-1)/2$ SVM classifiers.

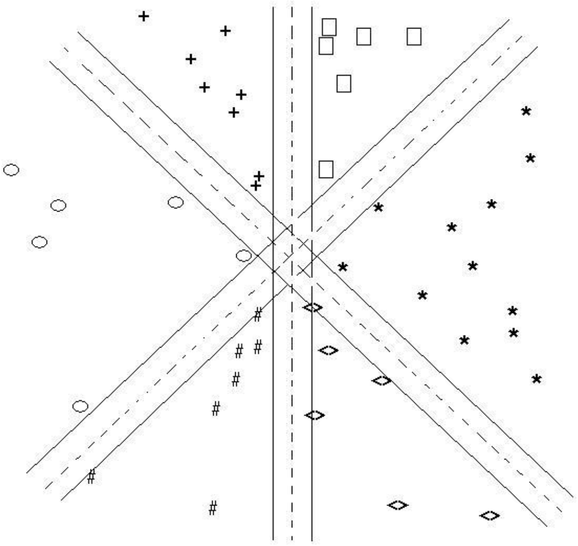


Figure 4.4 Multiclass SVM problem

The acknowledged drawback for OVO is it is more computationally intensive since it requires many SVM classifiers to be built. According to Weston and Watkins (1998), both approaches have the same accuracy. In this work, OVA is chosen for its simplicity, practicality and to avoid the intensive computation of OVO. However, software that is publicly free called SVM^{multiclass} (Joachim et al., 2009) is used in this work. A thorough explanation of the multiclass SVM implemented in SVM^{multiclass} can be found in Appendix A.

4.3 Experimental Technique

The aim of this experiment is to explore the potential of 3D facial surface normals in classifying six basic facial expressions. The experiment was done using 3–

fold cross-validation. This approach has been chosen as it will help to generalize the independent data set used in this study and thus reduce the risk of having overfitting problem. Overfitting occurs to the model that fits the data more than is warranted. It generally happens when a model is too complex (i.e. has too many parameters) and as a result, it will have poor predictive performance.

There are 64 subjects with 384 faces in total. 3-fold cross validation is chosen to maximize the number of subjects in one fold. With 3-fold cross validation, each fold have at least 21 subjects with six different facial expressions. Normally, most of the existing works used 10-fold cross validation. However, with 10-fold, each fold will only have at least 6 subjects which is too small for a sample size. All facial expressions of one subject belong to the same fold. This is to make sure when classification takes place, a facial expression of one particular subject will not be classified as another facial expression for the same subject. All facial expressions are used for both training and validation and each facial expression is used for validation exactly once.

The cross-validation process is repeated 3 times (based on the number of folds allocated). For each fold to be tested the remaining 2 folds belong to the training set. Two of these folds have 132 six-basic facial expressions and the remaining has only 126 expressions.

3D faces can be regarded as points in a high-dimensional space; they often lie on a subspace of much lower dimensionality, embedded in the high-dimensional image space. By employing dimensionality reduction, a more meaningful representation is established, therefore, addressing the issue of the ”curse of dimensionality” (Sharath et al., 2011). We used PCA as the dimensional reduction approach and the dimensional reduction goal is to find a low-dimensional representation of the data while still describing the data with sufficient accuracy. Dimensionality reduction techniques using linear transformations (i.e. PCA) have been very popular in determining the intrinsic dimensionality of the manifold as well as extracting its principal directions. For many datasets, the first several principal components explain most of the variance, so that the rest can be disregarded with minimal loss of information. As mentioned in the previous chapter, we chose to retain only 97% of the total variance. For 3D facial points, 97% of the total variance is equal to 31 out of 260 principal components and 22 out of 260 principle components for the 3D distance measurement feature 46 out of 260 principal components for 3D surface normals. 3D facial surface normals as the feature vector certainly capture a large amount of the facial expressions variation from different subjects compared to 3D facial points and 3D distance measurements.

3D facial surface normals of all training faces are used as the input to PCA and the probe 3D face is then projected to the face space defined by the eigenvectors of the PCA model. The shape weights from the PCA are used as the input to the classifiers. For classification, the nearest neighbour classifier and SVM are used.

For the purpose of evaluation, there are two feature vectors are used together with 3D facial surface normals: (1)3D facial points and (2) 3D distance measurements. The reason we chose 3D facial point is simply because 3D facial point is the raw information obtained from 3D space. 3D distance measurement feature is used as to duplicate Soyel's and Tang's idea. Soyel et al. (2010) used 83 facial features available in BU-3DFE to find distances between points. Tang et al. (2008) extracted a set of 96 features which consist of the normalized distances and slopes of the line segments connecting a subset of the 83 facial feature points. However, there are few facial landmarks that were used in Soyel and Tang's which are not provided in the database used in this work (i.e. facial points under the nose are not available for all 3D faces frontal profile) and therefore cannot be used as the feature vector. In our work, 3D distance measurements feature is the distance from each facial point to the nose tip.

The shape weights of 3D facial points and the 3D distance measurements must also be computed before the classification phase takes place. A total of 115 3D surface normals, 115 distance measurements vector and 115 3D facial points are used in this experiment.

Table 4.1 shows the results of using the simple nearest neighbour classifier with three different types of feature vectors as the baseline feature expressed by the confusion matrices. The emboldened numbers are the percentage of each expression correctly classified, along with where the misclassifications occurred. Since we are using a 3-fold cross–validation approach, the numbers in the table are the sum of the 3–fold cross–validation as a percentage.

Each confusion matrix shows where the main errors are introduced. For 3D facial points feature, the only expression that is correctly classified more than half the time is Happy expression. Two expressions that have lower than 30% correct classification are Disgust and Fear. The Sad expression has a slightly more than 30% correct classification. Surprise and Anger are roughly about the same percentage. The top two main confusions for Anger, Sad and Disgust come from incorrect classification as Anger or Sad or Disgust, showing the higher similarities of 3D facial points between these expressions.

Unlike 3D facial points features, the only expression that reaches more than 50% correct classification is Surprise when 3D distance facial point from nose tip are used as the feature vector. The results on Disgust, Fear and Sad are all lower than 30% whereas fear has the lowest percentage which is 20%. Anger and Happy have a similar result, which is 41–44% correct classification. Similar to 3D facial points, the main confusion for Anger, Sad and Disgust comes from incorrect classification as Anger or Sad or Disgust. However for Happy, three classes have the same misclassification rate which is Fear, Surprise and Sad.

However, using 3D facial surface normals as the baseline feature with nearest neighbour classifier definitely improves the facial expression classification rate. Disgust and Fear have the worst classification results where both are lower than 35%. Surprise and Happy are the expressions with correct classification of more than 50% and between both, Happy has the highest score. Anger and Sad are almost 50% correctly classified. The separation between expressions classes for 3D facial surface normals is quite a lot larger than those for 3D facial points and 3D distance measurements. Anger and Happy have never been classified as Surprise. However, the percentage for Surprise expression using 3D facial surface normals is slightly lower than using 3D distance measurements. We believe this is because for Surprise expression, the deformation most of the facial features are obvious, such as the eyebrows are raised, the upper eyelids are wide open, the lower eyelids relaxed and the jaw is opened (Pandzic & Forchheimer, 2002). Using 3D distance measurement feature, the deformation of the facial features is significant enough to be classified as Surprise.

Table 4.1 Confusion matrices of 3D facial expression classification using the nearest neighbour classifier Recall rates for each expression are shown in bold.

%	Anger	Disgust	Fear	Happy	Sad	Surprise
3D Facial Points						
Anger	**43.08%**	26.15%	15.38%	3.08%	20.00%	9.23%
Disgust	18.46%	**27.69%**	7.69%	10.77%	16.92%	3.08%
Fear	10.77%	9.23%	**23.08**	7.69%	9.23%	26.15%
Happy	3.08%	12.31%	6.15%	**61.54%**	12.31%	1.54%
Sad	20.00%	16.92%	9.23%	7.69%	**30.77%**	12.31%
Surprise	4.62%	7.69%	38.46%	9.23%	10.77%	**47.69%**
3D Distance Measurement						
Anger	**41.54%**	29.23%	4.62%	6.15%	20.00%	1.54%
Disgust	16.92%	**24.62%**	10.77%	9.23%	15.38%	4.62%
Fear	4.62%	7.69%	**20.00%**	13.85%	12.31%	21.54%
Happy	9.23%	7.69%	12.31%	**43.08%**	16.92%	9.23%
Sad	23.08%	23.08%	12.31%	13.85%	**27.69%**	6.15%
Surprise	4.62%	7.69%	40.00%	13.85%	7.69%	**56.92%**
3D Surface Normals						
Anger	**47.69%**	21.54%	15.38%	4.62%	21.54%	4.62%
Disgust	13.85%	**32.31%**	15.38%	12.31%	12.31%	6.15%

Fear	15.38%	12.31%	**26.15%**	3.08%	4.62%	26.15%
Happy	1.54%	3.08%	1.54%	**69.23%**	6.15%	1.54%
Sad	21.54%	15.38%	16.92%	10.77%	**43.08%**	6.15%
Surprise	0.00%	15.38%	24.62%	0.00%	12.31%	**55.38%**

Table 4.2 shows the results of using SVM with three different types of feature vectors as the baseline feature. For 3D facial points, no misclassification error for Happy and Surprise is the expression with the second highest correct classification. Fear has the lowest score in which most of it has been misclassified as Surprise. Disgust has less than 30% correct classification and it has been incorrectly classified as Anger and Happy. Most Anger expressions are misclassified as Sad and vice–versa.

3D distance measurement has three expressions which achieved more than 50% correct classification and they are Anger, Happy and Surprise. This result is slightly improved compared to using the nearest neighbour classifier. However, in contrast with other feature vectors in both classifiers, Disgust has the lowest rate of correctly classified expressions. Fear and Sad expressions also have low rates of correct classification which is 15.38% and 27.69% respectively. Again, as opposed to other feature vectors in both classification types, Happy is largely misclassified as Surprise. Fear and Disgust are never misclassified as Surprise.

Comparable with 3D facial points, 3D surface normals records a 100% correct classification for the Happy expression. Disgust expression has a slightly lower rate of correctly classified than using 3D facial points but it is still better than using 3D distance measurements. 3D facial surface normals have the lowest rate of correctly classified for the Surprise expression, however the classification rate is still more than 50%. For the rest of the expressions, 3D surface normals perform quite well compared to the two feature vectors. It is acceptable to confuse Disgust, Sad and Anger.

Table 4.2 Confusion matrices of 3D facial expression classification using SVM. Recall rates for each expression are shown in bold.

	Anger	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
3D Facial Points						
Anger	**56.92%**	27.69%	10.77%	0.00%	26.15%	3.08%
Disgust	6.15%	**27.69**%	9.23%	0.00%	4.62%	1.54%
Fear	3.08%	3.08%	**6.15**%	0.00%	6.15%	1.54%
Happy	0.00%	29.23%	3.08%	**100.00**%	9.23%	1.54%
Sad	30.77%	6.15%	7.69%	0.00%	**43.08%**	1.54%
Surprise	3.08%	6.15%	63.08%	0.00%	10.77%	**90.77%**
3D Distance Measurement						
Anger	**52.31%**	40.00%	3.08%	6.15%	40.00%	4.62%
Disgust	3.08%	**1.54%**	0.00%	4.62%	3.08%	0.00%
Fear	3.08%	7.69%	**15.38%**	9.23%	3.08%	4.62%
Happy	7.69%	27.69%	9.23%	**56.92%**	13.85%	7.69%
Sad	32.31%	15.38%	12.31%	1.54%	**27.69%**	4.62%
Surprise	1.54%	7.69%	60.00%	21.54%	12.31%	**78.46%**
3D Surface Normals						
Anger	**64.62%**	27.69%	7.69%	0.00%	23.08%	6.15%
Disgust	7.69%	**26.15%**	9.23%	0.00%	9.23%	3.08%

Fear	6.15%	4.62%	**29.23%**	0.00%	1.54%	9.23%
Happy	3.08%	26.15%	10.77%	**100.00%**	13.85%	20.00%
Sad	15.38%	12.31%	6.15%	0.00%	**47.69%**	7.69%
Surprise	3.08%	3.08%	36.92%	0.00%	4.62%	**53.85%**

4.4 Discussions

In general, most of the Fear expression examples are misclassified as Surprise and Surprise has mainly been incorrectly classified as Fear. This is an expected outcome due to the similarities in both expressions – mouth stretch, eyebrows raise and eyes open (Sandbach et al, 2012). However, the facial expression classification results of the three feature vectors using the simple nearest neighbour classifier are still not good enough compared to a recent study by Sandbach et al., (2012) based on 3D dynamic facial expression data. They used Gentleboost as the classifier and then a Hidden Markov Model (HMM) is employed in order to model the full temporal dynamics of the expression. Gentleboost is a binary classification for use with multilevel categorical predictors with a weighted least-square regression. The idea of Gentleboost is to put higher weights on the difficult images ().The highest score rate in Sandbach et al.'s 3D experimental work is Surprise and the lowest score rate belongs to Fear. In our work, using 3D facial surface normals with the nearest neighbour classifier, the highest score rate belongs to Happy whereas the lowest score is Fear. On the other hand, the highest correctly classified expression for 3D facial surface normals with SVM is also Happy, although the lowest correctly classified expression is Disgust. This is not a good comparison due to the obvious reason which is the data type difference. For dynamic data, an expressive sequence is expected to have its onset features followed by apex and offset features. This is like having a clue of what to anticipate from the data. Static facial expression data do not have these features and classification is made purely based on the facial features on a single image frame.

Soyel et al., (2007) used only a small set of 11 3D distance-vectors as the feature vector with neural network classifier. Nine distance-vectors were selected from the left side of the face as the repetitive selection on the right side is not needed due to symmetry. Tang et al., (2008) used images from the 3D facial expression database BU-3DFE and they are using features based on the ratio of distances between points which are based on properties of the line segments connecting a set of particular facial features. However, we cannot use both works (Soyel et al., 2007 and Tang et al., 2008) as a benchmark since they are using data from the BU-3DFE database. The different method to capture the 3D facial data for each database might influence the successfulness of the approach chosen. On the technical side, 3D facial data of Bosphorus database was captured with Inspeck Mega Capturor II 3D while for BU-3DFE database, 3DMD digitizer is used. In addition, Savran et al., (2008) mentioned that a special powder was applied on the subjects face to avoid specular reflections occurring on the face. To date, there is no work that has the same experiment set-up suitable for a valid comparison.

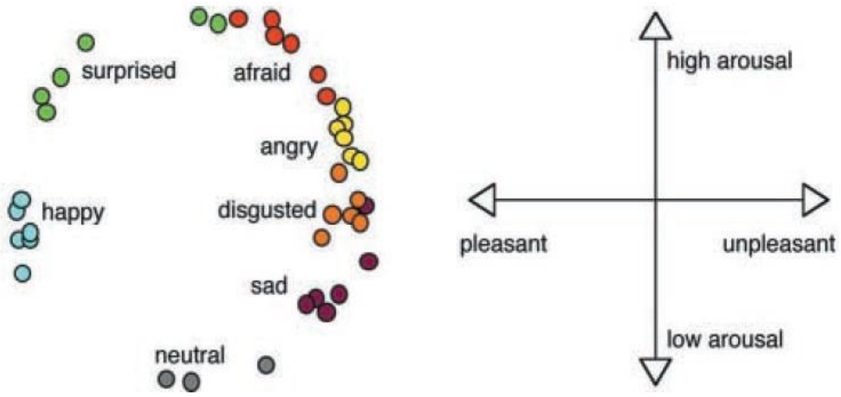
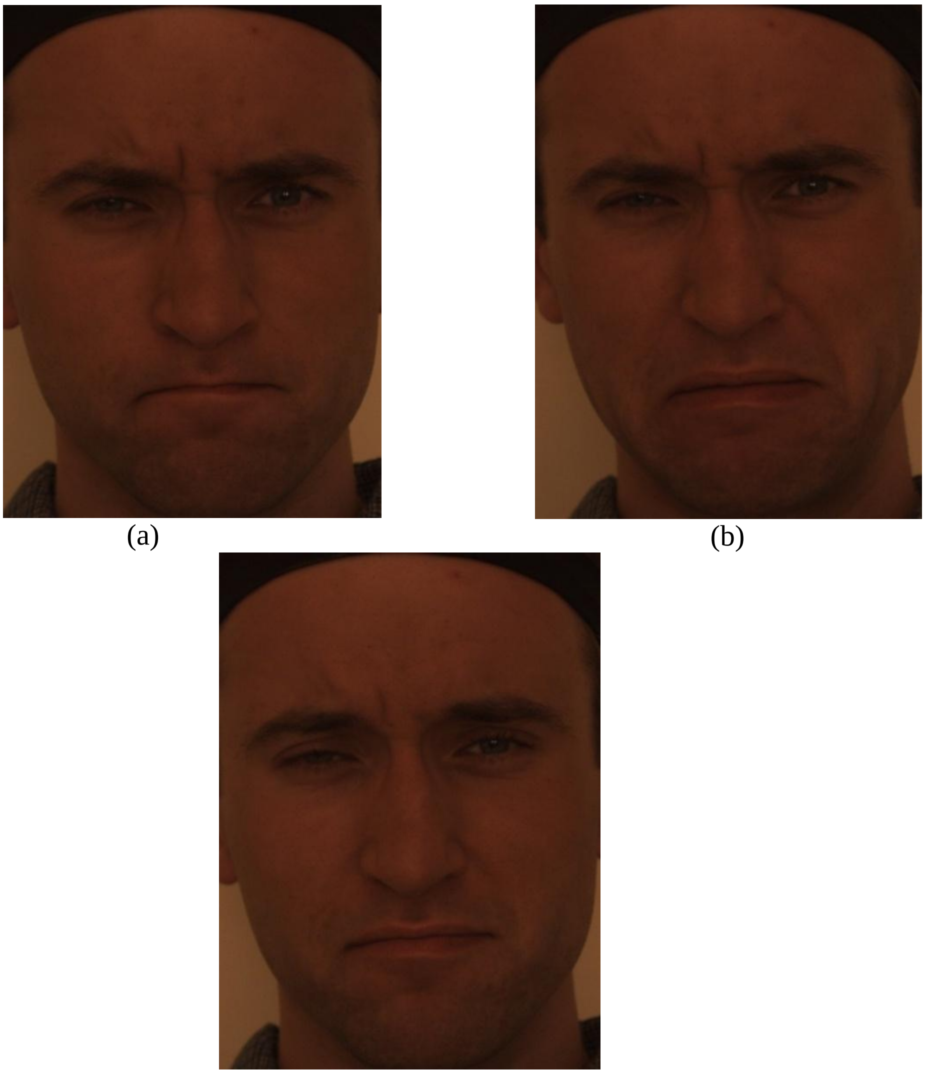


Figure 4.5 Dimensional structures of six basic facial expressions (Russell and Bullock, 1986)

Russell and Bullock (1986) stated that the facial expression categories have a fuzzy boundary between each other at the level of classification. Figure 4.5 shows the dimensional structure of six basic facial expressions. It is a two-dimensional structure where the *x*-axis is the pleasantness dimension and the *y*–axis is the arousal level dimension. Consistent with this claim, based on table 4.2 and table 4.3, Disgust is often mistakenly classified as Anger or Sad and vice versa.

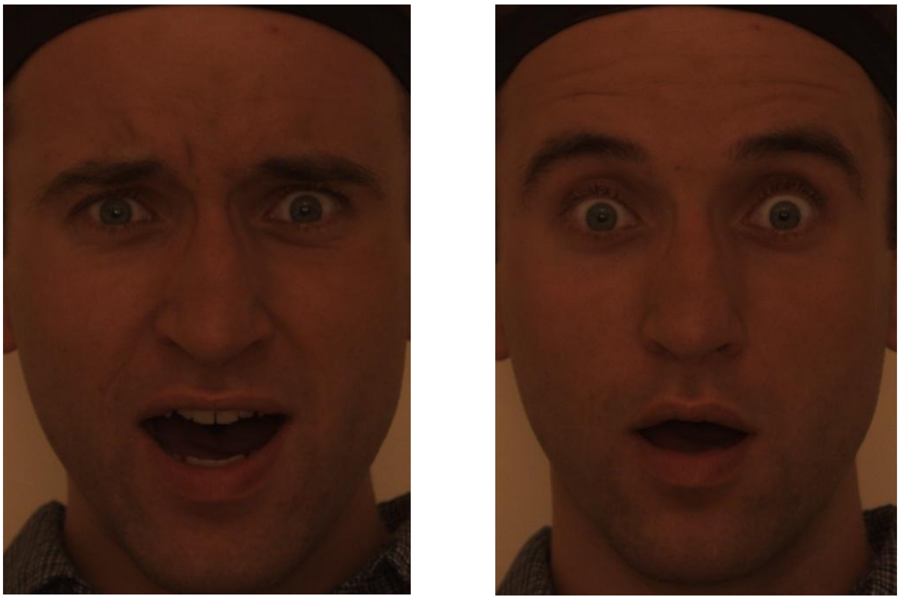


(a)

(b)

(c)

Figure 4.6 A subject showing (a) an Anger expression, (b) a Sad expression and (c) a Disgust expression



(a) (b)

Figure 4.7 A subject showing (a) a Fear expression and (b) a Surprise expression

Figures 4.6(a) - (c) show examples of Anger, Sad and Disgust expressions and we can clearly see similarities especially at the eyes, eyebrows, nose, cheek and forehead. We can see in figure 4.5, the gap between Disgust and Sad is not very clear and there is an overlap in classification space between Sad and Disgust (see the maroon-coloured dot among the orange-coloured dot). Therefore, Angry (Anger) could be wrongly classified as Sad. Surprise is the false positive of Fear/Afraid and vice versa. Figure 4.7 (a) and (b) show the similarities between the Fear and Surprise expressions which are obviously at the mouth and eyes area.

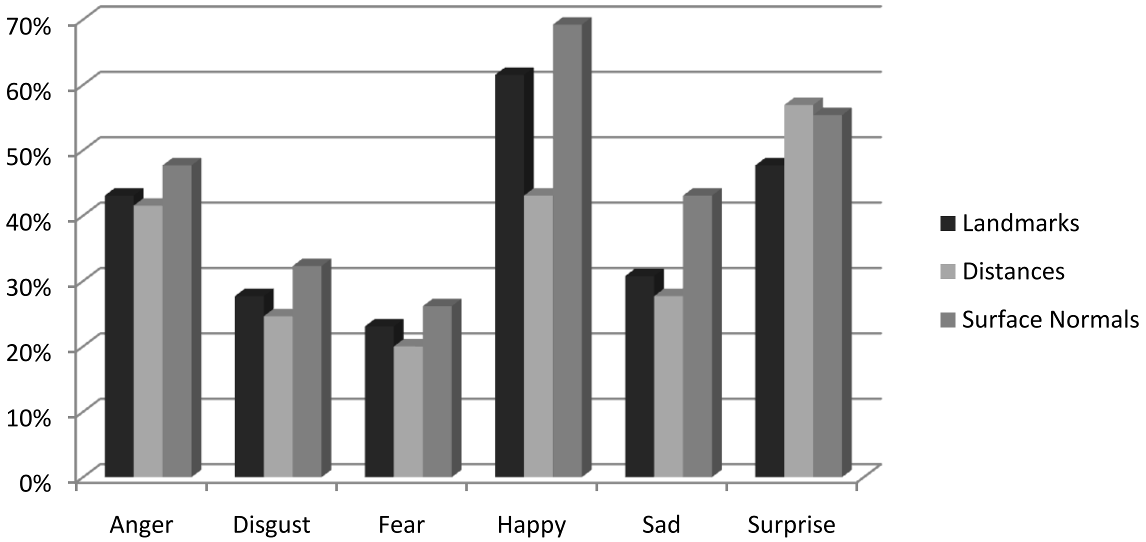


Figure 4.8 Overall 3D facial expression classification results for three types of feature vectors using the nearest neighbourhood classifier

Figure 4.8 shows the overall results on 3D facial expression classification results for three types of feature vectors using the nearest neighbour classifier. Evidently, 3D facial surface normals outperform 3D facial points and 3D distance measurement in all facial expressions except for the Surprise expression where it has the same results as 3D distance measurement. On the other hand, 3D facial points are better than 3D distance measurement in all expressions excluding the Surprise expression. This is because 3D distance measurements between each facial landmark and nose tip are able to encapsulate the Surprise expression which shows a significant distance between the mouth and cheek.

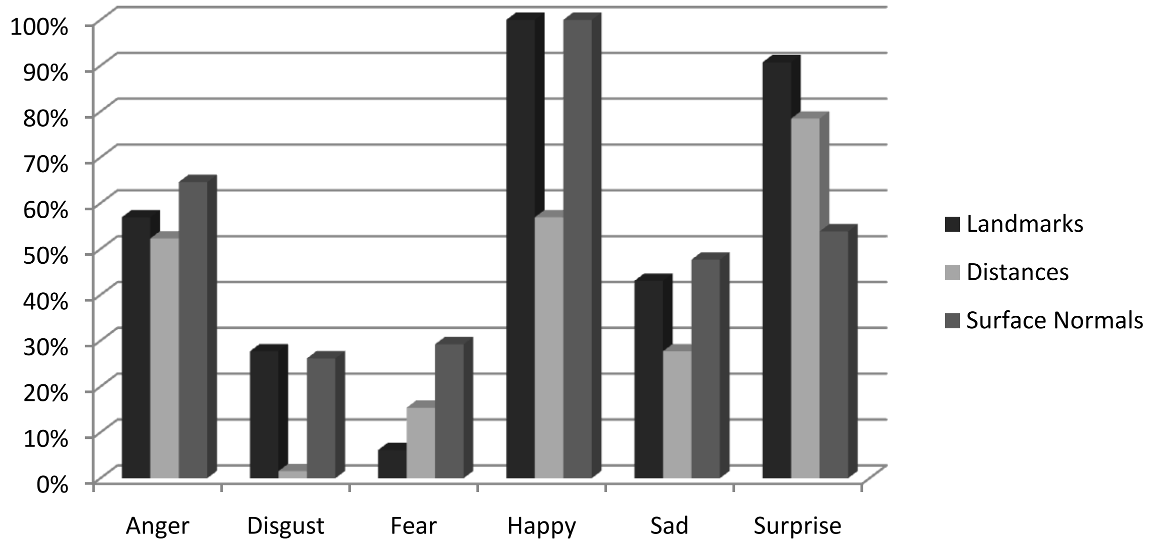


Figure 4.9 Overall 3D facial expression classification results for three types of feature vectors using simple SVM

Figure 4.9 shows the overall results on 3D facial expression classification results for three types of feature vectors using SVM. 3D facial points and 3D facial surface normals have the highest score for the Happy expression. Rather good results occur with Anger, Fear and Sad for 3D facial surface normals when compared against 3D facial points and 3D distance measurements. However, a large difference percentage on Surprise can be seen between 3D facial surface normals and the other two feature vectors. 3D facial surface normals achieve a better result for Fear expression but still fail to reach half of the correct classification.

Table 4.3 Average classification rates in our work

3D Facial Features	Nearest Neighbour Classifier	Support Vector Machine
3D Facial Points	39%	54%
3D Distance Measurement	36%	39%
3D Facial Surface Normals	46%	54%

Table 4.3 shows the average classification rates for the three feature vectors using nearest neighbour classifier and SVM. 3D facial surface normals obtained a good result though it still is slightly under half of the percentage. For SVM, 3D facial surface normals are similar to 3D facial points which is slightly over 50%. 3D distance measurements produce the worst classification for both classifiers.

4.5 Conclusions

Shape weights are computed and used as the input to a nearest neighbour and a SVM classifier. The probe 3D face is projected onto a sub space spanned by the PCA eigenvectors and its shape weights computed. A facial expression classification experiment using 3D facial points and 3D distance measurements was also carried out. Of all six experiments, the 3D facial surface normals approach performs well for both classifiers.

Even with only six basic facial expressions, the chance to mistakenly classify an expression is rather high. It is due to wide variations of facial expression between subjects as well as the differences between acted and natural examples. A facial expression is formed by a collection of facial features and to classify a facial expression from a whole face is like learning a global deformation of a face. Hence, we could not observe each of the facial feature deformations closely. We believe decomposing a face into several modules promotes the learning of a facial local structure and therefore the correlation between a facial feature and a facial expression is emphasised.

This motivates our work in the following chapter. In chapter 5, we focus on facial expression classification based on modular 3D surface normals.

CHAPTER 5

MODULAR 3D SURFACE NORMALS

In the previous chapter, we presented 3D facial expression classificationusing3D facial surface normal features using PCA. The outcomes of the experiments are only slightly improved compared to using 3D facial points or 3D distance measurements. We believe the reason for getting such results is the large variation of expression intensity between each facial expression that we have in the database which clearly affects the facial expression classification. This is due to the fact that human show their facial expression according to emotional level on which they experience. Intensity level of a facial expression is important as it will lead to a false impression of people's emotion if misinterpreted. Ekman and Friesen (1978) introduced five level of expression intensity while Yin et al. (2006) in their developed database used 4 different level of intensity.

A facial expression involves deformation of a collection of particular facial features and muscles. Classifying a facial expression from one whole face is like learning a global deformation of a face. The decomposition of a face into several modules promotes the learning of a face local structure and therefore the correlation between a facial feature and a facial expression is highlighted. Concurrently, the problem of the large variation of intensity for every facial feature might be solved as the face decomposition will help to put more weight on each facial feature.

The face decomposition approach has raised another question of how to determine the facial expression shown by the probe subject when one or more face modules do not agree on a facial expression classification. A facial expression must be determined based on a group of facial features. Certain facial features play a great role in deciding a facial expression classification. To solve this, we carried out a modular priority rank test using Adaptive Boosting (AdaBoost) which is a machine learning algorithm which works by combining several "weak learners". As a result, we have found the priority rank of each face module together with its weighting.

In our work, the framework begins with the face decomposition. When classifying the probe 3D face, the classification of each module was optimised independently and the results were then blended following a Weighted Voting Scheme (WVS) approach. Results from the modular priority rank test are used in the WVS. For the purpose of evaluation, we also carried out experiments using Majority Voting Scheme (MVS) in our work.

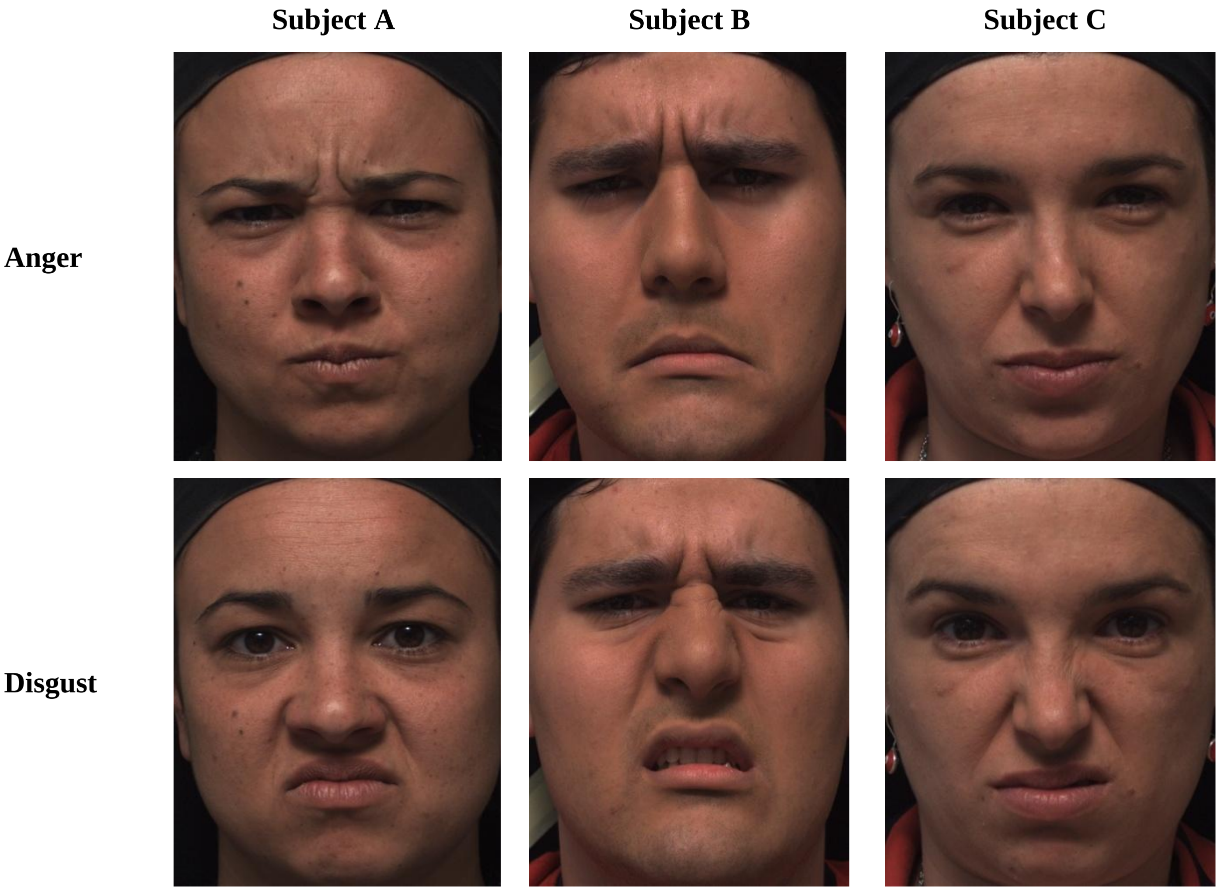
The remainder of this chapter is organized as follows: In section 5.1 we discuss the decomposition of a face into several modules. The next section explains the priority rank of each module. The integration of modules is dealt with using the WVS approach which is described in section 5.3. The results of this approach are given in section 5.4. In section 5.5, we discuss the results in general and section 5.6 concludes this chapter.

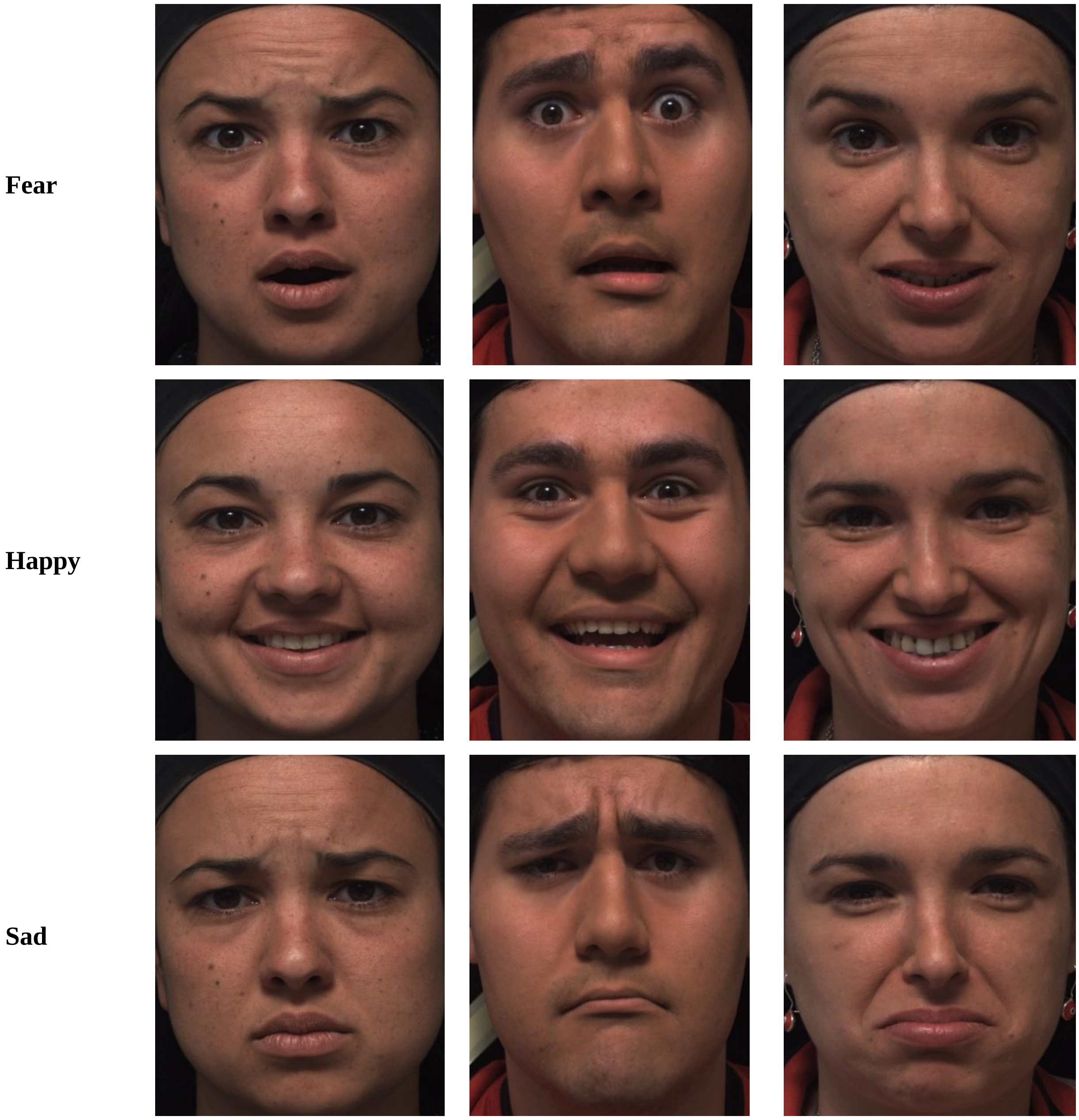
5. 1 Modular 3D Facial Surface Normals

When facial expression classification is carried out based on a whole face, the dependency between facial features is not being taken full advantage of. In other word, when a facial feature deforms due to facial expression changes, the other facial features which are connected to the facial feature are deforming as well. The level of intensity of the affected facial features is dependent to the intensity level of the facial feature that deform principally because it is the important deformation in a facial expression. We believe by decomposing a face into several modules, we are able to learn the local structure of each facial feature and their relationship between facial features and thus the classification of the facial expression should improve.

Different combinations of facial features and muscles produce different types of facial expressions. Facial expression varies from one person to another depending on their facial musculature, bone structure, facial features shapes, wrinkles, and so on. The intensity of facial features varies as well. Figure 5.1 shows three subjects taken from the Bosphorus database with six basic facial expressions and we can see the difference in their facial expressions by looking at each of the facial features. For

Anger and Disgust expressions, all subjects show a different type of mouth deformation. The size of the gap between the upper lip and lower lip in the Surprise expression is also dissimilar between subjects. Moreover, the deformations for eyebrows and mouth in the Sad expression for every subject are completely different. Hypothetically, by decomposing faces into modules, we could focus on capturing the variation as well as the intensity of facial features in each module.




Fear
Happy
Sad

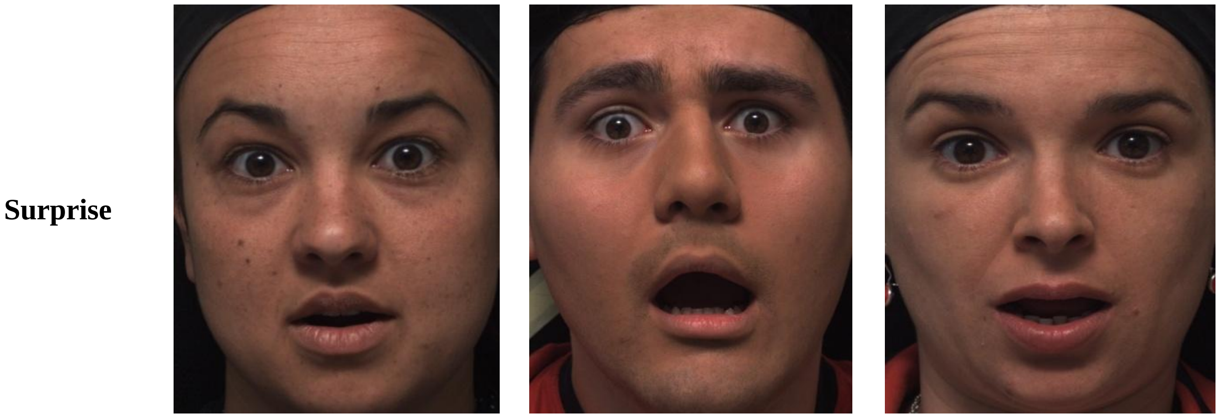


Figure 5.1 Three subjects with six basic facial expressions

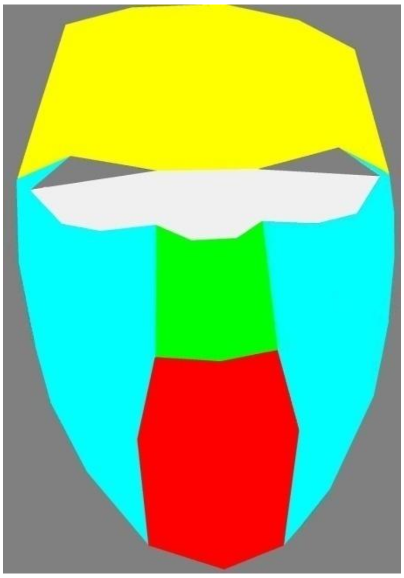


Figure 5.2 Face decomposition in our work

In this work, a face is divided into six modules as illustrated in figure 5.2 where one colour denotes one module. With this decomposition, we are considering that in any expressions, all facial features are involved regardless the importance of that facial features for a particular expression. This is simply because FER refers to the study of facial changes elicited as a result of relative changes in the shape and positions of the main components, such as eyebrows, eyelids, nose, lips, cheeks and chin (Rabiu et al., 2012). In addition, we could also measure the priority rank of the facial features. Tang's in their work made an assumption that the face is symmetrical, therefore only facial features on half of the face are considered. However, there are subjects who show an expression which was among the six basic expressions with unsymmetrical deformation of facial features, specifically the Eyebrows (refer to figure 5.2). Due to this kind of data, we have decided to include all facial features on both sides of the face.

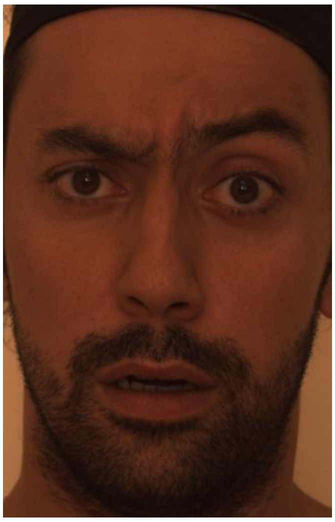


Figure 5.3 Subject shows the Fear expression with unsymmetrical deformation of the Eyebrows.

MPEG-4 is a method of defining compression of audio and visual (AV) digital data which also provides end users with a wide range of interaction with various animated object. Facial Animation Parameters (FAPs) are a set of parameters, used in animating MPEG-4 models that provides an alternative way of modelling facial expression and the underlying emotion (Zhang et al., 2008). FAPs give the measurement of muscular action relevant to the AUs. Each of the modules has a different set of facial features which are also associated with FAPs except for the forehead. No facial feature in the forehead is involved in the deformation in six basic facial expressions. However we also include the forehead module in this work because we wanted to see how it influences each of the expressions.

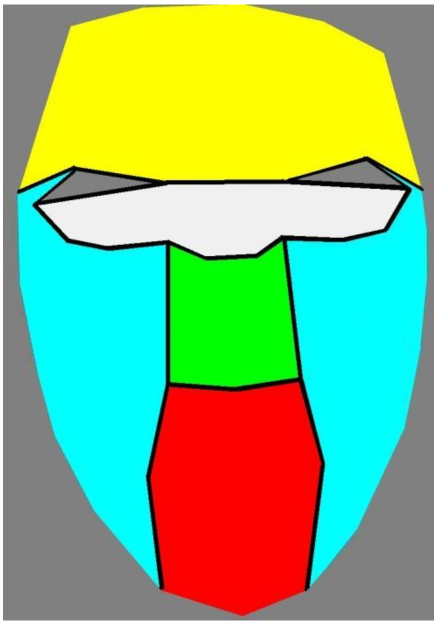


Figure 5.4 Facial features that belong to more than one module are on the black-coloured edge

Another interesting question is the facial features that are on the boundaries of a few modules. For example, the left and right nose saddles belong to the Nose and Cheeks module. We cannot just simply put these boundary facial features in only one module and completely ignore its effect on the other module. For that reason, we decided to include those boundary facial features in any modules that have them. Figure 5.3 shows the black-coloured edge where the boundary facial features that belong to more one module are positioned.

Table 5.1 FAPs in every face module

Module	Facial Animation Parameters (FAP)	Action Unit
Nose	stretch_l_nose, stretch_r_nose,	AU9
Cheek	lift_l_cheek, lift_r_cheek	AU6
Lips	open_jaw, raise_b_midlip, stretch_r_cornerlip, raise_l_cornerlip, raise_r_cornerlip, push_b_lip, stretch _l_cornerlip, depress_chin, raise_b_midlip_o, stretch_r_cornerlip_o, raise_l_cornerlip_o, raise_r_cornerlip_o, stretch _l_cornerlip_o	AU25, AU26, AU27, AU12, AU17
Eyes	close_t_r_eyelid, close_t_l_eyelid, close_b_r_eyelid, close_b_l_eyelid,	AU5, AU7
Eyebrows	raise_r_i_eyebrow, raise_r_i_eyebrow, squeeze_r_eyebrow , raise_l_i_eyebrow, raise_l_o_eyebrow, squeeze_l_ eyebrow,	AU1, AU2, AU4

Table 5.1 shows the FAPs in every face module. Several works in the facial expression analysis area state different Action Units (AUs) are involved in six basic facial expressions. The list of AUs for six basic facial expressions is shown in Chapter 2. According to Raouzaiou et al., (2002), the deformation of the cheek module only occurs in the Happy expression. On the other hand, Zhang et al., (2008) state that the deformation of the nose module occurs in Anger and Disgust. However, in Raouzaiou et al.'s work, they did not include those FAPs in the nose module to describe the Angry and Disgust expression. With face decomposition, the FAPs/AUs for each facial expression is able to be observe.

5.2 Modular Priority Rank

As mentioned before, in this work, the facial expression classification for each module is done independently. Therefore, each module is expected to produce a different classification from other modules. The problem arises when to decide which facial expression is being portrayed by the 3D probe in general when the results of each module classification are different. Each module has its own impact level to the six basic expressions. The simplest way to find which modules are affecting facial expression is to find their priority weighting.

Silapachote et al., (2005) select facial features using Adaptive Boosting (AdaBoost) and successfully single out the discriminative features. Consequently the discriminative image regions without relying on *a priori* domain knowledge are also determined. Their experiment shows that AdaBoost has successfully picked the mouth and the eyes as being informative and it also discarded irrelevant regions. However, they used 2D images as their data and only four expressions are considered in their experiments which are Neutral, Smile, Anger and Scream.

In this work, we also used AdaBoost to determine the important facial features. Our work differs from Silapachote et al.'s work in that we used 3D facial raw points as the feature vectors in AdaBoost whereas Silapachote et al (2005) used histograms of Gabor and Gaussian derivative responses as appearance features. We also aim to have a weighting priority for each of the modules which Silapachote et al do not provide in their results. 3D facial raw points is chosen because it is the feature vector that is obtained directly from the 3D acquisition device. Any discrepancy in the weighting results due to wrong computation of feature vector used will affect the classification result.

AdaBoost was first introduced by Freund and Schapire (1997). AdaBoost is easy to implement and provides feature selection on very large sets of features. Even though it offers a fairly good generalization, an over-fitting problem can occur in the presence of noise. Given $(x_1, y_1), \dots, (x_n, y_n)$ where $x_i \in \chi$ with its binary class $y_i \in \{-1,1\}$ and a weak classifier $h_t: \chi \to \{-1,1\}$. AdaBoost works by combining several "weak" learners$h_t(x)$ in a linear combination

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{5.1}$$

to construct a strong learner. It is suitable only for binary classification. Initially, all weights α_t are set equally and AdaBoost chooses the learner that classifies most data accurately. The process of AdaBoost maintains a distribution on the training samples. For the next T iterations, the weight of each weak learner is determined, the distribution is updated and the data is re-weighted to increase the "importance" of misclassified training samples. For every iteration, the performance of each single weak learner h_j on all training samples is assessed using the weighted error defined as

$$\epsilon_t = \sum_i w_t(i) 1_{\{h_j(x_i) \neq y_i\}} \tag{5.2}$$

At the end of each iteration, the learner h_u with the lowest error rate ϵ_t based on equation 5.2 is selected and stored as the best classifier h_t at iteration t. Parameter α_t is computed as follows:

$$\alpha_t = \frac{1}{2} \ln\left(\frac{1 - \epsilon_t}{\epsilon_t}\right) \tag{5.3}$$

AdaBoost later is extended to a multi-boosting classification by Schapire and Singer (1999) known as AdaBoost.MH. In our work, we directly used the MultiBoost software package developed by Benbouzid et al., (2012). MultiBoost implements AdaBoost with Multi-class Hamming Lost (AdaBoost.MH). AdaBoost.MH is a type of AdaBoost algorithm that converts the multi-class problem into multiple binary problems with an additional feature defined by the set of class labels. We set the iterations round to 200. At each iteration, a strong learner is determined together with its weight and here, a strong learner is referred to a 3D facial landmark. The weight of the 3D facial points that belong to the same module is stored and at the end of the iteration, the weights are summed up. Each module now has a weight and the weight must be normalized to ensure the total weight for all modules is equal to 1.

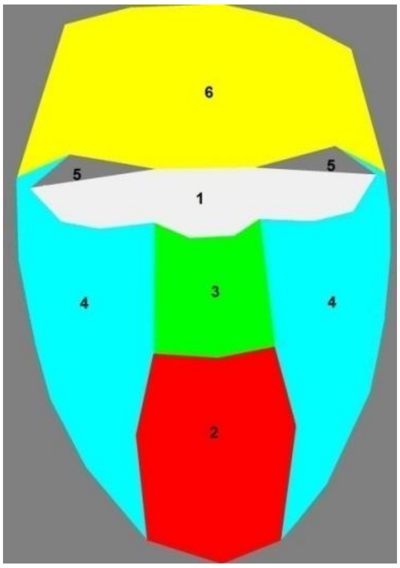


Figure 5.5 A face decomposition with priority rank for each module

Table 5.2 Face modules and its weight

Rank	Module	Weight
1	Eyebrows	0.1984
2	Mouth	0.1848
3	Eyes	0.1740
4	Cheeks	0.1719
5	Nose	0.1419
6	Forehead	0.1291

Figure 5.5 shows the face decomposition with the priority rank obtained from the multi-boosting result and Table 5.2 shows the weighting for each face module. The number on each module denotes their priority rank. In agreement with Silapachote et al., (2005), the eyebrows are the most important facial region in facial expression classification, followed by the mouth and the eyes region. The Nose module only significantly deforms in the Disgust expression, which explains it is a rather low weighting. As expected, the forehead module has the lowest weighting.

5.3 Weighted Voting Scheme

A weighted voting system (WVS) is based on the idea that not all voters are equal. In other words, one in which the preferences of some voters carry more weight than the preferences of other voters. In our work, WVS is used to determine the facial expression class for the 3D probe based on the class that has been determined for each face module. As mentioned in the previous section, each module (voter) has its own weight.

Figure 5.6 shows the framework of WVS for modular 3D facial expression classification. Initially, the face is decomposed into six modules and the facial expression classification for each module is done independently. The 3D geometrical feature for each module is passed to the PCA algorithm to generate the shape weights used as the feature vectors in the SVM as well as nearest neighbour classifier. Due to this independent mode, each module is expected to produce a different classification from other modules. Each module now has been classified as showing one of the facial expressions. The class information for each module together with its weight is passed into the Votes Counter algorithm. In the Votes Counter algorithm, the weight of the modules that belong to the same facial expression class is summed up. At this stage, each facial expression class has its accumulated weight and the facial expression class with the highest weight is considered as the facial expression shown by the 3D probe.

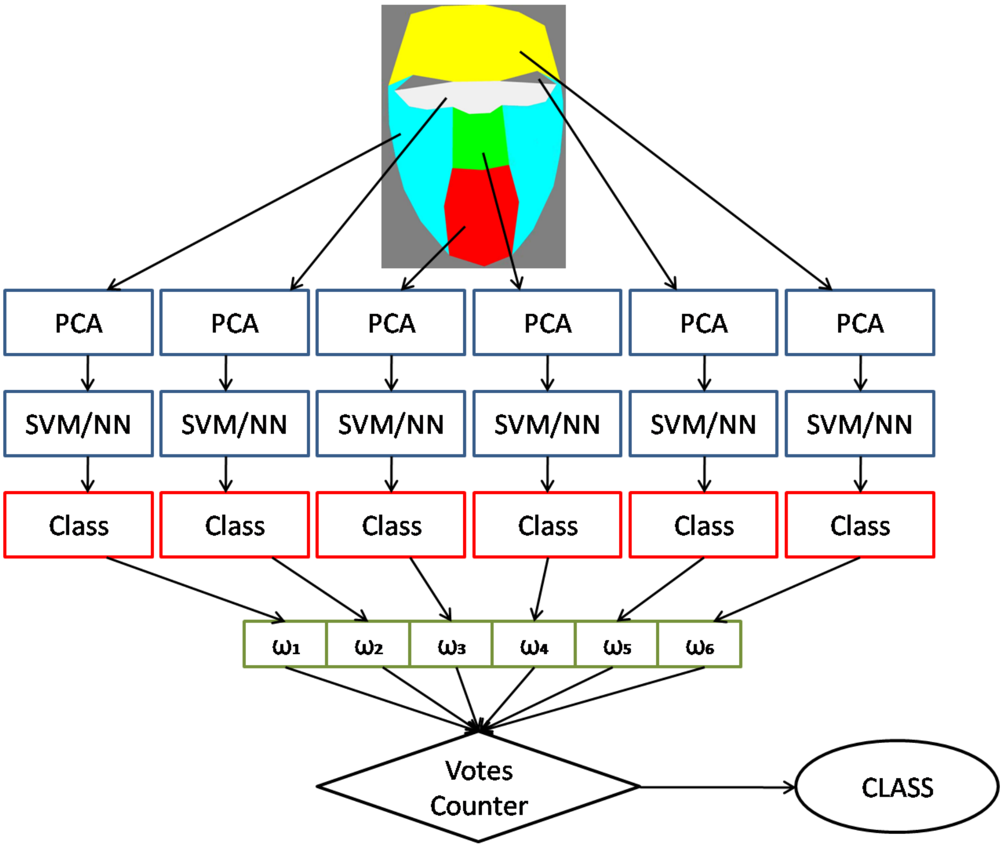


Figure 5.6 Weighted Voting Scheme for modular 3D facial expression classification

Table 5.3 An example of WVS problem

	Module	Weight	Facial Expression
1	Eyebrows	0.1984	Surprise
2	Mouth	0.1848	Disgust
3	Eyes	0.1740	Happy
4	Cheeks	0.1719	Happy
5	Nose	0.1419	Happy
6	Forehead	0.1291	Surprise

Table 5.3 shows an example of the results of the WVS algorithm. Eyes, Cheeks and Nose have been classified as Happy and the total weights for these three modules are 0.4878 while Mouth is the only module classified as Disgust which means its weight is solely 0.185. Forehead and Eyebrows are classified as Surprise with a total weight of 0.3275. The other three facial expressions namely Anger, Fear and Sad have a zero weight. WVS in this case voted Happy expression as the facial expression shown by the 3D probe.

In the case where each module has been classified as a completely different facial expression from all the others, WVS will vote on the final facial expression class based on the module with the largest weight.

5.4 Results

In this section, the results of the experiments are described and the analyses are presented. This section is divided into two sub-sections. The first sub-section discusses results for the modular 3D facial expression classification and the second sub-section discusses the WVS and MVS results.

5.4.1 Modular 3D Facial Expression Results

Similar to the previous experiments in chapter four, the experiments in this chapter were done using 3–fold cross-validation. The emboldened numbers are the percentage of each expression that was correctly classified, along with where the misclassifications occurred. Since we are using a 3–fold cross–validation approach, the numbers in the table are the sum of the 3-fold cross–validation expressed as a percentage.

Table 5.4 shows the modular 3D facial expression classification results using the nearest neighbour classifier with 3D facial points as the baseline feature in the confusion matrices. Not even one facial expression in the Eyebrows, Eyes and Forehead modules recorded more than a 37% rate of correct classification. The highest rate is the Happy expression for the Nose module with a correct classification rate of 73.85%. The Mouth, Cheeks and Nose modules achieved more than 50% success classification rate for the Happy expression only. In the Mouth module, Happy has never been misclassified as Anger and vice versa. The average success rates for all modules are between 22% - 39%.

Table 5.4 Confusion matrices of modular 3D facial expression classification using 3D facial points with nearest neighbour classifier. Recall rates for each expression are shown in bold.

Eyebrows Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**29.23%**	23.08%	13.85%	18.46%	13.85%	9.23%
Disgust	18.46%	**16.92%**	9.23%	15.38%	13.85%	7.69%
Fear	9.23%	16.92%	**16.92%**	16.92%	16.92%	15.38%
Happy	20.00%	24.62%	27.69%	**21.54%**	21.54%	18.46%
Sad	12.31%	15.38%	13.85%	9.23%	**15.38%**	13.85%
Surprise	10.77%	3.08%	18.46%	18.46%	18.46%	**35.38%**

Mouth Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**36.92%**	12.31%	12.31%	0.00%	35.38%	9.23%
Disgust	15.38%	**36.92%**	9.23%	24.62%	23.08%	15.38%
Fear	12.31%	7.69%	**20.00%**	6.15%	6.15%	18.46%
Happy	0.00%	12.31%	3.08%	**64.62%**	1.54%	1.54%
Sad	33.85%	23.08%	15.38%	1.54%	**27.69%**	7.69%
Surprise	1.54%	7.69%	40.00%	3.08%	6.15%	**47.69%**

Eyes Module	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**26.15%**	29.23	16.92	16.92	23.08	6.15
Disgust	21.54%	**21.54**	10.77	13.85	7.69	4.62
Fear	13.85%	10.77	**10.77**	10.77	15.38	23.08
Happy	20.00%	24.62	30.77	**26.15**	18.46	18.46
Sad	12.31%	10.77	12.31	12.31	**21.54**	20.00
Surprise	12.31%	10.77	12.31	12.31	21.54	**20.00**

Cheeks Module	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**29.23**	21.54	15.38	4.62	23.08	9.23
Disgust	23.08	**29.23**	10.77	13.85	24.62	6.15
Fear	12.31	15.38	**30.77**	9.23	9.23	24.62
Happy	3.08	4.62	1.54	**56.92**	9.23	6.15
Sad	23.08	20.00	13.85	10.77	**20.00**	7.69
Surprise	23.08	20.00	13.85	10.77	20.00	**7.69**

Nose Module	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**21.54**	13.85	12.31	1.54	23.08	9.23
Disgust	16.92	**30.77**	10.77	9.23	20.00	10.77
Fear	18.46	12.31	**35.38**	3.08	13.85	24.62
Happy	3.08	6.15	1.54	**73.85**	3.08	3.08
Sad	27.69	24.62	16.92	7.69	**24.62**	21.54
Surprise	12.31	12.31	23.08	4.62	15.38	**30.77**

Forehead Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**33.85**	15.38	10.77	9.23	27.69	7.69
Disgust	9.23	**27.69**	13.85	16.92	12.31	10.77
Fear	12.31	6.15	**23.08**	9.23	13.85	16.92
Happy	12.31	16.92	18.46	**36.92**	12.31	6.15
Sad	21.54	18.46	13.85	15.38	**18.46**	21.54
Surprise	21.54	18.46	13.85	15.38	18.46	**21.54**

Table 5.5 shows the modular 3D facial expression classification results using the nearest neighbour classifier with 3D distance measurements as the baseline feature in the confusion matrices. No expressions in any module except the Mouth module achieved a correct classification rate of more than 39%. In the Mouth module, two expressions have been 50% correctly classified namely Anger and Happy. The average success rates for all modules are between 18% - 37% that is lower than using 3D facial landmarks. In the Mouth module, Happy has never been misclassified as Anger and vice versa.

Table 5.5 Confusion matrices of modular 3D facial expression classification using 3D distance measurement with nearest neighbour classifier in percentage. Recall rates for each expression are shown in bold.

Eyebrows Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**18.46**	13.85	16.67	12.50	23.08	16.92
Disgust	20.00	**18.46**	18.18	17.19	16.92	13.85
Fear	16.92	15.38	**6.06**	12.50	23.08	10.77
Happy	12.31	21.54	18.18	**29.69**	7.69	16.92
Sad	15.38	20.00	19.70	10.94	**9.23**	15.38
Surprise	16.92	10.77	21.21	17.19	20.00	**26.15**

Mouth Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**50.77**	18.46	6.06	0.00	36.92	4.62
Disgust	10.77	**26.15**	16.67	17.19	23.08	12.31
Fear	7.69	18.46	**22.73**	12.50	6.15	27.69
Happy	0.00	13.85	10.61	**50.00**	4.62	7.69
Sad	27.69	18.46	7.58	7.81	**27.69**	3.08
Surprise	3.08	4.62	36.36	12.50	1.54	**44.62**

Eyes Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**30.77**	16.92	7.69	12.31	20.00	6.15
Disgust	16.92	**26.15**	9.23	18.46	24.62	3.08
Fear	6.15	7.69	**32.31**	18.46	10.77	26.15
Happy	16.92	18.46	12.31	**15.38**	18.46	7.69
Sad	20.00	20.00	7.69	13.85	**15.38**	12.31
Surprise	20.00	20.00	7.69	13.85	15.38	**12.31**

Cheeks Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**29.23**	29.23	12.31	12.31	27.69	7.69

Disgust	16.92	**27.69**	16.92	7.69	13.85	9.23
Fear	16.92	16.92	**29.23**	20.00	12.31	38.46
Happy	6.15	9.23	10.77	**38.46**	12.31	6.15
Sad	26.15	13.85	6.15	10.77	**30.77**	6.15
Surprise	26.15	13.85	6.15	10.77	30.77	**6.15**

Nose Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**27.69**	7.69	9.09	1.56	9.23	9.23
Disgust	10.77	**13.85**	15.15	23.44	20.00	20.00
Fear	15.38	20.00	**22.73**	17.19	21.54	21.54
Happy	6.15	23.08	15.15	**28.13**	4.62	16.92
Sad	29.23	20.00	15.15	7.81	**24.62**	15.38
Surprise	10.77	15.38	22.73	21.88	20.00	**16.92**

Forehead Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**16.92**	18.46	10.77	15.38	21.54	4.62
Disgust	20.00	**20.00**	7.69	12.31	16.92	16.92
Fear	10.77	10.77	**15.38**	15.38	16.92	16.92
Happy	18.46	21.54	23.08	**27.69**	13.85	10.77
Sad	24.62	10.77	18.46	15.38	**16.92**	18.46
Surprise	24.62	10.77	18.46	15.38	16.92	**18.46**

Table 5.6 shows the modular 3D facial expression classification results using the nearest neighbour classifier with 3D facial surface normal as the baseline feature in the confusion matrices. The Happy expression in all modules achieved more than 45%

success classification rate. In the Mouth module, another two expressions have been 45% correctly classified which are Sad and Surprise. The average success rates for all modules are between 28% - 43% which is better than using 3D facial points and 3D distance measurements. In the Mouth module, Happy has never been misclassified as Anger and Sad while Sad has never been misclassified as Surprise. In addition, Fear has never been misclassified as Happy in the Cheeks module.

Table 5.6 Confusion matrices of modular 3D facial expression classification using 3D facial surface normals with nearest neighbour classifier. Recall rates for each expression are shown in bold.

Eyebrows Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**36.92**	21.54	13.85	3.08	6.15	6.15
Disgust	15.38	**24.62**	12.31	10.77	18.46	6.15
Fear	12.31	13.85	**4.62**	7.69	18.46	16.92
Happy	9.23	18.46	18.46	**46.15**	12.31	27.69
Sad	9.23	9.23	20.00	6.15	**27.69**	13.85
Surprise	16.92	12.31	30.77	26.15	16.92	**29.23**

Mouth Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**38.46**	16.92	4.62	0.00	36.92	7.69
Disgust	12.31	**32.31**	21.54	9.23	13.85	7.69
Fear	6.15	9.23	**26.15**	9.23	1.54	30.77
Happy	1.54	16.92	6.15	**70.77**	3.08	6.15
Sad	40.00	18.46	7.69	0.00	**44.62**	3.08
Surprise	1.54	6.15	33.85	10.77	0.00	**44.62**

Eyes Module	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**38.46**	32.31	1.54	4.69	12.12	4.62
Disgust	13.85	**20.00**	7.69	9.38	10.61	1.54
Fear	10.77	6.15	**26.15**	7.81	13.64	24.62
Happy	10.77	23.08	18.46	**53.13**	21.21	9.23
Sad	15.38	13.85	10.77	18.75	**33.33**	13.85
Surprise	10.77	4.62	35.38	6.25	9.09	**46.15**

Cheeks Module	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**36.92**	20.00	10.77	6.15	27.69	9.23
Disgust	10.77	**29.23**	12.31	15.38	15.38	7.69
Fear	18.46	10.77	**27.69**	3.08	16.92	23.08
Happy	1.54	4.62	0.00	**55.38**	7.69	3.08
Sad	15.38	20.00	12.31	12.31	**21.54**	15.38
Surprise	16.92	15.38	36.92	7.69	10.77	**41.54**

Nose Module	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**40.00**	18.46	6.15	3.08	23.08	9.23
Disgust	15.38	**36.92**	7.69	15.38	10.77	10.77
Fear	10.77	12.31	**32.31**	7.69	7.69	35.38
Happy	4.62	13.85	7.69	**46.15**	9.23	1.54
Sad	23.08	13.85	20.00	18.46	**32.31**	12.31
Surprise	6.15	4.62	26.15	9.23	16.92	**30.77**

Forehead Module	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**41.54**	18.46	20.00	3.08	12.31	7.69
Disgust	15.38	**27.69**	10.77	9.23	12.31	10.77
Fear	9.23	15.38	**15.38**	4.62	23.08	21.54

Happy	4.62	12.31	7.69	**55.38**	9.23	13.85
Sad	23.08	10.77	15.38	13.85	**21.54**	10.77
Surprise	6.15	15.38	30.77	13.85	21.54	**35.38**

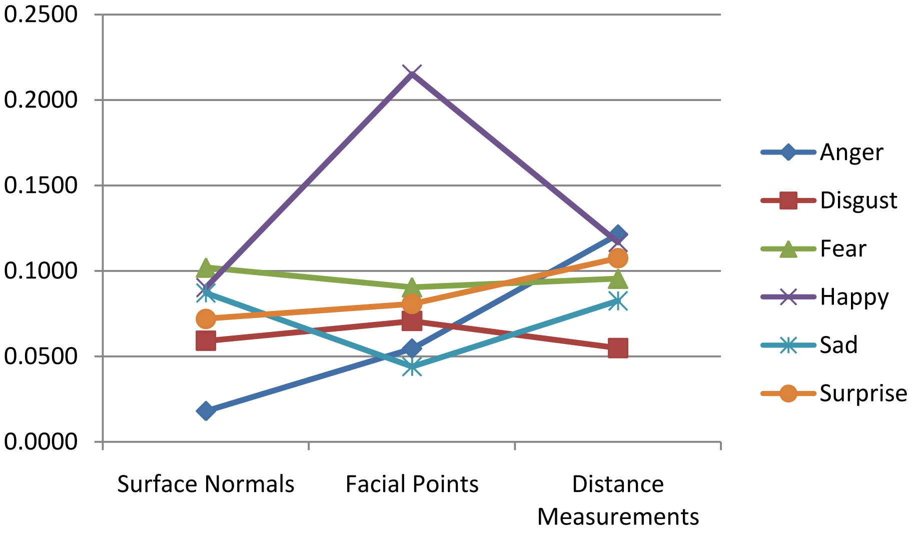


Figure 5.7 Consistency test using standard deviations of 3D facial features compiled from the success rate of the modules across facial expressions of nearest neighbour classifier.

Figure 5.7 shows consistency test using standard deviations of 3D facial features compiled from the success rate of the modules across facial expressions using nearest neighbour classifier. The following graphs (5.8-5.13) discuss the performance of the three 3D facial features for each expression and each of those graphs will refer to figure 5.7 in order to verify with the standard deviation which represents the consistency of the 3D facial feature across the modules.

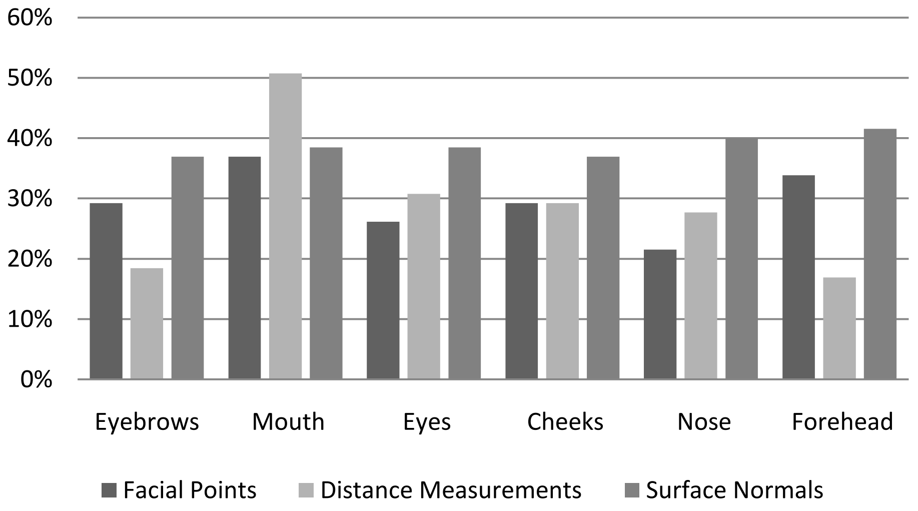


Figure 5.8 Success rate of modular facial expression classification for the Anger expression using nearest neighbour classifier.

Figure 5.8 shows the success rate of modular facial expression classification for the Anger expression using the nearest neighbour classifier. Throughout all the modules, 3D facial surface normals perform better than 3D facial points and 3D distance measurements except for the Mouth module. 3D distance measurements has a higher success rate in Mouth, Eyes and Nose modules compared to 3D facial points. While 3D facial points has a higher rate in Eyebrows, Cheeks and Forehead when compared with 3D distance measurements. The highest classification is achieved by the 3D distance measurements in the Mouth module while the lowest classification is also achieved by the same feature vector in the Forehead module. These clear difference of highest and lowest for the same 3D distance measurements results the bigger standard deviation for 3D distance measurements and it indicates that 3D distance measurement has an inconsistent performance across the modules (refer figure 5.7).

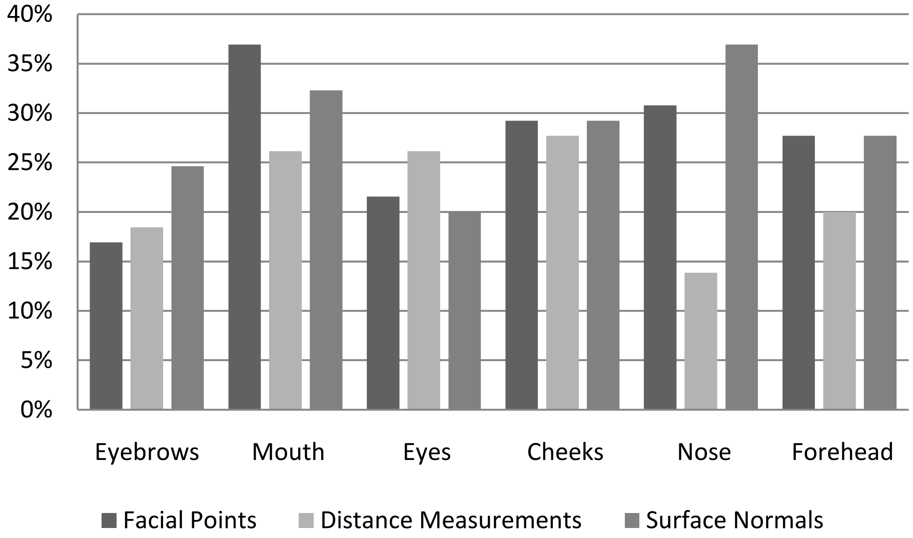


Figure 5.9 Success rate of modular facial expression classification for the Disgust expression using nearest neighbour classifier.

Figure 5.9 shows the success classification rate of modular facial expression classification for the Disgust expression using the nearest neighbour classifier. 3D facial surface normal is better than 3D facial points and 3D distance measurements in all modules except the Eyes and Mouth module. 3D facial points surpass the 3D distance measurements classification rate in the Mouth, Cheeks, Nose and Forehead modules. The highest classification is achieved by 3D facial points in the Mouth module as well as 3D facial surface normals in the Nose module. Meanwhile the lowest classification is achieved by 3D distance measurements in the Nose module. Based on figure 5.7, all three 3D facial features are regarded as having the similar performance across the modules.

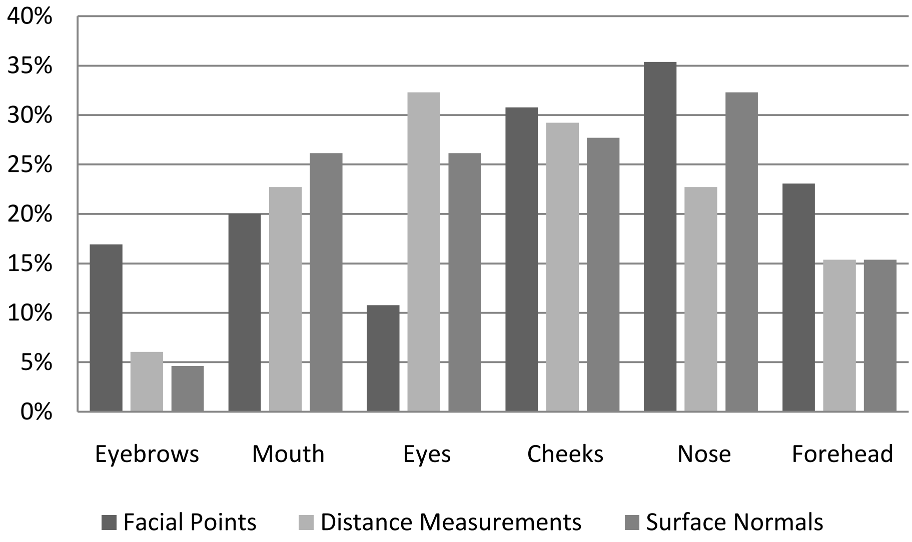


Figure 5.10 Success rate of modular facial expression classification for the Fear expression using nearest neighbour classifier.

Figure 5.10 shows the success classification rate of modular facial expression classification for the Fear expression using the nearest neighbour classifier. 3D facial points achieved a slightly higher classification rate than 3D facial surface normals and 3D distance measurements in Cheeks, Nose and Forehead. However, in the Eyebrows module, 3D facial points are significantly better than the two vectors. 3D distance measurements surpass the other two feature vectors in the Eyes module. The only module that 3D facial surface normals perform best in is the Mouth module. If we look closely at the Cheeks module, the success classification rates for the three facial features are very similar. The highest classification is achieved by 3D facial points in the Mouth module while the lowest classification is achieved by 3D facial surface normals in the Eyebrows module.

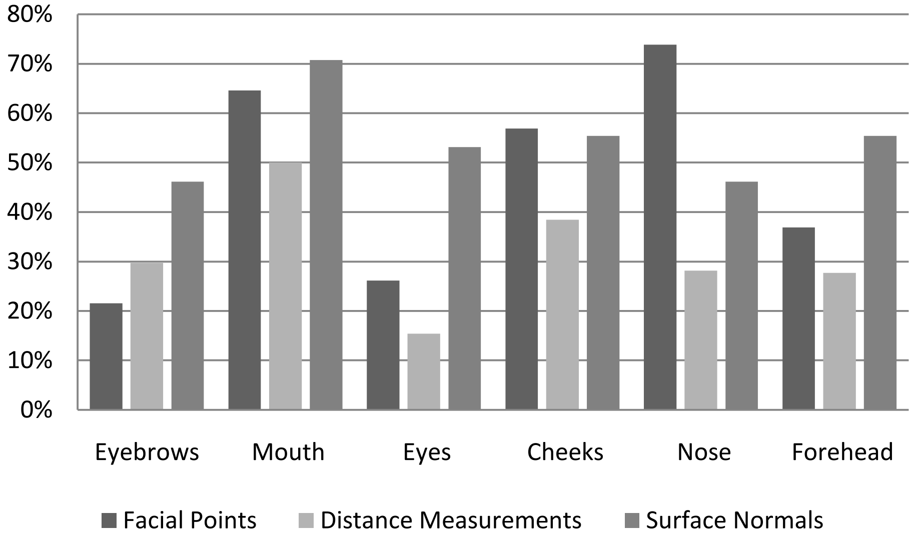


Figure 5.11 Success rate of modular facial expression classification for the Happy expression using nearest neighbour classifier.

Based on figure 5.11, the worst performance across the modules has to be 3D distance measurements in classifying the Happy expression. The best correct classification rate for 3D facial surface normals can be seen in the Eyebrows, Mouth, Eyes and Forehead modules. 3D facial points only perform better than the two other feature vectors in the Cheeks and Nose modules. The highest classification rate is achieved by 3D facial surface normals in the Mouth module while the lowest classification is achieved by 3D distance measurements in the Eyes module.

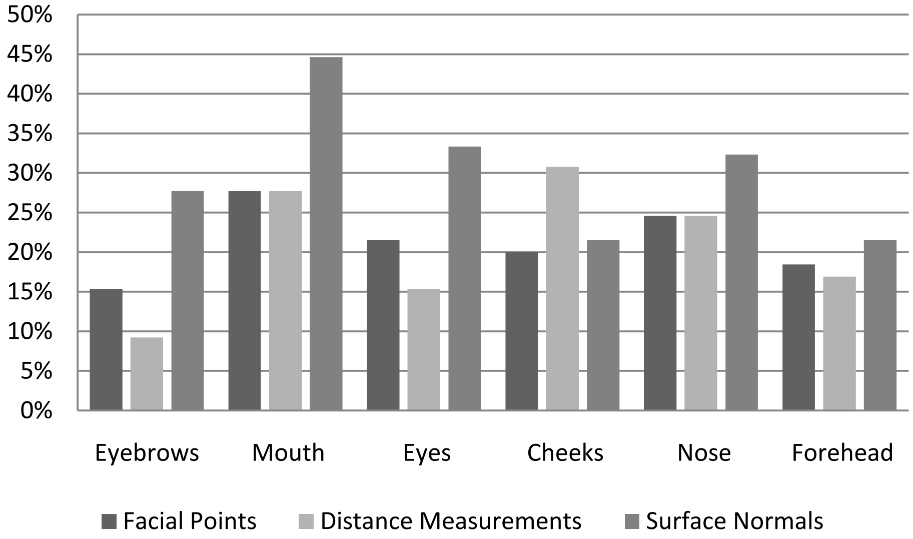


Figure 5.12Success rate of modular facial expression classification for the Sad expression using nearest neighbour classifier.

Figure 5.12 shows the success classification rate of modular facial expression classification for the Sad expression using the nearest neighbour classifier. 3D facial surface normals achieved a better classification rate than 3D facial points and 3D distance measurements across all modules except for the Cheeks module. A 3D distance measurement surpasses the other two feature vectors in the Cheeks module. The worst classification rate is 3D facial points. The highest classification rate is achieved by 3D facial surface normals in the Mouth module while the lowest classification is achieved by 3D distance measurements in the Eyebrows module.

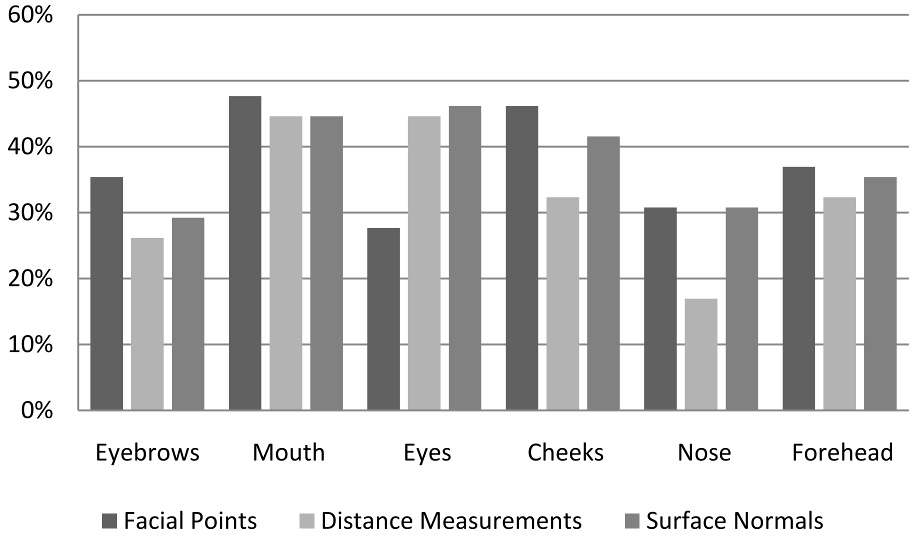


Figure 5.13 Success rate of modular facial expression classification for Surprise expression using nearest neighbour classifier.

Based on figure 5.13, the worst performance across the modules has to be 3D distance measurements for classifying the Surprise expression. A rather similar classification rate is obtained across the feature vectors in the Mouth and Forehead modules. However, 3D facial landmarks perform better in all modules except in the Eyes module. The only module that 3D facial surface normals perform best in is the Eyes module. The highest classification rate is achieved by 3D facial landmarks in the Mouth module while the lowest classification is achieved by 3D distance measurements in the Nose module.

Table 5.7 shows the modular 3D facial expression classification results using SVM with 3D facial points as the baseline feature in the confusion matrices. The Happy expression in the Nose, Mouth and Cheeks modules achieved a 100% success classification rate while for the Forehead module, the Happy expression records a 92% correct classification. The Surprise expression in all modules except Eyebrows achieves more than 56% correct classification and the Cheeks module has the highest rate of 92% correct classification. The Mouth, Cheeks and Nose modules have 52% correct classification for the Anger expression. In the Mouth module, Anger and Sad achieve approximately 50% correct classification. The average success rates for all modules are between 25% - 49% which is better than using 3D facial points with the nearest neighbour classifier. The module that has the lowest average success rate is Eyebrows and the Mouth module has the highest success rate.

Table 5.7 Confusion matrices of modular 3D facial expression classification using 3D facial points with SVM. Recall rates for each expression are shown in bold.

Eyebrows Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**33.85**	20.00	7.69	12.31	13.85	7.69
Disgust	20.00	**18.46**	27.69	24.62	13.85	20.00
Fear	7.69	6.15	**16.92**	6.15	15.38	13.85
Happy	9.23	13.85	4.62	**10.77**	6.15	6.15
Sad	12.31	21.54	32.31	26.15	**40.00**	24.62
Surprise	16.92	20.00	10.77	20.00	10.77	**27.69**

Mouth Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**53.85**	18.46	6.15	0.00	29.23	3.08
Disgust	1.54	**6.15**	7.69	0.00	3.08	6.15
Fear	1.54	1.54	**4.62**	0.00	1.54	1.54
Happy	6.15	52.31	16.92	**100.00**	13.85	3.08

Sad	32.31	12.31	6.15	0.00	**47.69**	1.54
Surprise	4.62	9.23	58.46	0.00	4.62	**84.62**

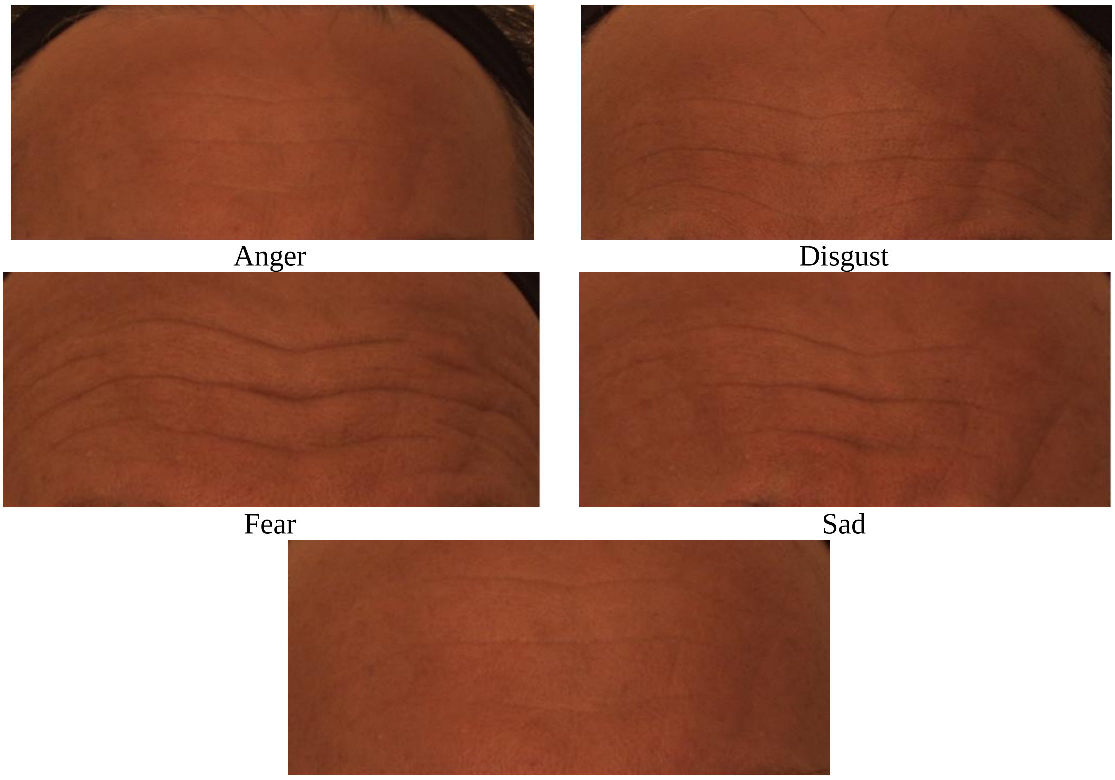
Eyes Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**41.54**	38.46	15.38	20.00	18.46	9.23
Disgust	21.54	**23.08**	7.69	15.38	10.77	6.15
Fear	10.77	4.62	**6.15**	4.62	9.23	1.54
Happy	10.77	13.85	10.77	**27.69**	10.77	10.77
Sad	6.15	13.85	13.85	13.85	**27.69**	3.08
Surprise	9.23	6.15	46.15	18.46	23.08	**69.23**

Cheeks Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**55.38**	35.38	9.23	0.00	36.92	6.15
Disgust	12.31	**15.38**	6.15	0.00	6.15	0.00
Fear	1.54	1.54	**3.08**	0.00	1.54	0.00
Happy	1.54	36.92	9.23	**100.00**	24.62	0.00
Sad	15.38	3.08	4.62	0.00	**13.85**	1.54
Surprise	13.85	7.69	67.69	0.00	16.92	**92.31**

Nose Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**52.31**	6.15	6.15	0.00	35.38	6.15
Disgust	7.69	**23.08**	15.38	0.00	4.62	13.85
Fear	3.08	4.62	**4.62**	0.00	3.08	12.31
Happy	9.23	58.46	18.46	**100.00**	27.69	9.23
Sad	23.08	6.15	13.85	0.00	**23.08**	3.08
Surprise	4.62	1.54	41.54	0.00	6.15	**55.38**

Forehead Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**41.54**	27.69	20.00	1.54	24.62	12.31
Disgust	13.85	**9.23**	9.23	4.62	6.15	3.08
Fear	0	0	**1.54**	0	4.62	1.54
Happy	20.00	41.54	30.77	**92.31**	41.54	18.46
Sad	7.69	3.08	1.54	0.00	**7.69**	0.00
Surprise	16.92	18.46	36.92	1.54	15.38	**64.62**

Based on table 5.7, Anger and Disgust have never been misclassified as Fear while Happy has never been misclassified as Fear and Sad in the Forehead module. Figure 5.14 shows Forehead deformation and we can see clearly that Anger and Happy have a mild Forehead deformation while Fear and Sad have a rather strong deformation of Forehead. In addition, Surprise has never been misclassified as Disgust, Fear and Happy in the Cheeks module. Figure 5.15 shows lower face deformation. The Cheeks module for Surprise expression has a strong deformation intensity compared to Disgust, Fear and Happy. Throughout all the modules, Disgust is always misclassified as Happy and Surprise is the false positive of Fear.



Anger

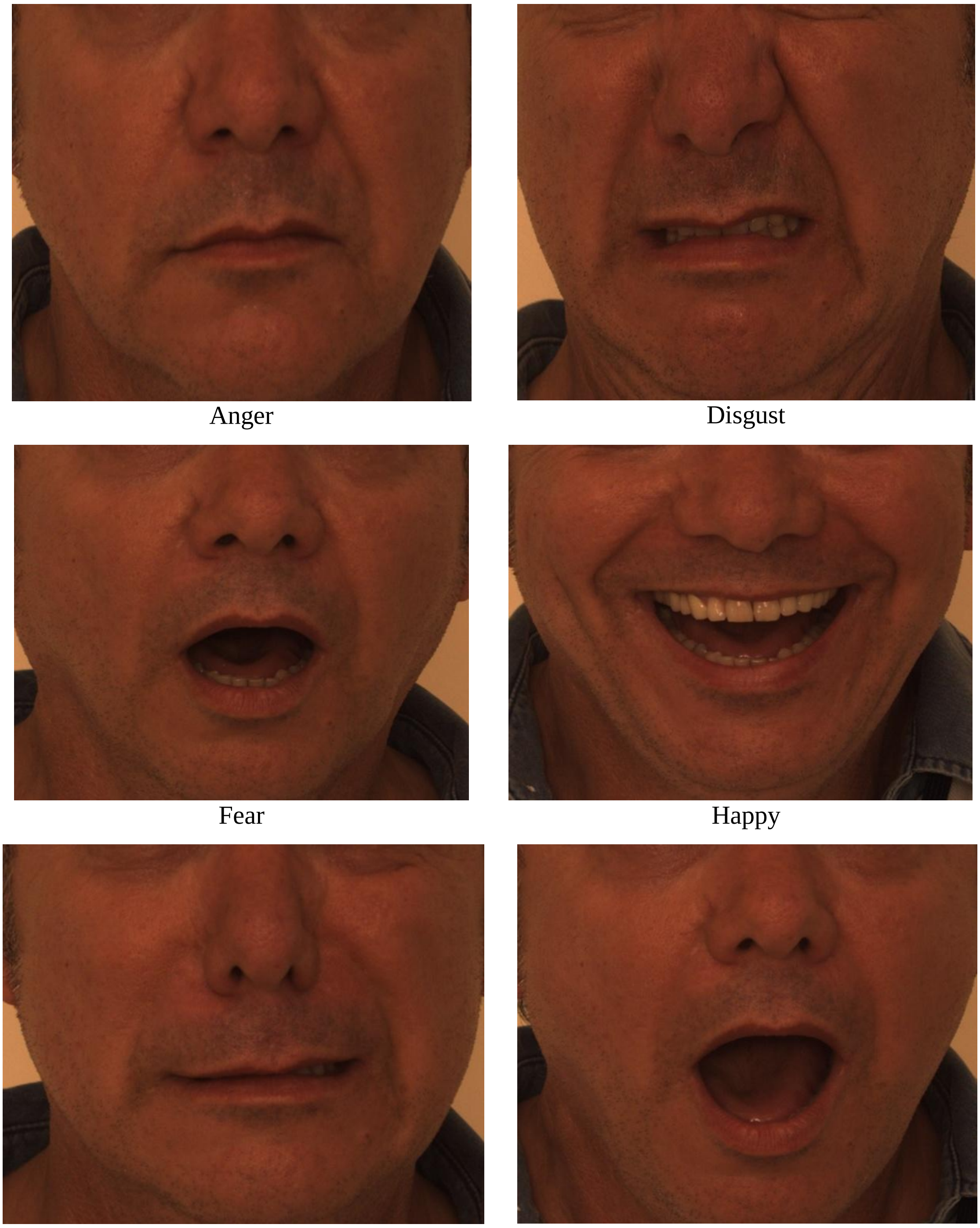
Disgust

Fear

Sad

Happy

Figure 5.14 Forehead deformations for Anger, Disgust, Fear, Sad and Happy expression.



Anger

Disgust

Fear

Happy

Sad

Surprise

Figure 5.15 Lower face deformations for Disgust, Fear, Sad, Happy and Surprise expression.

Table 5.8 shows the modular 3D facial expression classification results using the SVM with 3D distance measurements as the baseline feature in the confusion matrices. The Happy expression in the Mouth, Cheeks and Nose modules achieved more than 57% successful classification. The Surprise expression in the Mouth, Eyes and Cheeks modules perform better than 67% of correct classification and the Eyes module has the highest correct classification rate of 83%. In the Mouth module, Anger and Surprise also perform better than 50% correct classification. The Forehead module has the highest correct classification rate of 51% for the Sad expression. The average success rates for all modules are between 17% - 43% which is better than using 3D distance measurements with the nearest neighbour classifier. The module that has the lowest average success rate is Eyebrows and the Mouth module has the highest success rate. In the Eyebrows and Cheeks module, Fear has never been misclassified as Disgust. Happy has never been classified as Anger, Disgust and Sad in the Mouth module which is due to the obvious gap in the Happy expression (refer to figure 5.15). In addition, Surprise has never been misclassified as Disgust and Happy in the Eyes module due to the different deformation they have (refer to figure 5.16). Surprise in the Cheeks module has never been classified as Sad (refer to figure 5.15). Anger in the Nose module has never been classified as Happy and based on figure 5.15; the Nose is stretching due to a smile action in the Happy expression which causes an intense deformation compared to the Nose stretching intensity in the Anger expression.

Table 5.8 Confusion matrices of modular 3D facial expression classification using 3D distance measurement with SVM. Recall rates for each expression are shown in bold.

Eyebrows Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**4.62**	4.62	10.77	9.23	7.69	13.85
Disgust	3.08	**4.62**	0.00	4.62	1.54	4.62
Fear	23.08	24.62	**24.62**	26.15	24.62	35.38
Happy	18.46	16.92	20.00	**21.54**	18.46	29.23
Sad	29.23	35.38	33.85	20.00	**35.38**	6.15
Surprise	21.54	13.85	10.77	18.46	12.31	**10.77**

Mouth Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**53.85**	16.92	12.31	0.00	35.38	7.69
Disgust	1.54	**3.08**	0.00	0.00	1.54	3.08
Fear	3.08	3.08	**3.08**	3.08	1.54	6.15
Happy	6.15	43.08	21.54	**86.15**	16.92	10.77
Sad	32.31	26.15	7.69	0.00	**40.00**	1.54
Surprise	3.08	7.69	55.38	10.77	4.62	**70.77**

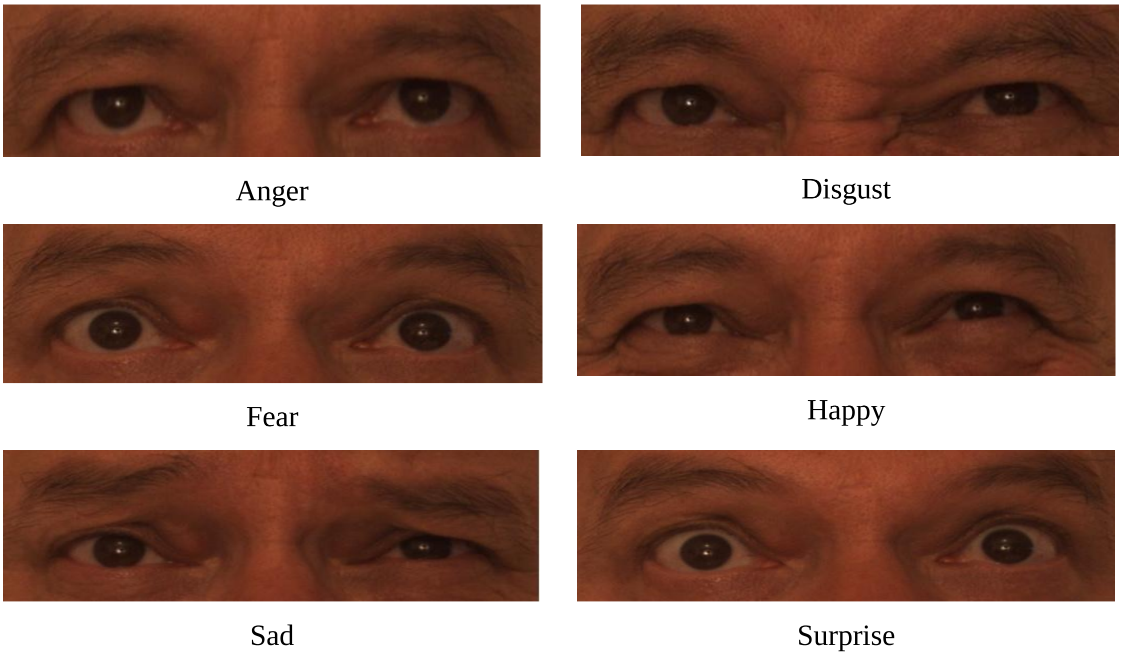
Eyes Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**38.46**	29.23	12.31	13.85	10.77	9.23
Disgust	15.38	**15.38**	1.54	18.46	6.15	0.00
Fear	10.77	13.85	**18.46**	1.54	6.15	4.62
Happy	4.62	10.77	3.08	**12.31**	12.31	0.00
Sad	13.85	13.85	12.31	20.00	**44.62**	3.08
Surprise	16.92	16.92	52.31	33.85	20.00	**83.08**

Cheeks Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**63.08**	46.15	13.85	6.15	49.23	6.15
Disgust	4.62	**7.69**	0.00	3.08	4.62	3.08

Fear	4.62	3.08	**9.23**	3.08	4.62	9.23
Happy	9.23	23.08	18.46	**56.92**	20.00	13.85
Sad	12.31	7.69	7.69	3.08	**7.69**	0.00
Surprise	6.15	12.31	50.77	27.69	13.85	**67.69**

Nose Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**86.15**	29.23	20.00	1.54	61.54	23.08
Disgust	6.15	**23.08**	15.38	6.15	13.85	7.69
Fear	3.08	7.69	**9.23**	12.31	6.15	20.00
Happy	0.00	26.15	44.62	**78.46**	10.77	35.38
Sad	1.54	6.15	1.54	0.00	**3.08**	6.15
Surprise	3.08	7.69	9.23	1.54	4.62	**7.69**

Forehead Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**16.92**	16.92	13.85	12.31	12.31	13.85
Disgust	10.77	**3.08**	3.08	9.23	3.08	7.69
Fear	6.15	9.23	**9.23**	15.38	3.08	7.69
Happy	21.54	21.54	9.23	**21.54**	10.77	13.85
Sad	27.69	30.77	38.46	18.46	**50.77**	18.46
Surprise	16.92	18.46	26.15	23.08	20.00	**38.46**



Anger

Disgust

Fear

Happy

Sad

Surprise

Figure 5.16 Eyes and Eyebrows deformation for all six basic facial expressions.

Table 5.9 shows the modular 3D facial expression classification results using SVM with 3D facial surface normal as the baseline feature in the confusion matrices. The Anger, Happy and Surprise expressions achieved more than 50% success classification across the modules. Only the Mouth module performs at more than 66% of correct classification for the Sad expression. The Fear expression has a zero classification rate in the Eyebrows module. The average success rates for all modules are between 33% - 54% which is better than using 3D facial surface normals with the nearest neighbour classifier. The module that has lowest average success classification rates is Eyebrows and the Mouth module has the highest success rate. In the Eyebrows module, Disgust, Happy and Sad have never been misclassified as Fear while Surprise and Anger have never been misclassified as Disgust. There are obvious differences of deformation in the Eyebrows module for each expression, (see figure 5.16). Happy has never been classified as Fear in the Eyebrows, Mouth, Cheeks and Nose modules (refer to figures 5.15 and 5.16 for the different deformation shown by the facial features in each module).In addition, Surprise has never been misclassified as Disgust in the Eyebrows and Forehead modules.

Overall, modular 3D facial expression classification using 3D facial surface normals with SVM performs better than the other two feature vectors used in both classifiers. However, 3D facial surface normals with SVM do not achieve 100% correct classification in the Happy expression which is in contrast to 3D facial points.

Table 5.9 Confusion matrices of modular 3D facial expression classification using 3D facial surface normals with SVM. Recall rates for each expression are shown in bold.

Eyebrows Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**53.85**	44.62	21.54	10.77	26.15	10.77
Disgust	0.00	**1.54**	4.62	1.54	4.62	0.00
Fear	1.54	0.00	**0.00**	0.00	0.00	1.54
Happy	18.46	21.54	26.15	**56.92**	26.15	24.62
Sad	6.15	9.23	10.77	6.15	**26.15**	1.54
Surprise	20.00	23.08	36.92	24.62	16.92	**61.54**

Mouth Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**60.00**	26.15	9.23	1.54	24.62	7.69
Disgust	3.08	**7.69**	4.62	0.00	1.54	1.54
Fear	1.54	10.77	**26.15**	0.00	3.08	20.00
Happy	1.54	15.38	4.62	**98.46**	3.08	6.15
Sad	29.23	29.23	10.77	0.00	**66.15**	1.54
Surprise	4.62	10.77	44.62	0.00	1.54	**63.08**

Eyes Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**50.77**	29.23	12.31	7.69	24.62	4.62
Disgust	13.85	**27.69**	3.08	4.62	9.23	3.08
Fear	3.08	0.00	**32.31**	1.54	6.15	16.92
Happy	15.38	27.69	6.15	**73.85**	16.92	6.15
Sad	12.31	7.69	6.15	9.23	**32.31**	0.00
Surprise	4.62	7.69	40.00	3.08	10.77	**69.23**

Cheeks Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**50.00**	23.08	15.38	0.00	26.15	9.23
Disgust	16.92	**33.85**	9.23	4.62	13.85	1.54
Fear	12.31	15.38	**23.08**	0.00	3.08	24.62
Happy	0.00	12.31	3.08	**90.77**	7.69	0.00
Sad	15.38	9.23	7.69	4.62	**40.00**	6.15
Surprise	6.15	6.15	41.54	0.00	9.23	**58.46**

Nose Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**66.15**	24.62	13.85	0.00	35.38	9.23
Disgust	7.69	**43.08**	7.69	6.15	15.38	9.23
Fear	6.15	3.08	**15.38**	0.00	6.15	18.46
Happy	1.54	18.46	9.23	**87.69**	6.15	6.15
Sad	15.38	7.69	13.85	1.54	**23.08**	10.77
Surprise	3.08	3.08	40.00	4.62	13.85	**50.00**

Forehead Module	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**66.15**	35.38	15.38	1.54	21.54	13.85
Disgust	6.15	**12.31**	6.15	1.54	4.62	0.00

Fear	4.62	3.08	**9.23**	1.54	4.62	10.77
Happy	6.15	35.38	13.85	**93.85**	23.08	12.31
Sad	4.62	4.62	10.77	0.00	**30.77**	3.08
Surprise	12.31	9.23	44.62	1.54	15.38	**60.00**

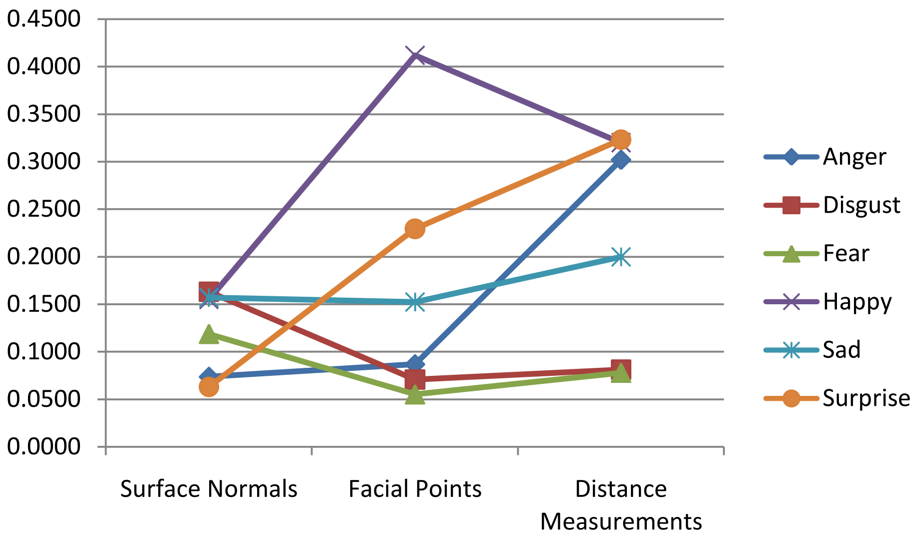


Figure 5.17 Consistency test using standard deviations of 3D facial features compiled from the success rate of the modules across facial expressions of SVM classifier.

Figure 5.17 shows consistency test using standard deviations of 3D facial features compiled from the success rate of the modules across facial expressions using SVM classifier. The following graphs (figure 5.18-5.23) discuss the performance of the three 3D facial features for each expression and each of those graphs will refer to figure 5.17 in order to verify with the standard deviation which represents the consistency of the 3D facial feature across the modules.

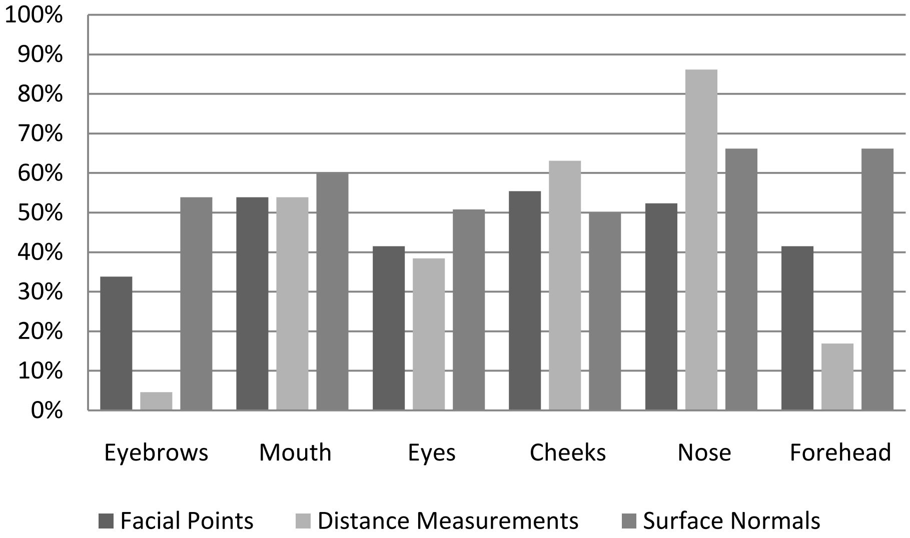


Figure 5.18 Success rate of modular facial expression classification for Anger expression using SVM

Figure 5.18 shows the success rate of modular facial expression classification for the Anger expression using SVM. Throughout all modules, 3D facial surface normals have a consistent performance compared to 3D facial points and 3D distance measurements as it has at least 50% for every module. This claim is supported with the consistency test using the standard deviations (refer figure 5.17). A bigger value of standard deviation for 3D distance measurements indicates the inconsistent of this facial feature across modules. 3D distance measurements surpass the other two feature vectors' performance only in the Cheeks and the Nose modules. The highest classification is achieved by 3D distance measurements in the Nose module while the lowest classification is also achieved by the same feature vector in the Eyebrows module.

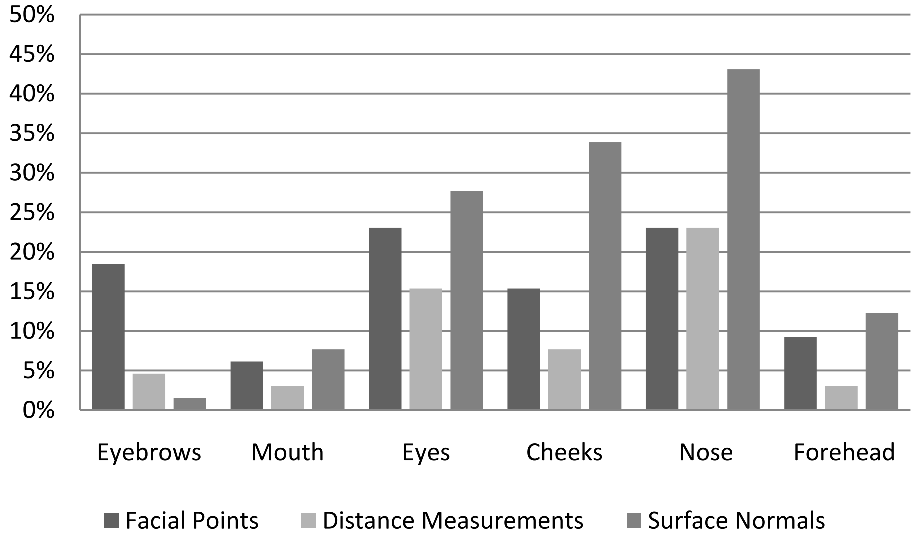


Figure 5.19 Success rate of modular facial expression classification for Disgust expression using SVM

Figure 5.19 shows the success classification rate of modular facial expression classification for the Disgust expression using SVM. Overall all feature vectors have the worst performance for this expression compared to results from the nearest neighbour classifier. We believe that this happens due to Disgust expression having a very high similarity with Anger which cause a high misclassification except for the Nose module (refer to figures 5.14, 5.15 and 5.16 for Anger and Disgust expression). However, 3D facial surface normals surpass the other two feature vectors in all moduleexcept for the Eyebrows module. The highest classification is achieved by 3D facial surface normals in the Nose module while the lowest classification is also achieved by the same feature vector in the Eyebrows module.

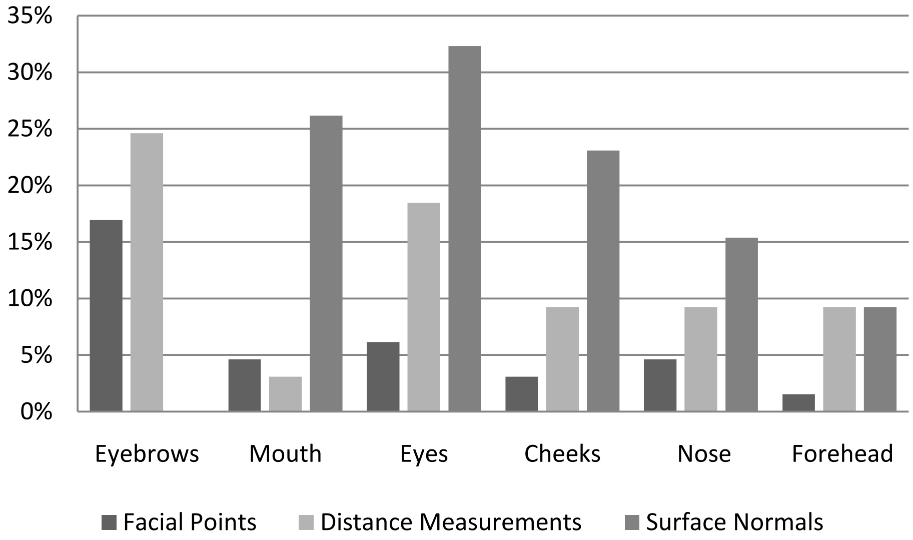


Figure 5.20 Success rate of modular facial expression classification for Fear expression using SVM

Figure 5.20 shows the success classification rate of modular facial expression classification for the Fear expression using SVM. 3D facial surface normals feature has the best result except for the Eyebrows module. 3D distance measurements perform better in the Eyebrows module. The highest classification is achieved by 3D facial surface normals in the Eyes module while the lowest classification is achieved by 3D facial surface normals in the Eyebrows module with zero correct classification. Overall all feature vectors have the worst performance for this expression compared to results from nearest neighbour classifier.

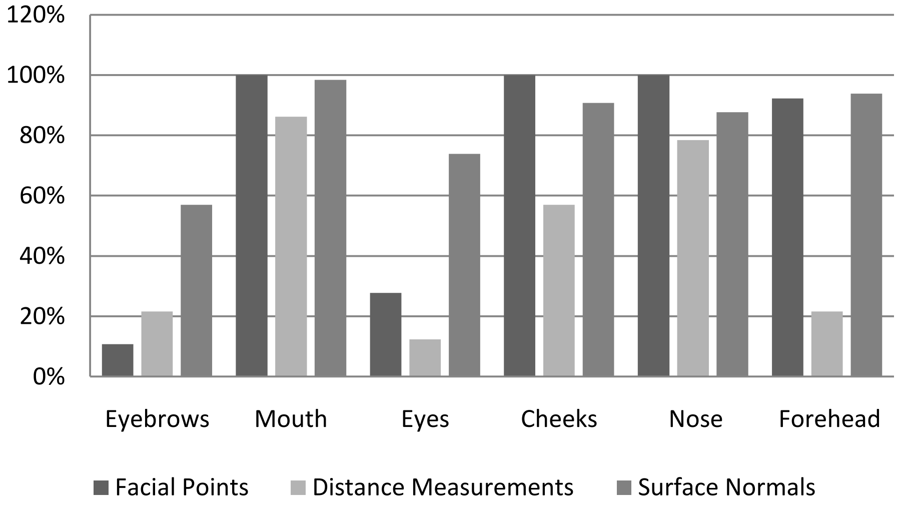


Figure 5.21 Success rate of modular facial expression classification for Happy expression using SVM

Based on figure 5.21, the worst performance across the modules has to be 3D distance measurements. The highest correct classification rate for 3D facial surface normals can be seen in Eyebrows, Eyes and Forehead while for other modules, 3D facial points perform better. We can see that 3D face surface normal have a consistent performance throughout the modules and it also proves by the standard deviation value in figure 5.17. The highest classification is achieved by 3D facial points in the Mouth module while the lowest classification is achieved by 3D facial points in the Eyebrows module.

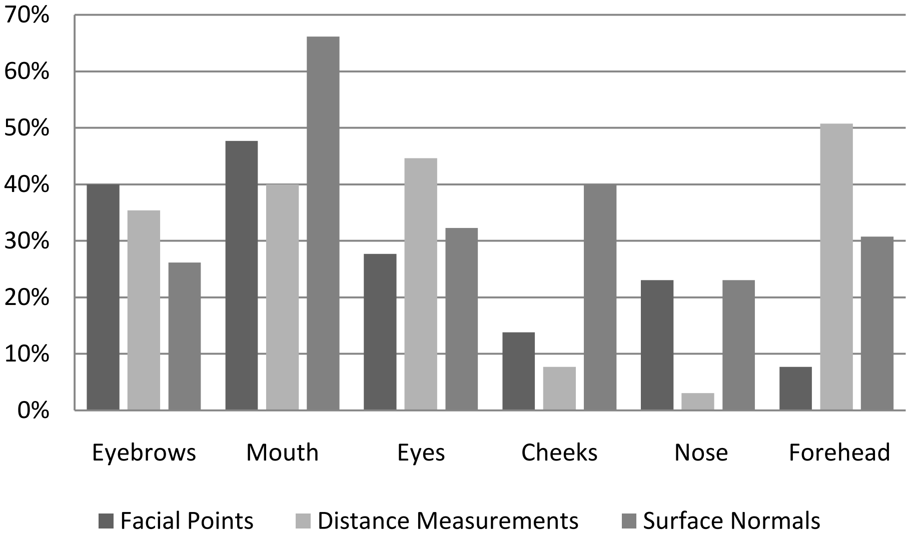


Figure 5.22 Success rate of modular facial expression classification for Sad expression using SVM

Figure 5.22 shows the success classification rate of modular facial expression classification for the Sad expression using SVM. The performances of all feature vectors are not consistent in this expression. Overall, the worst classification rate is 3D distance measurements even though it surpasses the rest of the feature vectors in the Forehead module. The highest classification rate is achieved by 3D facial surface normals in the Mouth module while the lowest classification is achieved by 3D distance measurements in the Nose module.

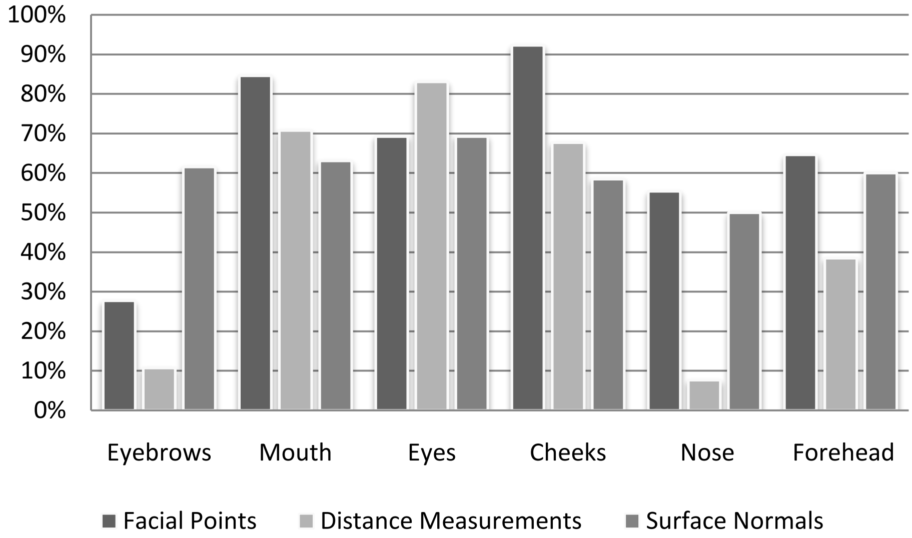


Figure 5.23 Success rate of modular facial expression classification for Surprise expression using SVM

Based on figure 5.23, the worst performance across the modules for classifying the Surprise expression using SVM has to be 3D distance measurements as the average classification rate for 3D facial points, 3D distance measurements and 3D facial surface normals are 65.54%, 42.32% and 60.39% respectively. 3D facial points perform better in all modules except in the Eyebrows and Eyes modules. The highest classification rate is achieved by 3D facial points in the Cheeks module while the lowest classification is achieved by 3D distance measurements in the Nose module. 3D facial surface normals result is consistent throughout this expression.

5.4.2 Weighted Voting Scheme Results

The results of modular facial expression classification experiments using the nearest neighbour and SVM classifiers are passed to the Votes Counter. n the Votes Counter algorithm, the weight of the modules that belong to the same facial expression class is summed up. Finally, the weight for each facial expression is computed and the final facial expression is the one with the highest vote.

With this weighted approach, the module with the largest weight plays an important role in classifying the final facial expression. In most cases, different modules vote for different expressions in classifying a facial expression. For example in the case of most Anger expressions, the first three modules (Eyebrows, Mouth and Eyes) vote for the Sad expression while the rest of the modules vote for the Anger expression. The final facial expression would be Sad as it has much more weight compared to the Anger expression.

Table 5.10 Confusion matrices of WVS on modular experiments using the nearest neighbour classifier results. Recall rates for each expression are shown in bold.

	Anger	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
3D Facial Points						
Anger	**36.92**	16.92	16.92	6.15	23.08	4.62
Disgust	16.92	**38.46**	10.77	4.62	21.54	9.23
Fear	13.85	12.31	**23.08**	6.15	12.31	16.92
Happy	3.08	13.85	10.77	**66.15**	9.23	7.69
Sad	24.62	16.92	12.31	6.15	**23.08**	6.15
Surprise	4.62	1.54	26.15	10.77	10.77	**55.38**
3D Distance Measurement						
Anger	**41.54**	16.92	13.85	7.69	24.62	4.62
Disgust	13.85	**23.08**	10.77	12.31	24.62	7.69
Fear	12.31	15.38	**21.54**	18.46	12.31	29.23
Happy	4.62	15.38	13.85	**41.54**	3.08	6.15
Sad	21.54	21.54	10.77	7.69	**26.15**	7.69
Surprise	6.15	7.69	29.23	12.31	9.23	**44.62**
3D Surface Normals						
Anger	**52.31**	18.46	4.62	1.54	16.92	6.15
Disgust	12.31	**40.00**	7.69	6.15	9.23	0.00
Fear	7.69	12.31	**30.77**	3.08	9.23	27.69
Happy	6.15	13.85	6.15	**78.46**	9.23	3.08
Sad	10.77	4.62	10.77	3.08	**47.69**	6.15
Surprise	10.77	10.77	40.00	7.69	7.69	**56.92**

Table 5.10 shows the confusion matrices for WVS based on a modular experiment using the nearest neighbour classifier results. For 3D facial points feature, only the Happy and Surprise expressions are correctly classified more than half the time. Two expressions that have lower than 30% correct classification are Fear and Sad. Anger and Disgust are roughly about the same percentage. Fear and Surprise are always misclassified as Surprise and Fear respectively.

Unlike for 3D facial points features, there are no expressions that reach more than 50% correct classification. Three expressions that managed to achieve more than 40% are Anger, Happy and Surprise. The results on Disgust, Fear and Sad are all lower than 30% whereas Fear has the lowest percentage which is 21.54%.

The average rate of classifications of the three feature vectors are 41%, 33% and 51% for 3D facial points, 3D distance measurements and 3D facial surface normals respectively. Again, using 3D facial surface normals as the baseline feature with the nearest neighbour classifier definitely improves the facial expression classification rate compared to the other two feature vectors. Fear has the worst classification results with only 30.77% correct classification. Angry, Surprise and Happy are the expressions with correct classification of more than 50% and Happy has the highest score. Disgust and Sad are more than 40% correctly classified. Surprise has never been misclassified as Disgust.

Table 5.11 Confusion matrices of WVS based on modular experiment using SVM results. Recall rates for each expression are shown in bold.

	Anger	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
3D Facial Points						
Anger	**63.08**	27.69	7.69	0.00	35.38	3.08
Disgust	10.77	**13.85**	12.31	0.00	1.54	4.62
Fear	3.08	0.00	**1.54**	0.00	4.62	0.00
Happy	3.08	44.62	10.77	**100.00**	13.85	1.54
Sad	15.38	6.15	7.69	0.00	**32.31**	0.00
Surprise	4.62	7.69	60.00	0.00	12.31	**90.77**
3D Distance Measurement						
Anger	**63.08**	30.77	7.69	1.54	29.23	7.69
Disgust	0.00	**6.15**	0.00	0.00	1.54	0.00
Fear	6.15	6.15	**4.62**	3.08	6.15	3.08
Happy	6.15	29.23	15.38	**80.00**	12.31	12.31
Sad	20.00	18.46	21.54	4.62	**44.62**	1.54
Surprise	4.62	9.23	50.77	10.77	6.15	**75.38**
3D Surface Normals						
Anger	**75.38**	26.15	12.31	0.00	16.92	4.62
Disgust	4.62	**43.08**	3.08	0.00	1.54	0.00
Fear	1.54	0.00	**21.54**	0.00	0.00	4.62
Happy	0.00	20.00	3.08	**100.00**	3.08	0.00
Sad	12.31	6.15	3.08	0.00	**67.69**	0.00
Surprise	6.15	4.62	56.92	0.00	10.77	**90.77**

Table 5.10 shows the confusion matrices of WVS based on a modular experiment using the SVM classifier. For 3D facial points, there is no misclassification error for Happy and Surprise is the expression with the second highest correct classification. Fear has the lowest score in which most of it has been misclassified as Surprise. Disgust has less than 15% correct classification and it has been incorrectly classified as Anger and Happy. Most Anger expressions are misclassified as Sad and vice-versa.

3D distance measurement has three expressions which achieved more than 50% correct classification and they are Anger, Happy and Surprise. This result is slightly improved compared to using the nearest neighbour classifier. In agreement with other feature vectors in both classifiers, Fear has the lowest rate of correctly classified expressions. Again, as opposed to other feature vectors in both classification types, Happy is largely misclassified as Surprise. Fear, Happy and Surprise are never misclassified as Disgust.

Similar to 3D facial points, 3D facial surface normals record a 100% correct classification for the Happy expression. The Disgust expression has a 29% higher classification rate than using 3D facial points. 3D facial surface normals have an equal rate of correct classification for the Surprise expression with 3D facial points. Overall, 3D facial surface normals perform quite well compared to the two other feature vectors where the average classification of 3D facial surface normals is 66% while 50% and 46% for 3D facial points and 3D distance measurements respectively.

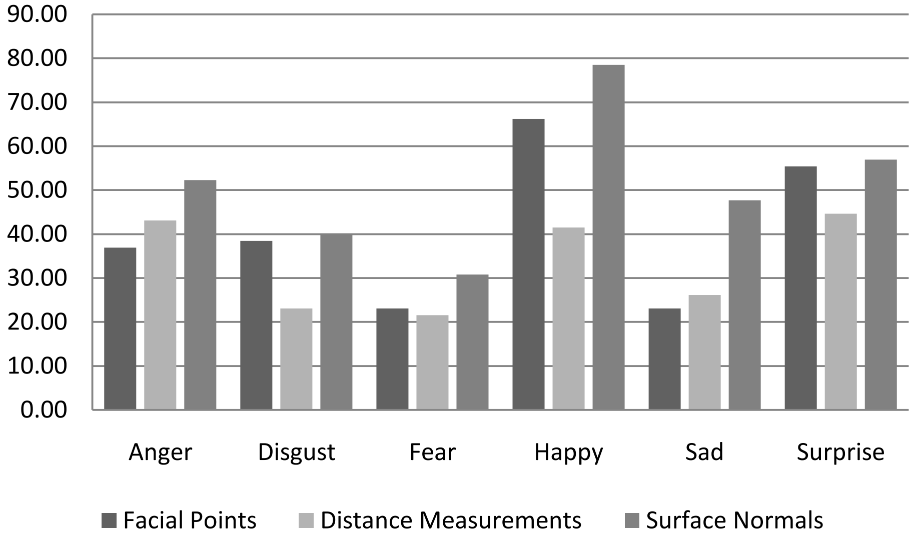


Figure 5.24 Success rate of WVS using the nearest neighbour classifier

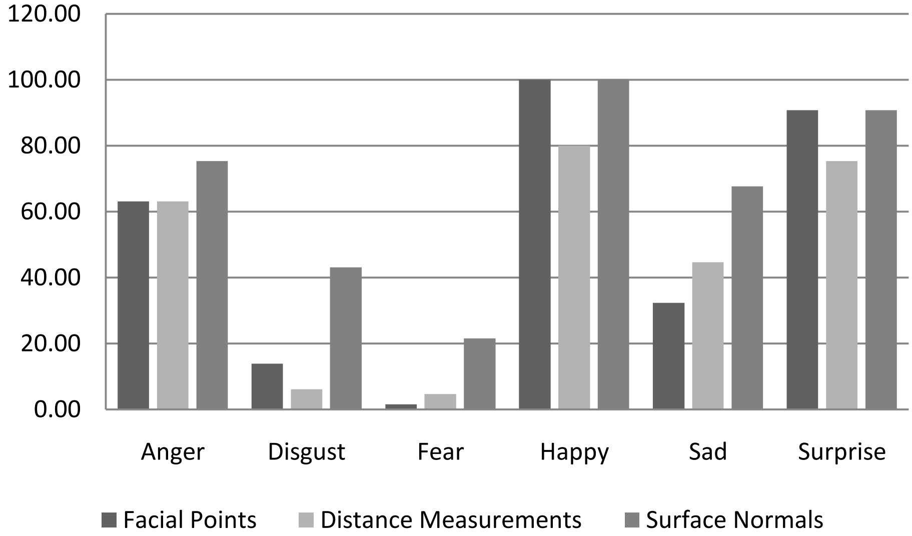


Figure 5.25 Success rate of WVS using the SVM classifier

Figure 5.24 and 5.25 show the success rate of WVS using nearest neighbour classifier and SVM. 3D facial surface normals record good results compared to the other feature vectors. The consistent performance of 3D facial surface normals across the modules is believed to be the main reason for having an improved result. The classification rate for Disgust and Fear expression using SVM for the three feature vectors are poorer than using nearest neighbour classifier. Despite the simple computation feature of the nearest neighbour classifier, the success classification rates are higher compared a more complicated computation performed by SVM and we see it as an advantage of this work.

Table 5.12 An example of equal votes in MVS

	Module	**Facial Expression**
1	Eyebrows	Surprise
2	Mouth	Surprise
3	Eyes	Fear
4	Cheeks	Happy
5	Nose	Happy
6	Forehead	Fear

For the purpose of comparison, we also carried out experiments using a Majority Voting Scheme (MVS), as opposed to WVS. In MVS, the final classification of multiple classifications goes to the class with the majority vote. However, in the case of two or more classes having equal votes, our algorithm will classify the final expression as False Positive (FP). Table 5.12 shows an example of MVS with equal votes.

Table 5.13 shows the confusion matrices for MVS based on a modular experiment using the nearest neighbour classifier results. For 3D facial points feature, only the Happy expression is correctly classified more than half the time. Three expressions that have lower than 30% correct classification are Anger, Disgust and Sad. The FP cases are higher in the Anger and Sad expressions and other expressions have approximately the same percentage.

Similar to 3D facial points features results in WVS, there are no expressions that reach more than 50% correct classification. The highest classification rate belongs to Anger whereas Fear has the lowest percentage which is 10.77%. The FP case is higher in the Sad expression and other expressions have approximately the same percentage.

Similar to 3D facial points only the Happy expression is correctly classified more than half the time when 3D facial surface normals are used. 3D facial surface normals also yield better results for the Anger, Happy and Sad expressions when compared to the other feature vectors. Similar to results in WVS, Fear has the worst classification results with only 12.31% correct classification.

The average rate of classifications of the three feature vectors are 32%, 26% and 36% for 3D facial points, 3D distance measurements and 3D facial surface normals respectively.

Table 5.13 Confusion matrices of MVS on modular experiments using the nearest neighbour classifier results. Recall rates for each expression are shown in bold.

	Anger	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
3D Facial Points						
Anger	**24.62**	10.77	10.77	3.08	18.46	4.62
Disgust	9.23	**33.85**	4.62	4.62	12.31	6.15
Fear	9.23	4.62	**20.00**	6.15	9.23	15.38
Happy	0.00	10.77	4.62	**56.92**	0.00	4.62
Sad	16.92	12.31	10.77	4.62	**12.31**	4.62
Surprise	3.08	1.54	23.08	3.08	7.69	**44.62**
FP	36.92	26.15	26.15	21.54	40.00	20.00
3D Distance Measurement						
Anger	**38.46**	10.77	6.15	0.00	13.85	1.54
Disgust	7.69	**23.08**	6.15	9.23	13.85	4.62
Fear	9.23	10.77	**10.77**	12.31	9.23	20.00
Happy	4.62	13.85	10.77	**30.77**	1.54	3.08
Sad	16.92	16.92	6.15	4.62	**18.46**	3.08
Surprise	1.54	7.69	24.62	7.69	3.08	**35.38**
FP	21.54	16.92	35.38	35.38	40.00	32.31
3D Surface Normals						
Anger	**44.62**	12.31	4.62	0.00	15.38	1.54
Disgust	9.23	**27.69**	0.00	4.62	6.15	0.00
Fear	4.62	10.77	**12.31**	1.54	6.15	20.00
Happy	3.08	9.23	3.08	**58.46**	4.62	1.54
Sad	10.77	3.08	10.77	0.00	**30.77**	1.54
Surprise	3.08	6.15	27.69	1.54	1.54	**43.08**
FP	24.62	30.77	41.54	33.85	35.38	32.31

Table 5.14 Confusion matrices of MVS on modular experiments using the SVM classifier results. Recall rates for each expression are shown in bold.

	Anger	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
3D Facial Points						
Anger	**53.85**	23.08	3.08	0.00	23.08	3.08
Disgust	9.23	**12.31**	9.23	0.00	0.00	4.62
Fear	1.54	0.00	**1.54**	0.00	1.54	0.00
Happy	1.54	41.54	6.15	**100.00**	15.38	1.54
Sad	10.77	1.54	3.08	0.00	**24.62**	0.00
Surprise	3.08	6.15	53.85	0.00	10.77	**83.08**
FP	20.00	15.38	23.08	0.00	24.62	7.69
3D Distance Measurement						
Anger	**53.85**	23.08	4.62	0.00	26.15	3.08
Disgust	0.00	**0.00**	0.00	0.00	0.00	0.00
Fear	1.54	6.15	**1.54**	1.54	3.08	3.08
Happy	4.62	21.54	10.77	**63.08**	7.69	4.62
Sad	10.77	15.38	15.38	0.00	**33.85**	0.00
Surprise	3.08	4.62	33.85	7.69	6.15	**66.15**
FP	26.15	29.23	33.85	27.69	23.08	23.08
3D Surface Normals						
Anger	**67.69**	27.69	9.23	0.00	16.92	4.62
Disgust	1.54	**24.62**	3.08	0.00	1.54	0.00
Fear	0.00	0.00	**10.77**	0.00	0.00	3.08
Happy	0.00	18.46	1.54	**98.46**	3.08	1.54
Sad	7.69	0.00	3.08	0.00	**40.00**	0.00
Surprise	4.62	3.08	47.69	0.00	10.77	**73.85**
FP	18.46	26.15	24.62	1.54	27.69	16.92

Table 5.14 shows the confusion matrices of MVS based on a modular experiment using the SVM classifier. Similar to WVS, for 3D facial points, there is no misclassification error for Happy and Surprise is the expression with the second highest correct classification. Fear has the lowest score in which most of it has been misclassified as Happy. Disgust has less than 15% correct classification and it has been incorrectly classified as Anger and Happy. Most Anger expressions are misclassified as Sad and vice-versa. The rate of FP cases is approximately the same in the Anger, Fear and Sad expressions. Clearly, the Happy expression does not have any FP cases.

3D distance measurement has three expressions which achieved more than 50% correct classification and they are Anger, Happy and Surprise. This result is slightly improved compared to using the nearest neighbour classifier. The Disgust expression has zero correct classification. Again, as opposed to other feature vectors in both classification types, Happy is largely misclassified as Surprise. All expressions are never misclassified as Disgust. The FP case is higher in the Fear expression and other expressions have approximately the same percentage.

3D facial surface normals record a 98.46% correct classification for the Happy expression which means only one case of the Happy expression is classified as FP. The Disgust expression has a 10% higher classification rate than using 3D facial points. The Surprise expression has a slightly lower correct classification rate when compared to using 3D facial points. Using the MVS approach, 3D facial surface normals still perform quite well compared to the two other feature vectors where the average classification of 3D facial surface normals is 53% while it is 46% and 36% for 3D facial points and 3D distance measurements respectively.

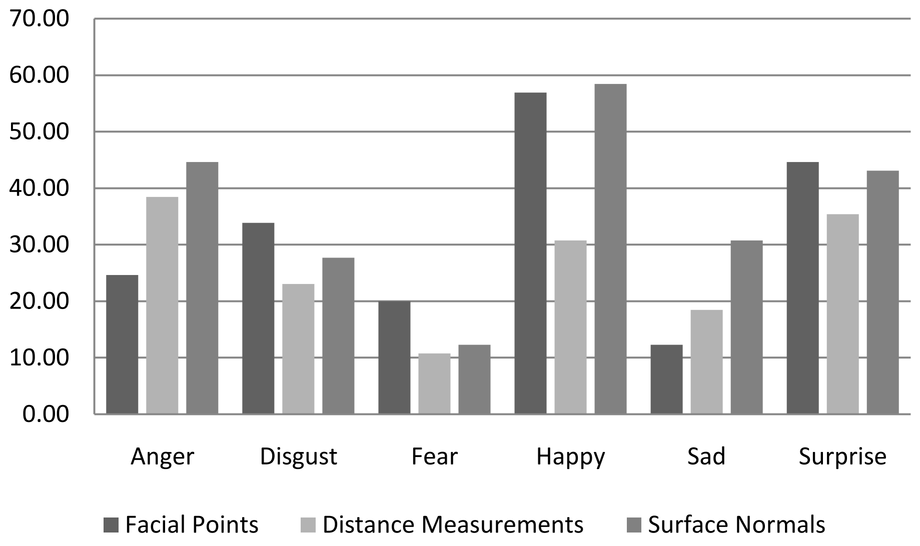


Figure 5.26 Success rate of MVS using the nearest neighbour classifier

Figure 5.26 shows the success rate of MVS using the nearest neighbour classifier. 3D facial surface normals record good results in the Anger, Happy and Sad expressions while 3D facial points has the highest correct classification in Disgust, Fear and Surprise. However, the correct classification rates for 3D facial surface normals in Disgust, Fear and Surprise expression only differs by a few percentage points from that obtained using 3D facial points. In addition, 3D distance measurements have the lowest correct classification across all expressions.

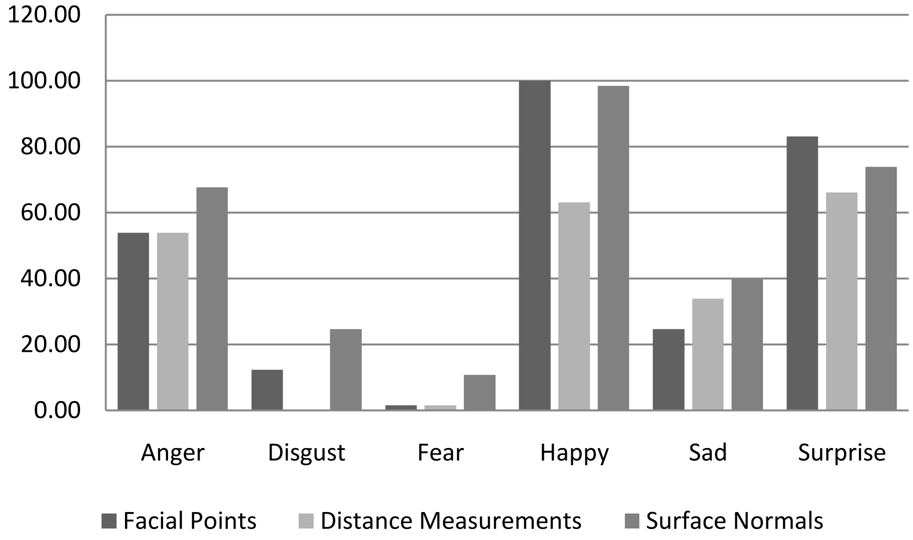


Figure 5.27 Successrate of MVS using SVM classifier

Figure 5.27 shows the success rate of MVS using SVM. 3D facial surface normals record improved results compared to the other feature vectors in all expressions except for the Happy expressions when SVM is used. Furthermore, 3D facial surface normals record higher correct classification rates using SVM in the Anger, Happy, Sad and Surprise expressions compared to using the nearest neighbour classifier. However, this is not the case in the Disgust and Fear expressions and it is even worse than WVS. This is due to poor results across the modules in the SVM experiments. Again, the consistent performance of 3D facial surface normals across the modules is believed to be the main reason for having an improved result.

5.6 Conclusions

The work in this chapter begins with the modularization of a face. We divided a face into six modules namely Forehead, Eyebrows, Eyes, Nose, Cheeks and Mouth. Subsequently, the modular priority rank is determined using AdaBoost and each of the modules is assigned a weight. Since the facial expression classification for every module is done independently, this means that for one face, each module could represent different facial expressions and therefore another approach is needed in order to determine which single facial expression is portrayed by the face. We used the WVS approach since we have the weight for each of the modules. The results of the modular experiments using both classifiers are passed to the votes counter and the expression which gets the vote will have the weight of that module. For the purpose of comparison, we also carried out experiments using Majority Voting Scheme (MVS). Finally, the weight of the vote for each of the facial expressions is computed and the final facial expression is the one with the highest vote. In the next chapter, several more key tables based on the results from this chapter are produced and discussed.

CHAPTER 6

ANALYSIS AND DISCUSSION

In the previous chapter, we presented 3D facial expression classification using3D facial surface normal features using Principal Component Analysis (PCA) along with the implementation of modular approach. The results of the experiments were briefly discussed. In this chapter, the key table is produced and a thorough analysis is carried out.

6.1 Comparison with a Non – Modular Approach

For the purpose of assessment of our modular approach, we carried out a non–modular facial expression classification using 3D facial surface normals. Table 6.1 shows average classification rates in our work which includes both non-modular and modular with WVS and MVS approaches. Across all classifiers and non-modular/modular approaches, 3D facial surface normals show improvement in the modular approach for both classifiers and shows the best overall results for both classifiers. Results using MVS records a poor result compared to WVS.

Table 6.1 Average classification rates in our work

3D Facial Features	Nearest Neighbour Classifier	Support Vector Machine
Non-Modular 3D Facial Surface Normals	46%	54%
Modular 3D Facial Surface Normals – WVS	51%	66%
Modular 3D Facial Surface Normals – MVS	36%	53%

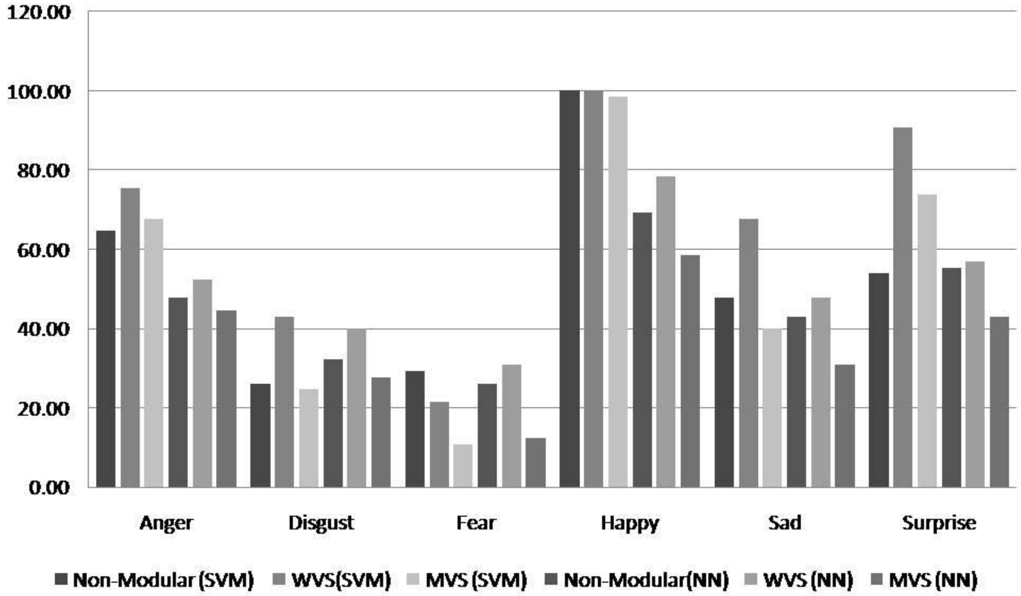


Figure 6.1 Success classification rates using 3D facial surface normals as the feature vector

Figure 6.1 shows the success classification rate using 3D facial surface normals as the feature vector. The SVM classifier with WVS produces the best results across all expressions except for the Fear expression whereas the non-modular approach using the SVM classifier shows an improved result.

6.2 Comparison with Other 3D Features

Table 6.2 Average classification rates in our modular work

3D Facial Features	Nearest Neighbour Classifier	Support Vector Machine
Modular 3D Facial Points – WVS	41%	50%
Modular 3D Facial Points – MVS	32%	46%
Modular 3D Distance Measurements – WVS	33%	46%
Modular 3D Distance Measurements – MVS	26%	36%
Modular 3D Facial Surface Normals – WVS	51%	66%
Modular 3D Facial Surface Normals – MVS	36%	53%

Table 6.2 shows average classification rates in our work using both WVS and MVS approaches. Across all classifiers and voting system approaches, 3D distance measurements classification performance is the worst. 3D facial points classification rate is slightly improved when the nearest neighbour classifier is used. A better result for 3D distance measurements is achieved when SVM is used. 3D facial surface

normals show improvement for both classifiers and show the best overall results for both classifiers regardless the voting system used. Results using MVS records a poor result compared to WVS. Based on these results, the following figures (6.2 and 6.3) describe the success rates for 3D facial features in each expression using specifically WVS approach.

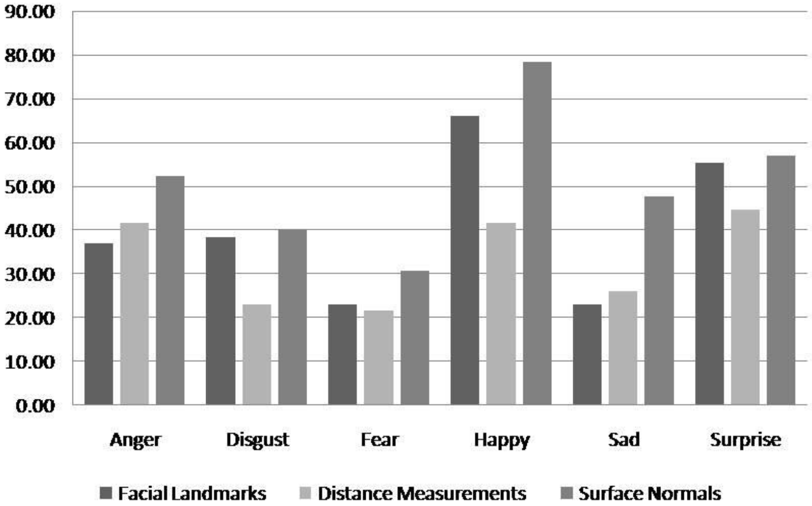


Figure 6.2 Success rate of WVS using the nearest neighbour classifier

Figure 6.2 shows the success facial gesture classification rate of WVS using the nearest neighbour classifier for the three 3D facial features. The average rate of classifications of the three feature vectors are 41%, 33% and 51% for 3D facial points, 3D distance measurements and 3D facial surface normals respectively. Using 3D facial surface normals as the baseline feature with the nearest neighbour classifier definitely improves the facial expression classification rate compared to the other two feature vectors. Overall, Fear has the worst classification results across the 3D features.

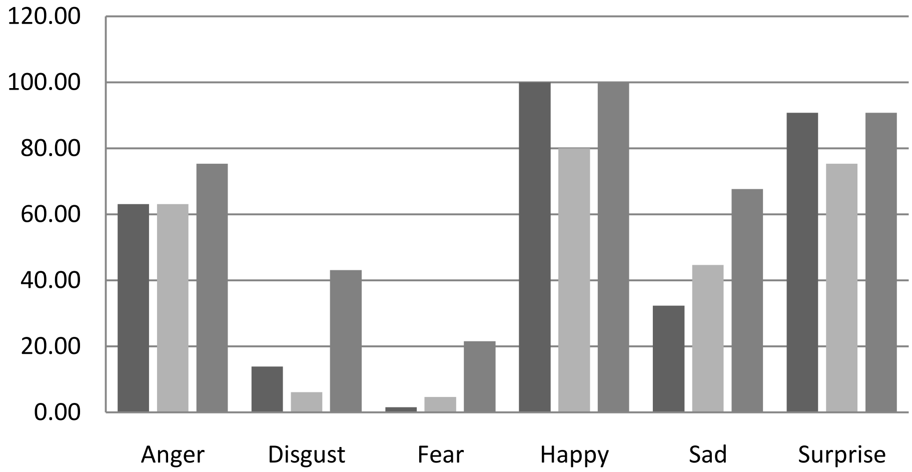


Figure 6.3 Success rate of WVS using the SVM classifier

Figure 6.3 shows the success classification rate of WVS using the SVM classifier across the 3D features. Similar to 3D facial points, 3D facial surface normals record a 100% correct classification for the Happy expression. For the Disgust expression, 3D facial surface normals has a 29% higher classification rate than using 3D facial points. 3D facial surface normals have an equal rate of correct classification for the Surprise expression with 3D facial points. Overall, 3D facial surface normals perform quite well compared to the two other feature vectors where the average classification of 3D facial surface normals is 66% compared to 50% and 46% for 3D facial points and 3D distance measurements respectively.

Table 6.3 Modules and voting scheme results for the Happy expression using 3D Facial Surface Normals.

Happy Expression	**Nearest Neighbour Classifier**	**Support Vector Machine**
Eyebrows	46.15%	56.92%
Mouth	70.77%	98.46%
Eyes	53.13%	73.85%
Cheeks	55.38%	90.77%
Nose	46.15%	87.69%
Forehead	55.38%	93.85%
Majority Voting Scheme (MVS)	58.40% (FP: 33.85%)	98.46% (FP: 1.54%)
Weighted Voting Scheme (WVS)	78.46%	100%

Based on figure 6.2 and 6.3, Happy expression recorded the highest correct classification compared to the other expressions across classifiers. Table 6.3 shows the modules and voting scheme results for the Happy expression using 3D facial surface normals. The results using SVM across modules are better than using the nearest neighbour classifier. Therefore, the results of both voting schemes are improved. However, MVS yielded one FP case. Even though there are four modules in the nearest neighbour classifier which achieves more than 50% correct classification, the MVS results still failed to achieve as least on par with WVS using SVM. In MVS, an expression class of x needs to have at least 4 votes to be voted as x. If it has less than 4 votes, it will be misclassified as another expression or it will be voted as an FP case. These results tell us that in most expressions, not all modules have the same classification results and that explains the poor results of MVS using the nearest neighbour classifier.

Nabatchian et al [36] divide the image into several sub-images and perform the training and classification process based on these sub–images in their face recognition system with illumination variation experiments. They used two types of voting scheme which is the Democratic Voting Scheme (DVS), also known as the Majority Voting Scheme and WVS to fuse the results of sub–images classification. In their use of WVS, the weights were set based on the illumination condition of each sub–image. We proved that our results are in agreement with theirs where WVS performs better than MVS.

6.3 Comparison with Other Studies

The results in this work still do not achieve at least the 83% correct classification rate which Wang et al., (2006) reported, despite the 3D database difference. Table 2.4 in Chapter 2 shows the average classification rates in other works using 3D facial static data. Although they are not directly comparable with our work, it shows the achievement in the 3D facial expression analysis area. Wang et al. (2006) carried the first experiment of 3D facial expression classification using BU-3DFE database. They recorded the average rate of 83.1%. The highest correctly classified expression was Happy (95%). There were four expressions under 81%, which were Fear, Sad, Disgust and Anger. The highest average recall rate for the experiment that used the BU-3DFE database was achieved by Maalej et al., (2010). Soyel et al. (2007) also include Neutral expression in their experiment along with six universal facial expressions. Anger and Neutral expressions have the lowest rate which is less than 90%. Tang et al., (2008) achieved an 87.1% average classification rate using a multi-class SVM classifier. The highest average classification obtained is 99.2% for the classification of Surprise. However, none of these works achieved the 100% correct classification of Happy expression which we attained in our study.

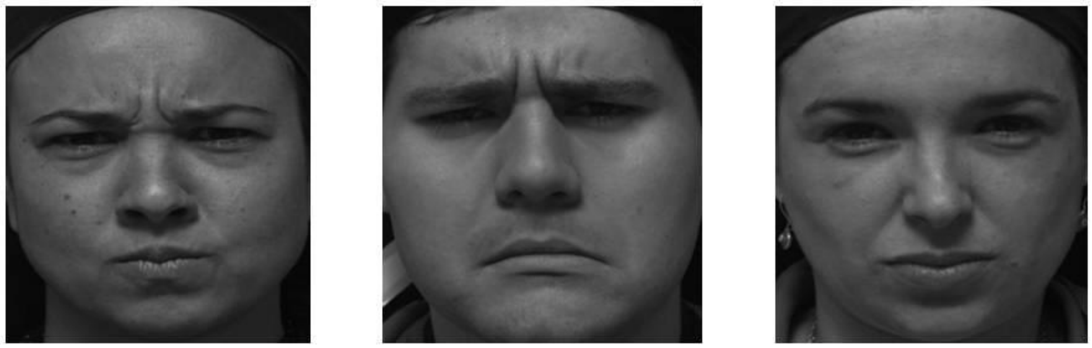


Figure 6.4 The subject in the first row show three different intensities of Anger expression, ranging from low (left) to high (right) intensity (taken from Frowd et al., 2009). The three different subjects from the Bosphorus Database portray Anger expression with no intensity information in the second row.

We also mentioned in Chapter 2, Wang et al., (2006), Soyel et al., (2007), Tang et al., (2008) and Gong et al., (2009) used BU-3DFE database in their study and they only include each of the facial expressions with the two highest levels of intensity. On the other hand, the Bosphorus database does not provide facial expression with intensity information. We believe this is the reason for the significant difference in the output as the facial expression with higher intensity means the deformation of each facial feature is obvious and easy to classify. Figure 6.4 shows the illustration of different intensity of Anger expression in the first row taken from Frowd et al. (2009) while the three different subjects from Bosphorus database portray Anger expression with no intensity information in the second row.

However, the Happy expression managed to achieve a 100% classification rate whereas none of the previous works have achieved this. This happened as each subject shows practically the same level of intensity for the Happy expression in the Bosphorus Database. We also believe the intensity of the Happy expression portray by the subjects from the Bosphorus Database are fairly high when compared to Frowd et al. (2009) facial expression intensity levels (see figure 6.5).

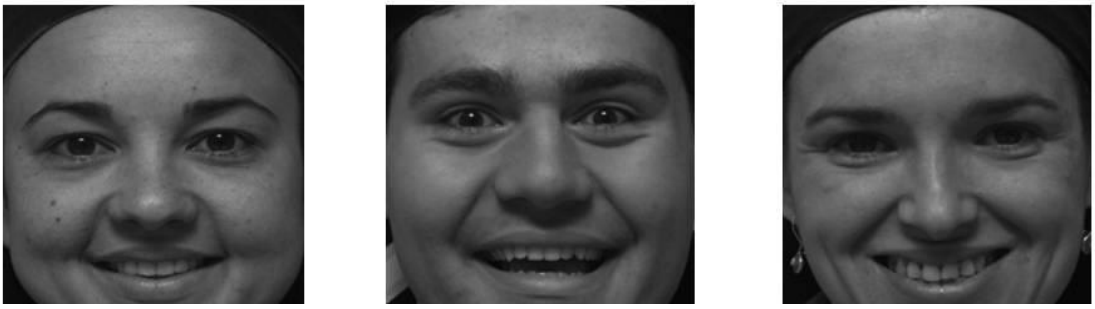


Figure 6.5 The subject in the first row show three different intensities of Happy expression, ranging from low (left) to high (right) intensity (taken from Frowd et al. (2009)). The three different subjects from the Bosphorus Database portray Happy expression with no intensity information in the second row

Hesse (2011) carried out facial expression classification utilizing 2D landmark coordinates, AAM shape and appearance parameters, SIFT and DCT appearance descriptors and combinations of AAM parameters and SIFT/DCT descriptors. An AAM contains a statistical model of the shape and grey-level appearance of the object of interest which can generalize to almost any valid example (Cootes et al, 1998). PCA is used in AAM to obtain both shape and appearance parameters. According to Hesse (2011), the worst classification performance is the shape parameters.

Table 6.4 A comparison of classification accuracies for different expressions between 2D based approach (Hesse, 2011) and 3D based approach (non-modular and modular with WVS)

Expressions	Hesse (2011)	Our work (Non-modular)	Our work (modular with WVS)
Anger	47.1%	64.6%	75.4%
Disgust	49.6%	25.1%	43.1%
Fear	38.1%	29.2%	21.5%
Happy	58.7%	100%	100%
Sad	49.4%	47.8%	67.7%
Surprise	67.2%	53.9%	90.7%
Average	51.7%	54%	66%

Table 6.4 shows a comparison of classification accuracies for different expressions between 2D and 3D approach. We used Hesse's results to represent the 2D approach. However, Hesse's result is averaged for all poses for 2D shape parameters. These results have highlighted the advantages of using3D modalities over 2D. Hesse used 2D images and 2D facial landmark points in AAM to generate shape and appearance parameters. In our study, we proposed 3D facial surface normals as the feature to be inserted into PCA and we used the obtained shape parameters as the feature vectors in the facial expression classification phase. If we look through this aspect (using shape parameters), we managed to improve Hesse's result especially in

Anger, Happy, Sad and Surprise expressions. However, the results on Disgust and Fear expressions are still worse.

A similar concept of surface normals is also used in the work of Ceolin’s (2007) and Sandbach et al., (2012). Sandbach et al (2012) introduced LBNP which uses the same concept of surface normals for only AUs classification. Ceolin (2012) used a 2.5D facial surface normals (or known as facial needle maps) which is acquired from 2D intensity images using Shape from Shading (SFS), referred to as Principal Geodesic Shape-From-Shading (PGSFS).In their work, the confusion matrix of facial expression classification results are not provided therefore any comparison with their works cannot be carried out.

6.4 Issues on Disgust and Fear Expression

We believe that the deformation of the Eyebrows is really significant in any facial expression which is in agreement with our AdaBoost experiment. Therefore, we decided to make it a separate module instead of combining it with the Eyes module. However it turns out that, the results for the Eyebrows module independently are not as good as the other modules. We also considered that the small number of 3D facial features computed in this module is the reason for the poor result.

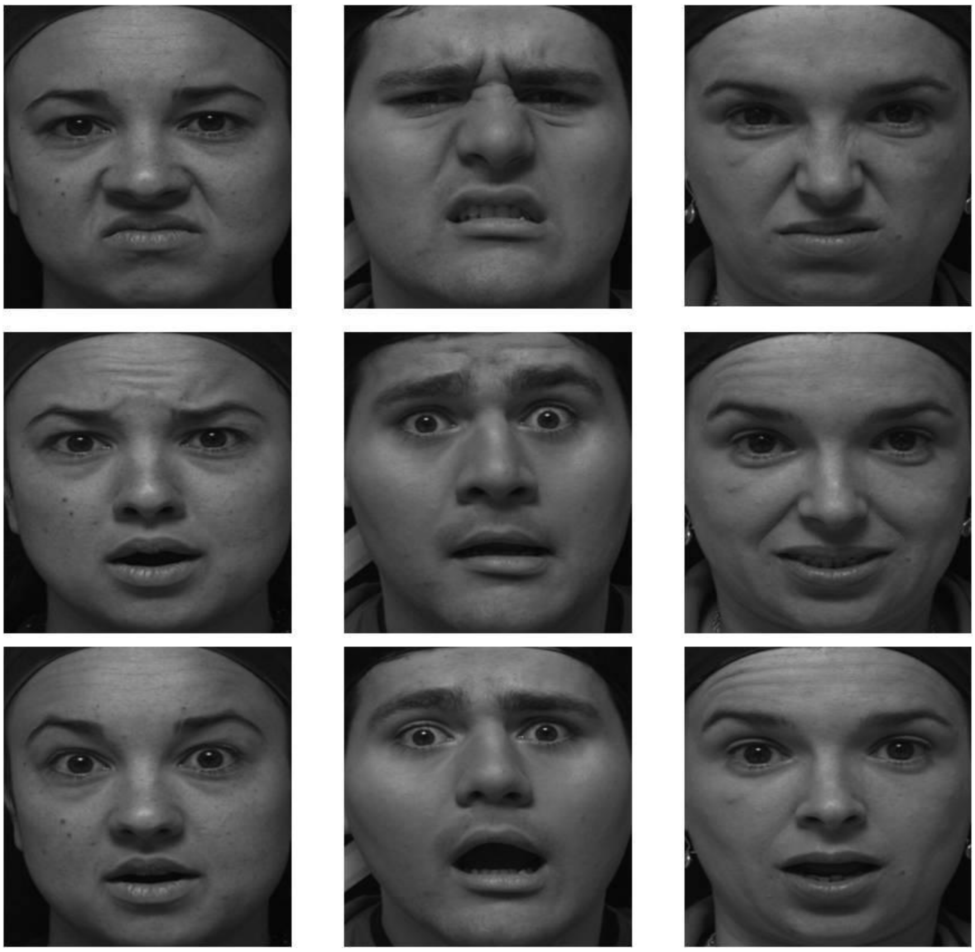


Figure 6.6 Three subjects A, B and C (from left to right) with Disgust (first row), Fear (second row) and Surprise (third row) expressions.

If we look closely at the Disgust and Fear expressions for the three different subjects in figure 6.6, we can see the differences between them specifically in the Eyebrows, Mouth and Eyes modules which are the modules with the largest weight. The large differences in those modules are significant in the classification phase, where for example, for the most intense Fear expression; there is an opened mouth as in the Surprise expression (see figure 6 for comparison). However, subject B and C did not show the same mouth deformation. To differentiate between Fear and Surprise, the Eyebrows for the Fear expression should be showing a different deformation in the Eyebrows to the Surprise expression. However, subject C is still showing the similar deformation in the Eyebrows in the Surprise and Fear expression; hence it is obvious we cannot differentiate the Surprise and Fear expression because of the same deformation in Eyebrows for both expressions.

Savran et al (2008) states that for the Disgust expression, only two AUs are involved namely AU9 (Nose wrinkler) and AU10 (Upper lip raiser) while in Ekman et al. [25], the Disgust expression is noted as having AU9 (Nose wrinkler), AU15 (Lip corner depressor) and AU16 (Lower lip depressor). Subject B in figure 12 showed the AU10 clearly while subject C is showing a bit less deformation of AU10. The eyebrows deformation for the three subjects is also different. Even though both Savran et al (2008) and Ekman and Friesen (1978) do include AU9 in the Disgust expression, it is still not enough to really distinguish the Disgust expression from other expressions since the weighting for the nose module is not large.

Table 6.5 Facial expression classifications of 3D facial surface normals for the Disgust and Fear expressions using nearest neighbour classifier and SVM

	Eyebrows	**Mouth**	**Eyes**	**Cheeks**	**Nose**	**Forehead**	**Modules Average**	**WVS**
Disgust-NN	24.62	32.31	20.00	36.92	29.23	27.69	28.46	40.00
Disgust-SVM	1.54	7.69	27.69	33.85	43.08	12.31	21.03	43.08
Fear - NN	4.62	26.15	26.15	27.69	32.31	15.38	22.05	30.77
Fear – SVM	0.00	26.15	32.31	15.38	25.08	9.23	18.03	21.54

Table 6.5 is an extract from the results presented before to summarize facial expression classifications of 3D facial surface normals for the Disgust and Fear expressions using nearest neighbour classifier and SVM in percentages. Even though the average rate of the modules for the Disgust expression using nearest neighbour is higher than using SVM, the WVS result using SVM is a slightly higher than the WVS result using nearest neighbour. This is because in the nearest neighbour case, for some of the test cases, the Eyebrows module was not supported by other modules, due to the combination of other modules having more weighting than the Eyebrows module alone. Our approach in this work is to project the feature vectors to the subspace. In the aspect of classifier difference, nearest neighbour classifier find a face that has the shortest distance to the "probe" face while the feature vectors have to undergo the training phase before being classified in SVM. The poor results for each module using SVM approach is due to the varying deformation of the facial feature in each module.

This will lead to indistinct separable gap between expressions, thus affecting the success rates using SVM. Furthermore, in SVM, if the number of feature vectors is much greater than the number of samples, the method is likely to give poor performances (Scikit-learn, 2013).In the Fear expression case, it is obvious that the success rate for each of the modules using SVM is not consistent enough due to the Eyebrows success rate which is 0%to produce a good WVS result. Furthermore, the higher value of surface normals standard deviation compared to two other features (refer to figure 5.17) means the success rate of surface normals across the modules are inconsistent.

Hesse (2011) also provided facial expression classification results for different facial expression intensities in which they have 4 levels of intensity. Based on their result, the classification accuracies are improved from level 1 up to level 4 of expression intensity. We also believe, in agreement with Hesse (2011), that facial expression classification for static data should be conducted according to the level of intensity. Thus, different deformation of facial features can be captured for every level of intensity. However, the comparison between our work and Hesse (2011) is not fair as there is a difference in terms of data modalities.

6.5 Conclusions

In this chapter, we produce several key tables based on the results taken from Chapter 4 and 6. Then, the analyses and discussions of the key tables are carried out. In general, the modular approach of WVS has a significant effect on 3D facial expression classification especially using 3D facial surface normals where a consistent result for the classification of the Anger, Happy and Surprise expressions are obtained. Also the Happy expression had a 100% success classification rate in both modular and non-modular approaches using the SVM classifier. All other expressions showed an improvement except the Fear expressions. The Disgust and Fear expressions have a low success classification rate in general. In the next chapter, we will summarize our study and propose several future works to improve these results.

CHAPTER 7

CONCLUSIONS

In this chapter, first we summarize the contributions of our work and then address the limitations of the developed methods. Following the analysis, we discuss some possible solutions and propose several suggestions for 3D facial expression classification. Section 7.1 restates the contributions. Section 7.2 addresses the limitations of our work and Section 7.3 proposes the directions for future works.

7.1 Summary of Contributions

The overall goal of this thesis comes in as package where a statistical modelling of 3D facial surface normals is used along with a modular approach. The resulting shape model on each module is used to perform six basic facial expression classifications with no expression intensity information provided.

The key results of our work is we have shown that 3D facial surface normals outperformed 3D facial points and 3D distance measurements as the feature vectors in 3D facial expression classification. In particular, we proved the feasibility of using 3D facial surface normals to capture face deformation produced by six basic facial expressions compared to the two other 3D facial features. In addition, we proved that each expression has a consistent distribution of surface normals which distinguish it from other expressions and therefore the facial deformation of each facial expression is easily monitored.

By using the modular approach, the discriminative variations of the facial features in each module are emphasised. We explored a modular approach and decomposed a face into six modules. We performed facial expression classification for each module independently. The classifications of each module are combined using WVS to determine the final classification of a facial expression. The priority rank experiment using AdaBoost has proved that the most important facial feature in six basic facial expressions is the eyebrows area while the less important is the forehead. The WVS used the modules priority rank result as the weighting factor. We also proved that WVS approach is better than using MVS to infer the expression from six different modules. The modular approach of WVS has a significant effect on 3D facial expression classification especially using 3D facial surface normals where a consistent result for the classification of the Anger, Happy and Surprise expressions are obtained. Also, the Happy expression had a 100% success classification rate in both modular and non-modular approaches using the SVM classifier.

Hesse (2011) used a similar approach which is using the shape parameters of PCA as the input to a classifier with 2D data. We compared our results with Hesse (2011) and it is improved by 14%. However, this comparison is indicated unfair due to data modalities difference. Furthermore, the Happy expression managed to achieve a 100% classification rate where none of the previous 3D studies have achieved this. On the other hand, Sandbach et al (2012) introduced LBNP which uses the same concept of surface normals for only AUs classification. Ceolin (2012) used a 2.5D facial surface normals (or known as facial needle maps) which is acquired from 2D intensity images using Shape from Shading (SFS), referred to as Principal Geodesic Shape-From-Shading (PGSFS). The PGSFS method is used to iteratively recover needle-maps that realistically capture facial shape and also satisfy the image irradiance equation as a hard constraint. Therefore, the recovered facial needle-maps both encode facial shape information and implicitly capture facial texture information. However, in their work, the confusion matrix of facial expression classification results are not provided therefore any comparison to with their works cannot be done.

7.2 Limitations

We performed facial expression classification for every module. The worst average classification rate was the Eyebrows module. This was due to the small number of 3D facial features computed in the Eyebrows module. The 3D facial expression classification produced by combining the results of individual modules using WVS was still not improved when compared with other 3D studies. The WVS method depends on the weighting factor. The Eyebrows had the largest weighting and because of the worst average classification rate it had, it affected the results of WVS.

The existing 3D studies only include the 3D facial expression data with the two highest levels of intensity. On the other hand, the Bosphorus database does not provide facial expression with intensity information. We believe this is the reason for the significant difference of the results between modules and the results with other 3D studies as the facial expressions with higher intensities mean the deformation of each facial feature is evident and easier to classify.

7.3 Future Directions

Having addressed the limitations of the method described in this thesis, in this section we put forward suggestions for future work to improve the results presented and propose a few suggestions for future research in 3D facial expression classification.

Even though the Bosphorus database does provide AU deformation data with intensity information, it is not the same in the case of 3D facial expression data. Our results are degraded due to the intensity differences in each subject shown in the Bosphorus database. For our future work, we would like to carry out facial expression classification experiments using 3D facial data with intensity information, specifically with the six basic facial expressions with the highest intensity level. We believe that our proposed approach will achieve a good result using this kind of data. Furthermore, we would like to train our system using facial expressions with different intensity levels and as a result, we will be able to classify the intensity level of facial expression.

According to Lucey et al., (2002), shape feature yielded higher classification for only certain AUs. For example AUs 1, 2 and 4 coincide with eye brow movement which can be easily picked up by the shape feature. However, for AUs 6 (cheek raiser), 9 (nose wrinkler) and 11 (nasolabial deepener), there is a lot of textural change in terms of wrinkles and not so much in terms of contour movement, which suggested that 2D appearance data performed better than 2D shape data for those AUs. In other words, certain AUs could only measure permanent features and the others only measure transient features. Their facial expression classification experiments also yielded that the combination of 2D shape and appearance features performed better than using shape or appearance features independently. We would like to further examine this theory with our approach in the future where we will fuse 3D facial surface normals with 2D appearance data using our modular approach.

To enable the 3D facial expression classification system to be use in unlimited types of facial deformation as well as to make a reliable AU monitoring system, the symmetrical assumption for face deformation needs to be relaxed. Furthermore, there are subjects who show an expression which was among the six basic expressions with unsymmetrical deformation of facial features, specifically the Eyebrows (refer to Figure 5.3). In future, we would like to extend our approach to two types of advances for the 3D face processing problem (i) to classify more than six basic facial expressions and (ii) to be able to monitor any AU deformation. In order to do this, the facial features will be assumed asymmetrical and therefore a face will be decomposed into more than six modules. We believe this will also solve the 3D face occlusion problem cause by hand, hair etc.

As systems aim to analyse more subtle facial expressions, it has emerged that dynamic information is very important. We believe that with dynamic information fused with our modular approach, we could monitor the AU/FAP deformation of each module easily and different intensities of each facial feature can be captured. In the future, we would like to use 3D facial surface normals as the feature vector in a modular approach with 3D facial dynamic data.

APPENDIX A

In chapter three, we discussed two ways of doing multiclass classification using SVMs: (i) one-versus-all classifiers (OVA) and (ii) one-versus-one classifiers (OVO). In this work, OVA is chosen for its simplicity, practicality and to avoid the intensive computation of OVO.

To mathematically describe multiclass SVM, we must begin with the binary SVM description. SVM map an input sample to a high dimensional feature space and try to find an optimal hyperplane that minimizes the classification error for the training data using the non–linear transformation function (Sebald, et al., 2001).

$$\mathrm{X} : x = (x_1, \dots, x_n) \rightarrow F\colon \Phi(x) = (\Phi_1(x), \dots, \Phi_n(x)) \qquad \text{(A.1)}$$

In a binary classification situation, let N be the number of training samples where each input $\mathbf{x}_\mathrm{i}$ is in one of two classes $y_i = -1$ or $+1$. The inputs to the training algorithm are the sets $\{\mathbf{x}_{\mathrm{i},}, \mathrm{y}_\mathrm{i}\}$ where $i = 1, \dots, N, y_i \in \{-1, 1\}$

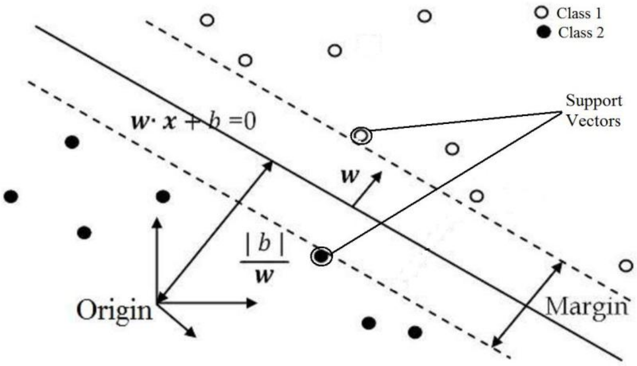


Figure A.1 Hyperplane through two linearly separable classes.

As described in figure A.1, the support vectors are the training samples closest to the hyperplanes and the SVM goal is to orientate the hyperplane to be as far as possible from the closest data for both classes.

The hyperplane can be described mathematically by $\mathbf{w} \cdot \mathbf{x} + b = 0$ where

- $\mathbf{w}$is the normal to the hyperplane
- $\frac{\mathrm{b}}{\|\mathbf{w}\|}$is the perpendicular distance from the hyperplane to the origin.

The crucial point in SVM is to select variables $\mathbf{w}$ and b, so that our training data can be described by:

$$\mathbf{w} \cdot \mathbf{x} + b \geq +1 \qquad \text{for } y_i = +1 \qquad \text{(A.2)}$$

$$\mathbf{w} \cdot \mathbf{x} + b \leq -1 \qquad \text{for } y_i = -1 \qquad \text{(A.3)}$$

The SVM algorithm makes a prediction based on a function of the form

$$f(x) = \sum_{i=1}^{N} w_i K(\mathbf{x}, \mathbf{x}_i) + b \tag{A.4}$$

where $\mathbf{w}^T = [w_1, w_2, \dots, w_i]$ are the weights. $K(.,.)$is a kernel function and in here we used a linear kernel which is $(\mathbf{x}, \mathbf{x}_i)$. b represents a bias term and trainable parameter.

Given a multiclass training set composed of P disjoint classes, we would like to use a linear kernel to train an OVA classifier for some arbitrary target class, j. The training dataset $\left(x_j, c_j\right)$ consists of N examples belonging to P classes. The class label is $c_i \in 1, 2, \dots, P$. Each SVM classifies samples into corresponding classes against all other classes in the OVA method. All N training examples are used in constructing an SVM for a class. The SVM for class j is constructed using the set of training examples and their desired output (x_i, y_i). The desired outputy_i for a training example x_i is defined as follows:

$$y_j = \begin{cases} +1 & if\ c_i = j \\ -1 & if\ c_i \neq j \end{cases} \tag{A.5}$$

REFERENCES

Al – Osami, F., Bennamoun, M. and Mian, A., 2009. An Expression Deformation Approach to Non – Rigid 3D Face Recognition. *International Journal of Computer Vision*, 81(3), pp 302 – 316.

Blanz, V. and Vetter, T., 1999. A Morphable Model for the Synthesis of 3D Faces.*26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pp 187 – 194.

Benbouzid, D., Busa-Fekete, R., Casagrande, N., Collin, F.-D. and Kégl, B., 2012. MultiBoost: a Multi-Purpose Boosting Package. *Journal of Machine Learning Research*, 13, pp. 549 – 553.

Belhumeur, P., Hespanha, J., Kriegman, D., 1997. Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7)**,** pp711–720.

Benedikt, L. Cosker, D., Rosin, P. and Marshall, D. Assessing the Uniqueness and Permanence of Facial Actions for Use in Biometric Applications, *IEEE Transactions System Man Cybernetics.* 40 (3) (2010) 449–460.

Beszédeš, M. and Culverhouse, P.F., 2007. Facial Emotions and Emotion Intensity Levels Classification and Classification Evaluation. *British Machine Vision Conference*.

Bettadapura, V. 2012. Face Expression Recognition and Analysis: The State of the Art. *Computing Research Repository Journal (CoRR)*.

Boiman, O., Shechtman, E. and Irani, M. 2008. In Defense of Nearest-Neighbour Based Image Classification. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, 24-26 June 2008, Anchorage, Alaska, USA

Ceolin.S.R., 2012. *Facial Shape Space using Statistical Models from Surface Normal.* PhD. University of York.

Chiang, C.C., Chen, X. and Yang, C., 2009. A Component-Based Face Synthesizing Method, *APSIPA Annual Summit and Conference*.

Cootes, T.F., Edwards. G.J. and Taylor, C.J., 1998. Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Springer-Verlag, pp. 484 – 498.

Carroll, J. M. and Russell.J., 1997.Facial Expression in Hollywood's Portrayal of Emotion. *Journal of Personality and Social Psychology*. 72:164–176.

DataFace, 2003.*Introduction to the DataFace Site: Facial Expressions, Emotion Expressions, Nonverbal Communication, Physiognomy*. [online]. Available at: http://face-and-emotion.com [Accessed 18 February 2013].

Dongcheng, S. and Jieqing, J., 2010. The Method of Facial Expression Recognition based on DWT-PCA/LDA, *3rd International Congress on Image and Signal Processing (CISP),* pp. 1970 – 1974.

Ekman, P. and Friesen, W.V. 1978. Facial Action Coding System: A Technique for the Measurement of Facial Movement. *Consulting Psychologists Press*, Palo Alto.

Ekman, P. and Rosenberg, E. L. 2005.What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS), second Ed., *Oxford University Press*.

Faltemier, T.C., Bowyer, K.W. and Flynn, P.J. 2007. Using a Multi-Instance Enrolment Representation to Improve 3D Face Recognition. First *IEEE International Conference onBiometrics: Theory, Applications, and Systems,(BTAS).*

Fasel, B. and Luettin, J. 2003. Automatic Facial Expression Analysis: a Survey. *Journal of Pattern Recognition*.

Fang, T., Zhao, X., Ocegueda, O.,Shah, S.K. and Kakadiaris, I.A., 2011. 3D Facial Expression Recognition: A Perspective on Promises and Challenges. *IEEE*

International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011),pp. 603 – 610.

Freund, Y. and Schapire, R. E., 1997.A Decision Theoretic Generalization of On-Line and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1), pp. 119 – 139.

Frowd, C. D., Matuszewski, B. J., Shark, L. and Quan, W., 2009. Towards a Comprehensive 3D Dynamic Facial Expression Database. *9th WSEAS International Conference on Signal, Speech and Image Processing*, pp. 113 – 119.

Gökberk, B., İrfanoğlu, M. O. and Akarun, L., 2006.3D Shape-based Face Representation and Feature Extraction for Face Recognition. *Journal of Image and Vision Computing*, 24(8), pp. 857-869.

Gong, B., Wang, Y., Liu, J., and Tang, X., 2009. Automatic Facial Expression Recognition on A Single 3D Face by Exploring Shape Deformation. *ACM Multimedia*, pp. 569-572.

Gottumukkal, R. and Asari, V. K. 2003. An Improved Face Recognition Technique based on Modular PCA Approach. *Pattern Recognition Letter*, 25(4), pp. 429 – 436.

Ghahramani, Z. 2004. Unsupervised Learning, Bousquet, O., Raetsch, G. and von Luxburg, U. (eds) Advanced Lectures on Machine Learning, *Springer-Verlag*.

Gupta, S., Markey, M.K and Bovik, A.C. 2010. Anthropometric 3D Face Recognition. *International Journal on Computer Vision, Springer*.

Heseltine^1, T., Pears, N. And Austin, J., 2004. Three-Dimensional Face Recognition: A Fishersurface Approach. *International Conference on Image Analysis and Recognition (ICIAR)*, pp.684 – 691.

Heseltine^2, T., Pears, N. And Austin, J., 2008. Three-Dimensional Face Recognition using Combinations of Surface Feature Map Subspace Components. *Journal of Image and Vision Computing*, 26(3), pp. 382 – 396.

Hesse, N., 2011.Multi – View Facial Expression Classification. Diploma Karlsruher Institutfür Technologie.

Hong, J., Min, J., Cho, U. and Cho, S., 2008. Fingerprint Classification using One-vs-All Support Vector Machines Dynamically Ordered With Naıve Bayes Classifiers. *Journal of Pattern Recognition*, 41(2), pp. 662 – 671.

Huenerfauth, M., Lu, P. and Rosenberg, A. 2011.Evaluating Importance of Facial Expression in American Sign Language and Pidgin Signed English Animations. *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS),* pp 99 - 106.

Hwang, B-W., Blanz, V., Vetter,T. and Lee, S-W. 2000. Face Reconstruction using a Small Set of Feature Points. *Lecture Notes in Computer Science of Biologically Motivated Computer Vision,* pp. 308-315.

Jaimes, A. and Sebe, N. 2007. Multimodal Human – Computer Interaction: A Survey.

Computer Vision and Image Understanding, 108(1-2): 116 – 134.

Ji, Q., Lan, P. and Looney, C. 2006. A Probabilistic Framework for Modelling and Real Time Monitoring Human Fatigue, IEEE Transactions on Systems, Man, and Cybernetics A, Vol. 36, No.35, p862-875.

Joachims, T., Finley, T., and Yu, C., 2009. Cutting-Plane Training of Structural SVMs, *Journal of Machine Learning*, 77(1), pp. 27 – 59.

Kapoor, A., Qi, Y. and Picard., R.W. 2003. Fully Automatic Upper Facial Action Recognition. *IEEE International Workshop on Analysis and Modelling of Faces and Gestures*.

Kapoor, S., Khanna, S. and Bhatia, R., 2010. Facial Gesture Recognition using Correlation and Mahalanobis Distance. *International Journal of Computer Science and Information Security (IJCSIS)*, 7(2).

Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z. And Matthews, I., 2002. The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset for Action Unit and

Emotion-Specified Expression. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 94 – 101.

Maalej, A., Ben Amor, B., Daoudi, M., Srivastava, A. and Berreti, S., 2010. Local 3D Shape Analysis for 3D Facial Expression Recognition. *International Conference on Pattern Recognition*, pp. 4129 – 4132.

Mathworks, 2012. Support Vector Machines (SVM). [online] Available at: http://www.mathworks.co.uk/help/toolbox/bioinfo/ug/bs3tbev-1.html [Accessed 2 August 2012].

Mazanec, J., Melisek, M., Oravec, M. and Pavlovicova, J. 2008. Support Vector Machines, PCA and LDA in Face Recognition. *Journal of Electrical Engineering*, pp. 203 – 209.

Moreno, A.B. and Sánchez, A., 2004. GavabDB: a 3D Face Database. *Workshop on Biometrics on the Internet* (March 2004), pp. 77-85.

Mpiperis, I., Malassiotis, S. and Strintzis, M.G. 2008. Bilinear Models for 3-D Face and Facial Expression Recognition. *IEEE Transactions on Information Forensics and Security*, 3(3), pp. 498 – 511.

Nabatchian, A., Abdel-Raheem, E. and Ahmadi, M., 2010.A Weighted Voting Scheme for Recognition of Faces with Illumination Variation.*11th International Conference onControl Automation Robotics & Vision (ICARCV),*pp. 896 – 899.

Pandzic, I. S. and Forchheimer, R., 2002. *MPEG-4 Facial Animation: The Standard, Implementation and Applications*. John Wiley & Sons, Inc.

Parrott, W.G. 2001. *Emotions in Social Psychology*. Psychology Press.

Philips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J. Worek, W. 2005. Overview of the Face Recognition Grand Challenge, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 947 – 954.

Pinto, S.C.D., Mena-Chalco, J.P., Lopes, F.M., Velho, L. and Cesar, R.M., 2011. 3D Facial Expression Analysis by Using 2D and 3D Wavelet Transforms. *18th IEEE International Conference on Image Processing (ICIP)*, pp. 1281 – 1284.

Pentland, A., Moghaddam, B. and Starner., T., 1994. View – based and Modular Eigenspaces for Face Recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 84 – 91.

Praseeda, L.V., Sasikumar, M. and Naveen, S., 2008.Analysis of Facial Expressions from Video Images using PCA.*World Congress on Engineering*.

Prabhu, U. and Seshadri, K. 2009. Facial Recognition Using Active Shape Models, Local Patches and Support Vector Machines. [online] Available at: http://www.contrib.andrew.cmu.edu/~kseshadr/ML_Paper.pdf . [Accessed 16 February 2013].

Rabiu, H., Saripan., M.I., Mashohor, S. And Marhaban, M.H. 2012. 3D Facial Expression Recognition using Maximum Relevance Minimum Redundancy Geometrical Features. *EURASIP Journal on Advances in Signal Processing, Springer Open Journal.*

Rady, H., 2011. Face Recognition using Principal Component Analysis with Different Distance Classifier. *International Journal of Computer Science and Network Security (IJCSNS)*, 10(10), pp. 134 – 144.

Raouzaiou, A., Tsapatsoulis, N., Karpouzis, K. and Kollias, S., 2002. Parameterized facial expression synthesis based on MPEG-4. *EURASIP Journal Applied Signal Process*.2002, 1 (January 2002), pp. 1021-1038.

Rose-Hulman, 2010. *Professional Development: Effective Communication*. [online] at: http://www.rose-hulman.edu [Accessed 23 November 2012].

Russell, J. A. and Bullock.M., 1986. Fuzzy Concepts and the Perception of Emotion in Facial Expressions. *Social Cognition*, 4(3), pp. 309 – 341.

Russell, J. A. and Fernández – Dols, J.M., 1997.*The Psychology of Facial Expression*. Cambridge University Press.

Sandbach^1, G.,Zafeiriou, S. and Pantic, M. and Rueckert, D. 2012.Recognition of 3D Facial Expression Dynamics, *Journal of Image and Vision Computing (in press).*

Sandbach^2, G.,Zafeiriou, S., Pantic, M. and Yin, L. 2012.Static and Dynamic 3D Facial Expression Recognition: A Comprehensive Survey, *Journal of Image and Vision Computing*. 30(10): pp. 683 - 697, *3D Facial Behaviour Analysis and Understanding*.

Sandbach^3, G.,Zafeiriou, S., and Pantic, M. 2012.Local Normal Binary Patterns for 3D Facial Action Unit Detection. *Proceedings of the IEEE International Conference on Image Processing (ICIP 2012). Orlando, FL, USA*, October 2012.

Savran, A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B. and Lale, A., 2008. Bosphorus Database for 3D Face Analysis. In: Schouten, B., Juul, N.C., Drygajlo, A. and Tistarelli, M., eds. 2008. *Biometrics and Identity Management*. Springer – Verlag, Berlin, Heidelberg, pp. 47-56.

Savran, A., Sankur, B., Bilge, M. T., 2012. Comparative Evaluation of 3D versus 2D Modality for Automatic Detection of Facial Action Units. *Pattern Recognition*, 45(2), pp. 767-782.

Schapire, R. E. and Singer, Y., 1999.Improved Boosting Algorithm using Confidence - Rated Predictions.*11th Annual Conference on Computational Learning Theory*, pp. 80 – 91.

Scikit-learn, 2013. [online]. Support Vector Machines. Available at: http://scikitlearn.org/stable/modules/svm.html.[Accessed 25 February 2013].

Sebald, D.J. and Bucklew, J.A., 2001. Support Vector Machines and the Multiple Hypothesis Test Problems. *IEEE Transactions on Signal Processing*, 49(11), pp.2865 - 2872.

Sharath, S.S. Murthy K.N. B. and Natarajan S., 2011, Dimensionality Reduction Techniques for Face Recognition. In: Corcoran, P.M. , eds., 2011. Reviews, Refinements and New Ideas in Face Recognition, *InTech Publishing.*[online] Available at:http://www.intechopen.com/books/reviews-refinements-and-new-ideas-in-face-recognition. . [Accessed 5February 2013].

Silapachote, P., Karuppiah, D. R. and Hanson, A. R., 2005. Feature Selection Using Adaboost for Face Expression Recognition. *United States Defense Technical Information Center OAI-PMH Repository*. [online] Available at: http://www.dtic.mil/docs/citations/ADA438800.[Accessed 2 August 2012].

Soyel, H. and Demirel, H., 2007. Facial Expression Recognition using 3D Facial Feature Distances. *LNCS Book Series, Image and Analysis Recognition, Springer*, pp.831 – 838.

Soyel, H. and Demirel, H., 2010. Optimal Feature Selection for 3D Facial Expression Recognition using Coarse-to-Fine Classification. *Turkish Journal of Electrical Engineering & Computer Sciences*, 18(6), pp. 1031-1040.

Stratou, G., Ghosh, A., Debevec, P. And Morency, L.-P. 2011. Effect of Illumination on Automatic Expression Recognition: A Novel 3D Relightable Facial Database. *2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011)*, pp. 611 – 618.

Sun, Y., Reale, M. and Yin, L., 2008. Recognizing Partial Facial Action Units Based on 3D Dynamic Range Data for Facial Expression Recognition. *8th IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 1 – 8.

Tang, H. and Huang, T.S. 2008. 3D Facial Expression Recognition Based on Properties of Line Segments Connecting Facial Feature Points. *8th IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 1-6.

Tena, J.R., De la Torre, F. and Matthews, I., 2011. Interactive Region-Based Linear 3D Face Models.*ACM Transactions Graph*. 30(4): 76.

Tchoukanski, I. 2012. Triangulated Irregular Network. [online] Available at: http://www.ian-ko.com/resources/triangulated_irregular_network.htm. [Accessed 17 February 2013].

Tian, Y., Kanade, T. and Cohn, J.F. 2001.Recognizing Action Units for Facial Expression Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.Volume 23(2). Pp. 97-115.

Tian, Y., Kanade, T. and Cohn, J.F. 2005.Facial Expression Analysis. Handbook of Face Recognition. *Springer-Verlag*.Volume 3 (5).pp 247-276.

Trivedi, S. 2009. Face Recognition using Eigenfaces and Distance Classifiers: A Tutorial. Available at: http://onionesquereality.wordpress.com/2009/02/11/face-recognition-using-eigenfaces-and-distance-classifiers-a-tutorial/.[Accessed19 February 2013].

Tsalakanidou, F.' and Malassiotis, S. 2010. Real-time 2D+3D Facial Action and Expression Recognition. *Journal of Pattern Recognition*, 43(5), pp. 1763 – 1775.

Turk, M., Pentland, A., 1991.Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), pp71–86.

Valdovinos, R.M. and Sánchez, J. S., 2009.Combining Multiple Classifiers with Dynamic Weighted Voting.*4th International Conference on Hybrid Artificial Intelligence Systems*, pp. 510 – 516.

Velusamy, S. Kannan, H. Anand, B. Sharma, A. Navathe, B., 2011. A Method to Infer Emotions from Facial Action Units.*IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2028 – 2031.

Vezzetti, E. and Marcolin, F., 2012. 3D Human Face Description: Landmarks Measures and Geometrical Features. *Journal of Image and Vision Computing*.

Vinitha, K.V. and Kumar, G.S. 2009.Face Recognition using Probabilistic Neural Networks. *IEEE World Congress on Nature & Biologically Inspired Computing*, pp. 1388 – 1393.

Vishwakarma, V.P., Pandey, S. and Gupta, M. N., 2007. A Novel Approach for Face Recognition using DCT Coefficients Re-scaling for Illumination Normalization.*15th International Conference on Advanced Computing and Communications*, pp. 535-539.

Wang, J., Yin, L., Wei, X., and Sun, Y., 2006. 3D Facial Expression Based on Primitive Surface Feature Distribution. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1399-1406.

Wikipedia[1], 2013. Principal Component Analysis. Available at: http://en.wikipedia.org/wiki/Principal_component_analysis [Accessed 19 February 2013]

Wikipedia[2], 2013.Statistical classification. Available at: http://en.wikipedia.org/wiki/Statistical_classification [Accessed 13March 2013]

Weston, J. And Watkins, C., 1998. Multi – Class SVM. *Technical Report CSD - TR98 – 04*, University of London.

Yin, L., Wei, X., and Sun, Y., Wang, J., Rosato, M.J., 2006. A 3D Facial Expression Database for Facial Behaviour Research. *7th International Conference on Automatic Face and Gesture Recognition (FGR06),* pp.211 - 216

Zhang, Y., Ji, Q., Zhu, Z. & Yi, B., 2008. Dynamic Facial Expression Analysis and Synthesis with MPEG-4 Facial Animation Parameters. *IEEE Transaction Circuits System Video Technology,*18(10), pp. 1383-1396.

Zhao, X., Dellandrea, E. and Chen, L. 2009. A 3D Statistical Facial Feature Model and its Application on Locating Facial Landmarks. ACIVS, pp. 686–697.

Zhao, X., Dellandréa, E. Chen, L. and Samaras, D., 2010. AU Recognition on 3D Faces Based on an Extended Statistical Facial Feature Model. *4th IEEE International Conference Theory Applications and Systems (BTAS).*

Zhong, C., Sun, Z. and Tan, T., 2007. Robust 3D Face Recognition Using Learned Visual Codebook. *IEEE Conference on Computer Vision and Pattern Recognition CVPR),* pp. 1 – 6.