Modelling Facial Dynamics Change as People Age

A thesis submitted in partial fulfilment

of the requirement for the degree of Doctor of Philosophy

KHTAM M. AL-MEYAH

December 2017

Cardiff University
School of Computer Science & Informatics

Declaration

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ..... (candidate)

Date .....

Statement 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed ..... (candidate)

Date .....

Statement 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed ..... (candidate)

Date .....

Statement 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate)

Date .....

I hide my tears when I say their names, but the pain in my heart is still the same. Although I smile and seem carefree, there is no one who misses them more than me. I dedicate this work to the spirit of my brother and sister. I also dedicate my work to my family. I have a special feeling of gratitude to my loving parents, and to my sisters and brothers, who have been a constant source of support and encouragement during the challenges of the PhD. I am truly thankful for having them in my life.

Abstract

In the recent years, increased research activity in the area of facial ageing modelling has been recorded. This interest is attributed to the potential of using facial ageing modelling techniques for a number of different applications, including age estimation, prediction of the current appearance of missing persons, age-specific human-computer interaction, computer graphics, forensic applications, and medical applications. This thesis describes a general AAM model for modelling 4D (3D dynamic) ageing and specific models to map facial dynamics as people age. A fully automatic and robust pre-processing pipeline is used, along with an approach for tracking and inter-subject registering of 3D sequences (4D data). A 4D database of 3D videos of individuals has been assembled to achieve this goal. The database is the first of its kind in the world. Various techniques were deployed to build this database to overcome problems due to noise and missing data. A two-factor (age groups and gender) analysis of variance (MANOVA) was performed on the dataset. The groups were then compared to assess the separate effects of age on gender through variance analysis. The results show that smiles alter with age and have different characteristics between males and females. We analysed the rich sources of information present in the 3D dynamic features of smiles to provide more insight into the patterns of smile dynamics. The sources of temporal information that have been investigated include the varying dynamics of lip movements, which are analysed to extract the descriptive features. We evaluated the dynamic features of closed-mouth smiles among 80 subjects of both genders. Multi-level Principal Components Analysis (mPCA) is used to analyse the effect of naturally

occurring groups in a population of individuals for smile dynamics data. A concise overview of the formal aspects of mPCA has been outlined, and we have demonstrated that mPCA offers a way to model the variations at different levels of structure in the data (between and within group levels).

Acknowledgements

I am deeply thankful to God for His amazing grace, which has kept me from the beginning of my life until now.

My sincere thanks go to my supervisors Prof. David Marshall and Prof. Paul Rosin for their invaluable guidance and support over the course of the time I spent pursuing my PhD. I am grateful to Dr Damian Farnell for his efforts and very much appreciated his participation in this work.

My appreciation also goes to Prof. Ralph Martin and Prof. Steven Schockaert as annual reviewers, to my colleagues and the staff at School of Computer Science and Informatics, and especially to the computer vision group, for useful research discussions in different events.

I have to allocate a space to say thank you to Jason for everything.

My biggest thanks go to my family; they have supported me unconditionally since I began this journey in a foreign land. I donât even have the words to thank my mother, father, sisters, brothers, nieces and nephews, and everyone who wished me success.

I am extremely grateful to all the individuals who voluntarily left their faces be scanned yearly .

Contents

List of Figures

List of Tables

List of Abbreviations

AAM	Active Appearance Model
AGES	Ageing Pattern Subspace
ASM	Active Shape Model
BIF	Bio-inspired Features
BSIF	Binarised Statistical Image Features
CDTW	Continuous Dynamic Time Warping
CEA	Conformed Embedding Analysis
CLBP	Completed Local Binary Patterns
DDTW	Derivatives Dynamic Time Warping
DTW	Dynamic Time Warping
DTWI	Dynamic Time Warping Interpolation
FACS	Facial Action Coding System
FERET	Face Recognition Technology Dataset
FGNET	Face and Gesture Recognition Research Network Dataset
GWT	Gabor Wavelet Transform

HE	Histogram Equalisation
IFDB	Iranian Face Database
KRR	Kernel Ridge Regression
LBP	Local Binary Patterns
LPP	Locality Preserving Projections
LBPV	Local Binary Patterns Volume
MAE	Mean Absolute Error
MFDA	Multi-Features Discrimination Analysis
MORPH	Craniofacial Longitudinal Morphological Face Dataset
mPCA	Multilevel Principle Components Analysis
OLPP	Orthogonal Locality Preserving Projections
PAL	Park Ageing Mind Laboratory Dataset
PCA	Principal Components Analysis
SIFT	Scale Invariant Features Transform
SVM	Support Vector Machine
TPS	Thin Plate Spline
UTM	Unified Texture Map
VLF	Vietnamese Longitudinal Face
WDDTW	Weighted Derivatives Dynamic Time Warping
WIT-DB	Waseda human-Computer Interaction Technology Database

List of Symbols

C	Covariance matrix
D	The signal amplitude
$D_{lip}(t)$	The value of mean amplitude signal of the lip corners in frame t
$I(x, y)$	The intensity of the image pixel
k	The number of principle components
P	The point in the first sequence A
$P(x, y, t)$	The ageing pattern of face image at time t
Q	The point in the second sequence B
T_k	Mesh frame
T_{new}	Registered frame
T_{ref}	Face template (reference)
x, y and z	A pixel coordinates in a 3D image
x_i	The vector of variables of x, y and z
X	The matrix of the column-wise vector x_i
ϕ	The eigenvectors of the Covariance matrix C

λ_i	The eigenvalue of the Covariance matrix C
p	The percentage of total variation are kept.
ϕ_s	The eigenvector of shape parameters
b_s	The parameters of shape model
ϕ^T	The transpose of feature vector of eigenvectors
$\overline{S}$	The reconstruction frame
g_k	The set of image texture vector
$\overline{g}$	The mean texture map
ϕ_g	The eigenvector of texture parameters
b_g	The parameters of texture model
$\mathbb{Q}_c$	A set of orthogonal modes and c is a vector of appearance parameters controlling both the shape and grey-levels of the model.
W_s	A diagonal matrix of weights for each shape parameter, allowing for the difference in units between the shape and grey models
$D_{lip}(t)$	The value of mean amplitude signal of the lip corners in frame t
l_i^{th}	The 3D location of the i^{th} point in frame t
ρ	The Euclidean distance between two points
D^+	The increasing segment in amplitude signal D
D^-	The decreasing segment in amplitude signal D
ω	The frame rate
V	The speed signal

A	The acceleration signal

Chapter 1

Introduction

In recent years, the computer vision community has become increasingly interested in addressing the various issues associated with facial ageing such as age estimation, appearance modelling and face recognition/verification. The temporal nature of these facial ageing problems pose numerous challenges in establishing a formal dataset compared with other challenges such as illumination differences, pose variations and facial expressions. The success of any modelling that is employed for facial ageing is always dependent on how effectively the facial ageing datasets capture the appearance variations that occur as a result of ageing. The first obstacle for modelling facial ageing dynamics and understanding patterns of ageing is the lack of relevant data. Over the past 10-15 years, many facial ageing databases have been created. These include numerous images and videos for individuals in different age groups. These databases may consist of 2D still images, 2D videos and 3D still images. The images are employed for a variety of research applications including age verification/identification, age synthesis and estimation, and age simulation and progression. The development of 3D scanners, however, is relatively recent. Therefore, collections of 3D scans of individuals over time are both scarce, and often incomplete. Waiting for individuals to age in order to re-scan them is challenging: it takes time. In addition, analysis of human facial expressions is an important task in developing automatic Human-Computer Interaction systems; a field that has garnered increasing attention over the past two decades [17, 215] . The dynamics of facial expressions convey important information (discriminating) and to date remain a relatively unexplored area. Automatic analysis

of facial expression investigates changes in facial appearance that result from changes in expression, as well as the dynamics of the changes. Changes in appearance are visible even in static images, while dynamics require temporal information, such as the timing, the duration and the speed of the temporal phases of various facial dynamics. While the primary focus of most work in facial expressions analysis is changes in appearance, both the dynamics and the configuration of facial parts are essential for the interpretation of the observed expressions [12, 159].

1.1 Research Motivation and Contributions

This study involves research to explore how facial dynamics change as people age using longitudinal 3D facial database which was collected over nine years.

The contributions of the thesis can be summarised as follows:

1. A 3D dynamic ageing database is the important product of this study. This study explores the changes in 3D facial dynamics over time by the capture of 3D videos of volunteers perform 3D dynamic face scans: two smiles, saying a sequence of single words and saying four sentences, which were collected over nine years. It is important to generate 3D dynamic ageing datasets to consider the problems of facial ageing. To date, no publicly available 3D dynamic databases exist. A pre-processing process using a meshlab script was developed to clean the data and fill the holes that occurred. Interpolation techniques were used to fill the gaps in the data for some people who had missing years due to absence, illness, etc. The methods were validated and tested by adding the interpolated data to the experiments in the next chapters (Chapter 3).

2. A novel method is developed to extract the dynamic features from the phases (onset, apex, and offset) of a dynamic smile from the first Principal Component (PC1) by building a general model of 4D (3D dynamic) sequences instead of

considering only static appearance features as in previous studies. The findings are displayed in the form of a graph to study the relationship between facial dynamics and age using linear regression model. This work aims to be certain has created a linear model of the dataset that shows the relation between progression of the dynamic features and age (Chapter 4).

3. When the automatic segmentation of a smile was carried out, three non overlapping phases are detected, namely, the onset, the apex and the offset and a set of dynamic features were extracted. Static features were computed for comparison, to investigate the influence of temporal facial information in different age groups. The experiments were carried out on the subjects who were divided in two age groups and each group was further subdivided by gender. A novel set of 3D geometric features is also presented in this study. These features are approximate geodesic/shortest path distances between the anthropometric landmarks (Chapter 5).

4. This thesis introduces for the first time the application of the multilevel Principal Component Analysis (mPCA) method to analyse 3D dynamic smile data. This data possesses a natural multilevel clustering of the subjects by both age and gender groups (between- group variation), and every subject has their own smile expression (within-group variation); mPCA uniquely exploits and is shown to offer an improvement over more traditional PCA methods (Chapter 6).

1.2 Thesis Outline

The remainder of the thesis is structured as follows. Chapter 2 conducts an overview of what has been done previously to investigate the techniques used and a critical examination of these studies indicates how further progress in the field can be made. Chapter 3 describes the 4D scanning system and the data acquisition protocol, as well as the pipeline that will be used for processing the data. Next, Chapter 4 describes the

construction of a generic and person-specific statistical face model as well as its use in synthesising a new face model. Then, Chapter 5 presents the dynamic features that were used to evaluate smiles in the different age groups and detect gender differences. Chapter 6 describes how the dynamics of smiles change in different age groups and the multivariate principle components used to analyse dynamical smiles. Finally, the thesis is concluded in Chapter 7 which highlights our contributions to knowledge in this field as well as giving directions for the future work.

1.3 Publications

Research conducted during the period of study has produced the following papers:

1. K. Al-Meyah, D. Marshall and P. Rosin, 4D Analysis of Facial Ageing Using Dynamic Features, Procedia Computer Science, 2017, Elsevier, Vol. 112, pp. 790-799.

2. DJJ Farnell1, J Galloway, A Zhurov, S Richmond, D. Marshall, PL Rosin, K Al-Meyah, P Pirttiniemi, and Raija Lahdesmaki. What's in a Smile? Initial Results of Multilevel Principal Components Analysis of Facial Shape and Image Texture. In *Annual Conference on Medical Image Understanding and Analysis*, pages 177-188. Springer, 2018.

Chapter 2

Literature Review

A variation that affects all personal characteristics (identity, gender, expression) is the variation caused by ageing. Although ageing affects almost all types of biometric features, to date, the topic of assessing facial dynamics over time has not received much attention regarding modelling these dynamics statistically. The main reason for this is the difficulty involved in collecting suitable training data for studying age-related effects on biometric features. The challenge of biometrics is the ability to construct a dataset and use the same subject over many years. Recent years, have seen an increased amount of research activity in the area of facial ageing simulation. This interest is attributed to the potential of using facial ageing modelling techniques.

This chapter surveys the relevant studies that have been reported, including age estimation, simulating ageing, and face verification across age progression. The chapter also discusses previous studies related to facial ageing, the statistical models and the extraction methods with reference to facial features that have been widely used for face modelling. They all depend on having suitable datasets for studying ageing variations. Therefore, this chapter reviews the publicly available datasets for ageing and the 3D dynamics for various applications related to age.

This chapter is organised as follows: Section 2.1 introduces the statistical modelling analysis and the techniques used in this study. A matching time series algorithm is presented in Section 2.2. Section 2.3 reviews the current literature on human ageing, while Section 2.4 discusses the feature extraction methods. Next, longitudinal related

data is described in Section 2.5, and finally, Section 2.6 concludes the chapter.

2.1 Statistical Modelling Analysis

Statistical methods for representing the facial form vary in how closely they model the structures of the face and how they exploit the shape information [166]. The following sections describe the main techniques used for the statistical analysis of the shape and the appearance of facial data, which are usually restricted to features.

2.1.1 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a supervised method and it principle consists of projecting the data linearly in such a way that the projected data are best separated. In supervised methods the data are labelled in order to identify to which to class they belong. Let us assume that we have c classes and that the data are categorised as $X_1, ..., X_c$, where $X_i \in \Re^{m \times n_i}$ and n_i represent the number of samples that belong to the i_{th} class, which means that the data are ordered in accordance with their class membership. The latent space in supervised methods is constructed in a discriminant way so as to assign the categorisation of the data into different classes. LDA has been successfully used as a dimensionality reduction technique with many classification problems, such as speech recognition, face recognition, and multimedia information retrieval [154, 212]. LDA, which was introduced by Belhumuer et al. [22], assumes that classes/labels are being used and that the features of the different classes/labels have a Gaussian distribution with the same covariance matrix but a different mean. The approach of LDA is to project all the data points into a new space, normally of a lower dimension, which maximises the between-class separability while minimising the within-class variability [143].

2.1.2 Principal Component Analysis (PCA)

Principal Component Analysis was originally formulated by Pearson [164] as a minimisation of the sum of squared residual errors between the projected data points and the original data [58]. PCA is a classical type of feature extraction and data representation; it is a standard dimensionality reduction technique widely used in the areas of pattern recognition and computer vision. The main idea of PCA is to extract the most important features that characterise the original information and manage the data in a low dimensional feature space to reduce data redundancy. At the same time, PCA can retain the most important information of the original data, and so PCA can solve the problem of high-dimensional data. The idea of applying PCA is not novel, but the advantage of PCA is that it reduces feature dimensionality and the amount of calculation and storage required [214]. In particular,PCA is a popular technique for parameterising shape, appearance and motion [30, 52, 147, 152, 197]. These learned PCA representations have proven useful for solving problems such as face and object recognition, tracking, detection and background modeling [29, 52, 147, 152, 156]. PCA has been extensively applied in data mining, pattern recognition and information retrieval for unsupervised dimensionality reduction [213]. Unsupervised methods such as PCA determine the statistical properties of the input data to discover relevant features. Thus, PCA is widely applied in computer vision applications, particularly in face processing studies. For example, when applied to face images [197], 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 [22, 197]. The Point Distributed Model (PDM), Active Shape Model (ASM), Active Appearance Model (AAM) and 3D Morphable Model are other examples of approaches that employ PCA to model

faces [197]. There has been a strong trend lately in face processing research away from geometric models towards appearance-based models. Appearance-based models employ dimensionality reduction to represent faces more compactly in a low-dimensional subspace which is found by optimising certain criteria [183]. These models represent faces as points in high-dimensional image spaces and employ dimensionality reduction to find a more meaningful representation, thus, addressing the issue of the problem of dimensionality [183]. Dimension reduction is a process of reducing the number of variables under observation. Though face images can be considered as points in a high-dimensional space, they often lie on a manifold (i.e., subspace) of much lower dimensionality, though 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 unsupervised (i.e, Eigenfaces) by Turk et al. [197] and supervised (i.e, Fisherfaces) by Belhumeur et al. [22].PCA has the advantage of capturing holistic features but ignores the localised features. In contrast, Fisher's faces which employ the LDA technique [22], extract discriminating features between classes, and perform better for large datasets.The main difference between LDA and PCA is that LDA is a supervised method. It searches for the project axes on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. When label information is available, e.g., for a classification task, LDA can achieve a significantly better performance than PCA. Given a number of independent features relative to which the data are described, LDA creates a linear combination of those which yields the largest mean differences between the desired classes. Even though LDA is said to perform better than PCA in classification, LDA requires more computation compared to PCA. Another important point to mention is that LDA is much more sensitive to different training sets than is PCA. The performance accuracy of the implementation of LDA could also be relative to the size of the training set [216]. A similar problem was identified and researched in detail by Martínez et al. [146], who concluded that when the training set is small, PCA

can outperform LDA.ASM and the AAM were proposed by Cootes et al. [51]. An AAM contains a statistical model of the shape and grey-level appearance of the object of interest, which can be generalised to almost any valid example. A year later, the 3D Morphable Model was introduced by Blanz et al. [31]; the 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 [31].

2.1.3 Active Shape Models (ASM)

Since its introduction by Cootes et al. [54], the ASM has become one of the most popular algorithms for analysing 2D and 3D images. In this work, we used the method as a feature-based approach in 3D video data. Using a set of landmarked training faces, we built an ASM to extract facial dynamics from the variations in the 3D shape [25]. Undoubtedly the most generic and, at the same time, the simplest method used to represent shapes is a set of points distributed across the surface, which can be extracted from volume data. Coordinates for all k points are concatenated to one vector x, which describes the shape [106].

$$X = (x_1, y_1, z_1, ..., x_k, y_k, z_k)^T$$

Essentially, constructing a statistical shape model involves extracting the mean shape and a number of modes of variation, via PCA, from a collection of training samples. Obviously, the methods employed depend greatly on the chosen representation of shape. An essential requirement for building shape models with PCA is that landmarks on all the training samples should be located in corresponding positions. Chapter 4 discusses this requirement in details.

2.1.4 Active Appearance Models (AAM)

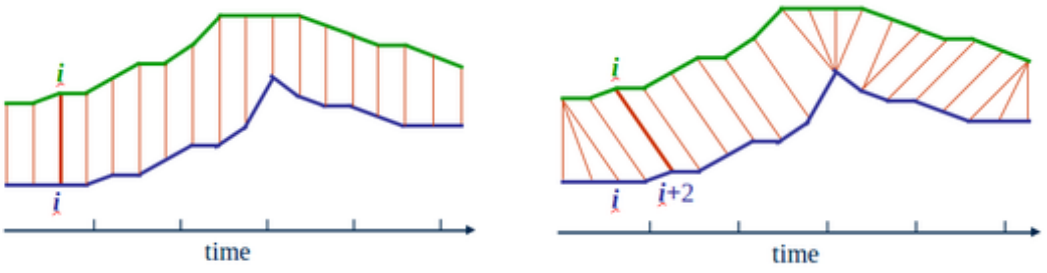
AAMs are generative models which use a compact set of parameters to describe the shape and texture variation of objects in images [160]. Parametric models of shape and texture, such as AAMs [52], are a widely used technique for object appearance modeling. The AAM has been employed in computer vision research for more than 15 years due to its ability to deal with a high level of information using a small number of parameters. These models of shape and appearance can generate instances of specific object class (e.g., faces) given a small number of model parameters that control shape and appearance variations. Such parametric models can be applied to both image synthesis and analysis problems. Representative applications are object tracking in video [60, 180], face recognition [32], face synthesis [120], and image stitching [191]. Many researchers have used the AAMs in modeling 2D and 3D faces [55]. AAMs build on the idea of the ASMs, which is a model-based statistical approach to locating shapes in an image [50]. AAMs use annotated images where key areas are labelled using landmark points. Landmark schemes vary by the number and location of points. AAMs include both shape and texture information by combing a model of shape and texture variations. The combining appearance model is computed by concatenating the resulting shape and shape normalised texture vectors, and performing PCA a final time to account for possible shape-texture correlation [52].

2.2 Matching time series

A time series is a sequential set of data measured over time. To compare one sequence with another, in some domains, a simple distance measure, such as Euclidean distance, is adequate (Figure 2.1 (a)). However, a problem that often arises is how to find the similarity between two time series with different lengths or different sample rates and that are variable along the time axis. A more robust technique to measure the distance for a time series is Dynamic Time Warping (DTW) [124]. DTW is sufficiently flexible

to match similar shapes even they vary in the length along the time axis; therefore, DTW is widely used in science, medicine, and industry [123].

Figure 2.1 (b) shows this in a simple example. In order to find the similarity between such sequences, warping the time axis of one or both sequences achieves a better alignment. DTW makes it possible to efficiently achieving this warping [124]. Further details about DTW and its derived forms are given in the next section.



(a) Euclidean distance.

(b) Time series processed by DTW.

Figure 2.1: Examples of comparing between sequences using, (a) Euclidean distance, (b) Dynamic Time Warping..

2.2.1 Dynamic Time Warping (DTW)

DTW is a familiar method for measuring the similarity between sequences that vary in time or speed. The technique was originally developed for speech recognition, but several researchers have applied it in other domains, in particular, those that are time-dependent such as data mining, gesture recognition and signature authentication; several variants have been developed, such as derivative DTW (DDTW) [124, 172, 175]. Figure 2.1 shows an example of the process of aligning two out of phase sequences using DTW. The DTW technique computes an optimal match between two sequences by allowing a nonlinear mapping of one sequence to another and stretching or shrinking the signal along its time axis. The sequences are warped nonlinearly to accommodate any nonlinear variations in the time dimension between two sequences [117, 118, 123].

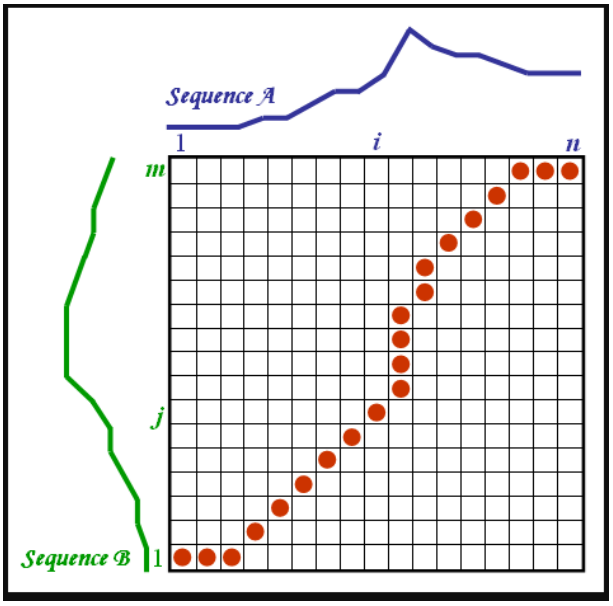


Figure 2.2: The optimal warping path through the distance matrix.

In this study, we used DTW, which has the flexibility to accept sequences with variable lengths; this is a useful property because videos are of varying lengths.

The methodology for standard DTW assumes a sequence A of length n and a sequence B of length m. A distance matrix of size m by n is created where the (i^{th}, j^{th}) element of the matrix consists of the distance between the two points a_i and b_j. Then, an optimal path through the distance matrix, from (1, 1) to (m,n) is found by moving one step in a horizontal, vertical, or diagonal direction [19, 175] (see Figure 2.2).

Classic DTW suffers from the limitation that it is defined on a sequence of discrete points and is therefore sensitive to the sampling rate. Where the sampling is sparse, there is a lack of resolution in the matching process due to the fact that the algorithm matches only discrete samples rather than continuous curves [79]. Continuous DTW (CDTW) is an extension of classic DTW that maps a sample point in one of the curves to a virtual point between two samples in the other curve [151]. The classic DTW algorithm [175] determines the warping path based on distances between data points, which can yield pathologic results in some particular cases. For instance, consider two data points P4 and Q3 in Figure 2.3 that are close, but one belongs to a rising slope and

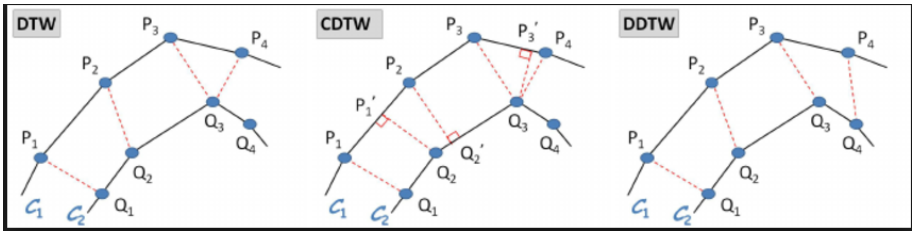


Figure 2.3: Curve matching using Dynamic Time Warping and its derived forms Continuous DTW (CDTW) and Derivative DTW (DDTW) [26].

the other to a falling slope. DTW considers a mapping between P4 and Q3 optimal, although it would be inaccurate to map a rising trend to a falling trend. By taking into account the derivatives, WDTW correctly maps P3 to Q3, and P4 to Q4.

Derivative DTW (DDTW) was developed to prevent this problem as a modification of DTW and CDTW. It does not consider the y-values of the data points, but rather the higher-level feature of shape. Information about shape is obtained by also considering the first derivative of the sequences, hence DDTW [124]. A further extension to DDTW called weighted DDTW (WDDTW) was proposed recently to include more than the first derivative in the algorithm [26] (See Figure 2.3). For example, the first derivative may give information on speed, the second on accelerations and decelerations including further derivatives; this theoretically provides a more accurate match. However since derivatives are noise sensitive, weighting factors also need to be incorporated so as their inclusion does not impair the performance of the algorithm.

2.3 Research activity related to human ageing

In recent years, there has been a growing interest in addressing the many problems related to facial ageing. This section discuss some of the interesting studies in computer vision on these topics of age estimation, appearance modeling, face recognition and verification as people get older.

2.3.1 Age estimation

Automatically estimating facial ageing is an important technique with a wide range of applications, it is still a challenging problem to estimate a person's age from a facial images [82]. Age estimation techniques often targeted the shape and texture-based features that were extracted from faces. While feature-based approaches to estimating age used anthropometric distances extracted from different facial regions, holistic approaches typically adopted subspace methods to reduce the dimensionality of faces and subsequently used regression techniques to estimate age from facial images. The mainly methods employed can be described as follows:

2.3.1.1 Anthropometric models

Face anthropometry is the scientific study of the measurements and proportions of the human face [62]. Such studies provide a quantitative description of the craniofacial complex using measurements taken between key landmarks on human faces at different ages [83, 84] and have played a critical role in surgical procedures employed on the faces of growing children [179]. In [83], Farkas provides a comprehensive overview of face anthropometry and its many significant applications. Kwon and Lobo in [131, 132] proposed an age classification approach that identifies the age group that a face image belongs to, based on the anthropometry of the face and the density of its wrinkles. In one study, they consider three age groups, namely, infants, young adults and senior adults, adopting craniofacial development theory [10]. Anthropometric method, which also considers the facial geometry, has limitations; including the fact that it is suitable for age estimation or modelling for young people only, because the shape of adults' faces does not change much.

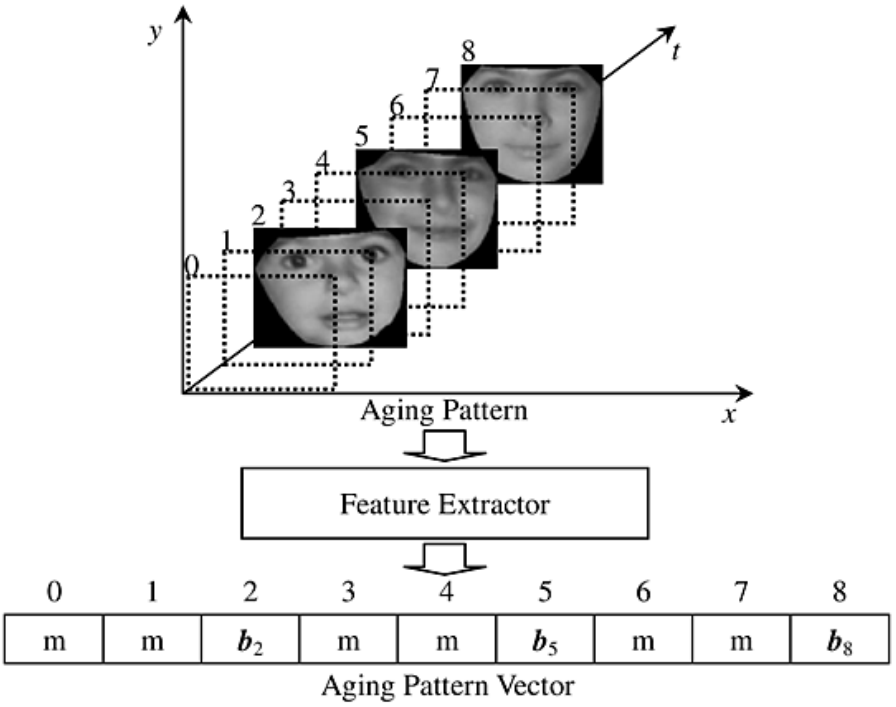


Figure 2.4: Vectorisation of the ageing pattern.The missing parts in the ageing pattern vector are marked by 'm'. Originally shown in [93].

2.3.1.2 Ageing pattern subspace

To model the ageing process, a sequence of face images in the course of an individual's ageing can be used together instead of dealing with each image separately, This is called the AGing pattErn Subspace (AGES) [93, 94]. An ageing pattern can be defined as a sequence of personal face images, coming from the same person and arranged in temporal order. If an individual's face images at all ages are available, the corresponding ageing pattern is called a complete ageing pattern, otherwise, it is called an incomplete ageing pattern. The AGES method can synthesise the missing ages by using an EM-like iterative learning algorithm. The age of a test face is determined by the projection in the subspace that can best reconstruct the face image [174].

2.3.1.3 Age regression

Facial features are often extracted using the Active Appearance Model (AAM) using regression. Adopting the AAM approach by [52], Lanitis et al. [136, 138] presented a combined shape and appearance model to represent the face images. Face images are then represented by a set of fitted model parameters which are the principal components of the eigenspaces that correspond to facial shape and appearance. Lanitis et al. [138] investigated three formulations for the ageing function; linear, quadratic and cubic, in turn, Deffenbacher et al. [61] debate whether the ageing factor has essentially sequential patterns which are quite similar to age morphing [168]. When a large number of ageing data are provided, ageing can present a significant trend of underlying sequential patterns.

Fu and Huang [88], built a low-dimensional manifold from a set of age-separated face images using such manifold learning approaches as Locality Preserving Projections (LPP), Orthogonal Locality Preserving Projections (OLPP) and Conformal Embedding Analysis (CEA), in estimating the age of a face, used linear and quadratic regression functions on the low dimensional feature vectors from the respective manifolds. On very similar lines, Guo et al. [101] adopted a manifold learning approach and used support vector regression to estimate the age of faces having low-dimensional representation. They reported that local adjustments made on the regression result lead to improved accuracy in age estimation.

Yang and Ai [209], proposed a learning-based approach to classifying face images on the basis of their age group. The Local Binary Pattern (LBP) is used as an image operator and the LBP histogram is extracted and subsequently used for texture characterisation. Adopting the Ada Boost technique [27], they identified a sequence of local features which when combined into a strong classifier performs the task of age classification successfully. Ahonen et al. [1], proposed for age estimation, multiple linear regression on the discriminate ageing manifold based on a quadratic function of face images for age estimation as in Figure 2.5.

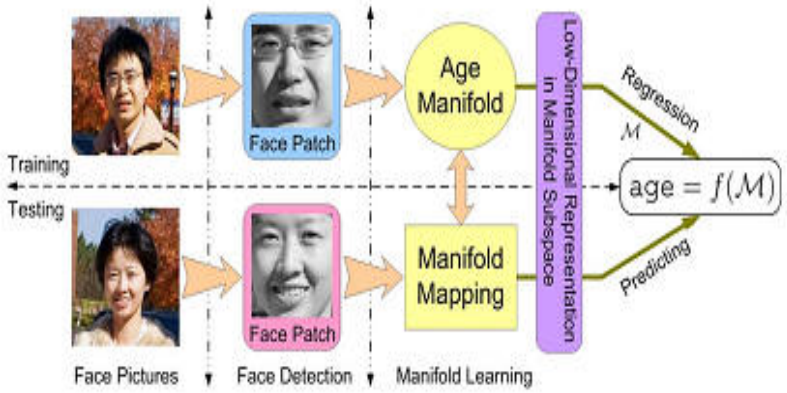


Figure 2.5: Framework of age estimation by face image analysis, originally shown in [89].

2.3.2 Modelling / Simulating the ageing process

For age synthesis (also called age progression) is often implemented by building a generic face model. Age progression is used to modify and improve the image in law

Figure 2.6: Age synthesis, originally shown in [190]

enforcement, by computer or manually, to identify the purpose of suspects/victims and missing people. This technique has been developed for applications in police investigative work. When images of missing family members (especially children) or wanted fugitives are no longer up to date, professionals can predict the natural ageing of the subjects' faces and produce updated face images, employing all the available individual information, such as facial features, lifestyle, occupation and genetics. Age synthesis by computer can significantly enhance the efficiency of a professional's work and at the same time provides more photo-realistic ageing effects that can satisfy aesthetic

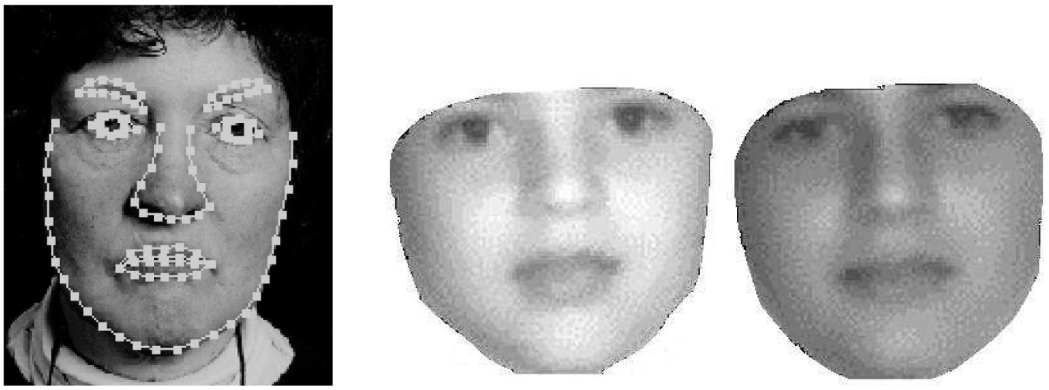
needs. From the perspective of computer vision, the ageing problem can be summed up as follows:

1. Shape vs. Texture: The effects of facial ageing appear in the shape of the face in the formative stages, while there are also variations in texture such as wrinkles. Other skin affects are clearly shown through the later stages of adulthood. The facial ageing problem of characterising facial shape and texture can be described as a functions of time. Due to the temporal nature of the induced variation, therefore, it is necessary to build models that have the same characteristics.

2. Feature Selection: To develop a model of facial growth or of facial ageing requires a convenient form of data to be identified that provides all the desired descriptions of the case. The data may be individual specific or population-specific. There are some forms of data that can help characterise the growth of a face, such as fiducial features (2D or 3D) extracted from age-separated faces, 2D facial imagery or 3D facial scans extracted from individuals across different ages and face anthropometric measurements extracted from a population.

3. Other factors: Regardless of the biological factors influencing the growth of facial bones, loss of elasticity in facial muscles and facial fat, many factors can be shown to influence the effects of facial ageing, such as ethnicity, gender, dietary habits and climatic conditions. Moreover, facial appearance changes with age, due to factors such as changes in hair-style. Therefore, all these factors should be taken into account when building a facial ageing model [83].

Parke et al. [162] built a 3D mesh model to generate cartoon faces. Animated facial expression is synthesised by analysing a typical pair of real face photographs. In addition, a quantity of face models, both 3D and 2D, photo-realistic and non-photo-realistic have been developed and reported for different applications. Burt and Perrett [37], developed an ageing simulation approach taking both shape and colour information into consideration. The age-related visual differences between the composite faces from



(a) (b)

Figure 2.7: Face ageing using active appearance model and principal component analysis. (a) Active appearance model [51, 52]. (b) Ageing appearance simulation result [138].

different age groups were added to the new test face images to produce the effects of ageing.

Tiddeman et al. [194] found limitations in the above method and proposed a new algorithm based on wavelets that captures the age information lost in the blending process. The age transformations are more accurate and reach the target age bracket. Further, Tiddeman et al. [195] achieved more realistic textures using a wavelet Markov random field method.

Lanitis et al. [138], presented certain ageing functions that define the relationship between an individual's age and the model parameters of the corresponding face images that are obtained on the basis of AAM. In particular, to account for the effects of individuals' lifestyles on their ageing process, these factors were considered in their algorithm and found to simulate more realistically the effects of ageing. Figure 2.7 shows an example of AAM and result of simulating the appearance of ageing by method of Lanitis et al.

Fu et al. [90] present a framework called Merging Face (M-Face) for appearance-based

photo-realistic facial modeling. An ageing ratio image is used to render texture of ageing and rejuvenating. The caricatured shape is changed by exaggerating individual distinctiveness.

Ramanathan and Chellapa [168, 171] propose a craniofacial growth model that characterised the shape variations in human faces at various ages. The authors observe that the growth parameter k for different facial features across ages can be adapted in the model to characterise facial growth.

Ramanathan and Chellapa [167] also propose a two step approach for modeling ageing in adults, which consists of a shape and texture variation model. The formulation of shape variations is performed by constructing physical models which characterises the functionalism of the facial muscles.

Suo et al. [190] propose a statistical model to present a dynamic ageing model of the face with multi-resolution and multilayer image representations. A hierarchical (AND, OR) graph representation is adopted with which faces are decomposed into different parts and organised in graph structures. A dynamic Markov process model is built on the graph structures of the training data. In particular, hair styles are processed as one kind of ageing effects. To simulate new ageing faces, the graph structures over different age groups are finally sampled regarding the dynamic model. Figure 2.8 shows this model and some ageing simulation results which exhibit highly photo-realistic results.

Ling et al. [142] propose a method to simulate the adult ageing effects on face images by means of super- resolution in tensor space and active appearance models. A super-resolution technique is used to construct high-resolution images based on the images in low resolution. Multi-linear analysis is introduced to deal with the problem that occurs when more than one kind of change needs to be considered and processed in the training and test face image set. An active appearance model normalises the faces in the images.

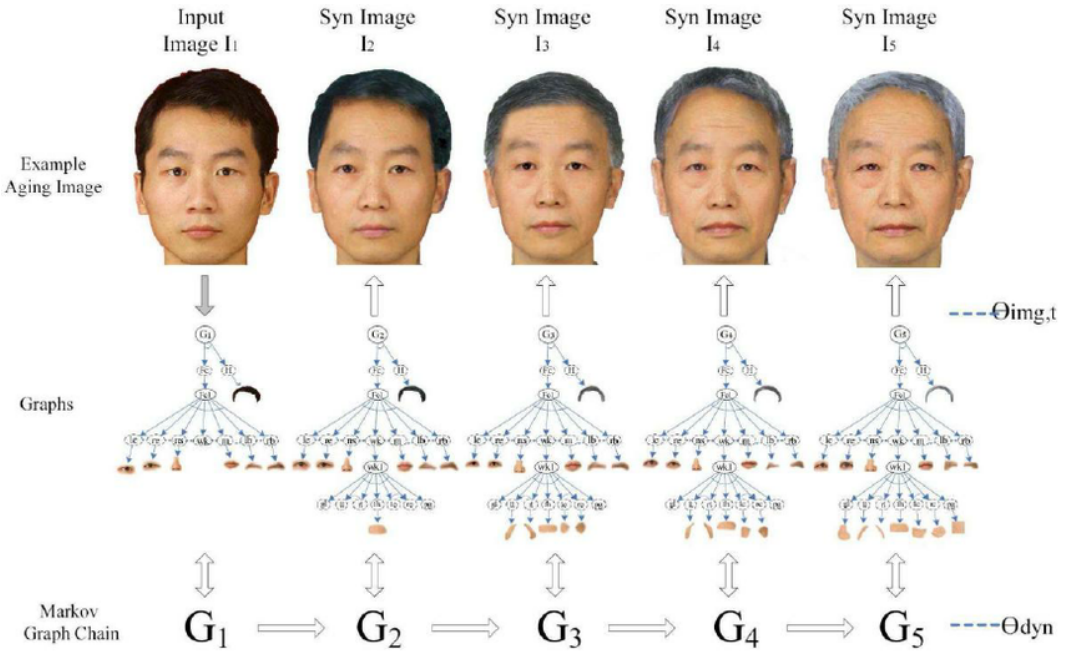


Figure 2.8: Multi-layer dynamic face ageing model and results[190]

2.3.3 Face verification across age progression

Face verification has a very wide range of applications in computer vision such as surveillance, human computer interaction and image retrieval. Human faces undergo significant variations with ageing. However, many researchers have detected the range of factors such as illumination variations, pose variations, facial expression and occlusions that affect face recognition. Still it is useful to study the role of natural factors such as ageing effects in face recognition.

In the past few years, Biswas et al. [28] have studied feature drifting in face images of people at different ages and applied it to the task of facial verification. Other studies and age transformation were used for verification [93, 185, 204, 161, 163]. Some previous research has been applied to age progression for the task of face verification. In a comparison of two photos, these methods transform one photo so that it seems to portray the same age as the other, or transform both to reduce effects of ageing. One of the earliest studies is that of Lanitis et al. [138], where a statistical model is used to

capture the variation of facial shapes over age progression. The model is then used in age estimation and face verification.

Ramanathan and Chellappa [168] use a face growing model for face verification tasks involving people under the age of eighteen. To perform face verification across time, two methods are proposed, the generative and non-generative. Since ageing needs to be modelled, the above methods can be classified as generative. Most of them use verification to evaluate the age modelling algorithm. Generative methods, however, have some limitations: for example, the construction of the ageing model usually requires the images being compared to carry additional information in the form of the actual age of the subject and the location of landmark points on each face image. The 68 landmark points that are used in this method are pre-labelled in each photo in the FGnet dataset [81].

To overcome these problems, non-generative approaches have been proposed for the ageing problem. The approaches proposed by Ling et al. [142] and Biswas et al. [28] are non-generative. Then derive an age-invariant signature from faces and use this to perform face verification in the course of age progression. Ling et al. [142] studied the problem of face recognition with age variation. They proposed a robust face descriptor, the gradient orientation pyramid, for face verification tasks at different ages. In comparison with previous descriptors such as image intensity, this descriptor is more robust and performs well on face images with wide age differences. The pyramid technique enables the descriptor to capture hierarchical facial information. The new method has demonstrated very promising results on two challenging passport databases.

Biswas et al. [28] analyse the coherence of the drifts in various facial features to verify whether or not two age-separated images belong to the same individual. The main point of this approach is to look for forward coherencies in facial feature drifts as people get older. These writers investigate whether the coherency of some selected facial features is greater on two different images of the same person at different ages. In this work, only the most frontal fiducial features were selected, because the features on the outer

boundaries of the face tend to change rapidly as the head pose changes and various expressions (laughing, even smiling) are performed.

Li et al. [141] develop a discriminate model that uses scale-invariant feature transform (SIFT) as a descriptor and a multi-feature discriminant analysis (MFDA) algorithm for dimensional reduction. The model can be recognised as a familiar face despite variations in illumination, pose and expression. Li et al.'s experiments were conducted on two public domain databases (MORPH and FGNET) to show the effectiveness of the proposed method.

2.4 Feature Extraction Methods

Extracting the features is one of the key functions which is reacquired according to the approach used. It involves reducing the number of items of information required to describe a large set of data while still describing the data with sufficient accuracy [134, 207, 208, 217]. Some of these approaches have been successfully employed in many static face recognition problems. The next sections presents the commonly methods in this area, whose the features can be divided into the following classes:

2.4.1 Static Feature Extraction

The recent methods of feature extraction can be split into three categories: local, global and hybrid approaches. The following sections describe the major facial feature extraction techniques.

2.4.1.1 Local Features

Guo et al. [102] proposed biologically-inspired features (BIF) of ageing for age estimation. These features are based on Gabor filter responses for different orientations

and scales. Alnajar et al. [11] introduced intensity and gradient-based features to adopt a learning-based encoding method for age estimation under unconstrained imaging conditions. For each pixel, neighbouring pixels are sampled in a ring-based pattern to form a low-level feature vector. Then, the features are encoded using a PCA-tree based codebook [211] to model the completed local binary patterns (CLBP) using an SVM regression. Choi et al. [43, 44] propose the extraction of wrinkles using a set of region specific Gabor filters in conjunction with the use of a local binary pattern (LBP) skin extractor. Bekhouche et al. [21] present a novel method to estimate the age from face images, using binarised statistical image features (BSIF) and local binary patterns (LBP) histograms as features, as performed by support vector regression and kernal ridge regression (KRR). This study evaluates a PAL database and finds better performance when combining BSIF and LBP features, but, this work does not consider gender, ethnic and pose variations. A new method to estimate possible ages for facial image is proposed in [133] with a novel measurement, where by three stages are described. The LBP model extracts local skin features. Jana et al. [119] describe a new methodology for estimating someone's real age by analysing the wrinkled areas of the facial images. Most significant features such as the distances between facial objects and the analysis of wrinkled areas considered. This study imposes some constraints, including the need to make all images a frontal views with uniform light on each part of the face and to contain only one face, with a forehead clear of hair.

2.4.1.2 Global Features

The global features that represent overall facial characteristics have largely been used to estimate detailed age. Methods used to determine global features include Active Appearance Models (AAM), Gabor wavelet transform (GWT), and subspace features using image intensity [211] and image frequency. Many previous ways of determining facial features based on the AAM model have been proposed. The AAM is a well-known method that represents faces in statistical terms appearance and shape models

using PCA. The first age estimation algorithm using AAM features and regression methods is proposed by Lanitis et al. [138]. In their work, the relationship between age and features is defined as quadratic ageing function and the facial age is then estimated using ageing features. At this point, Lanitis et al, used the AAM features to compare different classifiers for age estimation. Geng et al. [92] proposed the ageing pattern subspace (AGES) method, using the subspace representation of a sequence of individual ageing face images. In their work, an AAM is also used as a feature extractor. Yan et al. [41] designed a regression based on training samples with uncertain non-negative labels using AAM features. Though the AAM is frequently used in the above works, its features may not include information about wrinkles and skin, because of the dimensional reduction made by the PCA.

2.4.1.3 Hybrid approaches

Hybrid features are a combination of global and local features, proposed by Suo et al. [189], to solve the above problem with the features of AAM; they also design a minimal number of features consisting of the AAM, wrinkles, skin, hair and the configuration of the facial component features using the hierarchical face model. Topology, geometry, photometry and configuration features are four types that are extracted in each component. For example, the geometric features of wrinkles are described by a set of landmarks. However, extracting the landmarks of the wrinkles is more unstable than analysing their texture. The landmarks of the wrinkles represent the positions on the wrinkle edges, but, because facial wrinkles are include both fine and deep wrinkles, they are not always clearly portrayed in facial images. Extracting the landmarks of wrinkles is seen as a very challenging problem and a texture analysis of the wrinkle area is a more stable method of analysing various wrinkles. Hair features are extracted using a graph cut and generative hair model [190]. However, the problems associated with using this approach to an actual age estimation system remain unsolved, due to the inherent difficulties associated with hair modelling. Even when

complicated method is used, it does not improve the estimation of age as much as other research methods using only the AAM features [44]. Choi et al. [44] propose a new extraction method for wrinkle and skin features in order to improve performance. The wrinkle features are effectively extracted by a Gabor Filter set based on the direction of wrinkles on the face. Ali et al. [8] propose a combined PC and LBPV features to accommodate a range of age groups. Dibeklioglu et al. [64] propose a new method for verifying that combine spatio-temporal facial features and facial expression dynamics, which they call Completed Local Binary (CLBP-TOP) and combine with dynamics features. Dibeklioglu and his team have completed many studies in term of combining global and local features using 2D videos of facial expressions. We describe their work in Section 2.7 below.

2.4.2 Dynamic Feature Extraction

Most of previous studies regarded static face images as the most reliable evidence for conveying a significant amount of information such as the identity, emotional state, ethnic origin, gender, and age of a person shown in a face image [136]. Researchers have of late, however, attempted to quantify facial ageing effects through various other strategies. Ageing has been considered as the fourth dimension [177, 178]. The human face undergoes many skeletal and soft tissue cellular changes, due to age progression which consequently affects its functional behaviour [98, 140]. Studies have shown that lips become less elastic and less mobile with ageing [74, 127]. Oral structures such as teeth and periodontium also change with age. These changes may affect the smile. Moreover, facial movements differ between the genders (i.e. they have different smile behaviour), especially in adulthood [46, 158]. One of the most frequently studied facial expressions for the analysis of facial dynamics is the smile, because it is the most frequently used facial expression and the easiest facial expression to exhibit [80]. Thus, the focus of this study will be on the smile as a form of facial dynamic.

However, the earlier studies of smile aesthetics, these differences were not statistically

tested [63, 199]. Geld [199] and Desai et al. [63] by means of videos, have recently studied age related changes in smiles. The lack of information concerning gender and age differences has prevented gender differences in age-related changes in smiles from being explicitly identified. Chetan et al. [42] propose a cross-sectional study to determine the tendencies and patterns between the different age groups and check whether gender plays an important role in these tendencies. The findings from this study are as follows:

- There is a difference between male and female smiles.
- Smiles change with age. As individuals age, they are lose of muscle tone which leads to a reduction in the extent of their smiles.
- In smiling, females exhibit greater horizontal movement and males exhibit greater vertical movement.

Facial dynamics yields further information, allowing for more detailed analysis. This allows for greater discrimination across wider age differences. In the last decade, to discriminate between smiles their dynamic features (such as their duration, speed and amplitude), as opposed to morphometric cues, have received attention. Cohn et al. [48] analyse the correlations between lip-corner displacements, head rotations and eye motion during spontaneous smiles. In another study, Cohn and Schmidt [49] report that spontaneous smiles have a less onset amplitude of lip corner movement, but a more stable relation between amplitude and duration. Furthermore, the maximum speed of the smile onset is higher in the posed samples, and the posed eyebrow raises have higher maximum speed and larger amplitude, but they end sooner than spontaneous ones. The linear discriminant classifier were proposed to distinguish between spontaneous and deliberate enjoyment smiles using duration, amplitude and duration amplitude measures of smile onsets. Cohn and Schmidt analyse the significance of the proposed features and show that the amplitude of the lip corner movement is a strong linear function of duration in spontaneous smiles, but not in deliberate ones. Krumhuber et

al. [129, 130] study the effects of the dynamic attributes of smiles in human and synthetic faces and ask whether facial dynamics tell us something about the genuineness of an emotion. The findings demonstrate that an authentic smile was perceived as more likeable, attractive and trustworthy than a fake smile or a neutral expression.

Recently, Dibeklioglu et al. [66] have proposed a method of using facial expression dynamics to form a database for exploring the effect of dynamic features for age estimation. In another study the dynamic features of a person's smile are used for age estimation; using a person's smile and it was shown that the facial expression dynamics with appearance information are much more reliable for group classification tasks. Furthermore, experiments were carried out on disgust expression to evaluate the effectiveness of the proposed method in this study on a different expression and the results significantly improve the age estimation accuracy [65]. Also, Dibeklioglu et al. [67, 71] proposed using the dynamics of eyelid, cheek, and lip corner movements to distinguish between spontaneous and posed smiles, where distance-based and angular features are defined in terms of changes in eye aperture. One main aim of this study is to explore the effects of dynamic features in different age groups and gender differences. Despite the fact that the age of a person plays an important role during interaction, so far no researcher has been involved extracting dynamic features system based on 3D facial dynamic from feature vectors in low dimensional. For these reasons we need to select a feature extraction strategy that applies well to all these dynamics. ASM and AAM appear a suitable choice.

2.5 Longitudinal Related Databases

A large number of face databases have been collected over the years for the purposes related to face recognition [170]. However, it is difficult to compile large ageing databases that involve series of age images of individual. Very few publicly available datasets specifically address facial ageing in comparison to the datasets that address other

problems in face recognition. Here, we summarise well known public databases concerning to facial ageing.

2.5.1 2D Static Databases

The publicly available MORPH face database was collected by the face ageing group in the University of North Carolina at Wilmington, for the purpose of face biometrics applications. This longitudinal database records individuals meta-data, such as age, gender, ethnicity, height, weight, and ancestry, which is organised into two albums. Album 1 contains 1,724 face images of 515 subjects taken between 1962 and 1998. The ages range from the average of 27.3 to maximum 68 years. There are 294 images of females and 1,430 images of males. The age span is from 46 days to 29 years. Figure 2.9 show image examples for MORPH album 1. Album 2 contains about 55,000 face images which about 77% images are Black faces, 19% are White, and the remaining 4% includes Hispanic, Asian, Indian, and Other. Figure 2.10 show example images for MORPH album 2.

The FG-NET (face and gesture recognition research network) ageing database [42] comprises 1002 images of 82 subjects (6–18 images per subject) in the age range 0 – 69 years. The database also provides 68 landmark features that were identified manually, on all the face images. In addition, the following meta-information is available for all the images in the dataset namely: image size, age, gender, spectacles, hat, mustache, beard, horizontal pose and vertical pose. Since the images were retrieved from real-life albums of different subjects, aspects such as illumination, head pose, facial expressions, etc. are uncontrolled in this dataset. Nevertheless, this database is the only publicly available resource that provides multiple age-separated face images of individuals in the age range 0–18 years. YGA contains 8000 colour images of 1600 subjects, out of which 800 are females and 800 are males. Each subject is having on an average 5 frontal face images. Age range of this database is 0 to 93 [87].

Figure 2.9: Some sample images from MORPH album 1 [173]

The WIT-DB (Waseda human-computer Interaction Technology DataBase) comprises of images collected from more than 5000 different subjects of Japanese origin. It includes around 2500 females and 3000 male subjects. It has 1-14 images per subject. Age range of subjects is from 3 to 85 years. All images are further divided according to their actual age into 11 age groups. The faces in the database are frontal views and are captured in wide variety of lighting conditions [198].

The VLF (Vietnamese Longitudinal Face) database has 227 Vietnamese colour im-

Figure 2.10: Some sample images from MORPH album 2 [173]

Figure 2.11: Some sample images from FG-NET [174]

ages of 28 subjects, comprising of 12 males and 16 females. Age range of subjects is between 1 to 80 years of age. Each subject has around 8 frontal pose images provided along with ground truth age and annotated with 68 landmarks. This database was collected for analysing growth and development stage and facial heredity of familial

members [76].

The IFDB (Iranian Face Database) contains over 3600 colour images that were captured from 621 Caucasian faces. Subjects have age range from 1 to 85years. The database includes 4 different poses, 2 different expressions and one image with spectacles of every subject. Information regarding plastic surgeries, age, skin type, occupation, fingerprints is also collected from subjects [62].

2.5.2 3D Static Databases

Many 3D face databases have been produced in order to be used for face modelling and recognition in the past two decades. Chang et al. [38], presented the first 3D static facial expression dataset involves the six basic facial expressions. Using off-the-shelf NTSC video equipment and pair of cameras/projectors including in a custom-built system, also an active stereo using structured light projections the database is not publicly available and more details about 3D static databases have been discussed by Sandbach et al. [176]. The BU-3DFE database is available to the research community (e.g., areas of interest come from as diverse as affective computing, computer vision, human computer interaction, security, biomedicine, law-enforcement, and psychology). presently contains 100 subjects (56% female, 44% male), ranging from 18 to 70 years of age, with a variety of ethnic/racial ancestries, including White, Black, East-Asian, Middle-east Asian, Indian, and Hispanic Latino. Each subject performed seven expressions in front of the 3D face scanner. Except for the neutral expression, each of the six prototypic expressions (happiness, disgust, fear, angry, surprise and sadness) includes four levels of intensity [14, 210]. The 3D Morphable Database [87], this database contains the 3D scans of 200 adult (half males and half females) faces and 238 teenage faces (8-16 years), with 125 males and 113 females and a resolution of 76,800 vertices per scan (320×240).

2.5.3 2D Dynamic Databases

UvA-NEMO Smile Database, this database has been gathered in the Science Centre NEMO (Amsterdam). It was archived to analyse relationship between change in dynamics of smiles and progression of age. This database contains coloured videos which were recorded under a resolution of 1920×1080 pixels. Videos were recorded under controlled environment conditions. Around 400 subjects were recorded and from them 1240 smile videos were collected. Age (ranging from 8 to 76 years) and gender distribution is given for each individual. Both posed and spontaneous smile expression is recorded in each video. Each video segment starts and ends with a neutral expression [66]

2.5.4 3D Dynamic Databases

The first database including of 4D faces is BU-4DFE in [210], which is presented by the same continued effort of institution that produced the static database BU-3DFE using the DI3D capturing system of dynamic face, and including 101 subjects and temporal segments (onset, apex and offset) of sequences of the six prototypical facial expressions. The resolution of temporal and spatial are 25 frames/s. In order to evaluate the individualism of facial motion for person verification a database was created in [25]. Using the 3DMD Face Dynamic System, the dataset was collected, and also involving of 94 participants uttering a word. Smiles were recorded from about 50 subjects. Another database which is called the D3DFACS created in [57]. It involves 10 subjects, consisting 4 FACS experts. The 3dMD Face Dynamic System used to capture the database. As well as, it is the first database to contain coded of dynamic 3D AUs and will allow research into dynamic 3D AU recognition and analysis.

2.5.5 Overview of Cardiff Longitudinal 3D Face Database

The longitudinal 3D dynamic face data is a novel longitudinal database that offers 3D dynamic data of expression and spoken sentences of different individuals over a period of nine years. Over the years, numerous face databases have been collected to help study the different problems related to face recognition. Given the difficulty in compiling face datasets that comprise of age-separated face images, there are very few publicly available datasets that specifically address facial ageing in comparison to the datasets that address other problems in face recognition (discussed in details above). Three publicly available databases comprise of age separated face image samples namely, the MORPH Database, the FG-NET Ageing Database and the FERET Database [6]. Here, we provide a list of existing databases in relevance to facial ageing as shown in Table 2.1. Table 2.1 shows in the last column on the right side of the table the longitudinal 3D data used in this study. However, there are three dynamic databases (UvA-NEMO, BU-4DFE, D3DFACS) but, UvA-NEMO is a 2D video, and others a dynamic 3D facial expression data set based on the Facial Action Coding System. The rest group of databases consist 2D static face images. Overall, none of these existing databases contain face images in all possible 4D (3D plus time) of the human face; all face images in front view, all individuals have static and dynamic of 3D face scans with video rate 60f/s and special conditions (dark room and no glasses). We used the Cardiff Longitudinal 3D Face database in this study which is discussed in further details in Chapter 3.

Table 2.1: Overview of the existing facial ageing databases

DataBase	Format	Exp.	Longi-tudinal	No.Subjects	age (year)	Gender	Resolution (pixel)	Video rate	Controlled lighting	Profile view
MORPH-Album1	2D static	✓	✓	515	$1-14$	✓	64×64	×	×	✓
MORPH-Album2	2D static	✓	✓	4000	$0-29$	✓	64×64	×	×	✓
FG-NET	2D static	✓	✓	82	$0-69$	✓	Vary	×	×	✓
YGA	2D static	✓	×	1600	$0-93$	✓	Vary	×	×	✓
WIT-DB	2D static	✓	×	5000	$3-85$	✓	Vary	×	×	✓
VLF	2D static	✓	×	28	$1-80$	✓	Vary	×	×	✓
IFDB	2D static	✓	×	621	$1-85$	✓	Vary	×	×	✓
FERET	2D static	✓	×	438	18 months	✓	320×240	×	×	✓
UVA-NEMO	2D dynamic	✓	×	400	$8-76$	✓	1920×1080	50f/s	✓	×
BU-4DFE	3D dynamic	✓	×	101	×	✓	1040×1329	25f/s	✓	×
D3DFACS	3D dynamic	✓	×	10	$23-41$	✓	1024×1280	60f/s	✓	×
Our Data	**4D dynamic**	✓	✓	**316**	**4-61**	✓	**1024×1024**	**60f/s**	✓	×

2.6 Conclusion

Facial Ageing is a challenging area of research as different persons age in different ways under the influence of many biological and environmental factors. Facial ageing is a complex process that causes significant variation in both shape and appearance. From previous research [1, 52, 55, 62, 82, 83, 87, 89, 90, 92, 93, 94, 101, 135, 136, 137, 138, 162, 167, 168, 169, 171, 174, 190, 209], it can be seen that a number of studies have been conducted in the tasks of age estimation, simulation and face verification across age progression. All of this research was carried out for different purposes using various techniques and using different datasets. The common aim for these studies can be described as an attempt to determine the role of natural factors such as ageing effects in face recognition and to increase accuracy as the ageing process itself differs from one person to another for a myriad of reasons, which makes the accuracy of the estimation and prediction rather difficult. It can also be noted that these previous studies have not concentrated on the modeling of the facial dynamics over time.Recent studies [64, 65, 66, 67, 68, 69, 70, 71] have shown that the smile changes with age and there is a difference between male and female smiles. These studies used facial dynamics for more detailed analysis and for greater discrimination across large age differences. The dynamic features of a smile, such as duration, speed, and amplitude, have been compared to static features to determine the trends and patterns between the different age groups and whether gender plays an important role. Some work in studying facial expressions have extracted dynamic features in different face regions, for example the work in [64, 65, 66, 67, 68, 69, 70, 71]. However, these approaches are currently limited to 2D due to the unavailability of 3D dynamic face data. In this thesis, for the first time, we use 3D video of faces and demonstrate how this can be used for modelling facial ageing. This allows for greater realism when compared to previous 2D video approaches.

For the facial ageing, the size and quality of an ageing database is a major challenge

as mentioned in previous sections. Table 2.1 summarises the various facial ageing databases that were used by researchers, including the details on each database. The vast majority of the previous and current research, in both facial ageing estimation and facial ageing simulation, adopts two-dimensional approaches, mainly owing to the limitation of available 3D data resources. 3D facial ageing has the potential to achieve better accuracy than its 2D equivalent by explicitly measuring the geometry of rigid features on the face.Currently almost all research efforts in facial ageing aspects deal with static images. However, temporal information that includes both face movements and expressions can also provide important age-related clues. In order to support the studies stated above, there is a clear need for the development of new ageing datasets that contain 3D face images and video sequences. In summary, this chapter has reviewed previous work in the area of facial age estimation, ageing simulation and face verification across age has been reviewed. The statistical models were briefly discussed to prepare the reader for Chapter 4 in which statistical models were applied for modelling facial dynamics as people age. In the last decade, the study of dynamic features of smiles has received some attention to discriminate between individuals or smile types. These features have been adopted in this study. Also, the most common of the databases related to age have been reviewed. Note that the FG-NET Ageing database and MORPH database are both publicly available. The other databases are potentially available by contacting the owners. The MORPH collected databases (MORPH1 and MORPH2) are large-size ageing face databases which are promising for data-driven statistical algorithms of age estimation, such as AAM and age manifold with the regression model. The FG-NET is a baseline database for comparisons with many existing age synthesis and estimation techniques, such as AGES, the 3D morphable database specifically collected for 3D age synthesis. The other dynamic databases mentioned here were widely used in facial analyses and recognition, but might be useful for some specific applications. We will use the Cardiff Longitudinal 3D Face database in this study which is discussed in further details in Chapter 3.

Chapter 3

4D Longitudinal Data and Pre-Processing

3.1 Introduction

As stated in the literature review (Chapter 2), a longitudinal database can be used in a variety of applications, such as in biometrics, medicine and psychology. Such a resource also provides the research community with natural data that could be used in a variety of research areas. For example, in the field of analysing facial movement in 3D over time, there is a lack of available data. To date, there is no 3D facial dynamic database available for facial ageing and if there is any plan to create a new database, it will take decades to develop one. Thus, we provide the first database of its kind in the world. This database was captured because existing longitudinal databases did not meet our requirements for the 4D capture of the faces of individuals in terms of ageing. In this chapter, the hardware and the software of the two generations of 4D captured systems are described in Sections 3.2 and 3.3 respectively. Sections 3.4 and 3.5 produce the longitudinal database and the process of the 3D frame. This data was validated and used to build statistical models for creating and analysing a general model of the facial dynamics as people age. This model's creation is described in detail in Chapter 4. To remove unwanted mesh issues and create a single-view, that is, a Unified Texture Map (UTM), this chapter describes the preprocessing pipeline in the Section 3.6. This pipeline was developed by Vandeventer [201] in a previous PhD

at Cardiff university and used in this study. This chapter also discusses the problems that occurred during the capturing process. The modifications methods that are used to tackle these problems, such as smoothing the meshes, colour matching, and filling the gaps (interpolation), are discussed further in the following sections.

3.2 3dMD Dynamic Face Capture System

Two generations of a commercially available 3dMD 4D (3D temporal) system were used for our data captures; both are different generations of the 3dMDface Dynamic System (3dMD, Atlanta, GA). The first 3dMDface Dynamic System (old system) is based on active stereophotogrammetry using a random infrared speckle projection. This generation of system employs six camera, as shown in Figure 3.1, that is, a two pods system comprising six 1.3 mega pixel cameras: two colour (Figure 3.2) and four grey scale (Figure 3.3). It can capture static images as well as sequential imaging up to 60 fps. The output files (OBJ) are 3D geometry definition files that are generated as a one point cloud produced from the two stereo camera viewpoints. The point cloud can theoretically eliminate the data errors associated with merging/stitching the left and right sides from separate cameras.

Distortion of the infrared speckle as it projects onto a subject's face allows the generation of 3D coordinates of the face through complex algorithms; several thousand points make up the vertices for a 3D image of the face. White light is used to capture colour texture images of the face simultaneously with the infrared pattern produced by the 3D data. However, small infrared components of the white light may have the potential to drown some of the infrared speckle pattern and therefore compromise image capture [57, 166]. We used the old generation system to capture 3D faces over the first four years, and the last generation of the 3dMD face capture system (new system) was used to capture the remaining data for our research. The system uses structured light to capture a textured 3D mesh and capture the front and sides (beyond the ears) of

Figure 3.1: Figure shows the old generation of 3dMD capture system used to capture the 3D data for our research over the first four years employing two pods system comprising of six cameras; two colour and four grey scale.

.

the subjects' faces. The system consisted of seven cameras, three LED light panels, two speckle projectors (Figure 3.4), and a PC with 3dMD capturing and processing software. The PC was a 64-bit, Windows 7, 3.60Ghz, Intel Core i7-3820 with RAM of 64 GB, with a GeForce 620 GPU, and seven 250GB Samsung EVO SSDs (one for each camera). The system consisted of 7 Allied Vision Technologies Prosilica GT cameras: 3 colour (Figure 3.5) and 4 monochrome (Figure 3.6). These cameras were 2 megapixel cameras with a gigabit Ethernet interface, Truesense KAI-02050, 2/3 CCD Progressive sensors, and Auto Iris, having a resolution of 1600×1200, a bit depth of 14-bit (momo) and 12-bit (colour), and a maximum frame rate of 62fps [201]. The layout of these cameras consisted of two pods containing two mono cameras on the

Figure 3.2: The 2 colour images of a single frame using old generation 3dMD capture system.

Figure 3.3: The 4 mono images of a single frame using old generation 3dMD capture system.

top and bottom, with a colour camera in the middle. These pods were on the left and right side of the subject, with the seventh camera, a single colour camera, in the centre of the subject. This camera was placed in front and slightly above the subject. The speckle projectors were modified Moritex Schott LLS 3 LED light sources, which have an optical output of 550 lumens and fast triggered strobe capabilities.

The mono cameras and speckle projector were slightly offset with the colour cameras and 3 LED light panels. The LED panels oscillate at 120 Hz. When the LED panels are on, each colour camera captures a texture frame (Figure 3.5). When off, the

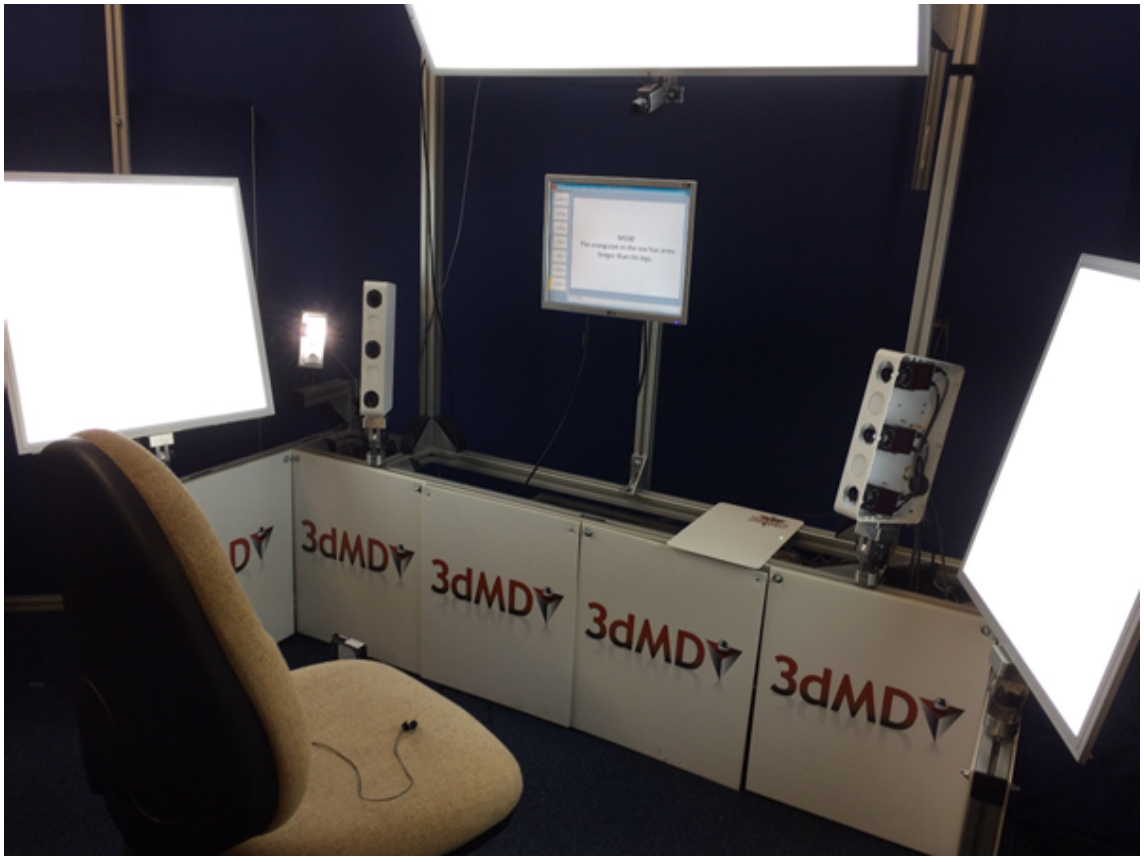


Figure 3.4: Figure shows the last generation of 3dMD capture system used to capture the remaining data for our research over the last five years, consisted of two pods containing two mono cameras on the top and bottom, with a colour camera in the middle.

projectors project a speckle pattern onto the subject and the mono cameras capture a frame (Figure 3.6).This results in seven images per frame, which are used with the camera calibration information to create a 3D frame using 3dMD's proprietary *MStereo* software program. The two file formats provided by *MStereo* are TSB (proprietary) and OBJ (open-source format). A lapel microphone was used for audio capture (44.1 KH_Z). Both systems can be considered as true three-dimensional motion capture systems, as the output data obtained are three dimensional in nature, and therefore changes in surface topology, shape, and volume can be analysed. Both systems output geometry definition file formats (OBJ), which are simple data-formats that represents

Figure 3.5: The 3 colour images of a single frame using new generation 3dMD capture system.

Figure 3.6: The 4 mono images of a single frame using new generation 3dMD capture system.

three-dimensional geometry alone, namely, the position of each vertex, the texture coordinate associated with a vertex, the normal at each vertex, and the faces that make each polygon. The new system captured facial scans at a higher mesh density compared to those acquired with the old system. Both systems essentially provided the same mesh data; the difference was in the accuracy of the mesh, and the last one improved the texture resolution.

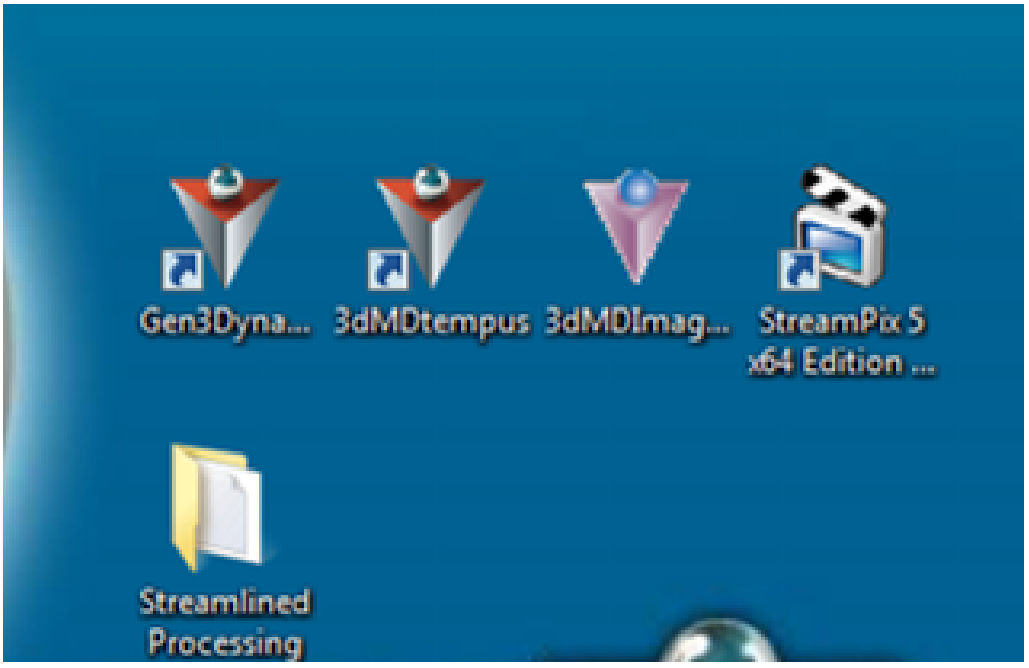


Figure 3.7: Figure shows the software provided by 3dMD capturing system

3.3 The Data Capture System Software

The 3dMD hardware consisted of 7 cameras and two projectors which were all used simultaneously to capture 3D video of an individual's face. There were three colour cameras and four cameras that only recorded in black and white. The capture system came with a third-party software tool, Norpix's *StreamPix* [113] for the actual capture process. This assists in viewing the scene and saving the captured data to disk. In order to start utilising the system, the user first needs to open up the software Gen3dynamic and Streampix5, which can be located on the desktop (Figure 3.7). Once they have started up, the user needs to go on to Streampix5 and select the 'tools' tab and select the 'load' icon (Figure 3.8). Before the user can begin capturing video of the participant, they need to ensure that these cameras are calibrated correctly. In order to calibrate the cameras, the user will firstly need to get the calibration board into position in front of the cameras (Figure 3.9). Once the file has been opened, the user then needs to make

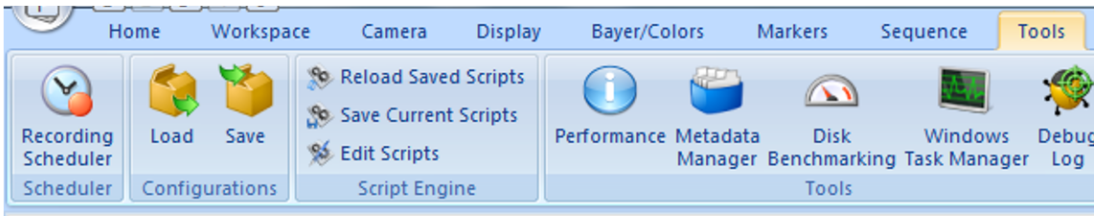


Figure 3.8: Figure shows the first window of Streampix5 that were started up: Select the tools tab and select the load icon..

Figure 3.9: Figure shows the calibration board; calibrate before using the system

sure that the cameras and LED projector have been switched on. The user will then need to click the trigger pulses on the button, which should activate the camera views on the streampix5 software showing the user four views of the cameras (Figure 3.10). Appendix C shows the guide to the main steps of the 3D imaging system. As we had to collect data, we needed volunteers whom we could scan. Previous scans were recorded in a spreadsheet. From this spreadsheet, contact was carried out previous volunteers by email for a re-scan every year and, sometimes volunteers were booked in person. Prior to recording participants, we gave them an information leaflet about the scan we

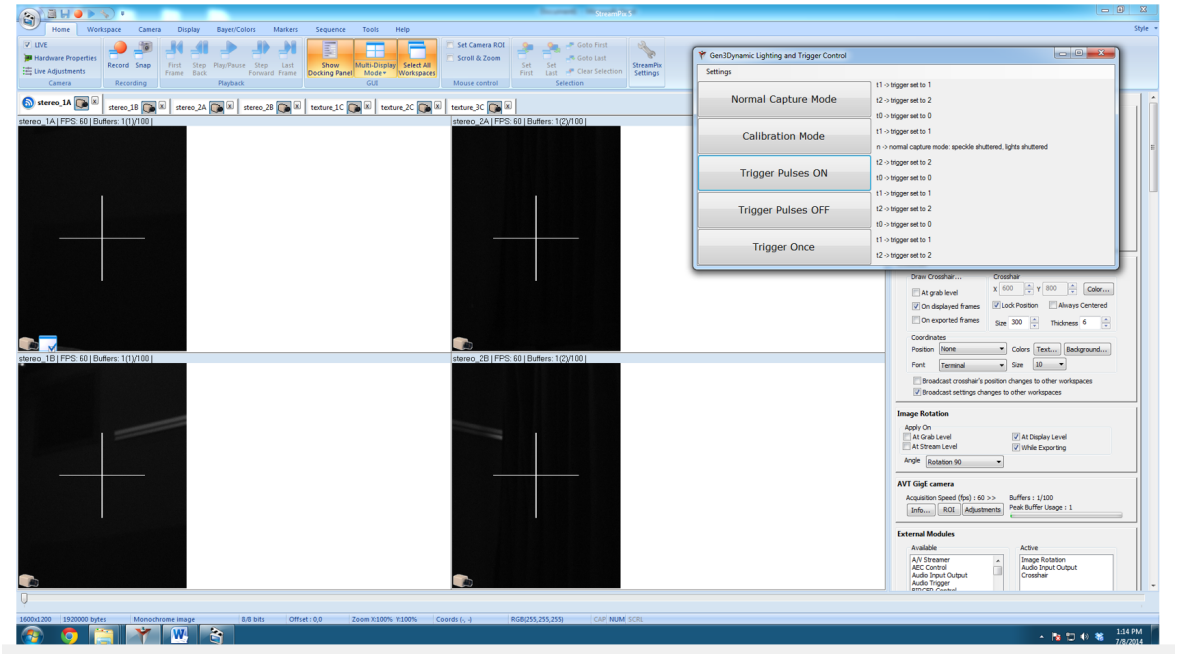


Figure 3.10: The figure shows cameras views activate on the streampix 5 software showing the user 4 views of cameras and Gen3 Dynamic software showing lighting and trigger control.

were doing and a consent form asking for permissions for follow up scans, and asking if their data could be used for research and online (see Appendix C). This software system was originally capturing the data in a proprietary format, for each of the seven cameras. One BMP for each frame was captured for each of the seven cameras. A Python [114, 200] script was used for moving these BMPs to a destination folder and renaming each file as appropriate to follow the 3dMD file format guidelines. Along with these BMPs, system calibration information and other relevant files were copied to the destination folder. From here, it was be possible to run 3dMD's proprietary *MStereo* program to produce the 3D frames. 3dMD's *MStereo* program, which uses the seven BMPs and calibration information to create the 3D frames, was used.

For each frame, the seven BMPs along with calibration information was fed into the *MStereo* program, and as output, a TSB file was created. This process included a simple multiple-processing approach and bypassing the on screen displaying feature,

as shown in Figure 3.11.

The TSB file is a 3dMD proprietary format and is useful in a variety of 3dMD software. However, because of its closed format, it would not be useful for reading in viewing programs, such as Blender [115] or Meshlab [45], nor would it allow for editing of information. Using a 3dMD program, the TSB files were converted to OBJs (with a three-view texture map BMP file and MTL file) (Figure 3.13 (a) and 3.14 (a)).

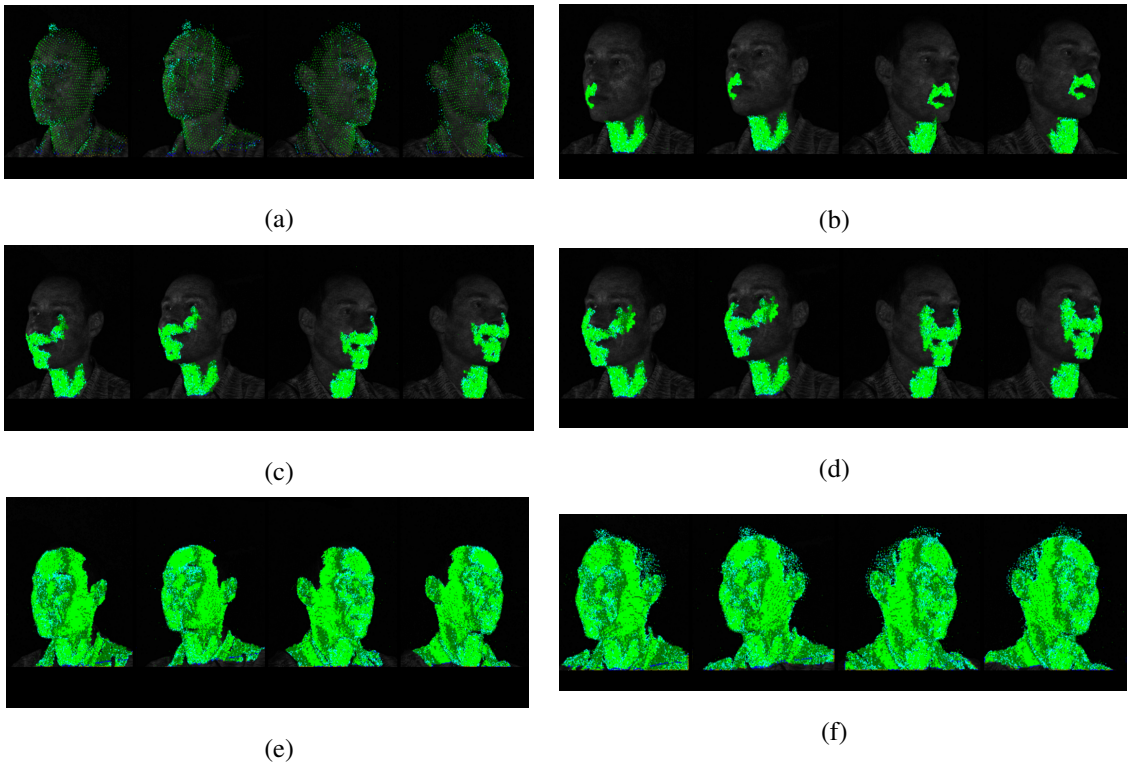


Figure 3.11: Visualisation of MStereo 3D reconstruction process

3.4 Longitudinal Studies

A large amount of data from a relatively small number of subjects is provided by our longitudinal study, which included taking repeated performances of the same individuals over time. The data aims to highlight individual changes, in particular those caused by time such as the individual ageing, which may not be detected in a cross sectional

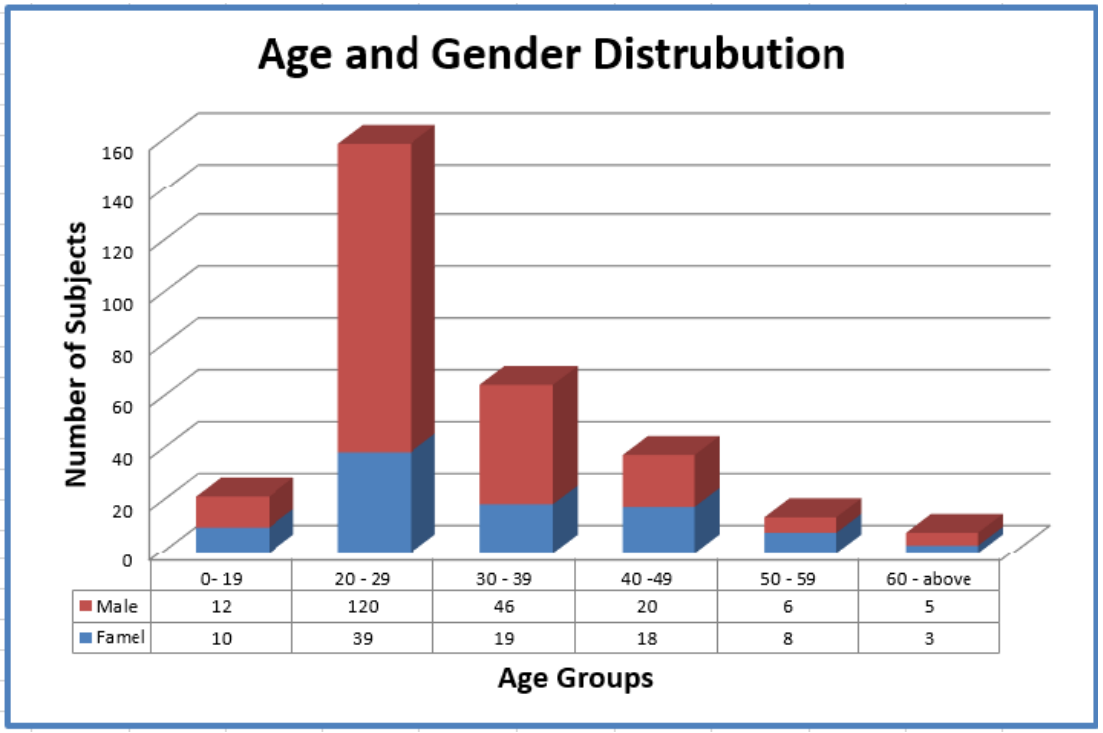


Figure 3.12: The age and gender distribution of the subjects in the Cardiff longitudinal 3D face database.

study. Longitudinal analysis gives more consistent and meaningful results because growth changes are often subtle and not always observed when data is evaluated cross sectionally. However, longitudinal studies have many disadvantages. This type of study is more expensive and takes longer to carry out than cross sectional studies.

Our study required the data to be collected over a long period of time. Often several factors influence these studies; for example, the sample size of longitudinal studies is often small because of the loss of subjects due to absence, migration, retirement, or withdrawal from the study. Further details about this the data is given in Section 3.5.

3.5 Cardiff Longitudinal 3D Face Database

Over the last nine years, the 3D longitudinal database of 3D videos was collected for our research. Prof. David Marshall from the Cardiff School of Computer Science and Informatics carried out the supervision of this project, and we were able to recruit 306 unpaid volunteers (209 male and 97 female) of different ethnicities, and ranging from approximately 4 to 61 years of age which were distributed in different age groups as shown in Figure 3.12. The participants were mainly staff and students of the Schools of Computer Science and Informatics and Engineering, and also students of the summer school at Cardiff University. Approximately 15 minutes were needed to capture every session. Each session consisted of four sentences, a sequence of single words, and two smile sequences (open and closed mouth). Participants were given approximately a minute before each capture session to allow them to become comfortable with the environment. They were given an indication of when recording began, and then eight sequences captures were recorded. Many participants were asked to come back yearly for a further scan to see the differences in ageing, and whether the facial dynamics had changed over time. Thus, nine consecutive years of data has been gathered for some participants, while some have gaps in the data because of the loss of subjects due to absence at the specific time of data collection (see Table 3.1). Table 3.1 shows the number of subjects over scan years -starting any year and allowing a one year gap. The first column denotes the number of scan years, with data captured over nine years. The table provides information on the number of subjects for both males and females. We excluded 23 subjects from the total number of volunteers in the Table 3.1. This information is also clearly labelled in the table and takes into account the gaps for some participants. 3D video face scans of volunteers were acquired using the available 3dMD face dynamic system.

Table 3.1: Number of subjects over scan years; Start any year and allow 1 gap.

No.of Subjects/Scan years	Years overall		Years Cumulative		One year gap		After filling gap		Total After filling gap
	Male	Female	Male	Female	Male	Female	Male	Female	
9 years	5	2	5	2	1	6	7	7	14
8 years	3	6	8	8	4	1	12	9	21
7 years	8	1	16	9	1	2	16	11	27
6 years	6	4	22	13	3	1	25	14	39
5 years	5	4	27	17	2	2	29	19	48
4 years	14	9	41	26	5	2	46	28	74
3 years	15	10	56	36	7	2	63	38	101
2 years	32	15	88	51	-	-	88	51	139
1 years	102	42	190	93	-	-	190	93	283

3.5.1 3D Acquisition Protocol

Subjects were asked to perform two non-verbal facial expressions,that is, smile with lips closed and smile with lips open and to say six verbal facial gestures:"puppy", "rope", "baby", "bob", "password" and "bubble" in a normal relaxed manner. This was carried out at every session and repeated yearly. The participants were also asked to say four sentences, that is, two used by a Ormond Street Hospital, cleft lip study [181], and the others from Messiah corpus [192, 193], which were phonetically rich. These sentences were designed so that the facial muscles would be pulled to extremes and to make the participant move their face in certain controlled directions. The sentences were spoken in a neutral speaking style (no expression), and the lighting was adjusted to ensure the visible articulators were clear in the video. The sentences were recorded to add to a vast database containing useful speech for use in the future. They were recorded to define normal values for lip movements based on consonant-vowel-consonant (CVC) sequences and for tongue movements based on vowel-consonant-vowel (VCV) sequences. The sentences taken from the Great Ormond Street Sentences [181] were "The puppy was playing with the rope" and "Bob is a baby boy". The Messiah sentences were "Look out of the window and see if it's raining" and "He scooped a large spoonful from the foaming vat". The sentences were designed to capture all the consonants which extent pressure lip strong, lip lateral stretch, and lip vertical aperture and lips touching.

3.5.2 3D Frame Processing

The database consists of data captures (eight sequences) at 60fps of every individual. Therefore, each sequence consists of approximately 150-500 frames, with seven camera images for each frame: three colour (Figure 3.5) and four mono (Figure 3.6). The camera calibration information and seven images are used to create 3D frames. The frames are 3D surface object OBJs (OBJ), with a three-view texture map (BMP) (Fig-

ure 3.16). Each OBJ consists of approximately 40,000 vertices, 80,000 faces (polygons), texture coordinates and normals. The total size per 3D frame (OBJ and BMP)

(a) Multi-View Texture Map (b) Single-View, Unified Texture Map (UTM)

Figure 3.13: The figure shows (a) multi-view texture map and (b) unified texture map.

is typically around 20 MB. A cleaned OBJ is then produced using the tools that were developed by Vandeventer [201]. The cleaning process removes non-manifold vertices and edges, isolated vertices, and small components, and it then produces a unified (single-view) texture map (PNG) (Figure 3.16(c)). Further details of this process can be described in the Section 3.6.1. The total size of the new OBJ and PNG is typically around 6 MB.

3.6 Pre-processing Stage

The sections below describe the cleaning process for removing unwanted mesh issues (e.g., non-manifolds) and the process for creating a single-view, Unified Texture Map (UTM). This new texture map allows us to modify a mesh (i.e., modified mesh vertex locations) without creating mesh texture problems. However, the UTM cannot be

created until the unwanted mesh issues are removed by the cleaning process. The following sections explain the issues that occur with the data and why, these issues occur, and the methods that are implemented for fixing these issues.

3.6.1 Mesh Cleaning and Unified Texture Map (UTM) Creation

To use the meshes for modelling the facial dynamics in this research, we used a fully automatic cleaning process developed by Vandeventer [201] for cleaning the meshes and creating a single-view texture map (UTM) from the raw data provided by the capture system, which produces multi-view texture maps and meshes with duplicate vertices. The cleaning process involves removing unwanted mesh artefacts (e.g., non-manifolds) and creating a single-view (Unified Texture Map (UTM)). One of the main issues with the captured meshes is the existence of duplicate vertices. A duplicate vertex arises when a single vertex location is a member of multiple faces and, thus, refers to multiple texture coordinates. When the vertices of a mesh are modified, their corresponding textures are modified, which may result in texture UV coordinates in separate images of a multi-view texture map (Figure 3.16 (b)). Modifying the location of the vertices is necessary for the registration approach, and these registered data is necessary for building the statistical models we used in this research (Chapter 4). Therefore, it was necessary that we perform the pre-processing step to use these data with cleaned meshes and a new texture map (UTM).

The cleaning steps include removing non-manifold edges and vertices, along with small components and isolated vertices, and the holes created from the removal of these imperfections are identified and filled. Further details of cleaning process are described in Vandeventer [201, 202]. The UTM is created by simply flattening the surface mesh onto a plane (2D planar parameterisation), filling the mesh faces with the appropriate texture values, and rendering the mesh as an image. This new image becomes the texture map used by the cleaned 3D mesh. In order for this flattening to occur, the mesh needs to be cleaned of problematic issues, such as non-manifold

(a) Three Wide Texture Map

(b) Stacked Image Texture Map

Figure 3.14: The figure shows 3dMD texture map layouts of the new generation system.

vertices and edges, isolated (floating) vertices, and small components (separate mesh areas disconnected from the main mesh).

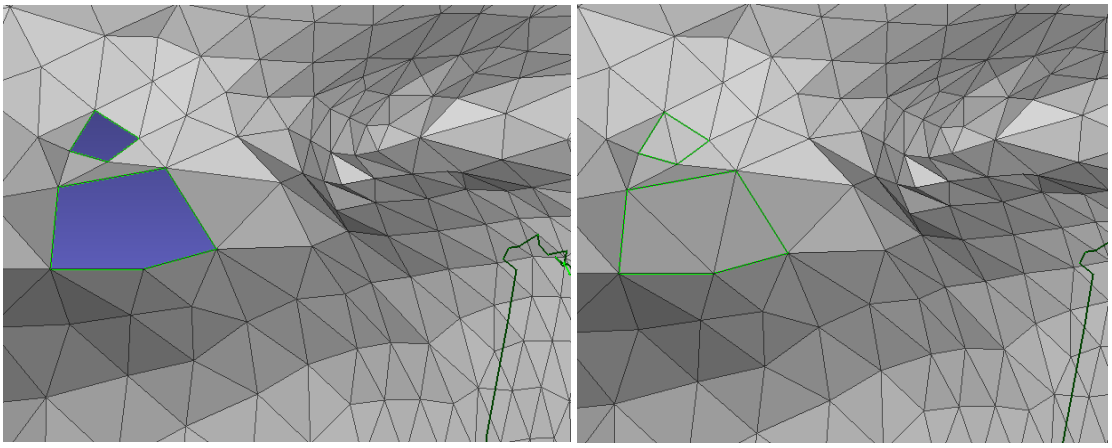
The issue of small components occurs in messy areas of the mesh capture, such as the neck and shoulders. These issues are often caused by clothing. Isolated vertices are vertices which are not members of any mesh faces. Using the vertex structure and faces structure, any vertex structure indices that do not exist in the faces structure are identified as isolated vertices. These vertices are deleted and the update and shifting process is performed.

The other issue that occurs is non-manifold edges which arise when a single edge is occupied by more than two faces. A non-manifold edge therefore consists of two vertices. To find non-manifold edges, a vertex face adjacency matrix is computed. This provides information about which faces are connected to which vertices.

If any pair of vertices are connected to more than two faces, they contain non-manifold edges. A non-manifold vertex can also be identified and removed by this process. A

non-manifold vertex exists when a vertex connects two faces that do not share an edge.

The most important approach of the cleaning process is the filling of holes which are created from the removal of these imperfections; these are identified and filled. This approach, of deletion and hole filling was found to be the most generalised but successful way of dealing with the variety of complex issues arising from non-manifolds in a dense mesh. In many instances the holes created by removing non-manifolds consist of



(a) Example of holes to be filled (b) The figure shows the result of filling a holes

Figure 3.15: Example of holes resulting by removing Non-manifolds

new or not-yet-repaired non-manifold vertices or edges that are part of the hole in some way. Figure 3.15 shows an example of two holes in a surface mesh. There are many types and configurations of issues and resulting holes that occur when non-manifold edges are removed. Further interesting information about the whole cleaning process can be found in Vandeventer [201].

(a) Original mesh

(b) Multi-view texture map

(c) Unified Texture Map (UTM)

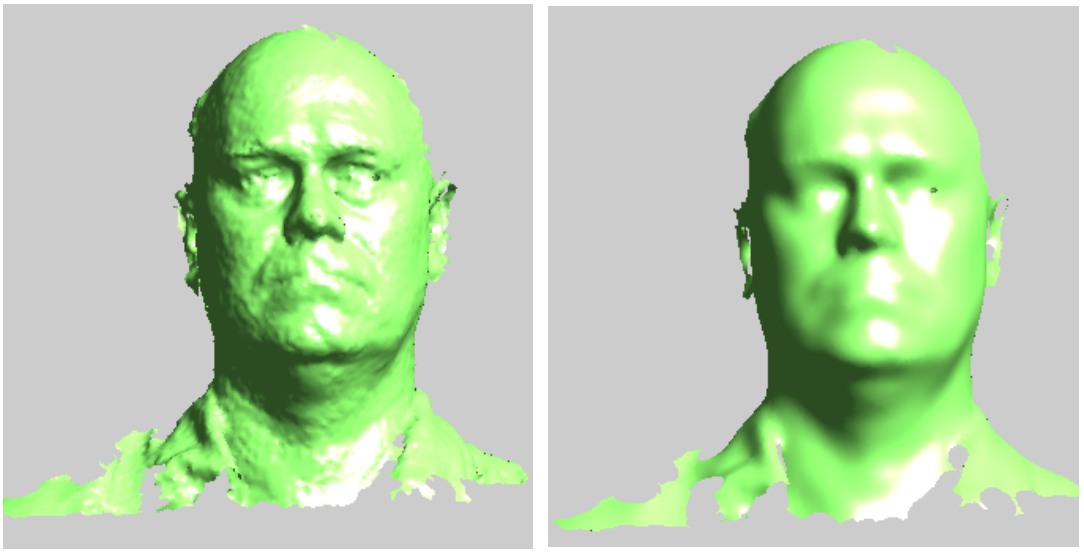
(d) Cleaned mesh

Figure 3.16: The figure shows 3dMD dynamic scanner example

3.6.2 Mesh Smoothing

The data from the scanner can also contain errors, e.g., noise and missing data, which manifests itself as holes on the face. Noise can be dealt with using a smoothing operator, but holes are a more serious problem, requiring detection and interpolation. Even when using high accuracy 3D scanners to create 3D face models, the acquired 3D models are occasionally noisy. Thus, the facial surfaces need to be smoothed to reduce the noise using a suitable software technique. In order not to distort the source mesh, a fast and robust mesh smoothing method is required to remove the undesirable noise before further processing is carried out and to preserve the geometric details. Therefore, the Laplacian smoothing method with shape and volume preservation is adopted in this study [196]. A wide variety of mesh smoothing algorithms have been proposed in recent years [121].

Laplacian smoothing is a method to smooth a polygonal mesh. For each vertex in a mesh, a new position is chosen based on local information (such as the position of neighbors), and the vertex is moved there. Where a mesh is topologically a rectangular grid, each internal vertex is connected to four neighbors, and this operation produces the Laplacian of the mesh [108, 188]. In our study, we used a simple method, inspired by existing methods, which computes vertex positions as a Laplacian. This is a high pass filter operator that computes second order derivatives. This method is supported by [111]. The geometry captured during the scanning process consists of a 3D surface with a very high resolution. To register the vertices, it is necessary to filter the geometry. This pre-processing is achieved by reducing the noise and smoothing the data, that is, removing the noise (the high frequencies) and keeping the overall shape (the low frequencies). Compute mesh Laplacian method [111] is used, and the result is as shown in Figure 3.17.



(a) Original mesh (b) Smoothed mesh

Figure 3.17: Smoothing: original mesh (a), and the smoothed mesh (b)

3.6.3 Colour Equalisation

A standard problem in the field of image processing and computer vision is image contrast enhancement. It is widely used in many image processing applications, such as object recognition and medical image processing, and as a pre-processing step in video processing applications [153]. Image enhancement can be defined as the improvement of the quality of an image using a variety of techniques. These include contrast enhancement, colour enhancement, and edge enhancement. In this study, due to using two generations of 3dMD face systems and different light conditions, the resulting image from these systems varied in the contrast of the texture image map. We used the histogram equalisation function to adjust the contrast of the image. Histogram equalisation (HE) is a very popular method for image enhancement in the image processing of contrast adjustment using the image's histogram. It shows a good performance in almost all types of images due to its simplicity. The histogram matching function [110] was used to adjust an image to match its histogram to that of another image, B = IMHISTMATCH (A, REF). Thus, this function transforms the input grayscale or truecolor image A, so that the histogram of the output image B approximately matches the his-

togram of the reference image REF, when the same number of bins are used for both histograms. For truecolor images, each colour channel of A is matched independently to the corresponding colour channel of REF (see Figure 3.18).

(a) Input image (b) Reference image (c) Output image

Figure 3.18: The figure shows the results of colour equalisation function (a) the input image (A), (b) the reference image (REF) and (c) the histogram matched image (B) .

3.6.4 Tracking and Registration

To analyse the facial dynamics and to build statistical models for 4D facial ageing data, facial feature points for each frame are required. This section describes the tracking and registration approach used in this research. These feature points provide a sparse correspondence between frames and facilitate dense registration. In recent years, a semi-automatic 4D tracking and registration method has been developed by Vandeventer [201, 202], and this approach was adopted in this study, as it keeps the 3D information for each frame and results in high-resolution, accurate, inter-subject densely registered 3D frames. In this work, the tracked feature points were used for our registration method. This registration method produces the output that was used as an input to build the statistical models. In total, 41 manually placed landmarks were used to analyse the face surface shape and appearance: 4 points to the corners of each eye, the nose bridge, the nose tip; 7 points around the nose; 12 points at each

lip corner; 1 point over the chin; 8 points for the contour of the face; and 2 points on the left and the right cheek. The approach used in this study requires the user to landmark a single mesh from the sequence, clicking on the ordered sequence of 41 facial feature points: eye corners, centre of upper eyelids, nose bridge and tip, nason nose, lip corners, contour of face, and cheek centre. To produce realistic and accurately tracked 4D sequences a tracking approach is required that uses actual 3D information (3D texture and shape), where the 3D texture comes from the texture information from the 3D mesh faces (not from the 2D texture map) and the 3D shape comes from the mesh surface curvature information [201].

3.6.5 3D Landmarking Method

The approach used in this study requires the user to landmark a single mesh from the sequence, clicking on the ordered sequence of the facial feature points: eye corners, the centre of upper eyelids, nose bridge and tip, nason nose, lip corners, contour of face, and cheek centre. The Matlab function interface with Blender software was used to help annotate the first frame of each sequence; Blender is a 3D computer graphics software program. The 3D annotation tool consists of a Matlab function to select the files and launch a Blender program to load the mesh file to be annotated; it also provided the view of two Python scripts as shown in the screen-shot of this tool in Figure 3.19. The landmark selected by the user writes out to file using the Python script. This file has the same name as the mesh file, and it is used to create the file needed for the automatic tracking process. This tool allowed us to manipulate the 3D mesh to place the landmark quickly and accurately. The tool also helps as well to remove and add the landmark points, and the speed of annotation is according to the annotator's ability. These points are then automatically tracked following the approach proposed by Vandeventer [201, 202], and this enables us to analyse the facial dynamics subsequently, as shown in Figure 3.20. A tracking method is required that uses the 3D information to produce realistic and accurately tracked 4D sequences.

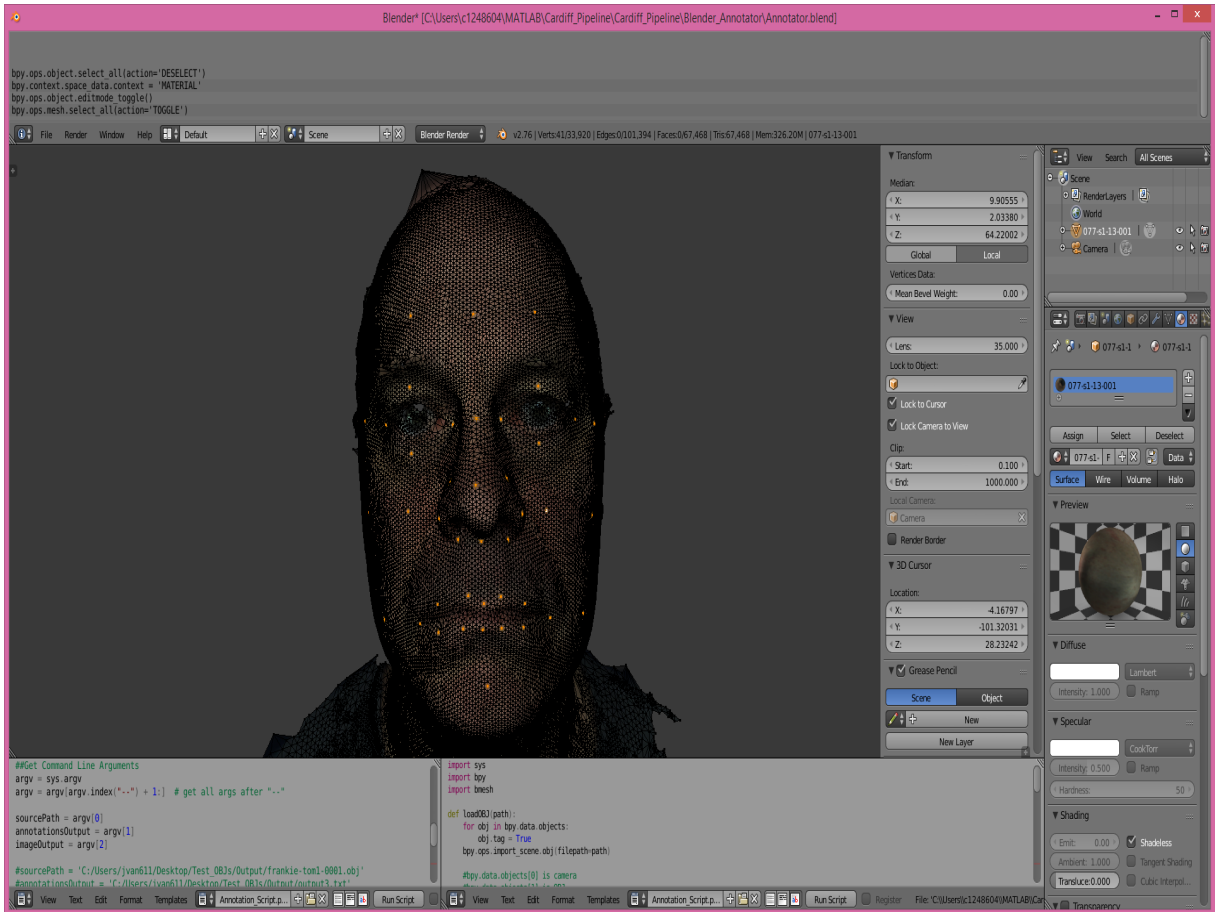


Figure 3.19: The figure shows the window of the 3D Blender annotation tool

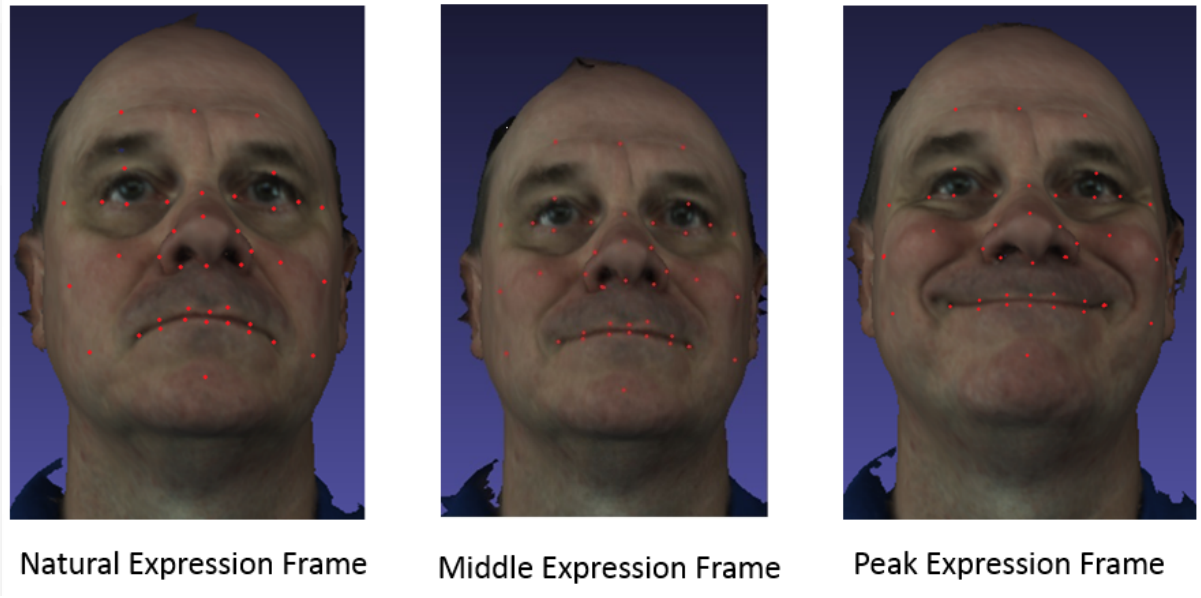


Figure 3.20: The figure shows the tracked smile sequence of three frames

3.6.6 Inter-Subject Registration

After the feature points have been tracked automatically, then these feature points are used as control points to create a dense correspondence between meshes. These control points can be used to correspond any sequences in the same order, which makes it possible to carry out both intra-subject and inter-subject registration. The facial area of the individual is the region of interest for our research. Thus, the unwanted areas, such as the chest, shoulder, neck, and hair, are removed from the selected frame of the subject to be the reference mesh. This mesh is called the registration mask, and all other frames are registered to that mask. Using Blender software, the vertices are selected from the section of face that needs to be kept, and the other areas are deleted. This process can be performed manually and is easily achieved with Blender. The same landmarking technique was selected to annotate the registration mask, which was used as a reference mesh for the registration process. We can summarise this process in several steps [24, 201, 202]

- Select a mesh frame, T_{ref} as a face template (registration mask or reference

mesh) for registering any other mesh frame, T_k. This can be a noise free frame selected from the training set. Every vertex in T_k must have a corresponding vertex in T_{ref}.

- Using Thin Plate Splines (TPS) [33, 202], T_{ref} is warped to the other mesh T_k and this creates the registered vertices for T_k which are called T_{new}.

- The interpolated mesh resulting from step 2 produces new vertices and is approximately identical to the original shape of T_k. The remaining displacement is between the surface of T_{new} and T_k is removed by snapping each vertex of T_{new} onto the surface of T_k.

TPS is an effective tool for interpolation, as it has multidimensional data points and involves minimising the bending energy function [33]. TPS has been widely used for the non-rigid transformation of data for image alignment.

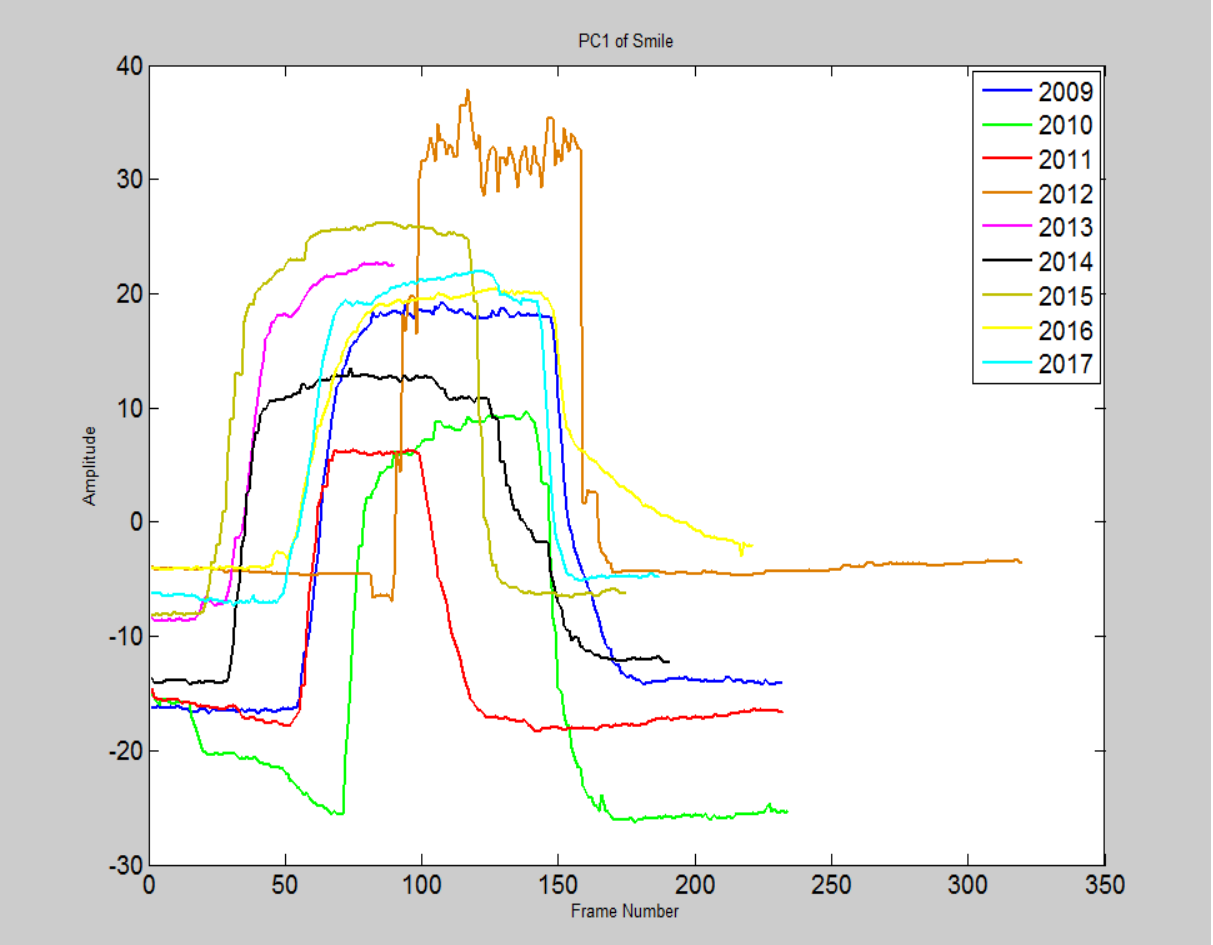


Figure 3.21: Temporal variation of the ASM for full landmark of smile of the first mode of variation from the same subject over nine years.

3.7 Performance of 3D Specific-Person ASM and AAM

One of our goals in this work is to analyse the 3D dynamic ageing data; as we mentioned previously, no such 3D dynamic data has been available to date. Therefore, we built a specific AAM for each subject to analyse these dynamics to depict how facial dynamics change over time. Figure 3.23 shows the variation of AAM for one subject over nine years; the issue with texture is because the scanner is not accurate with texture, and there is a large variation in colour in the first mode. However, we are looking how the dynamic changes, and so the texture is not in the scope of our research. We

validate this issue by building the ASM using the full landmark and mouth landmark of a smile, and the results were improved, as shown in Figure 3.21 and Figure 3.22.

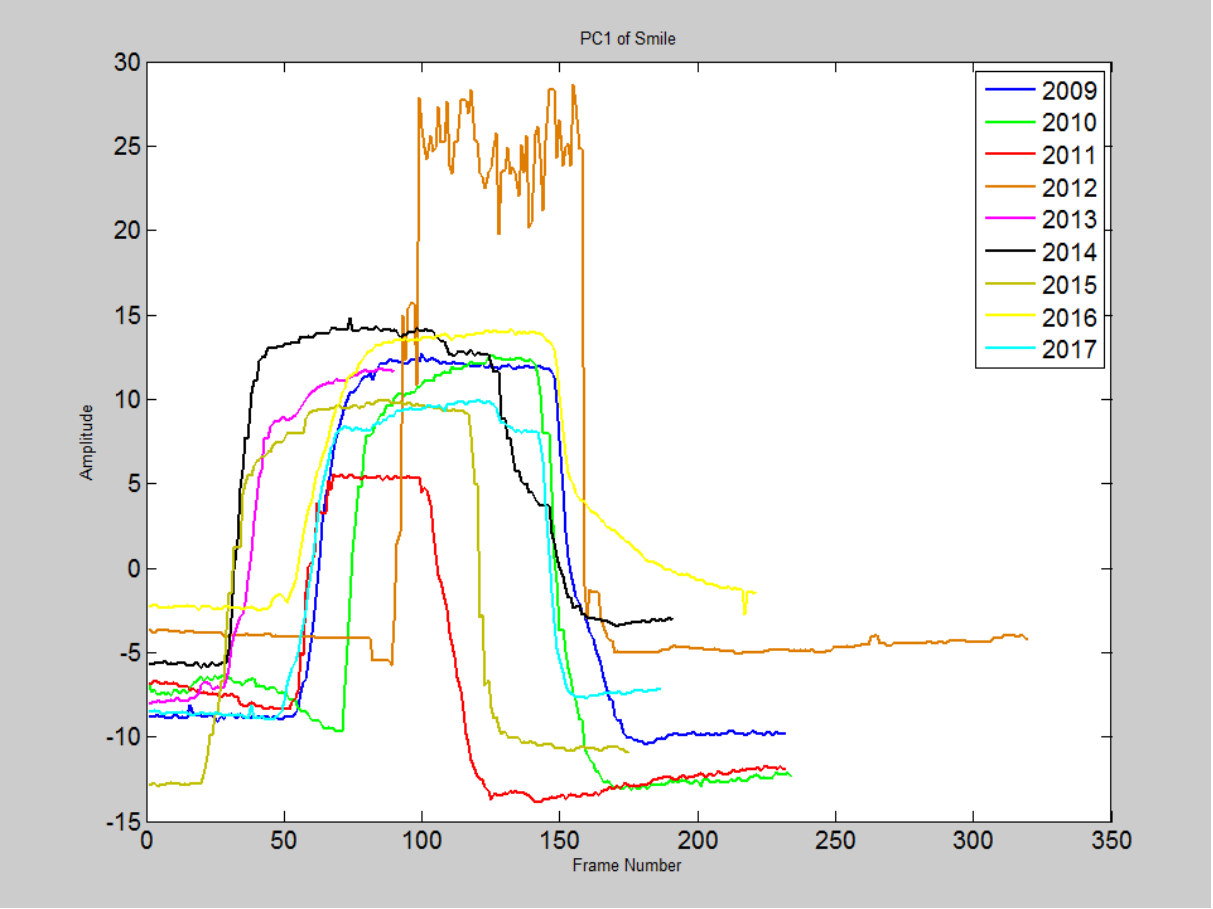


Figure 3.22: Temporal variation of ASM for mouth landmark of smile of the first mode of variation from the same subject over nine years.

3.8 Ageing Pattern

The interesting problem is that people age differently, so the concept of the 'age' is specific to each person. An ageing pattern can be defined as a sequence of personal

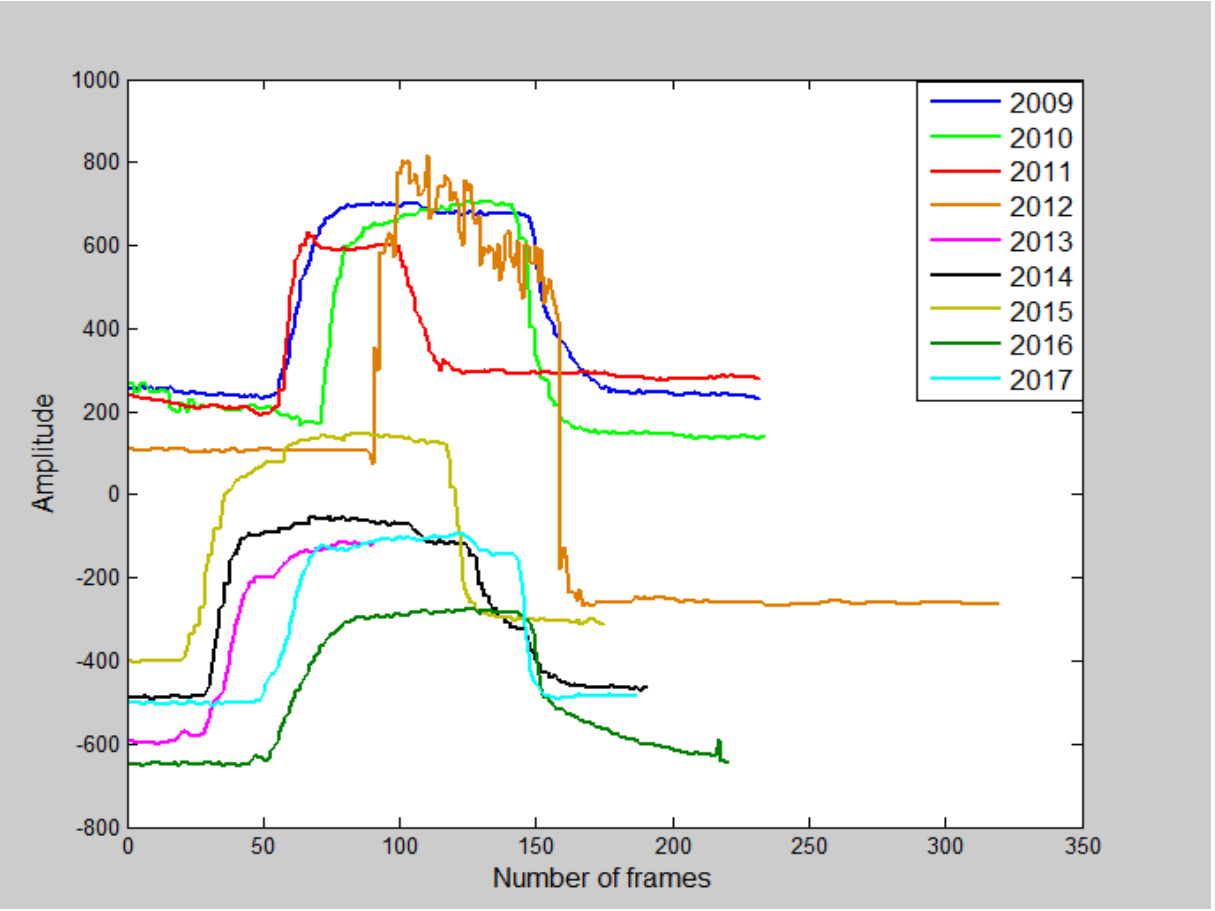


Figure 3.23: Temporal variation of AAM of smile of the first mode of variation from the same subject over nine years.

face images sorted in time order. In an ageing pattern, all face images must come from the same person and be ordered by time [87, 93]. The 2D matrix I represents the face image, where $I(x, y)$ is the intensity of the image pixel (x, y). Thus, the 3D matrix P represents the ageing pattern of the face image at time t, where the $P(x, y)$ is the intensity of the image pixel (x, y). Figure 3.24 is an example of the ageing pattern, where each image represents the scan year of the subjects. Over nine years (2009 to 2017), each age is allocated one position. If the scan year is available for certain ages, they are filled with the corresponding positions. If not, the positions are left blank (squares). If positions are filled over nine years, the ageing pattern is called a complete ageing pattern. Otherwise, it is called an incomplete ageing pattern. The missing scans

for people are a problem for our research.

Figure 3.24: Sample 3D scans of the same subject from the longitudinal database over 9 years, blank square means the missing year.

Incomplete datasets may lead to results that are different from those that would have been obtained from a complete dataset. One approach to solving incomplete data problems is the adoption of interpolation techniques [122]. To prepare enough data for our project, nine consecutive years of data were gathered for some people, while some had gaps due to the loss of subjects because of absence, illness, migration, retirement, and graduation or withdrawal from the study. In addition, the data was collected at a specific time, and some of these people were not available at that time. However, we can fill the gaps in the database using the interpolation techniques. To evaluate the missing years, which can be estimated from the existing years, we used the performance indicators to measure the difference between the prediction and the actual value.

3.8.1 The Interpolation Techniques

Different interpolation techniques were used to estimate the missing values. We used the linear interpolation function, as supported by [109], and specified an alternative interpolation method: 'linear', 'nearest', 'cubic', and 'spline'. The linear interpolation function returns interpolated values of a function of n variables at specific query points using linear interpolation. We used the interpolation specified by the method to generate the missing data. We tested our data with two methods (linear and cubic), which are described in the following sections.

3.8.2 Linear Interpolation

The simplest form of interpolation is to connect two data points with a straight line. This technique is called linear interpolation. If the two known points are given by the coordinates (x_0, y_0) and (x_1, y_1), the linear interpolant is the straight line between these points. For a value x in the interval (x_0, x_1), the value y along the straight line is given from the equation of slopes [91].

$$\frac{y - y_0}{x - x_0} = \frac{y_1 - y_0}{x_1 - x_0}, \tag{3.1}$$

Solving this equation for y, which is the unknown value at x, gives

$$y = y_0 + (x - x_0)\frac{y_1 - y_0}{x_1 - x_0} = \frac{y_0(x_1 - x) + y_1(x - x_0)}{x_1 - x_0}, \tag{3.2}$$

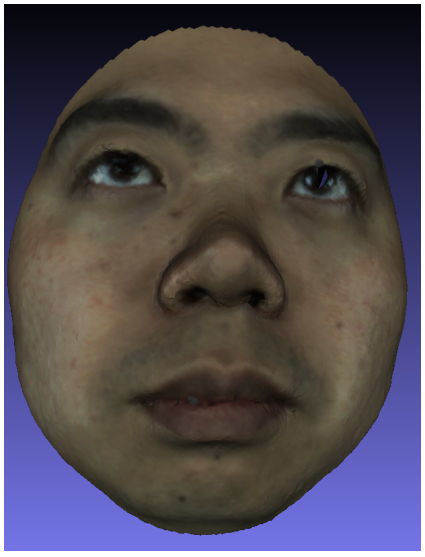
3.8.3 Cubic Interpolation

When four data points are available, a third-order polynomial (also called a cubic polynomial) can be applied. The cubic interpolation formula has the form:[18]

$$f_3(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1) + b_3(x - x_0)(x - x_1)(x - x_2) \tag{3.3}$$

(a) Target Mesh

(b) Interpolated Mesh



(c) Target Mesh

(d) Interpolated Mesh

(e) Target Mesh

(f) Interpolated Mesh

Figure 3.25: The figure shows the examples of interpolated mesh from different individuals.

The equation of the linear interpolation function is: [39]

$$f_1(x) = b_0 + b_1(x - x_0) \tag{3.4}$$

where x is the independent variable, x is a known value of the independent variable and y is the value of the dependent variable for a value x of the independent variable.

$$b_0 = f(x_0) \tag{3.5}$$

$$b_1 = \frac{(f(x_1) - f(x_0)}{x_1 - x_0} \tag{3.6}$$

$$b_2 = \frac{(f(x_2) - f(x_1)}{x_2 - x_1} \tag{3.7}$$

The procedure to determine the coefficients of b_0, b_1 and b_2 are the same as in Equations (3.5) to (3.7). b_3 is given by

$$b_3 = \frac{\frac{f(x_3)-f(x_2)}{x_3-x_2} - \frac{(f(x_2)-f(x_1)}{x_2-x_1} - \frac{f(x_1)-f(x_0)}{x_1-x_0}}{x_3 - x_0} \tag{3.8}$$

From the equations above, x is the vector representing vertex position of known values and the $f(x)$ represents the interpolated vector of missing year.

3.8.4 Performance Indicator

The performance indicator was used to describe the goodness-of-fit or the distribution when missing values are estimated using interpolation techniques. Both the root mean square error (RMSE) and the mean absolute error (MAE) are regularly employed in model evaluation studies. We used MAE, the most commonly used measurement, to compute the error between the interpolated and the actual values of the data.

3.8.5 Mean Absolute Error (MAE)

MAE is a measure of difference between two continuous variables. Assume X and Y are variables of paired observations that express the same phenomenon; examples

of Y versus X include comparisons of predicted versus observed, subsequent time versus initial time, and one technique of measurement versus an alternative technique of measurement. Consider a scatter plot of n points, where point i has coordinates (x_i, y_i). MAE is the average vertical distance between each point and the Y=X line, which is also known as the one-to-one line. MAE is also the average horizontal distance between each point and the Y=X line.

The Mean Absolute Error is given by:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \tag{3.9}$$

In our experiment, we considered the MAE to be the average of the difference between the predicted and the actual values of the data. MAE is conceptually simpler, more interpretable, and fundamentally easier to understand than the square root of the average of the sum of squared deviations. MAE does not require the use of squares or square roots.

The mean absolute error (MAE) is evaluated by the equation:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - y_i^{'}|}{n} \tag{3.10}$$

Where $y_i^{'}$ is the prediction and y_i the true value. Mean absolute error (MAE) ranges from 0 to infinity and a perfect fit is obtained when MAE equals to 0.

3.8.6 Experiment 1: Performance of Interpolation Technique Using MAE

Before we worked on real data, we wanted to generate some interpolation sequences and test our method on them. In this section, we used 10 complete ageing patterns (10 subjects), removed one year scan from the existing data, and then interpolated the missing year. The test data was the data after registration, where the frames in the sequence have the same topology (i.e., the same vertices). For example, the data for

Table 3.2: The table shows the performance of interpolation technique using MAE.

MAE		
Subjects	Linear Method	Cubic method
Subject 1	**0.0131**	0.0135
Subject 2	**0.0159**	0.0176
Subject 3	**0.0091**	0.0195
Subject 4	**0.0204**	0.0222
Subject 5	**0.3006**	0.3137
Subject 6	**0.0113**	0.0155
Subject 7	**0.0244**	0.0265
Subject 8	**0.0149**	0.0172
Subject 9	**0.0128**	0.0137
Subject 10	**0.0152**	0.0195

each subject was represented by matrix ($v \times 3 \times m$), where v is the number of vertices, and m is the number of frames in each sequence. Using existing years as a test data and taking out one year, we then applied the interpolation function on this year to interpolate the year that was taken out. We used the MAE to evaluate the results (see Table 3.2) and obtained good results when computing the error between the target year in the real data and the interpolated year. All landmark measurements were expressed in pixels.

Table 3.2 shows the MAE between the actual values and the interpolated values using different methods. Clearly, the performance of linear interpolation was better compared with the cubic method. Figure 3.25 shows the frames from different subjects in the missing year interpolated from the existing years.

3.8.7 Experiment 2: Dynamic Time Warping Interpolation (DTWI)

In Chapter 2, we presented DTW, a technique introduced in 1978 by H. Sakoe and S.Chiba in a study about spoken word recognition [175]. DTW is a well-known technique to find an optimal alignment between two given (time-dependent) sequences under certain restrictions (Figure 3.26). Intuitively, the sequences are warped in a non-linear fashion to match each other. Originally, DTW was used to compare different speech patterns in automatic speech recognition [150].

In fields such as data mining and information retrieval, DTW has been successfully applied to cope automatically with the time deformations and different speeds associated with time-dependent data. In order to find the similarity between such sequences, or as a preprocessing step before averaging them, we must "warp" the time axis of one (or both) sequences to achieve a better alignment. DTW technique achieves this warping efficiently [124]. The following search for an optimal warping path between the elements in position $(1, 1)$ and (n, m) of that matrix led us to find the similarities between the two sequences. We introduced the interpolation strategy based on DTW to solve the problem of incomplete data. Specifically, we applied the DTWI to interpolate the unknown sequences that existed between two known sequences. Considering two time-dependent sequences:-

$X = x_1, x_2, ...x_N$,

$Y = y_1, y_2, ...y_M$

where N and M are the dimensions of each sequence. The new sequence represents the DTWI result where the interpolated sequence is the middle of X and Y. Let us suppose there are two feature vectors, which are, respectively, depicted by two sequences of discrete data points of a smile (Figure 3.27 and Figure 3.28); we compute the average of the warping sequence (Figure 3.29) and the second sequence to obtain the interpolated sequence (Figure 3.30). The number of frames of the interpolated sequence was determined by averaging the warping and the second sequence. The proposed method for interpolating the missing year includes building a PCA model

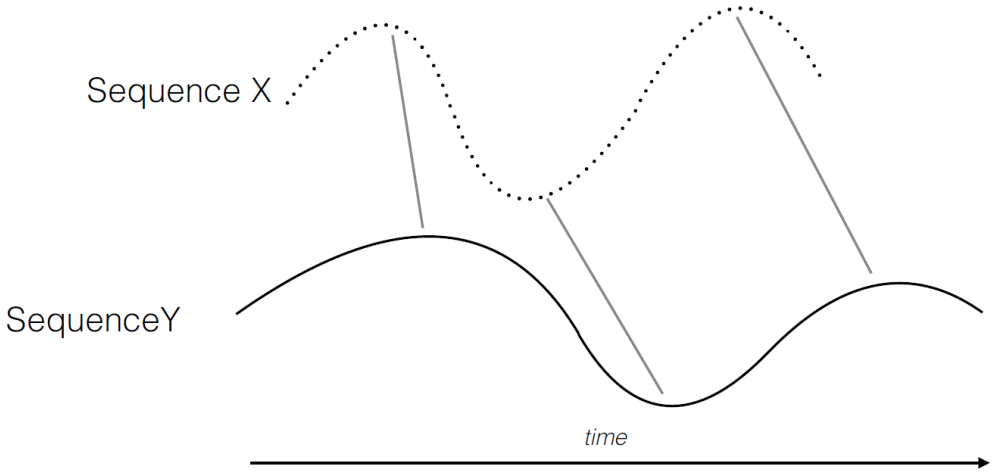


Figure 3.26: The figure shows an alignment of two time-dependent sequences. The grey lines represent the correlations between the series, originally shown in [157].

of two sequences (smile videos) and projecting the data (sequences) into the model to obtain the feature vectors of the two sequences, which are used to interpolate the new sequence (missing year). The DTW technique, described in Chapter 2, is used to warp the first sequence $(t-1)$, and the second sequence $(t+1)$ is used to generate the missing sequence (t). The average between these sequences presents a new sequence (interpolated sequence). This sequence is projected into the PCA model to construct the missing sequence. In this experiment, the temporal variation of the ASM/AAM component(s) have been interpolated. This data was validated and used to build the models in the next chapters.

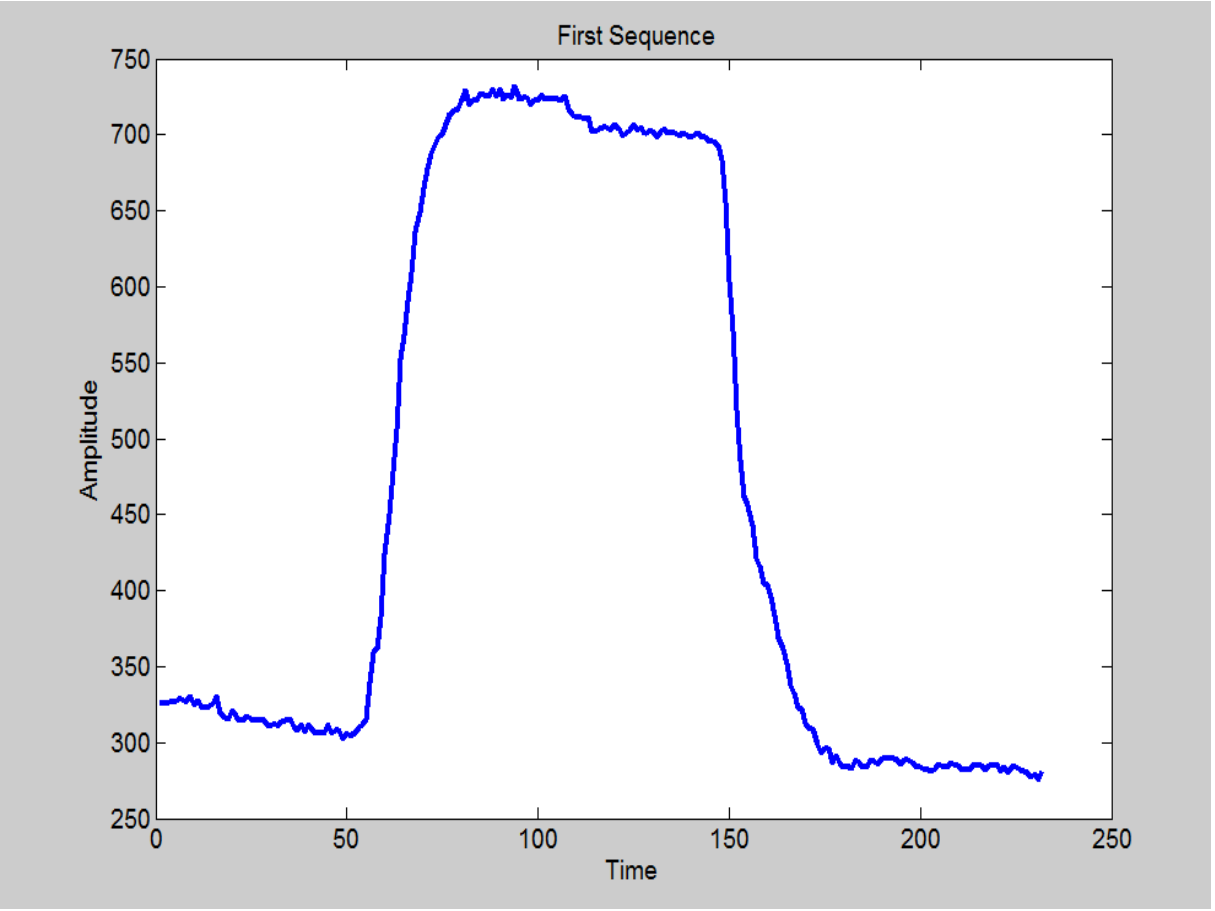


Figure 3.27: The figure shows example of the first sequence (smile: Year 2009).The unit of time is the number of frames, and the amplitude is the measured eigenvalue of PC.

3.9 Summary

In this chapter, the 3dMD 4D capture system to collect data for a 3D video scan database. The number of participants that have taken part in this study to date is 306 subjects. This database is the first in the world and will assist researchers in many applications in the fields of medicine, psychology, and computer vision research for analysing facial movement in 3D over time. A robust pre-processing technique was described for cleaning 3D meshes and creating a single-view texture map (UTM) from systems which provide multi-view texture maps and meshes with duplicate vertices. To

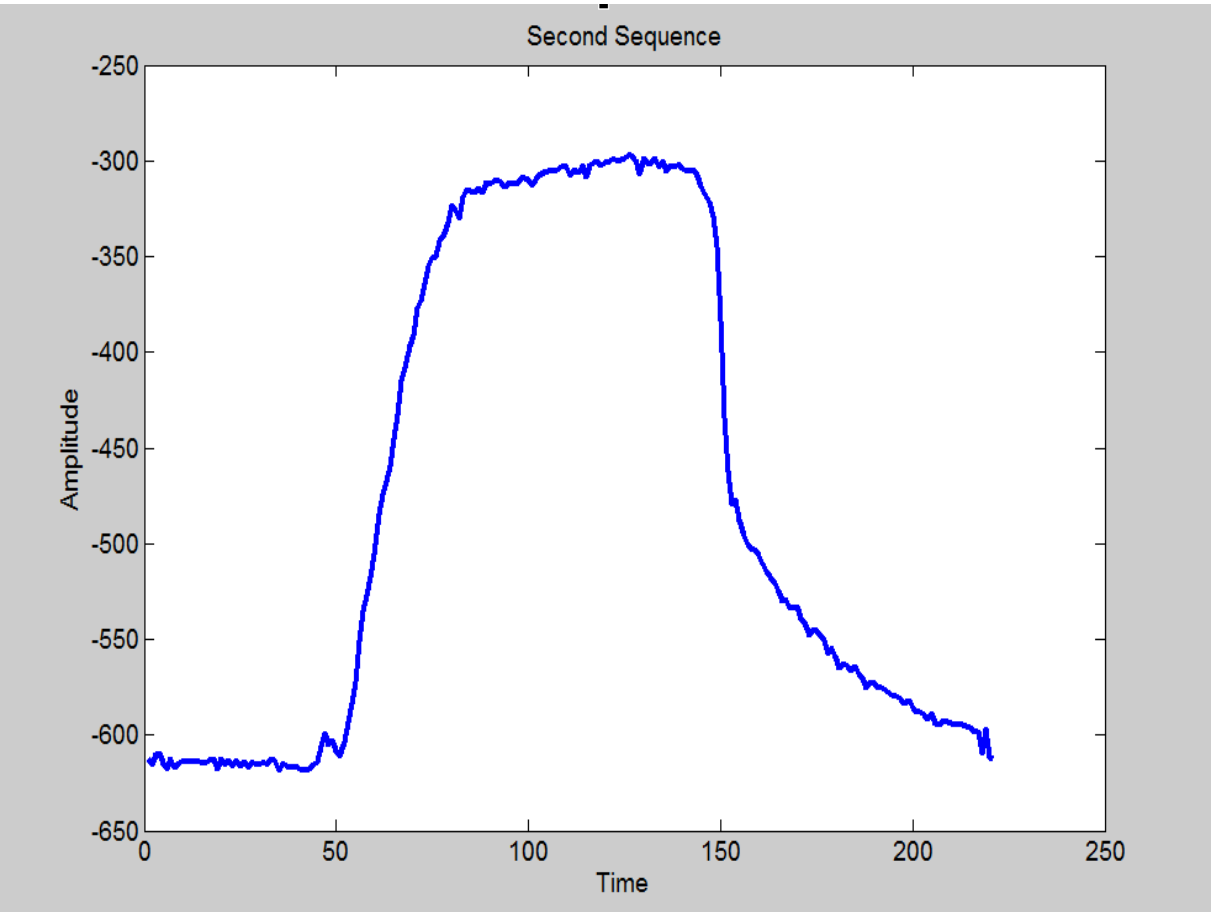


Figure 3.28: The figure shows example of the second sequence (smile: Year 2011). The unit of time is the number of frames, and the amplitude is the measured eigenvalue of PC.

build 3D AAMs models of smile expressions as people age, the 4D tracking and inter-subject registration techniques are performed on the database. These registered meshes are used for building statistical models and are described in the following chapter. Smoothing and equalisation steps were used to filter noise from the meshes. Two experiments were used to interpolate missing data. In the first experiment, the linear interpolation technique was used to fill the missing years in our data; we used the MAE as the criteria to compute the error between the interpolated and the actual year. The second experiment was the interpolation strategy based on DTW was used for missing data interpolation.

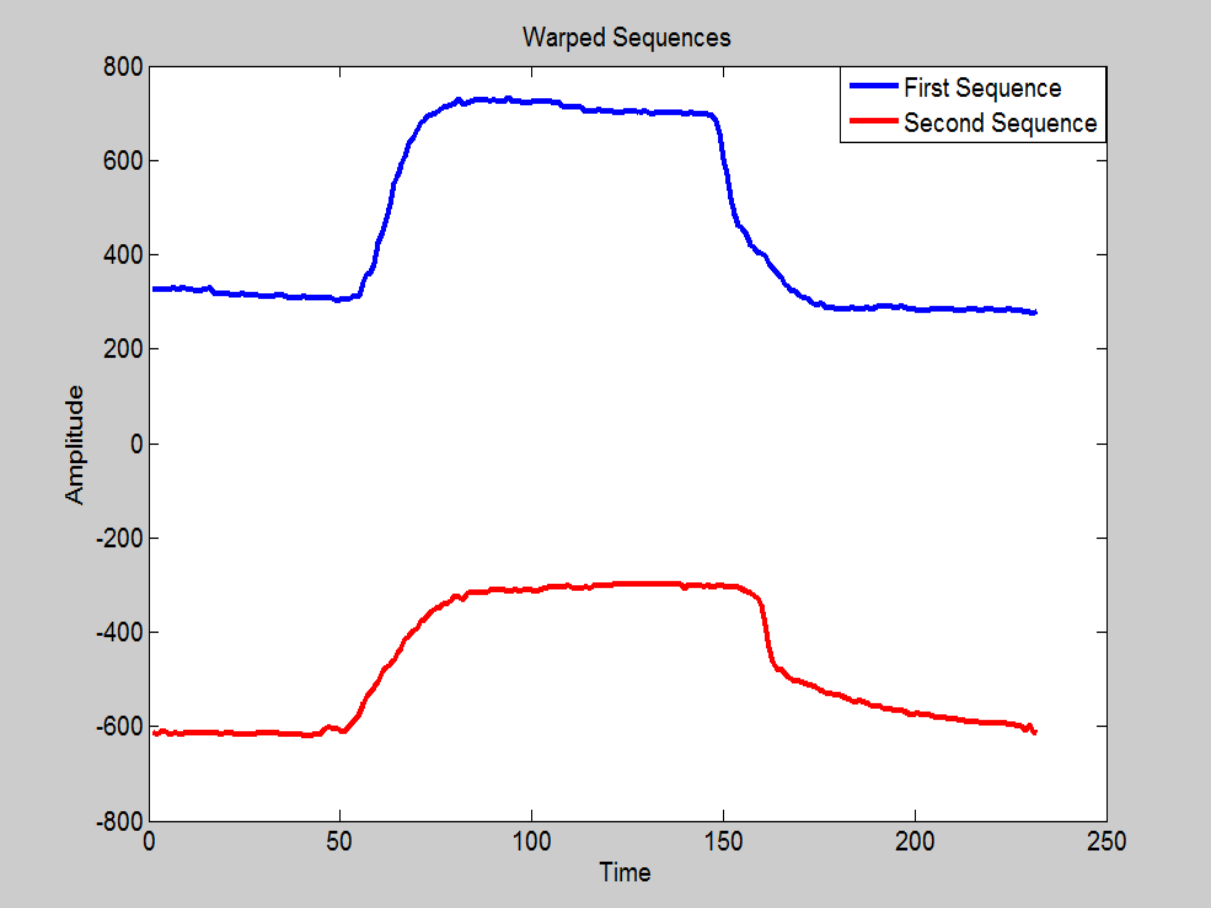


Figure 3.29: The figure shows warping sequences (smile: Year 2009 and Year 2011). The unit of time is the number of frames, and the amplitude is the measured eigenvalue of PC.

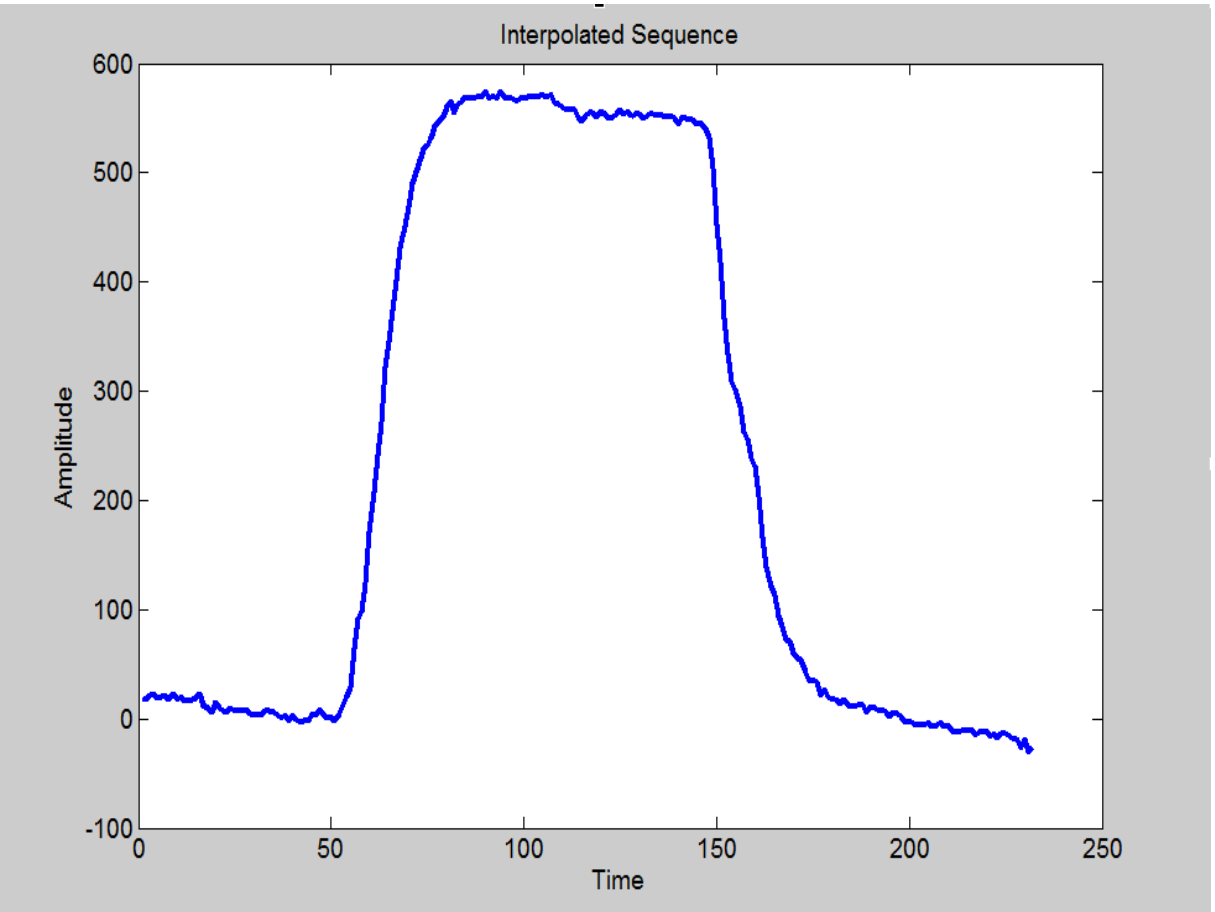


Figure 3.30: The figure shows interpolation smile of missing year (smile 2010). The unit of time is the number of frames, and the amplitude is the measured eigenvalue of PC.

Chapter 4

3D/4D Statistical Modelling of Age Data

4.1 Modelling Facial Appearance

In this chapter, we introduce the principal methodology of this thesis which is covered in Sections 4.1.1, 4.2 and 4.5. The method discusses the approach for building both a general and a specific model of 4D (3D dynamic) sequences and for synthesising new frames by modifying the model parameters. This method is used for creating a 4D, highly realistic model of facial ageing using a training set of 3D dynamic face scans of a smile expression to train a statistical ageing model. The contribution of this chapter is to build a general and a person-specific AAM model and to extract the dynamic features from the first Principal Component (PC1) of a smile sequence to analyse the facial dynamics over years. This chapter presents a method to determine how the facial appearance models can be generated. Following the approach of Edwards et al. [78], our statistical appearance models were generated by combining a model of face shape variation with appearance variations. Active Appearance Models (AAMs) are a well-known method that have been successfully used for face modelling in a variety of applications such as face alignment, face tracking, face recognition, and real time modification of facial expressions and for conversation expressions, analysis, synthesis appearance models, and the synthesis of highly realistic facial expression sequences [20, 184, 201].

In order to build 3D/4D statistical models of facial dynamics over age, a database of facial dynamics over age is required. Such a database was compiled over nine years for this research. Full details can be found in Chapter 3. To build an Active Appearance Model (AAM) of facial dynamics over age (e.g.,smile sequences) for all subjects, the data must be inter-subject registered. We adopted the 4D tracking approach that was developed by Vandeventer [201]. For this approach, it is necessary to landmark a single mesh from the each sequence, and the feature points must be tracked automatically,so they can be used them to create a dense correspondence between meshes. Meshes are aligned using Procrustes Analysis, then warped into close proximity using thin-plate splines. The Thin Plate Spline (TPS) based algorithm is used to achieve a dense correspondence (see Chapter 2). The registered data is used as input to an AAM. The Principal Component Analysis (PCA) and 3D Procrustes Analysis were performed on shape, texture, and combined weighted shape and texture by building an AAM model. The core technique of the AAM to analyse the statistical properties of the shape and texture is the PCA.

PCA is performed, and the end result is a shape and texture model that reduces the dimensionality of the mesh data down to comparatively few modes of variation. An AAM model with five modes of variation was created, which described 95% of the variance. The section below explain this technique and describes the steps for calculating the PCA.

4.1.1 Principle Component Analysis for Feature Extraction

PCA is a statistical analysis technique used by the majority of scientific specialities to reduce the dimensionality of a data while preserving as much as possible of the variation existing in the data set [77, 186, 206]. The main goal of PCA is to identify patterns in data or to detect the correlation between variables. PCA is used for finding the directions of maximum variance in high-dimensional data and projecting this data onto a smaller dimensional subspace while retaining the most salient information. PCA

is a linear transformation method which yields the directions (principle components) that maximize the variance of the data.

In our work, we used PCA to extract the important information from the data and to regard this information as a set of new variables called principal components. PCA is performed using the following process as described by Cootes et al. [55, 186].

In the first step, the data is arranged in a matrix X of vectors x_i, where each element of that vector is a variable of the x, y, and z values. Each vector is placed into the $m \times n$ matrix X, where m is the number of variables in each observation, and n is the number of frame in the matrix.

The next step required is to calculate the mean of the data. The mean is calculated for every variable as follows. To create a vector of the mean values:

$$X = \frac{1}{N} \sum_{i=1}^{n} x_i \tag{4.1}$$

Calculate the covariance matrix C of the data as follows:

$$C = \frac{1}{N-1} \sum_{i=1}^{n} (x_i - \overline{x})(x_i - \overline{x})^T \tag{4.2}$$

Where x_i is a data point in x and N is the number of elements in x.

The eigenvectors Φ and λ_i eigenvalues of covariance matrix C are calculated. An eigenvector of C is a non-zero vector Φ such that C multiplied by Φ yield λ_i. λ is the eigenvalue corresponding to the eigenvector Φ. The eigenvalues λ_i are sorted in descending order (highest to lowest) variation, so that the first Principal Component (PC1) depicts the highest variation by performing the AAM. To reduce the number of principle components k, a percentage $p\%$ of the total variations are kept, which is equivalent to retaining the k first mode of variations. The higher order of eigenvectors are considered as noise. k can be calculated as follows:

$$\sum_{i=1}^{k} \lambda_i \geq \frac{p}{100} \sum_{i=1}^{3M} \lambda_i \tag{4.3}$$

By applying the square root of the eigenvalues, their standard deviations can be obtained. We can approximate new shape x by varying b in

$$x \approx \overline{x} + \Phi_s b_s \tag{4.4}$$

where $\overline{x}$ is the mean shape and Φ is the matrix of the k largest eigenvectors depicting $p\%$ of the shape variation, and b_s is the $k-$dimensional vector of Eigencoefficients corresponding to the shape x. whereas b is constrained by $b_i = \pm 3\sqrt{(\lambda_i)}$. By solving Equation 4.4 with respect to b, it is possible to calculate the set of model parameters corresponding to a given example.

$$b_s \approx \Phi^T(x - \overline{x}) \tag{4.5}$$

In our research, these steps are required to describe a new 3D frame in relation to the model. By projecting the unseen data into the AAM model, we can obtain the vector b of parameters corresponding to the model (Equation 4.5), which shows the trajectory of the facial dynamics depicted in the PCA space spanned by the first three modes of variation.

Using the b vector and the AAM, we can project out the data to reconstruct the 3D frame (Equation 4.6).

$$\overline{S} \approx \overline{x} + \Phi_s b_s \tag{4.6}$$

4.1.2 Combining Shape and Appearance Models

Following the original AAM approach described by Cootes et al. [55], we built separate PCA models of the shape variation and then built combined weighted shape and appearance models into a single PCA by merging the input vectors. The texture map computed by systems captured comes from multi-viewpoint images. These images need to be merged into a continuous texture map in order to build the AAM.

Suppose a set of image texture vectors:

$$g_k = (r_1^k, g_1^k, b_1^k, ..., r_i^k, g_i^k, b_i^k, ..., r_M^k, g_M^k, b_M^k)^T$$

Where (r_i, g_i, b_i) are the RGB component values of i^{th} pixel in the k^{th} image. A texture model can be built:

$$g \approx \overline{g} + \Phi_g b_g \tag{4.7}$$

Where $\overline{g}$ is the mean texture map, $\Phi_g = (\Phi_1, \Phi_2, ..., \Phi_t)$ is the matrix of the k largest eigenvectors depicting $p\%$ of the shape variation, and b_g is the matrix of the eigenvalues of the texture variation.

In order to build a combined AAM model, we can concatenate the shape and shape normalised texture into a single model. W_s is a matrix that needs to be introduced to be appropriate to expand the difference in units between the shape and texture models so that they can be represented in the same coordinate space (for further details, see Cootes et al. [55]).

$$b = \begin{pmatrix} W_s b_s \\ b_s \end{pmatrix} = \begin{pmatrix} W_s \Phi_s^T (x - \overline{x}) \\ \Phi_g^T (g - \overline{g}) \end{pmatrix} \tag{4.8}$$

PCA is applied to b, then we can obtain $b = Q_c$. Where c is the vector of appearance parameters that describe both the shape and texture variations of the combined model. Q contains the first t eigenvectors that describe both shape and texture variations, and may be separated into two matrices Q_s and Q_g:

$$Q = \begin{pmatrix} Q_s \\ Q_g \end{pmatrix} \tag{4.9}$$

Thus, s and g can be written as a function of c using equation

$$x = \overline{x} + \Phi_s W^{-1} Q_s c \tag{4.10}$$

$$g = \overline{g} + \Phi_g Q_g c \tag{4.11}$$

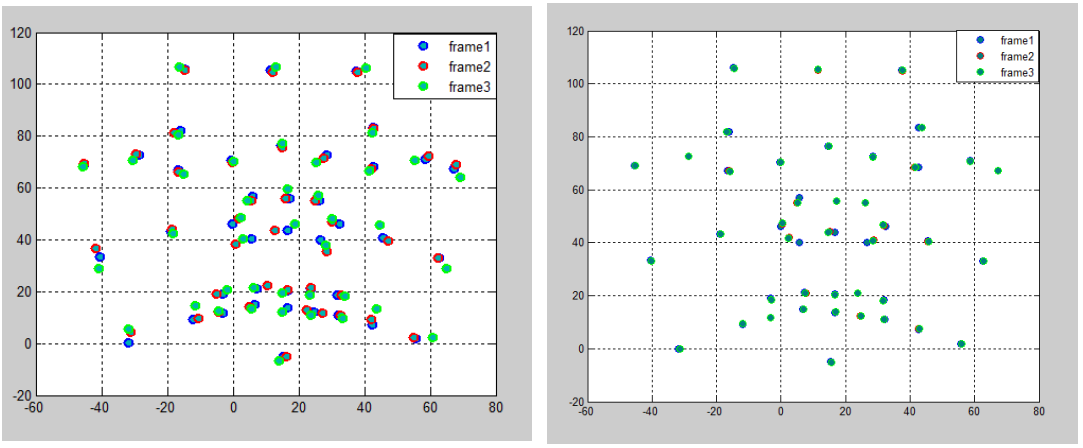
For example, we carried out the first experiment on a $1,908$ frames video sequence of a subject performing a smile over 4 years. An AAM was built from these data, which allows a decomposition of the complex facial action sequence into several modes of variation. Figure 4.3 shows some faces synthesised by varying the first four modes of variation of 3D ASM built from a training set of $N = 1,908$ face images, depicting a subject smile over 4 years. Inverting Equation 4.9 and equation 4.10 and applying them (respectively) to shapes and texture maps, the appearance parameters c can be computed corresponding to each frame.

4.1.3 Implementation of an ASM

In 3D, the first step in building an ASM is to register the shapes to eliminate the variations due to rigid movements of the head, such as translation rotation, and scaling. This can be performed by applying the procrustes analysis as follows [24]:

1. Translate each face so that its centre of gravity is at the origin.
2. Choose one face as an initial estimate of the mean shape and the scale where $s = 1$.
3. Record the first estimate as s_0 to define the default reference frame.
4. Align all the shapes with the current estimate of the mean shape.
5. Re-estimate the mean from the aligned shapes.
6. Apply constraints on the current estimate of the mean by aligning it with s_0 and scaling where $s = 1$.
7. If not converged, return to 4. (Convergence is achieved if the estimate of the mean does not change significantly after an iteration).

Figure 4.1 and Figure 4.2 show examples of the shapes taken from the original training set, which are not aligned, and the corresponding shapes after alignment using the Procrustes Analysis. Figure 4.3 shows some face smiles synthesised by varying the



(a) Faces with different poses

(b) After Alignment

Figure 4.1: The figure shows the position of the landmark (x,y) and the shapes with different poses (a), and after alignment with Procrustes (b).

first four modes of variation of a 3D ASM built from a training set of face shapes, depicting a subject smiling.

As an example, let us consider a 1,882 frames video sequences $S_k, k = 1, ...1,882$ (corresponding texture maps $g_k, k = 1, ..., 1,882$) of a subject smiling. An AAM is built from this data, which allows a decomposition of the complex facial poses corresponding to the variations of the first four modes of variation. Frames from a neutral-peak-neutral smile expression were used to build an AAM model. Figure 4.4 shows the AAM of intra-subject variations of texture of the first four modes. The individual images in Figure 4.5 show the deviations from the mean of the AAM model. The intra-subject variations can be observed in Figure 4.5 where smiles of the same subject have been recorded over nine years. Standard deviations of the range -3 to +3 are applied as this is a commonly used range for observing realistic deviations from the mean. Figures 4.6, 4.7 and 4.8 show the dynamics of the first three modes of variation of a smile expression. Tests conducted on $1,882$ frames show that 98% of the shape variations

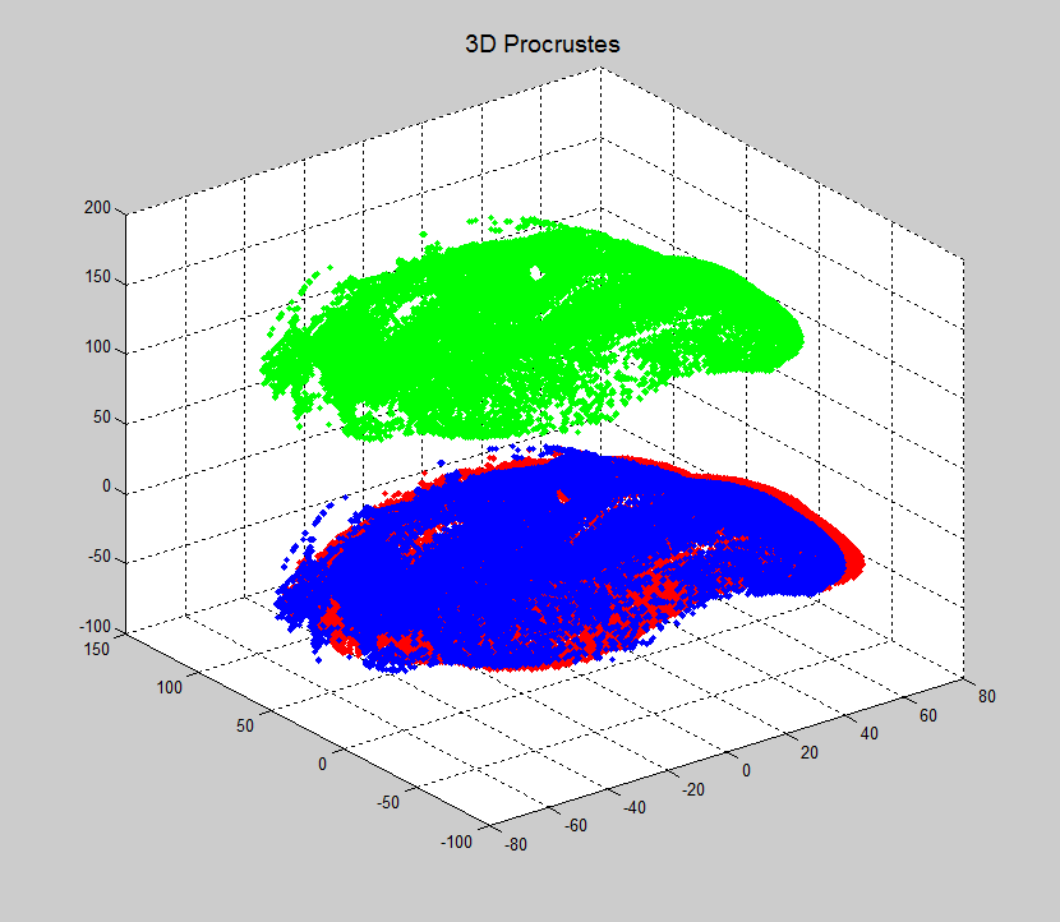


Figure 4.2: 3D Procrustes, reference (red), original (green), registered (blue).

are contained in the first five modes of variation. The dynamics exhibit familiar patterns, e.g., the same trends of the onsets and offsets with similar duration and signal magnitudes.

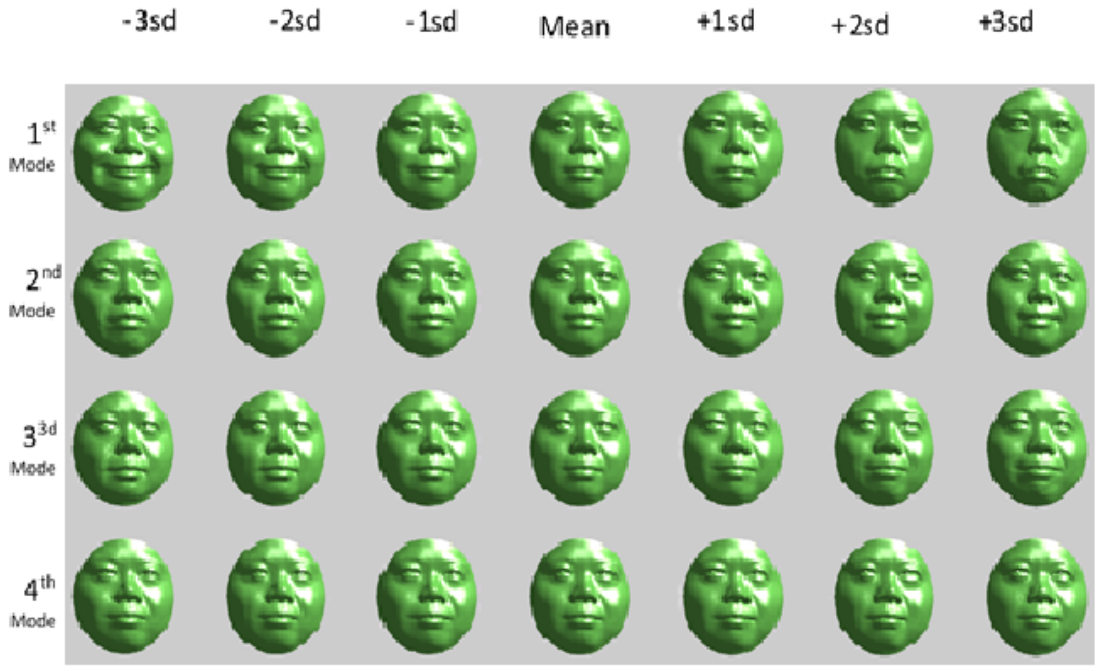


Figure 4.3: ASM model: -3 to +3 standard deviation for first four modes.

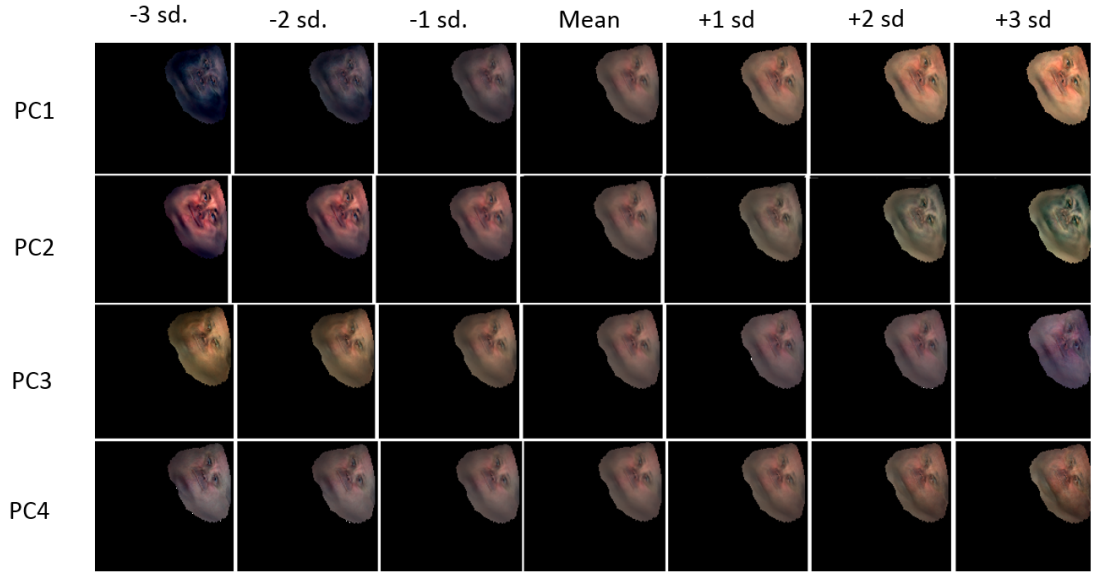


Figure 4.4: AAM model: Variation of texture, 3 to +3 standard deviation for first four modes..

Figure 4.5: AAM model: Variation shape with texture, -3 to +3 standard deviation for mode 1..

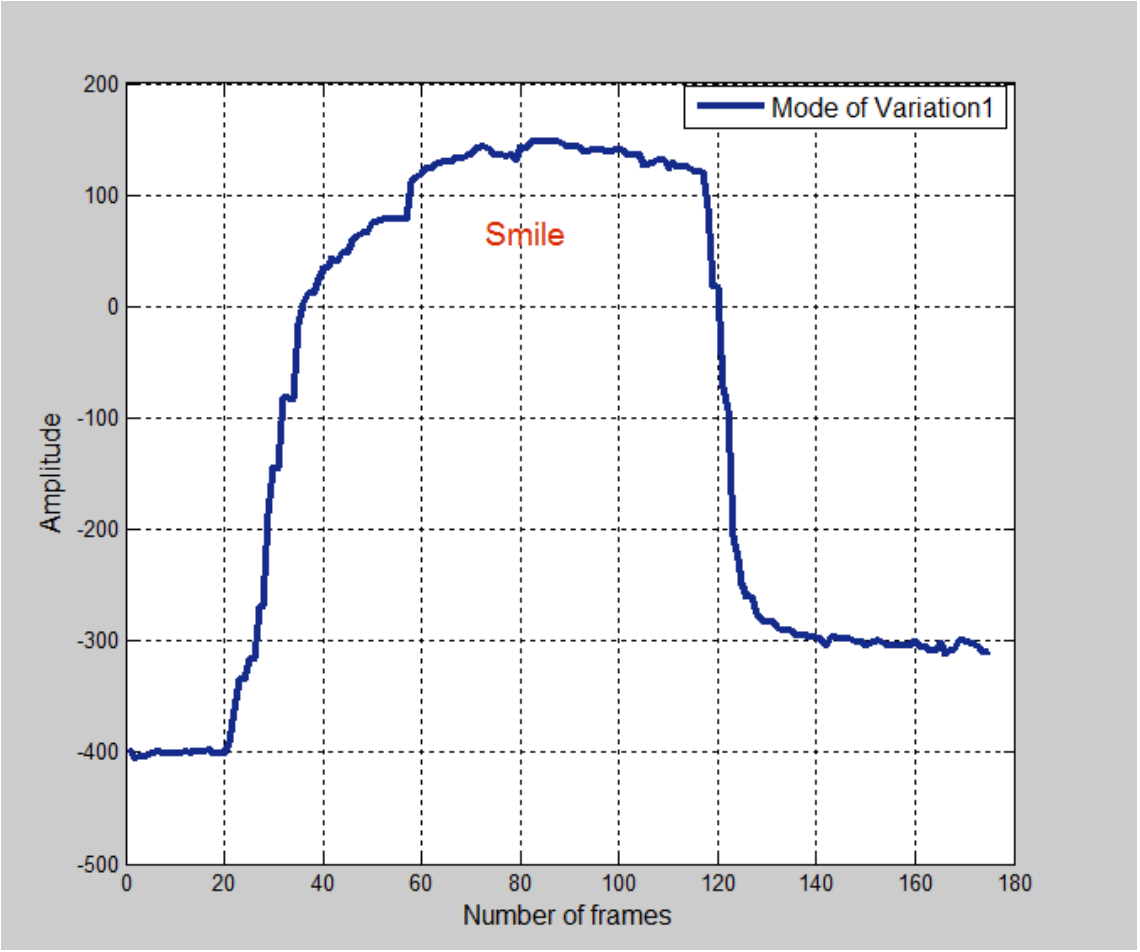


Figure 4.6: Temporal variations of the first mode for a subject smiling

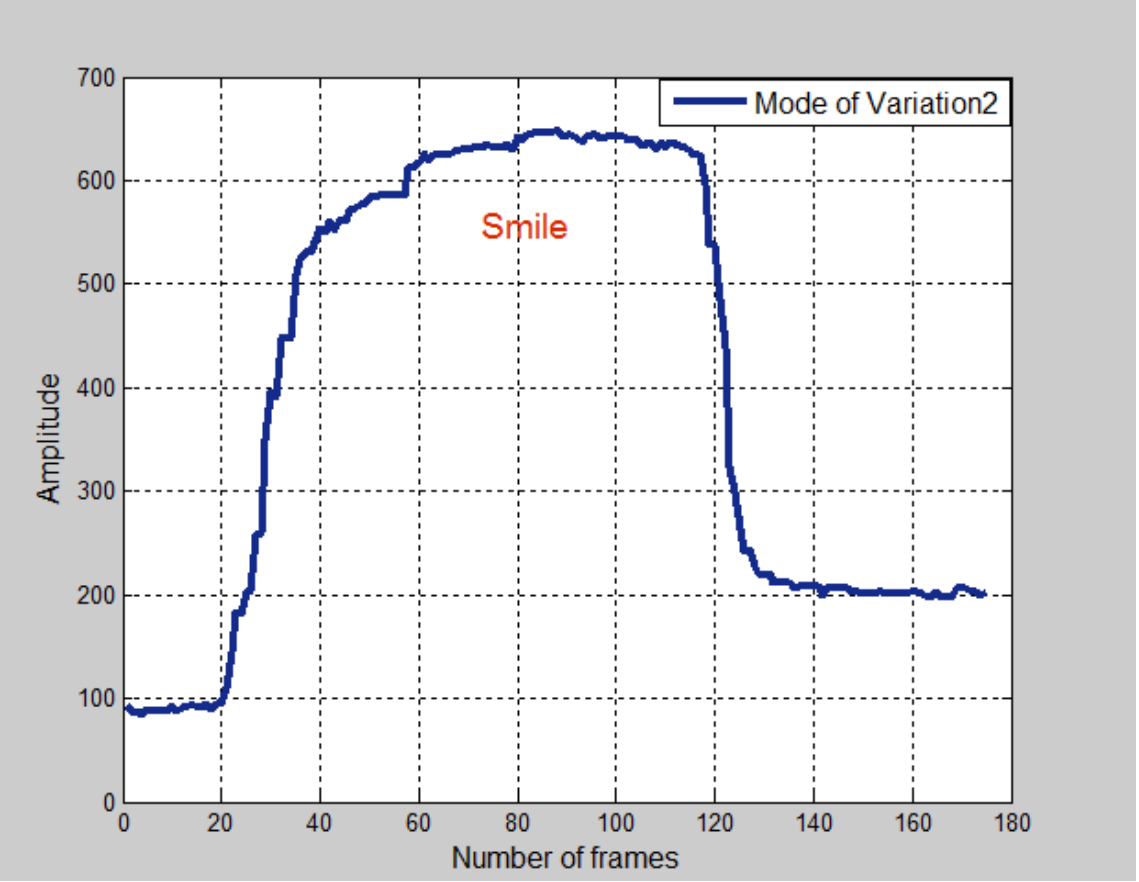


Figure 4.7: Temporal variations of the second mode for a subject smiling

4.2 Modelling Age

Recognising that age is a confounding influence in modelling human faces, ageing effects need to be examined in greater depth to gain an understanding of how age might be more effectively modelled. Hutton [116], reported that the effects of age are important to various specific applications in many domains, such as medicine, where many genetic conditions exhibit symptoms that alter with age, or as facial characteristics develop over time. During mid-life, bone structure varies little, but cartilage growth continues, particularly in men, most significantly altering the shape of the nose. In addition, in later life, reductions in muscle tone and skin elasticity dramatically alter the outer shape of the face. Modelling these intricate variations necessitates a significant

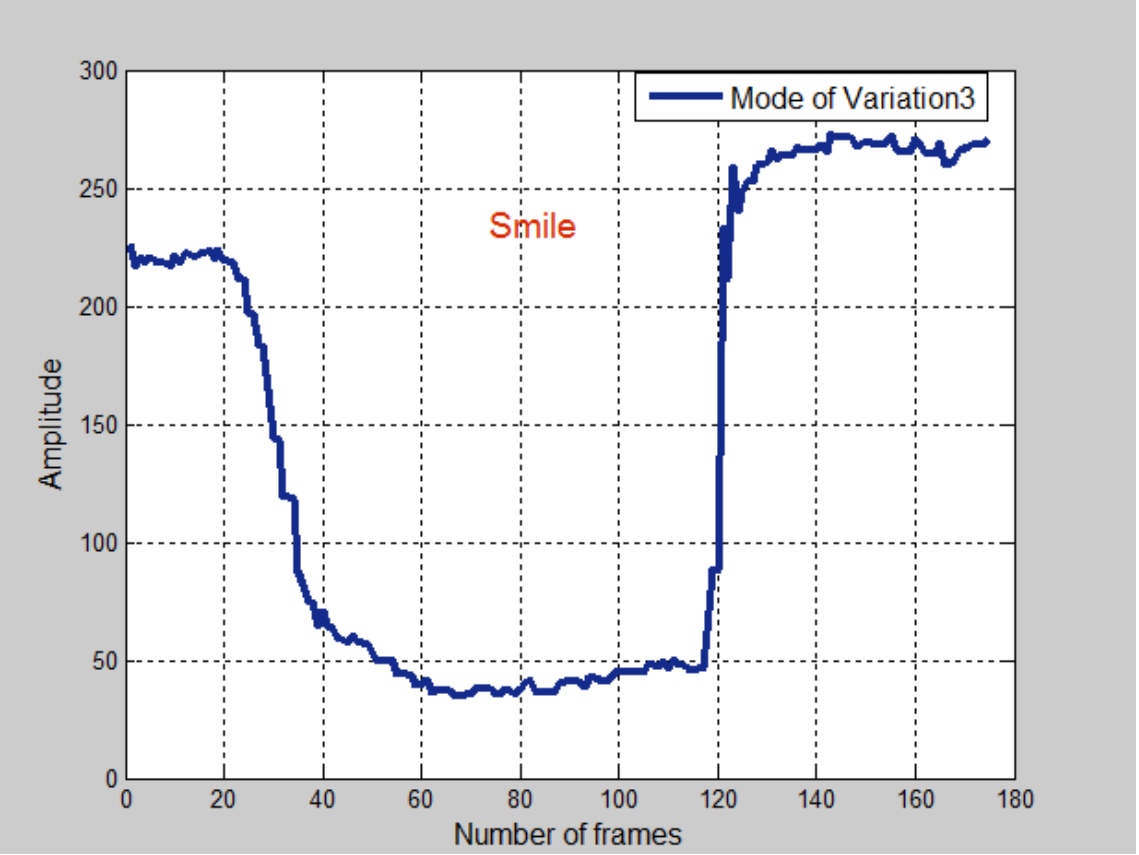


Figure 4.8: Temporal variations of the third mode for a subject smiling

quantity of raw input data and more advanced models.Using longitudinal data would be ideal by conducting 3D scans of the same subjects at various ages throughout their lives. Their ageing path could then be plotted for each subject through the shape-space correlated with the ageing process their *ageing trajectory*. However, 3D scanning technology has only become widely available in recent years, and therefore, such longitudinal data is only now becoming readily available. In addition, even while such data is available, many subjects are still required so that a robust estimate of general growth patterns could be obtained, taking individual differences into consideration. Longitudinal studies of facial ageing employing retrospective photographs have been conducted previously in [138] with ten time samples for each individual; their trajectory through the parameter space of an appearance model was modelled as a linear, quadratic, or cubic curve. They suggested that ageing could be simulated for a new example by

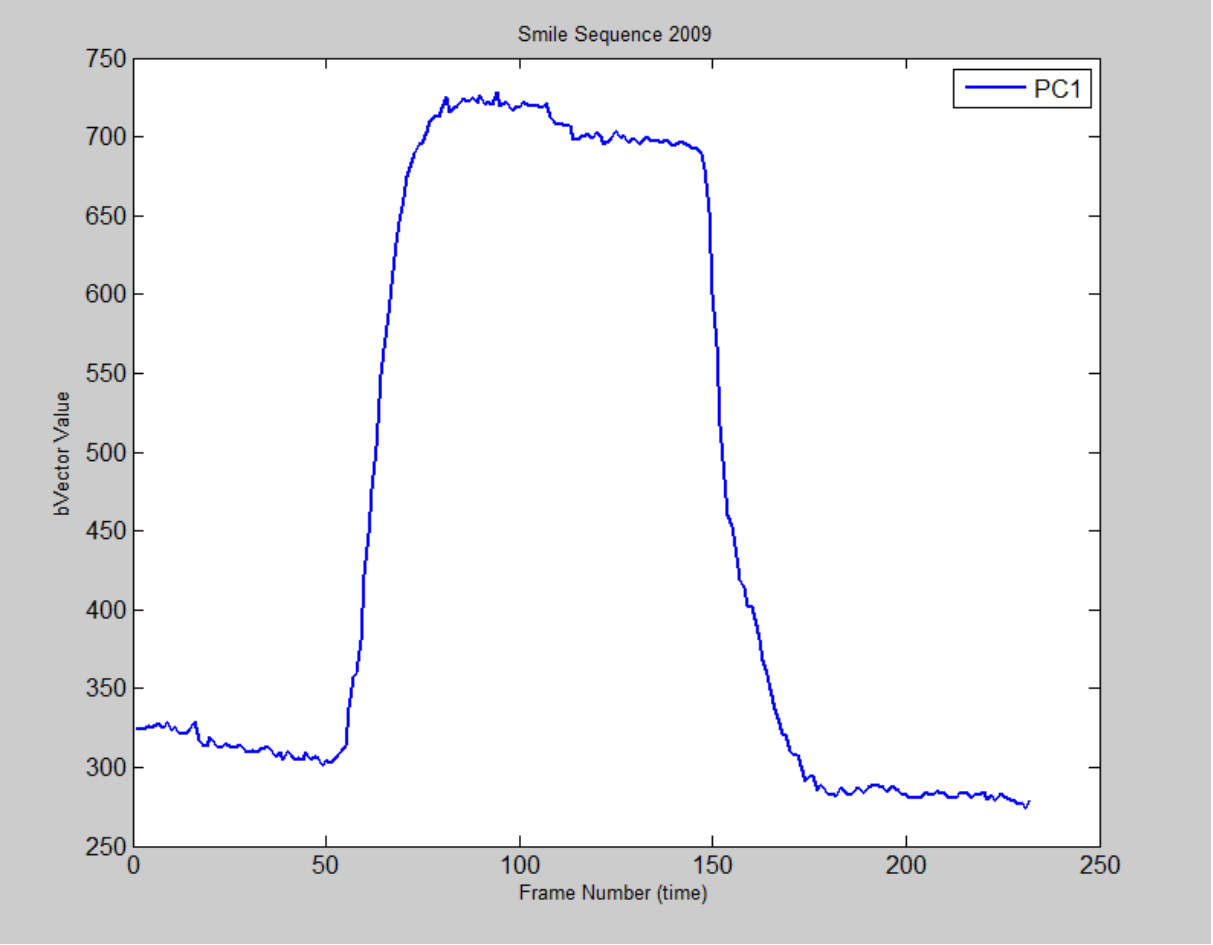


Figure 4.9: Dynamics of the smile sequence of subject over 2009; the first mode of variation (PC1).

using a weighted combination of the ageing trajectories for the subjects in the dataset, based on appearance or lifestyle similarities. It was shown that a subject's age could be used to improve the performance of face recognition systems. In addition, several longitudinal studies on landmark positioning have utilised morphometric techniques, including work on faces from laser scans by [59, 149, 155, 203]. A dense correspondence to examine the growth of the human mandible has also been reported by Andresen et al. [13]. This investigation demonstrated that growth of all subjects followed a linear relationship. However, a major problem affecting nearly all previous studies has been

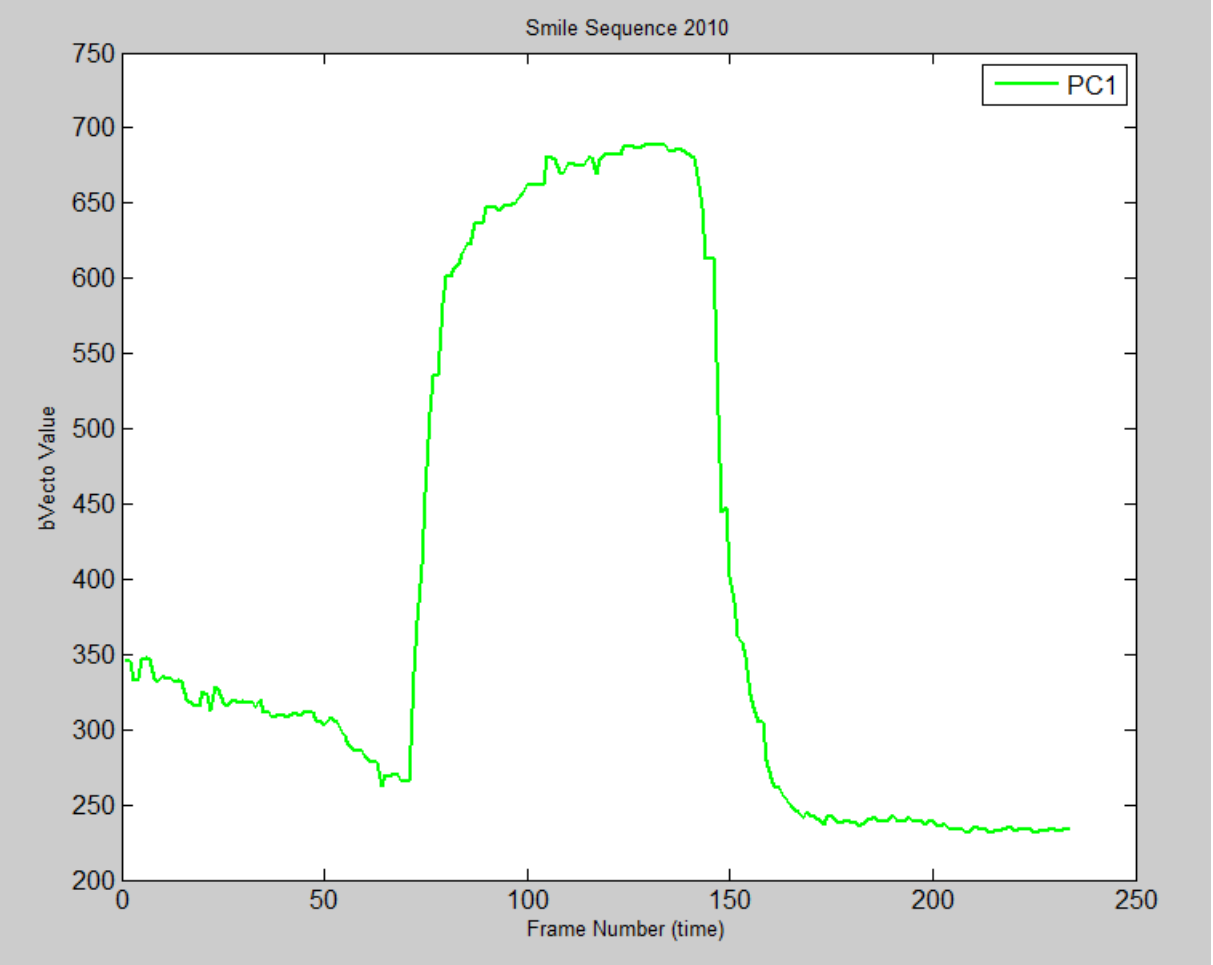


Figure 4.10: Dynamics of the smile sequence of subject over 2010; the first mode of variation (PC1).

the paucity of available longitudinal information. Growth analysis would then become a technique of identifying models for ageing that conformed well with the data, despite the incorporation of confounding factors, such as gender, lifestyle, ethnic background, diet, and genotype. Statistical analytical methodologies for inferring underlying trends from within datasets are generally referred to as "regression" techniques. A variety of polynomial regression techniques have been reported by Lanitis et al. [138], from linear to cubic. However, modelling more complex curves in this way requires high numbers of data points in order to obtain robust estimates.

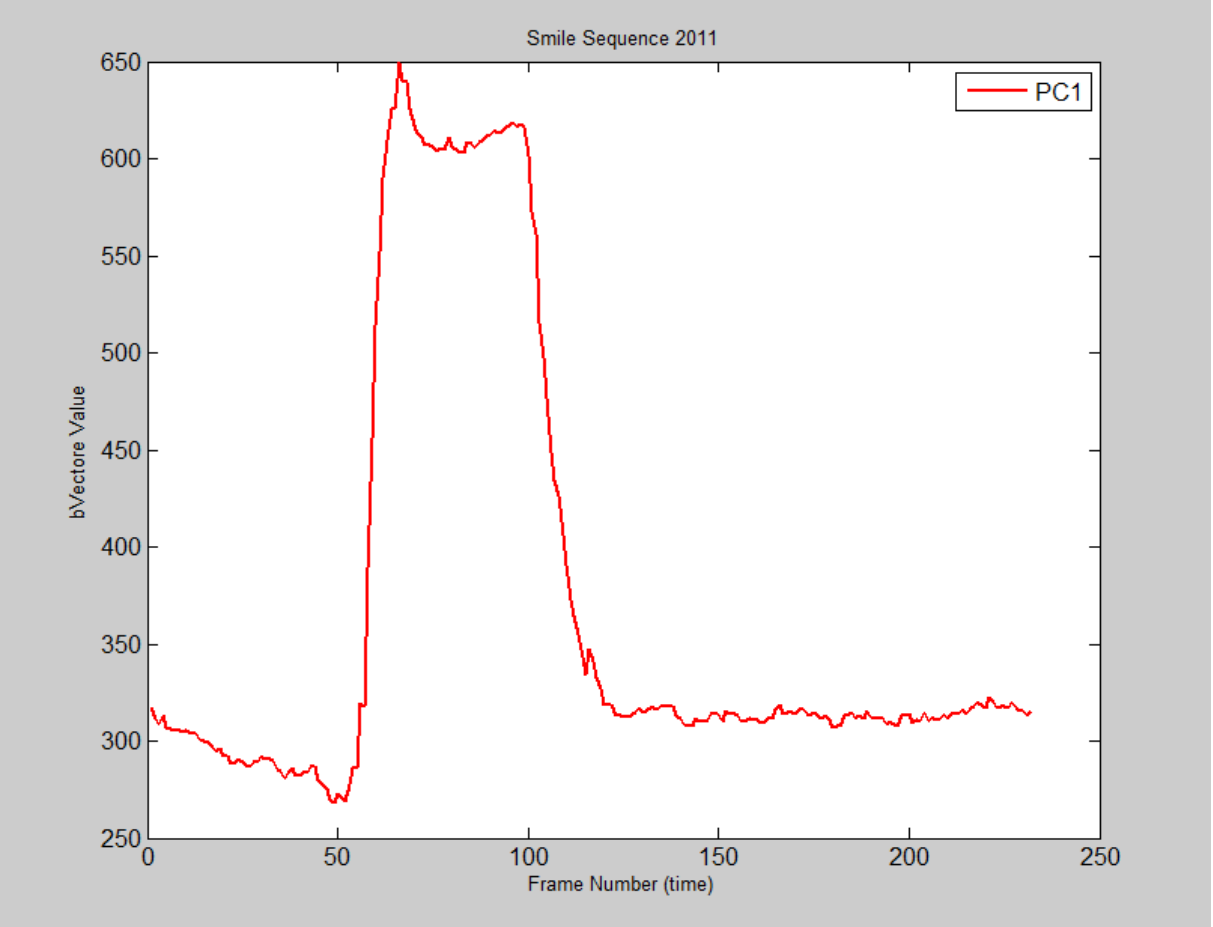


Figure 4.11: Dynamics of the smile sequence of subject over 2011; the first mode of variation (PC1).

4.2.1 Expression Sequence Modelling

Ageing data included dynamic facial expressions. These expressions can differ greatly in intensity and length, and their variations may be encoded different. Since our interest is to build statistical models of facial dynamics expression as people age, the sequences will need to be comparable while maintaining their expression characteristics. As first described in Section 4.1, by projecting a 3D frame into the AAM, we obtained parameters that described the 3D frame in relation to the AAM. These values are known as b vectors and are the Principal Component (PC) weights for the projected frame. Fig-

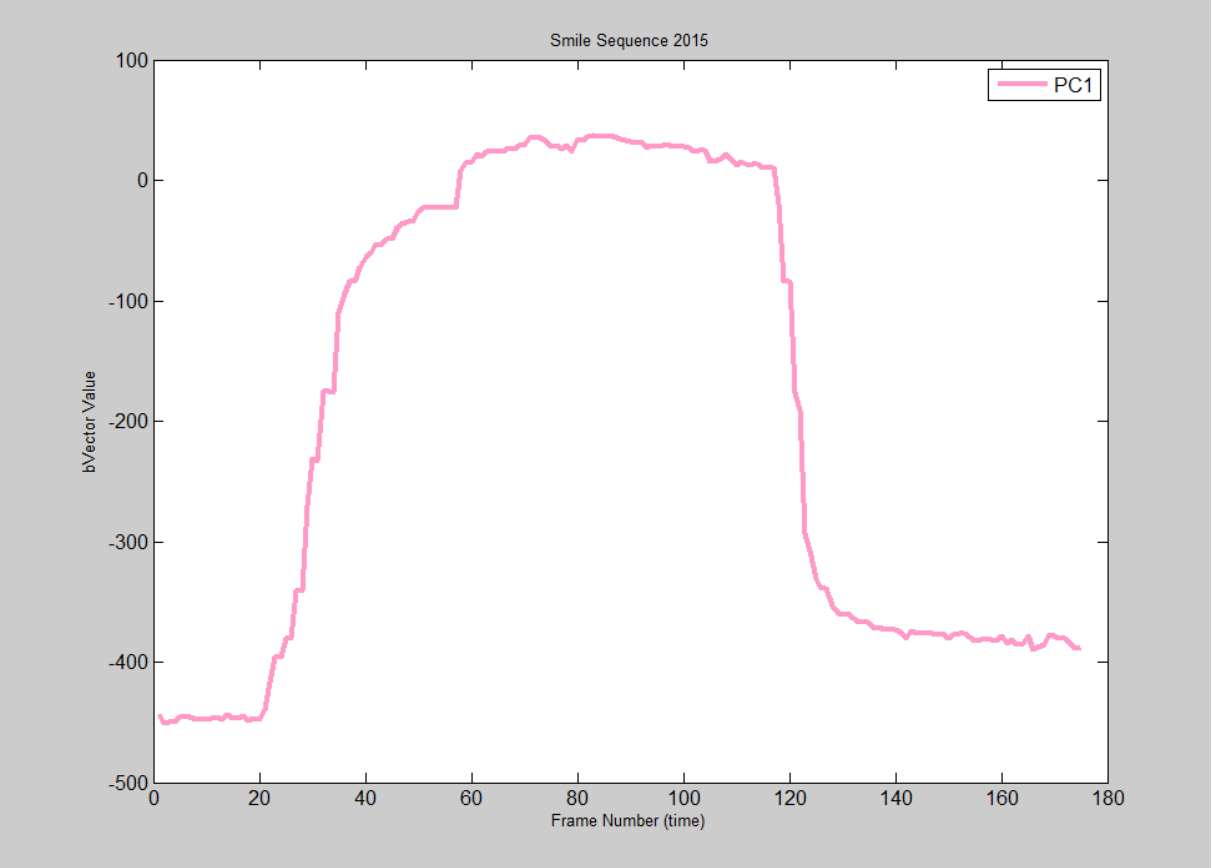


Figure 4.12: Dynamics of the smile sequence of subject over 2015; the first mode of variation (PC1).

ure 4.6 shows the b vector values for the first three PCs for each frame in a sequence. In this example, the sequence is a controlled smile expression (neutral-peak-neutral).

4.2.2 Similarity measuring for dynamic signals

Faces are represented as discrete points in a low-dimensional subspace in a static face recognition. The distance between the points is measured to determine whether two faces belong to the same individual, using Euclidean distance or the cosine of the projection angles[197]. Facial performance in dynamic expressions is represented by trajectories in a low-dimensional subspace. Measuring the similarity between two mul-

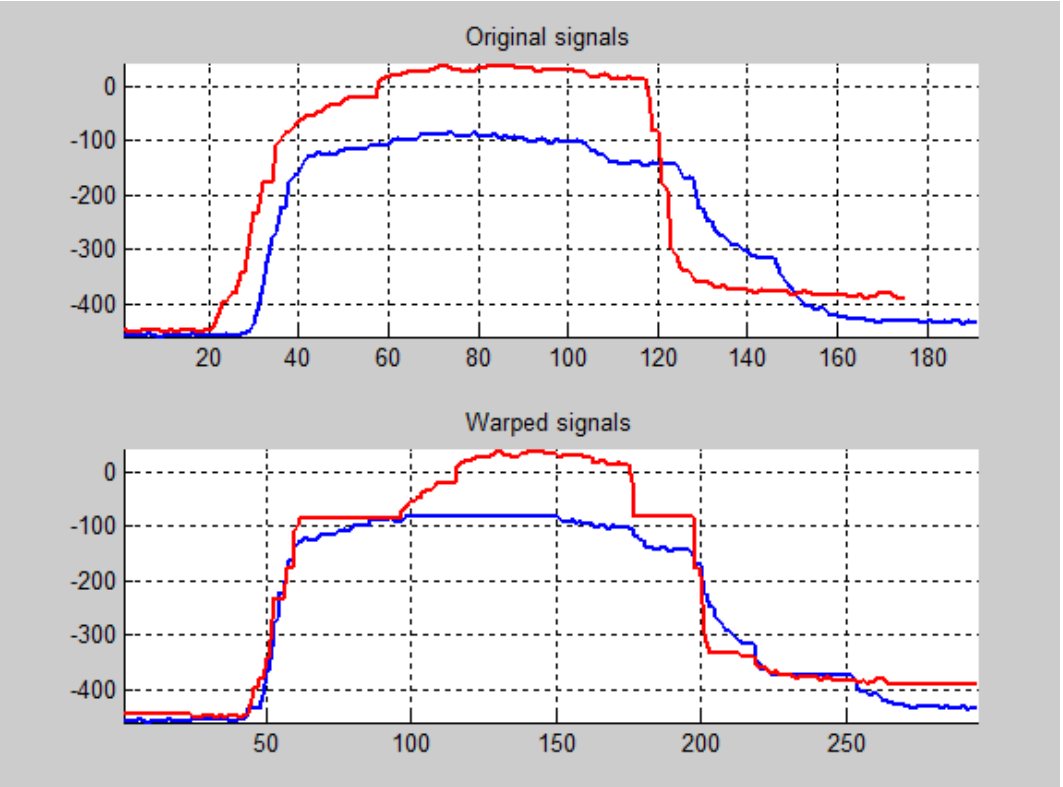


Figure 4.13: The two smiles before and after alignment using DTW; x axis represents the number of frames and y represents the amplitude of smile.

tidimensional time series of different lengths is challenging and presents several problems. For example, Figure 4.6 shows the temporal variations of the first mode of variation depicting repetitions of a subject's smile over a nine year period, and shows these nine trajectories in the space distribution by the first mode of variation. Modelling temporal facial information requires a model that can handle temporal information. In this study, we considered a Dynamic Time Warping (DTW), because temporal facial information can be represented as a sequence of symbols. DTW has the advantage that it accepts sequences with variable lengths, which is useful because the videos are of varying lengths, as shown in Figures 4.9, 4.10, 4.11, and 4.12. DTW is a robust method for measuring the similarity between sequences that vary in time or speed. To accommodate such variations, DTW computes the optimal alignment between two time series by nonlinearly warping one signal to the other, stretching or shrinking the

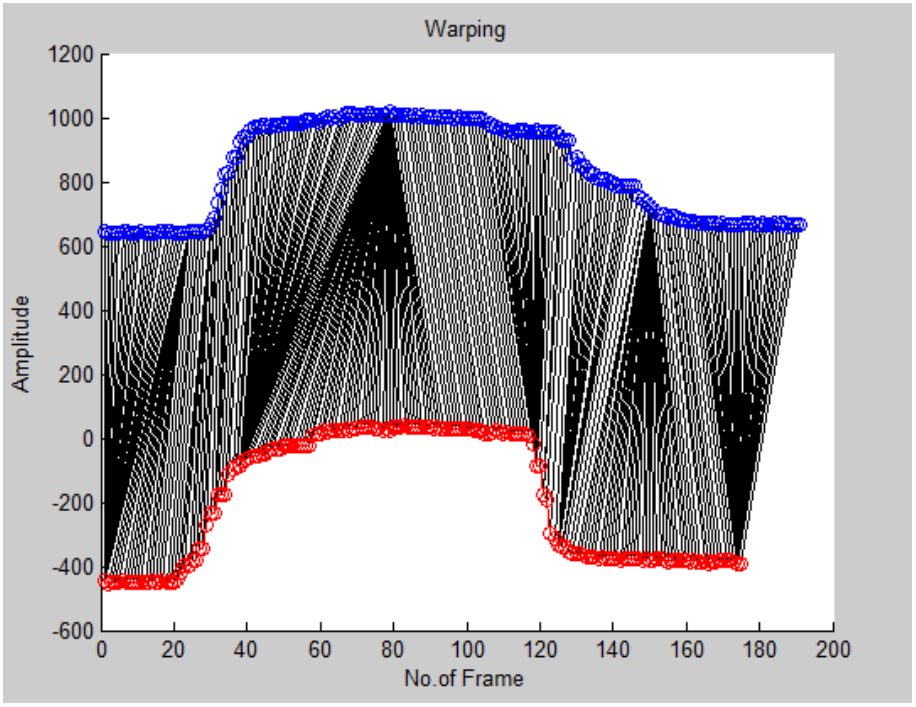


Figure 4.14: The Weighted Derivative Dynamic Time Warping of two smiles

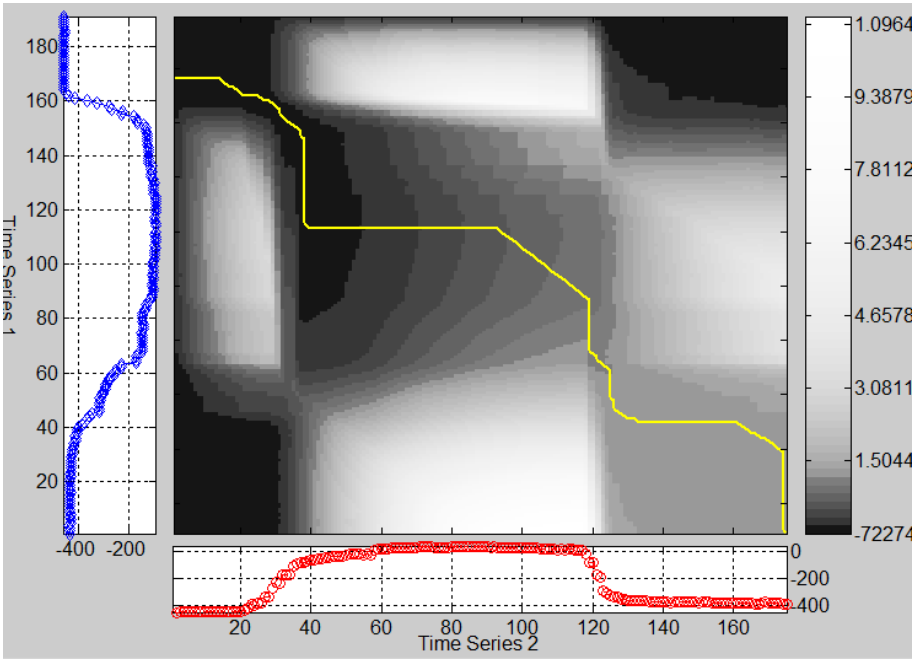


Figure 4.15: The distance matrix showing the optimal path along diagonal

signal along its time axis. Figures 4.13, 4.14, and 4.15 show the alignment between sequences that vary in time using DTW method and the distance matrix showing the optimal path respectively.

4.3 Experiments

In order to evaluate the tracking, inter-subject registration, and modelling used in this study, two experiments were devised. These experiments were conducted using data from the Cardiff longitudinal 3D face database, which is the main data used in this work and which is described in detail in Chapter 3. The first involved is generic and person-specific ASMs as described in the next subsection. In the second experiment, a linear regression model was used to find the best fit for the ageing model based on dynamic features. The details for each experiment will be described in the next subsections.

4.3.1 Generic and Person-Specific ASMs

We performed this experiment to validate the Cardiff longitudinal 3D face database for creating a 4D, highly realistic model of facial ageing using a training set of 3D dynamic face scans of smile expressions to train a statistical ageing model. An AAM was built from these data, which allows a decomposition of the smile sequence into several modes of variations. Figure 4.16 shows the facial poses corresponding to the variations of the first five modes of variations. For example, the first mode of variations that depicts the dynamics of a smile (neutral-smile-neutral). AAMs can be built by a variety of methods; for example, an AAM might be required to model the variation in the appearance of a single person over pose, illumination, and expression. Person-specific AAMs may find an application in interactive user interfaces requiring an estimation of head pose, gaze, or expression. On the other hand, an AAM might be constructed in an attempt to depict any face, including unseen subjects not in the

Table 4.1: Number of eigenvectors required to retain a given percentage of the variations of smile dynamics.

Number of Eigenvectors	Eigen-energy	Cumulative Energy	Variation	Cumulative Variation
PC1	1420.209	1420.209	**88.005**	88.005
PC2	107.009	2840.418	6.631	94.6367
PC3	35.589	2947.428	2.205	96.842
PC4	15.728	2983.018	0.974	97.816
PC5	13.134	2998.746	0.813	98.630
PC6	11.647	3011.880	0.721	99.352
PC7	10.450	3023.528	0.647	100

training set. The most obvious use of such a generic AAM is face recognition. As previously mentioned, evidence indicates that person-specific AAMs have been shown to be significantly more effective than generic AAMs. This effectiveness of an AAM depends on the ability of the AAM to generate 3D faces in the class under consideration of modelling and the success of the AAM in fitting a new input face. AAMs comprise a shape element and an appearance element.

This study so far has given consideration only to the configuration where a single AAM is constructed for the entire dataset in terms of ageing. However, we construct person-specific AAMs to investigate how facial dynamics change over years. We empirically compared generic and person-specific AAMs. The results showed that building a generic shape model, as shown in Figure 4.16, is comparatively easy, whereas building a generic appearance model requires far more training data, as shown in Figure 4.17. The merging of different faces into a single model for building an ageing model enables the traditional AAM to capture and encode not just the behavioural data, but also the physiological anomalies between different individuals. For example, if one person has

a long nose whilst another possesses a large forehead, this increased amount of data requires encoding via a larger number of modes of variation. Alternatively, a person-specific AAM will only encode the behavioural data. For example, a common AAM for 40 sequences from different subjects performing a smile expression over time typically needs a high number of modes of eigenvectors to preserve 98% of the variations, whilst a person-specific AAM requires just 4 to 5 eigenvectors (see Figures 4.3, 4.4 and 4.5). The results in Table 4.1 show the number of eigenvectors required to retain a given percentage of the variations of generic AAM. Thus, the first column of the table represents the number of eigenvectors, the second column shows the energy of each eigenvector and the percentage of the variations of each eigenvector was described in the fourth column.

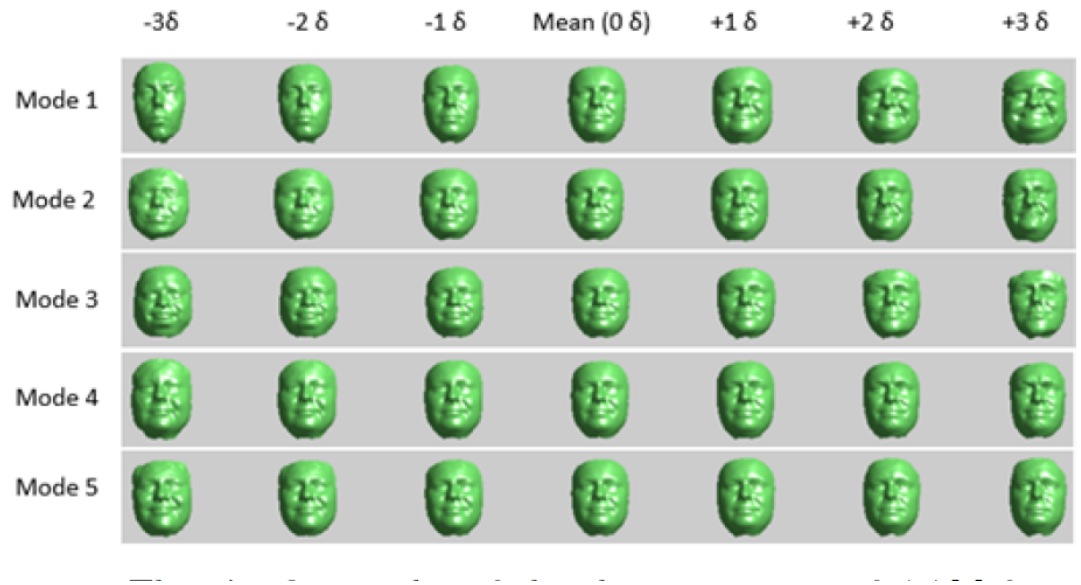


Figure 4.16: The first five modes of the shape variation of AAM for multi-subjects.

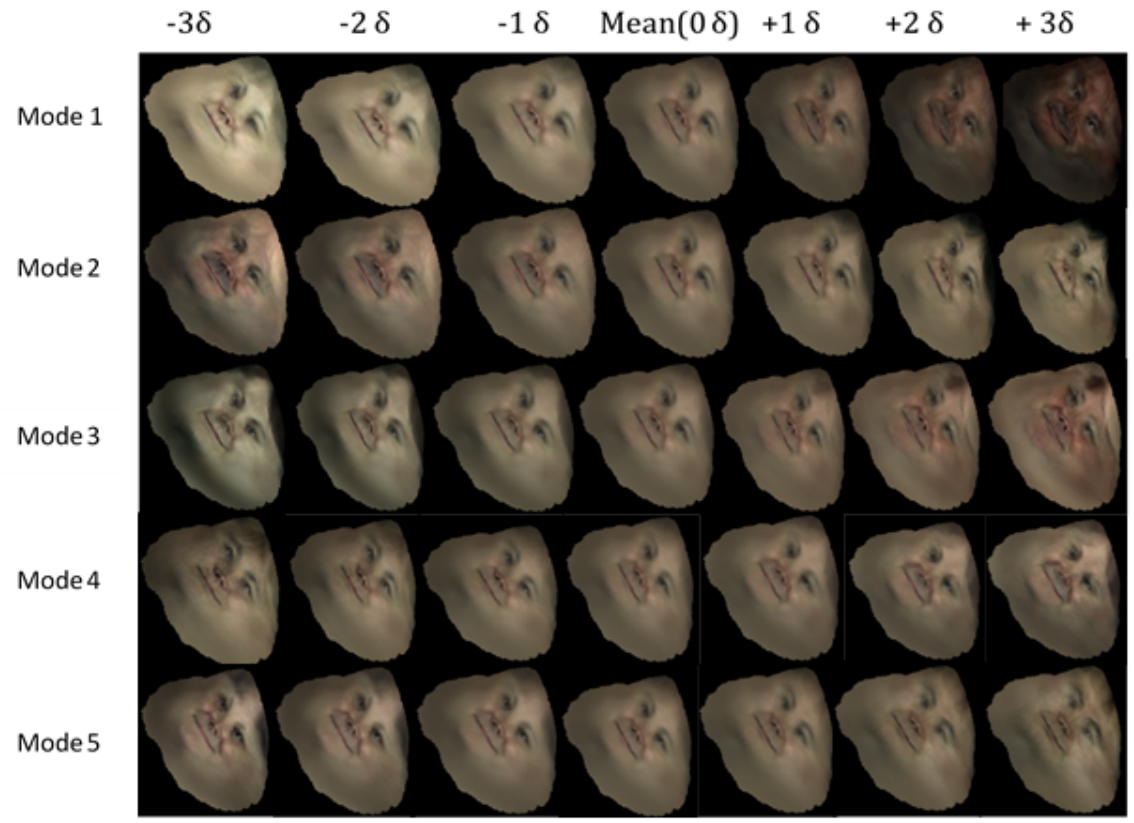


Figure 4.17: The first five modes of the texture variation of AAM for multi-subjects.

4.3.2 Determining the Relationship between Facial Dynamics and Age via Linear Regression

In this experiment, we model the relationship of age and facial dynamics to predict which dynamic features change due to ageing, and therefore, which features might be used to estimate the individual's age. Regression expresses the relationship in the form of an equation. In our case, we show the change in facial dynamic features (the predictor or x variable) with the age (the response or y variable). We want to estimate the underlying linear relationship so that we can predict the age from the invariant dynamic features. Regression can be used to find the equation of this line, which is

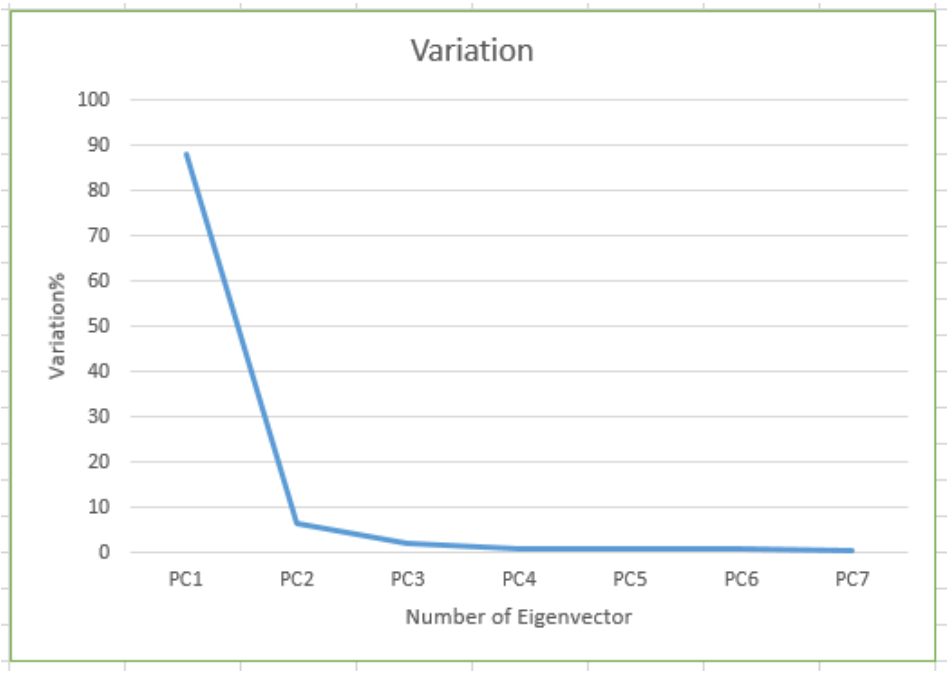


Figure 4.18: The distribution energy of the eigenvectors of the generic AAM

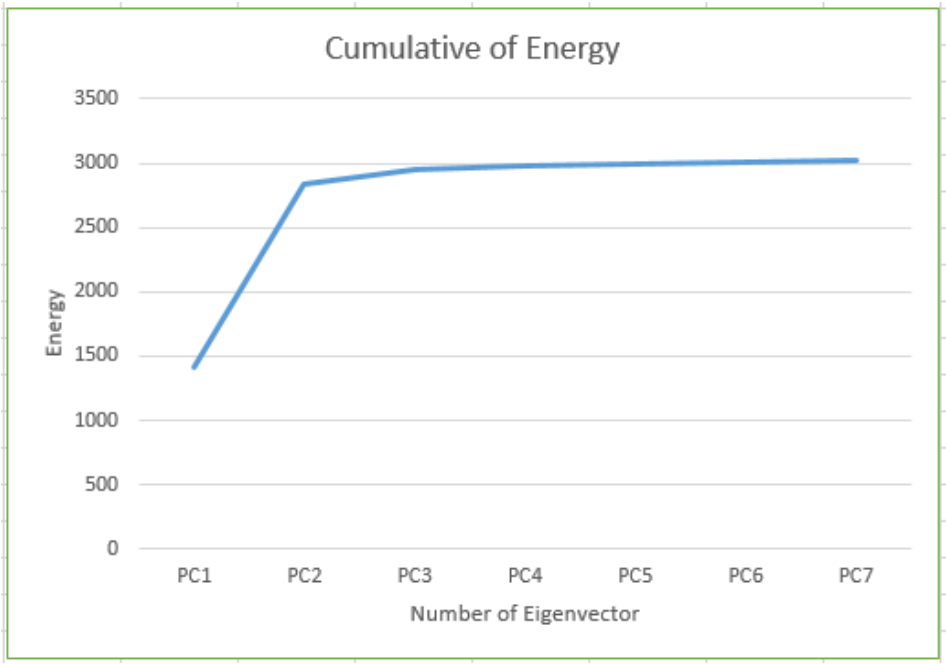


Figure 4.19: The cumulative energy of the Generic AAM

usually referred to as the regression line. In dynamic face recognition, facial performances are represented as trajectories in a low-dimensional subspace. However, the problem here is how to find the relationship between the two variables. For example, Figures 4.9, 4.10, and 4.9 show the temporal variations of the first mode of variation (PC1) depicting the dynamic of a subject performing a smile over a 3-year period. These values are known as b vectors and are the Principal Component (PC) weights for the projected frame. Smile amplitude is estimated as the trajectories computed from the PC1. A temporal segmentation of the smile is used to analyse the facial dynamics using the dynamic features of temporal phases (onset, apex, and offset) as used in Section 5.3 and Equation 5.1.

In our experiment, we extract a set of dynamic features from three phases of the PC1. Further details of the dynamic features can be seen in the next chapter.

There is no quantitative information on the progression of facial dynamics over the years, which might be because individuals age in a different manner under different conditions. However, in this work, we are trying to find the best possible modelling of the relationship between the variable dynamic features against age.

Because the values of the dynamic features change over different genders and age groups, it should be noted that a straight "best fit" line can be drawn across the points (Figure 4.20). In mathematics, or more precisely, in statistics, this whole operation is called a Linear Regression Model [23, 138], and the "best fit" line is called the Least Squares Regression Line [16]. The linear regression model was chosen because the linear regression method is useful as a primary tool for modelling data due to its effectiveness and completeness. Our finding demonstrates that the linear regression cannot model the dependence between the selected feature and the age as shown in Figure 4.20, since the black line shows the best fit across the points of dynamic features between different subjects. It is clear that the line differentiates the data into two age groups: young people (group 1: ages 15-30 years) and adults (group 2: ages 31-60 years). Undoubtedly, there is a variation between individuals in age groups, but there

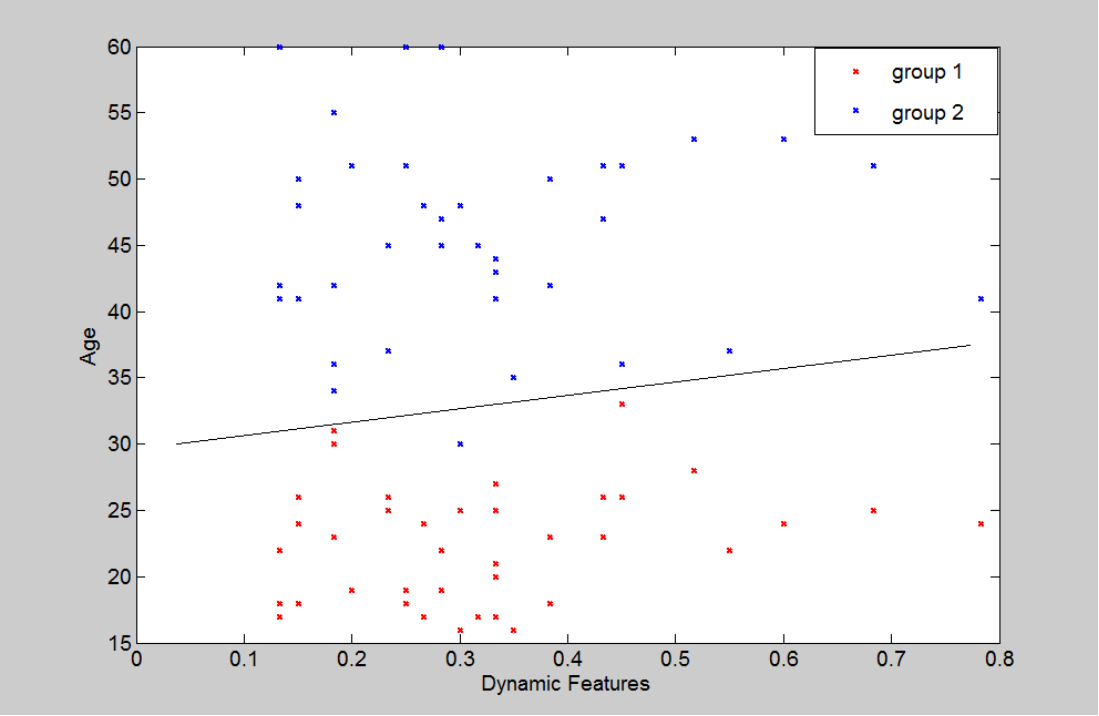


Figure 4.20: The regression line. The scatter diagram shows the dynamic features and two age groups for a representative sample of 80 subjects age 15-60.

are no significant differences using dynamic features. The result might be improved by including more age groups. Figure 4.20 shows a scatter plot of the example data.

Linear regression consists of finding the best-fitting straight line through the points, which is called a regression line. The black line in Figure 4.20 is the regression line, and it consists of the predicted score on y for each possible value of x. The distance from the points to the regression line represents the errors of prediction. As we can see, some points are very near the regression line; thus, its error of prediction is small. In contrast, the majority of points are much higher than the regression line, and therefore its error of prediction is large. A sample graph is presented in Figure 4.21, for person 'A', which shows the dynamic features progression over nine years. The different patterns (Asterisk, Fill Circle, Squares, Downward-triangle, Plus sign, Diamond, Upward-triangle, and Star) correspond to the dynamic features, which are displayed in the legend.

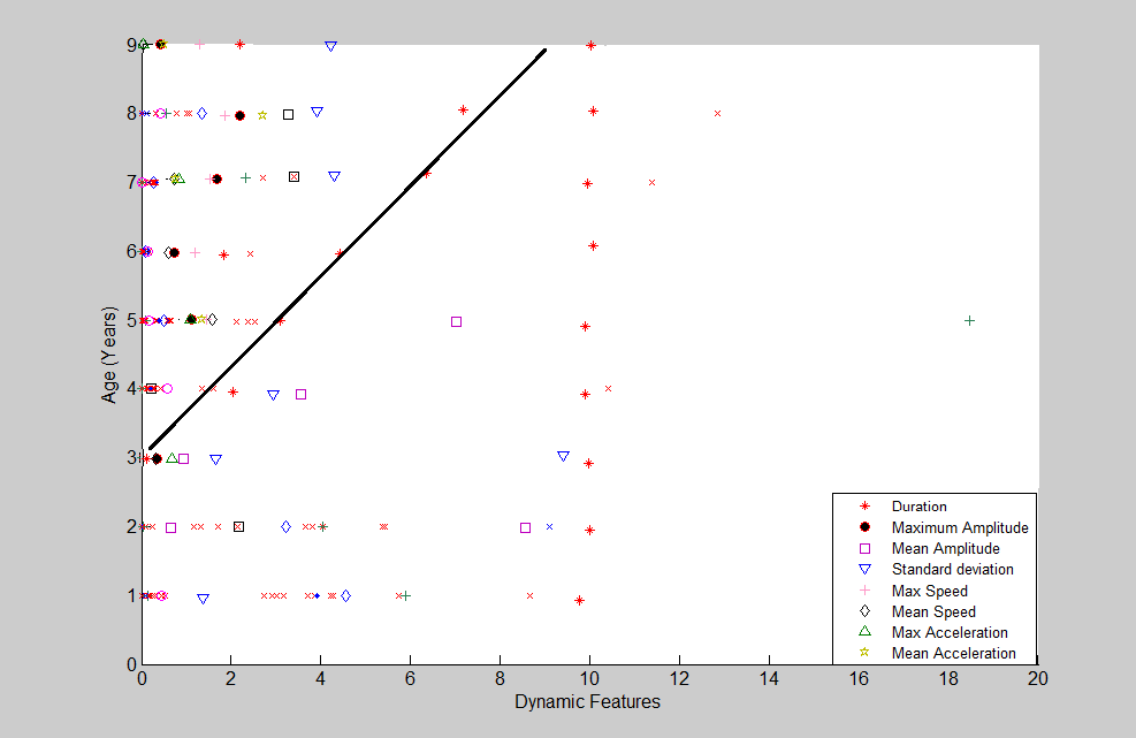


Figure 4.21: A sample graph presenting the change in facial dynamic features over nine years of person A and the "best fit" line is drawn across the points.

In this work, we start by assuming a simple linear model given by,

$$y = a + bx \tag{4.12}$$

Where a is the point of interception and b is the gradient of the line, and they can be found by the following:

$$a = \overline{y} - b\overline{x} \tag{4.13}$$

$$b = \frac{\sum(x - \overline{x})(y - \overline{y})}{\sum(x - \overline{x})^2} \tag{4.14}$$

If n is the number of the scattered points, then:

$$\overline{x} = \frac{\sum(x)}{n} \tag{4.15}$$

$$\overline{y} = \frac{\sum(y)}{n} \tag{4.16}$$

Figure 4.21 depicts the best fit line of the relationship drawn between the points of dynamic features of the smile of person 'A' over nine years and shows which of these dynamic features change with age. The visual results presented in Figure 4.21 demonstrate that there is a small change in some features over age due to the ages being close and there being no long span.

4.4 Summary

In this chapter, generic and specific ageing models were presented. We demonstrated the steps of modelling the facial dynamics and analyse these dynamics using an ASM and an AAM. The percentage of the variation of smile dynamics, the cumulative energy, and the distribution of the energy of the eigenvectors of the general model were computed. The dynamic features from the PC1 were extracted by projecting the data that is used to build the model into the model. The results of the experiments for the temporal alignment of pairs of videos from our data were provided; the selected pairs had the same expression. Finally, we attempted to find a relationship between two variables (dynamic features against age) by projecting a 3D frame into the AAM, and we obtained parameters that describe the 3D frame in relation to the AAM.

Chapter 5

The Dynamics of A Smile in Different Age Groups

5.1 Introduction

In this chapter, we present the dynamic features used to evaluate smiles in different age groups and detect gender differences. The effects of age and gender related facial changes are recorded using dynamic 3D facial scans. The source of temporal information that is investigated is lip movement in a closed mouth smile. In order to investigate the influence of temporal facial information on different age groups, data was collected and processed. To achieve this, experiments were carried out on the subjects, who were divided into two groups by age (15-30 years and 31-60 years). Each group was further subdivided by gender. Facial features were extracted and then automatically tracked through the duration of the data capture. The smile was automatically segmented into three phases: onset, apex, and offset. Subsequently, a set of dynamic features were computed from these facial features, as well as from the static features for comparison. A two-way multivariate analysis of variance (MANOVA) of these features demonstrated that statistically significant age and gender-related differences could be detected. We show that 3D facial dynamics provide more useful information than static features for the characterisation of smiles. In this thesis, we focus on dynamic facial behaviour analysis and in particular on the analysis of facial dynamics (smile expression). Facial expressions encoded in terms of smile activation are manifested by the

motion of individual facial parts or facial muscles.Therefore, facial dynamics can be modelled as temporally evolving deformations of local facial parts (e.g., the mouth in the case of a smile). It is widely argued that the temporal dynamics of a smile expression are typically described by the following temporal segments: neutral- there is no facial motion; onset- the facial motion starts until it reaches an apex; apex- the point of the strongest possible facial deformation. The person that displays the expression usually stands still for some moments; and offset- the reverse path from the apex to the relaxed neutral position (see Figure 5.1).

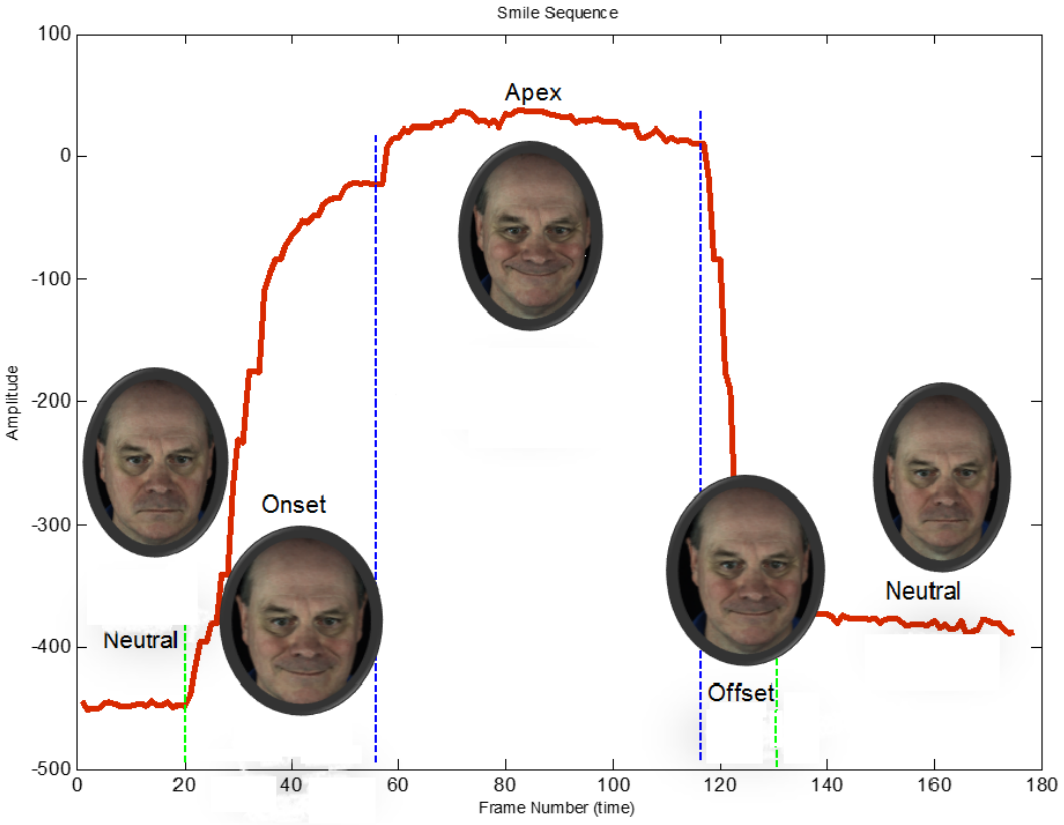


Figure 5.1: Ground truth of the temporal dynamics of smile

5.2 Facial Ageing Analysis

The human face conveys important information related to personal characteristics, including identity, gender, ethnicity, and age. Age can play an important role in many applications, such as age estimation in crime investigation, age-adaptive targeted marketing, and age-invariant person identification [87], and is also the main risk factor for many complex diseases [144]. As facial ageing is one of the most prominent and accessible phenotypes of human ageing, it is important for assessing the risks of age-related diseases and for designing individualised treatments [40]. Ageing is an inevitable process that leads to many soft-tissue changes, and this process particularly affects the lips, causing many changes, such as thinning and an increase in length [56, 75]. Since the ageing process can change the characteristics of the smile, it is critical to acquire knowledge of age-related facial changes to inform the above applications. However, to date, gender differences in age-related changes in the 3D dynamics of the smile have remained unexplored. A lack of suitable datasets in particular has proved to be the limiting factor. To address this issue we have conducted a study using 3D videos of closed mouth smiles for 80 subjects to determine whether ageing affects smile expressions.

5.3 Proposed Method of 3D Smile Analysis

One of the main contributions of this chapter is a method for the temporal segmentation of a smile to analyse the facial dynamics using the the dynamic features of temporal phases (e.g., neutral, onset, apex, and offset). In this section, details of the proposed method are summarised. Figure 5.2 shows the proposed method followed in this chapter.

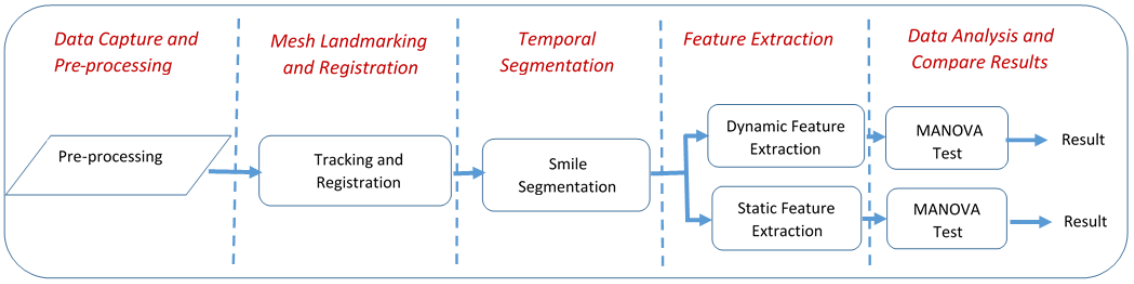


Figure 5.2: The proposed method for smile segmentation

5.3.1 Data Collection

We have recently processed a 3D smile video to analyse the dynamic of smiles. Data collection was carried out as a part of research in the 3D innovative research program of computer science at Cardiff University to analyse facial movement. This data was captured using the 3dMD capture system over nine years. Videos were recorded with a high resolution at a rate 60 frames per second. Full details can be found in Chapter 3. The database has 316 subjects, both male and female, and the age of subjects ranges from 4 to 61 years. For posed smiles, each subject was asked to pose a smile expression as realistically as possible after being shown in a picture the proper way to make a smiling face. A Closed mouth smile is formed from a neutral expression into a closed mouth smile and back to neutral. Figure 5.5 shows sample frames of this. In this work, we have used 80 subjects who performed closed mouth smiles; they were subdivided into two groups according to age, specifically, young people (Group 1: ages 15-30 years) and adults (Group 2: ages 31-60 years). Both groups contained 40 subjects split equally between males and females. The data underwent several pre-processing steps that were necessary to remove artefacts and improve the subsequent processing steps. These steps were performed to clean the 3D meshes and create a single-view texture map (UTM) (Section 3.5).

Figure 5.3: The Unified Texture Map (UTM) and cleaned mesh for 4 subjects.

5.3.2 Facial Tracking and Registration

The tracking approach we used requires the user to annotate a single mesh from the sequence, clicking on the ordered sequence of 41 facial feature points: eye corners, centre of upper eyelids, cheek centre, nose tip, lip corners, and contour of face. These points were then automatically tracked following the approach proposed by Vandeventer [201, 202], and they enabled us to subsequently analyse the facial dynamics as shown in Figure 5.4. In each frame of a sequence, the face should be aligned before feature extraction to eliminate any effects of head movement. This is carried out using Procrustes analysis to determine the rigid transformation.

5.3.3 Temporal Segmentation

Typically, dynamics are related to temporal segmentation, in which the speed and the duration of temporal phases (e.g., neutral, onset, apex, and offset) have to be analysed from the stream of the facial expression [215]. This allows a more detailed analysis of the feature dynamics to occur. Most of these features were originally proposed to analyse the smile expression [66], and a similar set has been employed for automatic kinship estimation through smile dynamics [64].

Figure 5.4: View of the 3D landmarked frame with annotated points

A smile can be defined as the upward movement of the lip corners, which corresponds to Action Unit 12 in the facial action coding system (FACS) [80]. Most facial expressions are composed of three non-overlapping temporal phases, namely, onset, apex, and offset. Onset is the initial phase of a facial expression, and it defines the duration from the neutral to the expressive state. The apex phase is the stable peak period (which may be short in duration) of the expression between onset and offset. Offset is the final phase from expressive to neutral state. Following the registration step, the onset, apex, and offset phases of the smile are detected using the approach proposed by Dibeklioğlu et al. [65]. Smile amplitude is estimated as the mean amplitude of the right and left lip corners, normalised by the length of the lip. Let $D_{lip}(t)$ be the value of the mean amplitude signal of the lip corners in frame t.

$$D_{lip}(t) = \frac{\rho\left(\frac{l_{18}^1+l_{24}^1}{2}, l_{18}^t\right) + \rho\left(\frac{l_{18}^1+l_{24}^1}{2}, l_{24}^t\right)}{2\rho\left(l_{18}^1, l_{24}^1\right)} \tag{5.1}$$

where l_i^t denotes the 3D location of the i^{th} point in frame t, and ρ is the Euclidean distance between two points. Temporal smoothing is applied to l_i^t using a robust local

regression smoothing method [47]. Then, the onset phase is determined by the longest continuous increase in D_{lip}. Similarly, the offset phase is detected as the longest continuous decrease in D_{lip}. The phase between the last frame of the onset and the first frame of the offset defines the apex. Smile amplitude is computed, and polynomials are fit to the resulting function for smoothing purposes. Subsequently, onset, apex, and offset are selected on the smoothed curve [65, 66, 67].

Segmentation of the smile is performed automatically by detecting the main positive and negative peaks in the first derivative of the smile amplitude. The first positive peak in the first derivative is selected as the first frame of the onset phase. Likewise, the last negative peak determines the last frame of the offset phase. The phase between the onset and offset contains the peak of the smile. It is determined by finding the end of the onset and the start of the offset. The former is estimated by starting at the first frame of the onset and then repeatedly moving to the following frame while the smile amplitude decreases until the first derivative drops below an experimentally determined threshold. The end of the peak smile phase is found in an analogous manner. Each phase is divided into increasing and decreasing segments to provide a more detailed analysis of the feature dynamics. See Figure 5.5 and Figure 5.6 for examples of automatic segmentation and smoothed smiles; the left blue line is the first frame of the smile onset, the section between the two red lines is the apex phase, and the right blue line delimits the last frame of the smile offset.

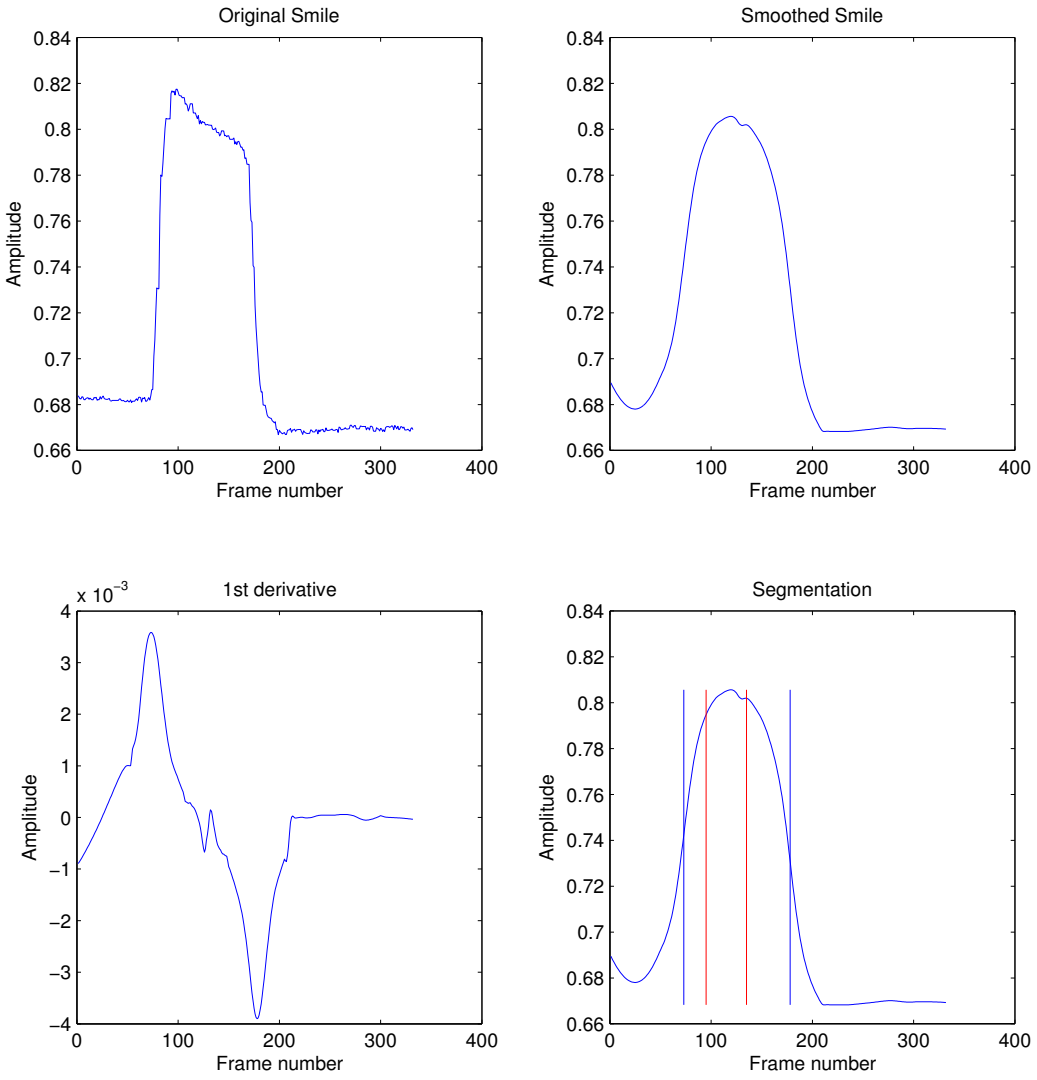


Figure 5.5: Temporal segmentation of smiles (a female subject)

5.3.4 Dynamic Feature Extraction

To provide more insight into the patterns of smile dynamics, a set of dynamic features were extracted from three phases of the mouth region. In fact, in our procedure, we used the same dynamic features that were adopted by Dibeklioğlu et al. [65]. The dynamic features extracted from the lip region were grouped by the temporal phases

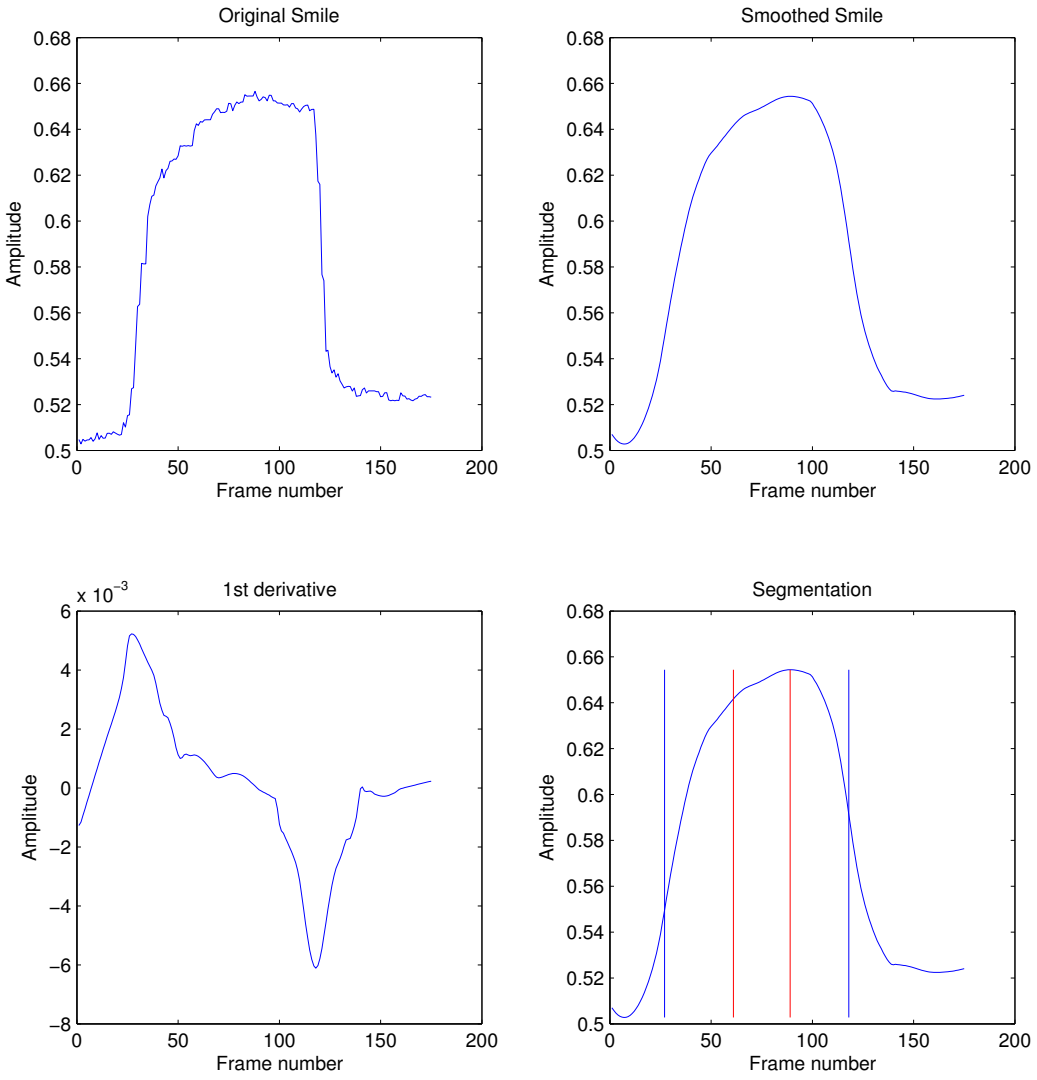


Figure 5.6: Temporal segmentation of smiles (a male subject)

(onset, apex, offset), as can be seen in Table 5.1. In addition to smile amplitude, speed V and acceleration A signals were extracted by computing the first and second derivatives of amplitude, respectively.

$$V(t) = \frac{dD}{dt} \tag{5.2}$$

$$A(t) = \frac{dV}{dt} \tag{5.3}$$

Dynamic Features	Description
Duration	$\left[\frac{n(D^+)}{\omega}, \frac{n(D^-)}{\omega}, \frac{n(D)}{\omega}\right]$
Maximum Amplitude	$max(D)$
Mean Amplitude	$\left[\frac{\sum(D)}{n(D)}, \frac{\sum(D^+)}{n(D^+)}, \frac{\sum(D^-)}{n(D^-)}\right]$
Standard deviation of Amplitude	$std(D)$
Max Speed	$\left[max(V^+), max(\mid V^- \mid)\right]$
Mean Speed	$\left[\frac{\sum(V^+)}{n(V^+)}, \frac{\sum(V^-)}{n(V^-)}\right]$
Max Acceleration	$\left[max(A^+), max(\mid A^- \mid)\right]$
Mean Acceleration	$\left[\frac{\sum(A^+)}{n(A^+)}, \frac{\sum(A^-)}{n(A^-)}\right]$

Table 5.1: Dynamic features used in the study

The features were extracted separately from each phase of the smile. In order to obtain a more detailed analysis of the feature dynamics, each phase was further divided into increasing and decreasing segments; symbols $(+)$ and $(-)$ are used to denote the type of segments. For example D^+ (respectively D^-) represents the increasing (respectively decreasing) segments in D, where D refers to amplitude signals. The number of frames is represented as n, and the frame rate of the video by ω.

5.3.5 Static Features Extraction

The medical literature describes studies to evaluate smiles in different age groups. Chetan et al. [42] considered the perioral zone at rest and smiling when analysing smiles in different age groups by using video records of these subjects. A cross-sectional study was performed to measure the characteristics within groups of different

Static Measurements	Description
1. Upper lip length	Distance measured between subnasale and stomion superius.
2. Upper lip thickness	Distance measured between labrale superius and stomion superius.
3. Outer intercommissural width	Distance measured between right and left outer lip corner.
4. Commissural height	Distance measured from horizontal line passing through subnasale to outer commissure.

Table 5.2: Static measurements used in the Chetan et al. [42]

Static Measurements	Male (*P value=0.05*)	Female (*P value=0.05*)
Intercommissural width	0.985	0.168
Commissural height	0.960	0.605

Table 5.3: Comparisons of static measurements between *Group 1 vs Group 2* within males and females.

ages and determine whether they were significantly affected by gender. The four linear measurements listed in Table 5.2 were computed from rest and smile photographs.

In our approach, we automatically extracted two of these measurements (3 and 4, see Table 5.2) as static features, and their values were computed for one rest and one smile frame and then the average was calculated. The percentage change from the corresponding distance in the rest position can be considered as a normalisation to avoid anatomical variation due to physical size differences between the individuals. The results obtained are shown in Table 5.3.

5.3.6 Geodesic Path and Distance

Finding shortest paths and shortest distances between points on a surface S in a 3D space is a well-studied problem in differential geometry and computational geometry. The shortest path between two points on S is called a *geodesic path*, and the shortest distance between two points on S is called a *geodesic distance* [34]. Given the set of corresponding points and their connections, the second step is to determine the shortest paths between each pair of connected surface points [205]. Figure 5.7 illustrates the difference between the geodesic distance or path and the Euclidean distance in neutral and smile expressions respectively. The Euclidean distance $d(p, q)$ between points p and q in three dimensions, as defined by their coordinates (X_p, Y_p, Z_p) and (X_q, Y_q, Z_q) is calculated using the classic formula:

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(X_p - X_q)^2 + (Y_p - Y_q)^2 + (Z_p - Z_q)^2} \tag{5.4}$$

A geodesic curve is the shortest curve joining two points $\mathbf{x_s}$, $\mathbf{x_e} \in S$. The geodesic distance between two points $\mathbf{x_s}$, $\mathbf{x_e}$ is the length of γ^*.

where $\mathbf{x_s}$ and $\mathbf{x_e}$ are the starting and ending points, and γ^* is the geodesic curve or path [165].

$$\gamma^* = \arg \min_{\gamma \in \rho(x_s, x_e)} L(\gamma) \tag{5.5}$$

where $L(\gamma)$ is called geodesic distance between $\mathbf{x_s}$ and $\mathbf{x_e}$.

$$d(\mathbf{x_s}, \mathbf{x_e}) = \min_{\gamma \in \rho(x_s, x_e)} L(\gamma) = L(\gamma^*) \tag{5.6}$$

Figure 5.8 shows an example of the surface, together with a set of geodesics joining pairs of points and geodesic paths. There are many methods to solve the problem of finding the shortest surface path. The algorithm we used is based on Kimmel's two methods [125, 126], and is almost the same as his extended Fast Marching Method [126]. The Fast Marching Method [182] is an extremely fast numerical algorithm for solving the Eikonal equation $\nabla T = F(x, y)$ on a rectangular orthogonal

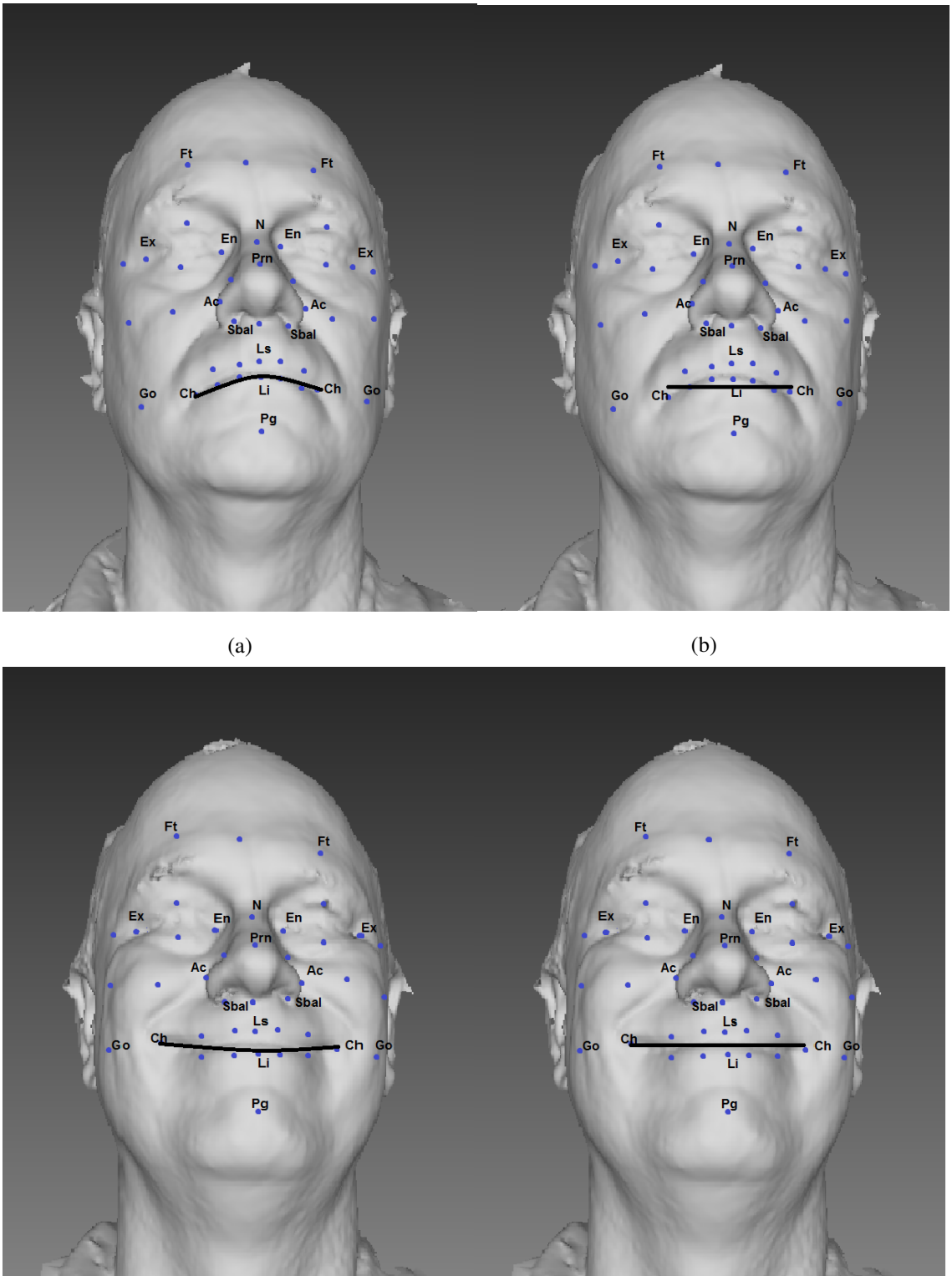
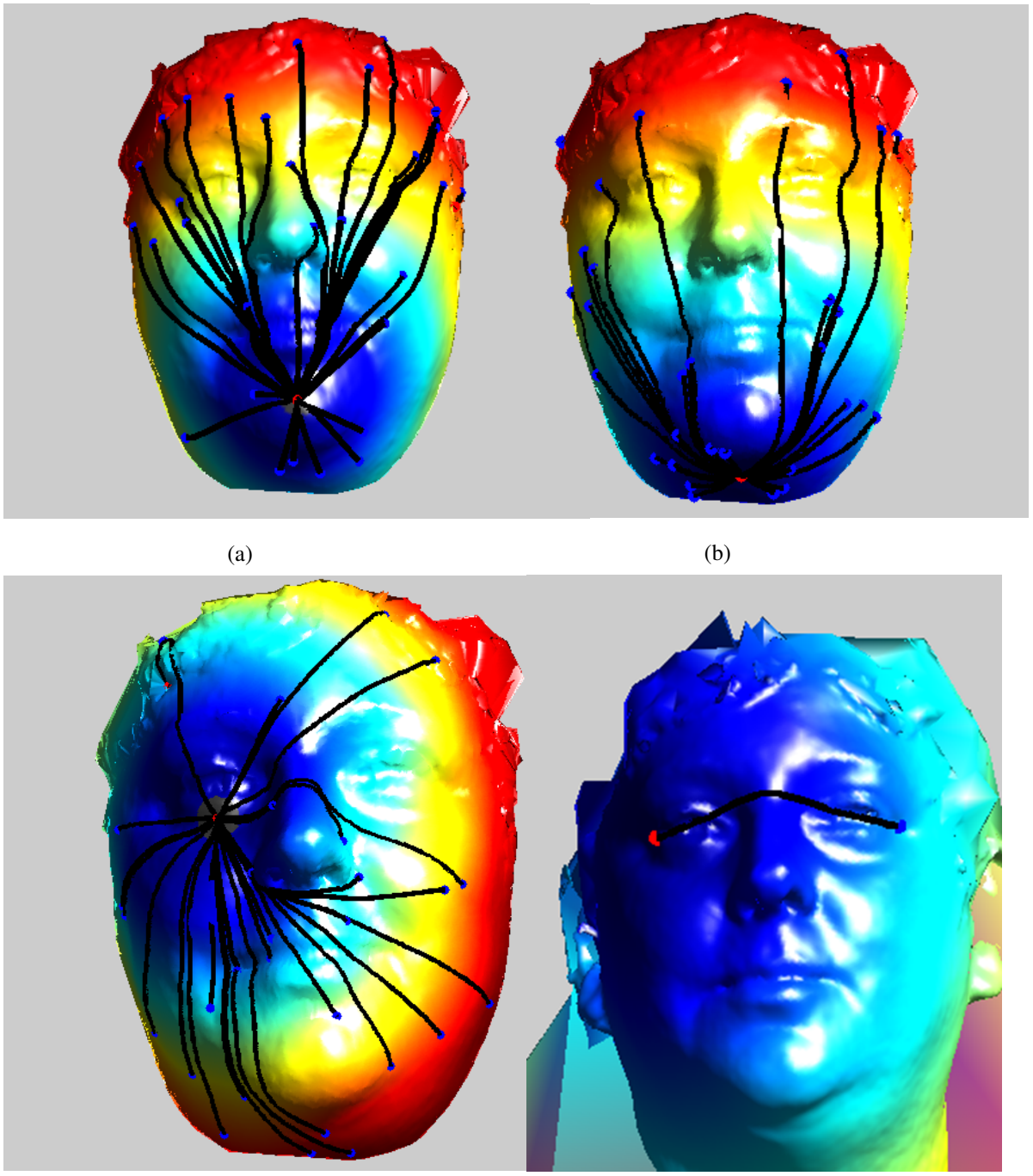


Figure 5.7: Geodesic distance (a) and (c) is the shortest surface distance between any two landmarks on the mesh, while Euclidean distance (b) and (d) are the straight-line distance between the two landmark in neutral and extreme expression between mouth corner (Ch-Ch) landmarks on the mesh of smile for the same individual.



(a) (b)

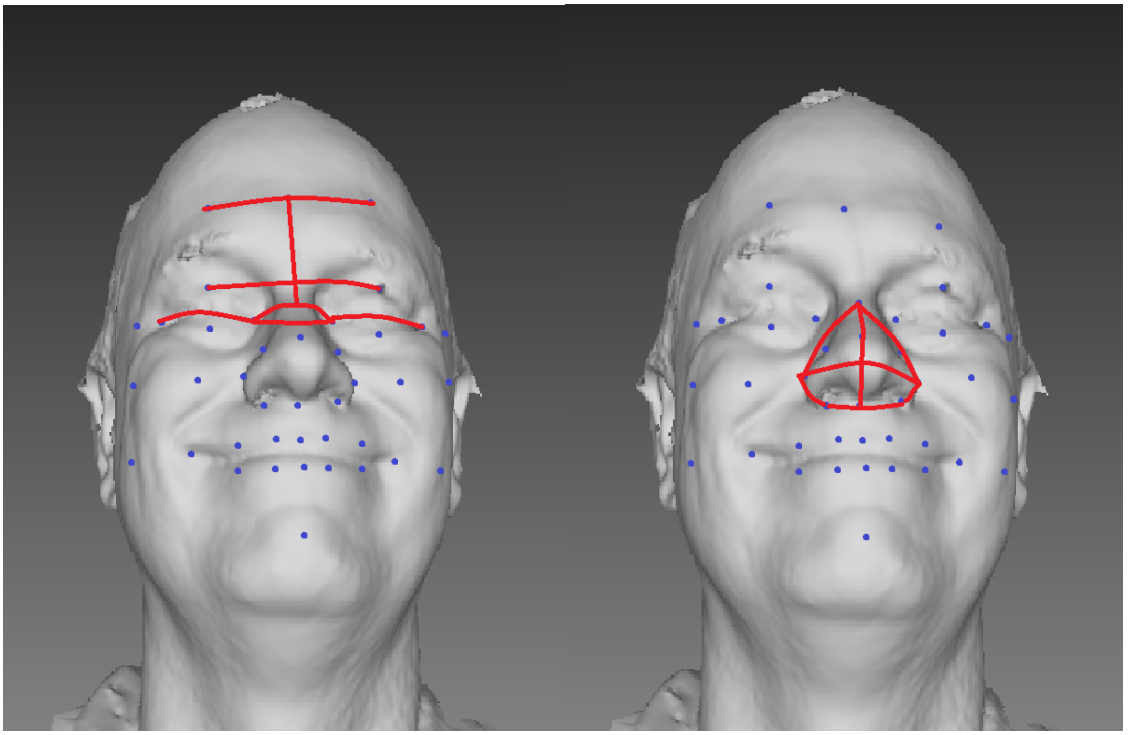
(c) (d)

Figure 5.8: The fast marching algorithm was used to compute geodesic paths and geodesic distance: The figure shows examples of using specific start points; (a) geodesic distance and paths, start point in Chin (b) geodesic distance and paths, start point in neck (c) geodesic distance and paths, start point in right lower eye (d) geodesic distance and paths, start and end points in external eyes corners.

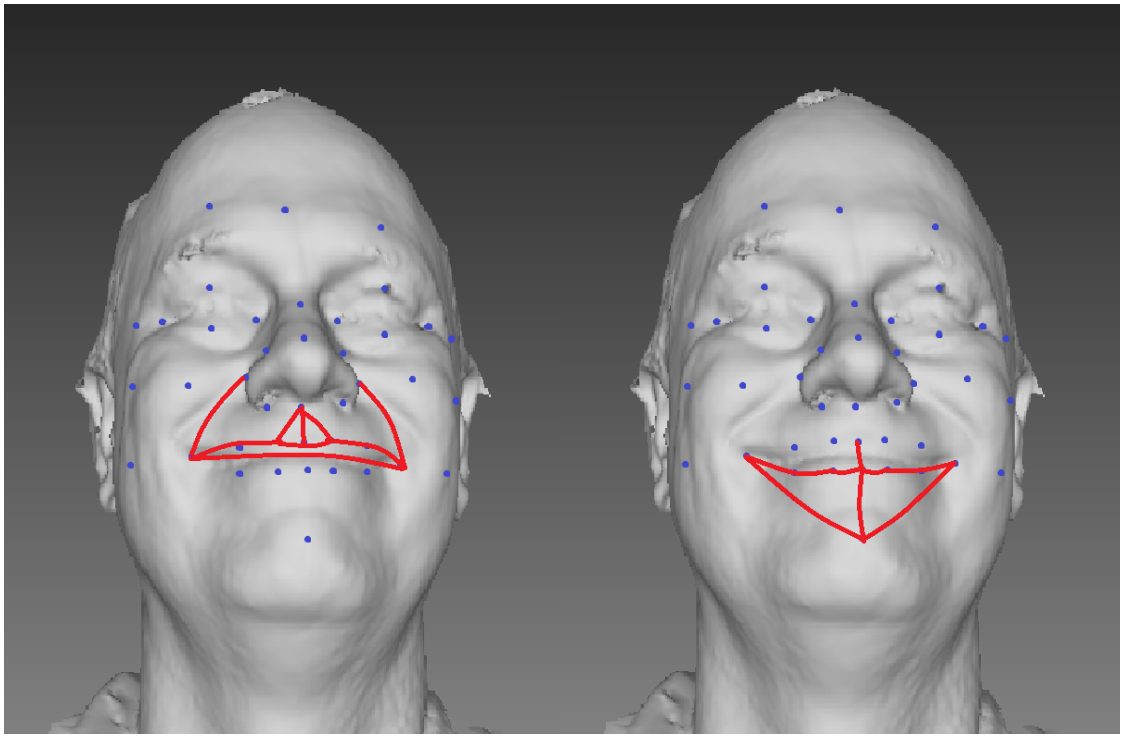
mesh in O(M log M) steps, where M is the total number of grid points. 3D facial data can provide a promising way to understand the features of the human face in 3D space and has the potential to improve the performance of the face recognition system. There are some distinct advantages of using 3D information: sufficient geometrical information, in variance of measured features relative to transformation, and a capture process by laser scanners that is immune to illumination variation [105]. For automated human face recognition, 3D facial images have some advantages over 2D facial images. They provide structural information about the face (e.g., surface curvature and geodesic distances), which cannot be obtained from a single 2D image during image acquisition [103]

Geodesic distance is an important geometric measure for understanding complex shapes. One of the first uses in brain analysis was by Griffin [100], who used the mean geodesic distance to characterise cortical shapes. We computed geodesic distances along the facial surface using the fast marching path algorithm [112] as shown in Figure 5.9. Besides 3D Euclidean distances, the motivation for employing geodesic distances was that previous studies have shown that geodesic distances are better than 3D Euclidean distances at representing 'free-form' 3D objects [104]. Furthermore, a recent study has suggested that changes in facial expressions may be modelled as isometric deformations of the facial surface [36]. When a surface is deformed isometrically, the intrinsic properties of the surface, including Gaussian and mean curvature and geodesic distances, are preserved [73]. Hence, algorithms based on geodesic distances are likely to be robust to changes in facial expressions [103].

Peyre's MATLAB fast marching toolbox [112] was used to determine geodesic paths and geodesic distances between landmarks. To this end, a number of landmark pairs were selected; the same landmarks were used to calculate the respective the dynamic features that are used in this study, as shown in Figure 5.10. Figure 5.9 illustrates examples of the paths that will be used for further analysis in four facial regions: forehead/eyes, nose, and mouth. For each face, nine paths were used for the forehead/eyes



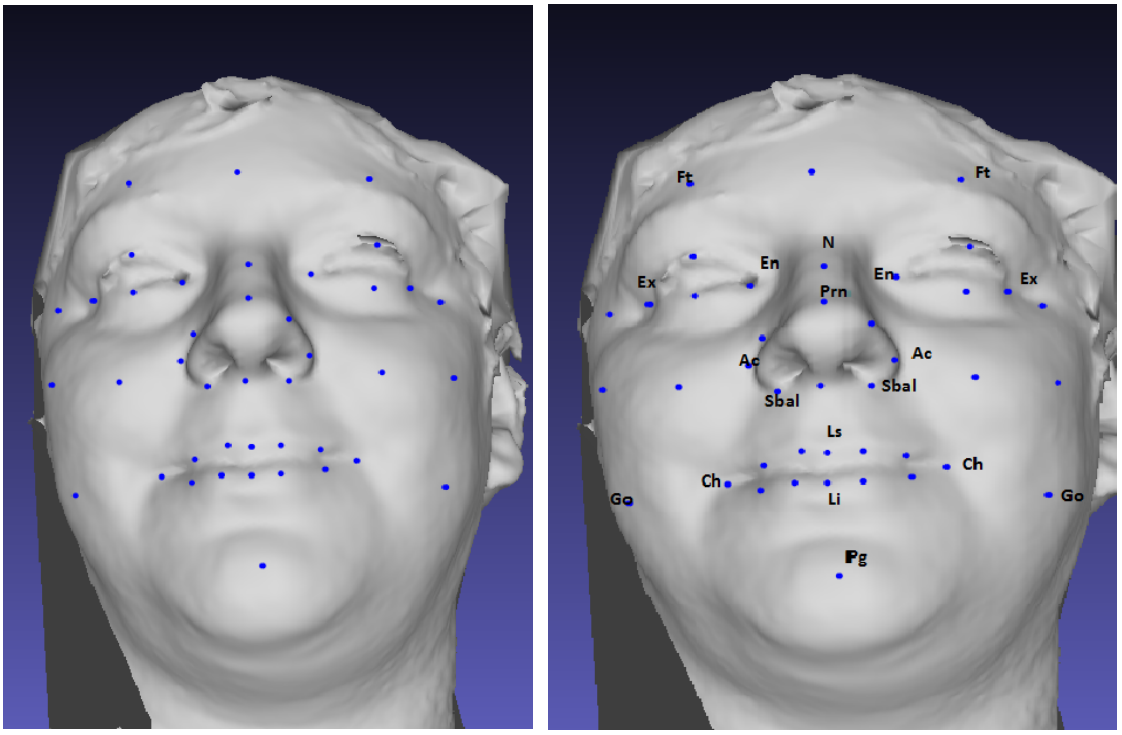
(a) (b)



(c) (d)

Figure 5.9: Figure shows geodesic paths used in the algorithm. The curvature features were extracted for these paths' surface points. Each face trait or region has a different number of geodesic paths. (a): forehead/eyes paths; (b): nose paths; (c): upper lip paths; (d): lower lip/chin paths.

region, nine paths for the nose region, and sixteen paths for the upper and lower lip region (mouth). These paths were selected following our scheme of the landmark. Our second experiment used geodesic distances to evaluate smiles in different age groups and detect gender differences. For each face, 34 geodesic distances were calculated between the same landmarks that were utilised to extract the geodesic paths. To ex-



(a) (b)

Figure 5.10: Landmarks on 3D face. (a) 41 automatically annotated landmarks shown on a shaded 3D model. (b) 19 biologically significant landmarks [83]. Labelled points are 19 landmarks selected out of the 41. Note that the distance between Ch in left side and Ch in right side increases significantly in expressions involving a gap between the lips.

tract the geodesic path distance from the meshes in our data, it was necessary perform extra pre-processing process for cleaning the meshes from non-manifold vertices and non-manifold edges. The holes which are created after cleaning pre-processing should be closed. Using MeshLab, which is an open source, portable, and extensible system

for the processing and editing of unstructured 3D triangular meshes, we extended the pre-processing pipeline including the MeshLab script to perform mesh cleaning and hole filling as described in Appendix A.

5.4 Experimentation and Results

To demonstrate the effectiveness of our proposed methods, experiments were conducted on real data. Regarding 3D dynamic data, we employed the available data, as described in detail in Chapter 3. We performed a number of experiments to search for the best features that evaluate the influence of temporal facial information in different age and gender groups, and to investigate the robustness of these features in different concepts. Experiments 1 to 2 aimed to searching for the feature types that are possibly used by the face recognition system in discriminating the individuals. A detailed explanation of these experiments is given in the following subsections.

5.4.1 Experiment-1: Analyse Age and Gender Differences

In order to investigate the age and gender differences of dynamics,the close smile expressions of 80 subjects were used as a sample in this experiment. Experiments were carried out on the balanced data of the 80 videos of smile expression (neutral-to smile-to neutral) for different subjects, who were divided into two groups by age (15-30 years and 31-60 years) and further subdivided by gender. A two-factor (age groups and gender) analysis of variance (MANOVA) test was conducted for the dynamic and static features by using age groups (Group 1 and Group 2) as the between groups factor with the dependent variables.To analyse the differences between groups in several variables, we performed a multivariate ANOVA (MANOVA). The MANOVA or multivariate analysis of variance tests the hypothesis that one or more independent variables, or factors, have an effect on a set of two or more dependent variables. The null hypothesis, H_0, is the commonly accepted fact, and it is the opposite of the alternate hypothesis H_a. The

null hypothesis H_0 of MANOVA is that the multivariate means of all groups are equal. We were interested in testing the null hypothesis that group mean vectors are all equal to one another. Mathematically this is expressed as:

$$H_0 = \mu_1 = \mu_2 = ... = \mu k. \tag{5.7}$$

Here, we work to reject the null hypothesis and then work to test and come up with an alternative hypothesis that there is a difference between at least one pair of samples on at least one variable, or:

$$H_a = \mu_i \neq \mu_j \tag{5.8}$$

for at least one $i \neq j$ and at least one variable k.

The significance level p-value, also denoted as alpha or α, is the probability of rejecting the null hypothesis when it is true. For example, a significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference. To investigate the differences in smile with age and gender, a two-factor (age groups and gender) analysis of variance (MANOVA) using general linear models was performed. All statistical tests were performed with the SPSS software package with consideration for a significance level of 5% and a confidence interval of 95%; descriptive statistics (means and standard deviations) were obtained for the groups. The groups were then compared to evaluate the effect of age in men and women separately by the analysis of variance. For analysis of the main hypothesis, an alpha level of 0.05 was used.

The results shown in Table 5.5 analyse differences between gender and age groups. There are significant differences $(p < 0.05)$ between the male and female groups of young people for the maximum and mean amplitude of their smile. Also, the maximum acceleration feature in Group 1 was more significant $(p < 0.05)$ than in Group 2 between genders. The mean acceleration feature is significant $(p < 0.05)$ in Group 2 between genders (Table 5.5). Interestingly, there were also significant $(p < 0.05)$ differences in gender in *all* dynamic features of the smile.

Further investigation has shown that although there are clear differences between age groups they are not statistically significant. This is probably because of the relatively

Static Measurements	Groups	Gender				(*P value=0.05*)
		Male		Female		
		Mean	SD	Mean	SD	
Commissural width	Group 1	0.010277	0.000176	0.010216	0.000111	0.202
	Group 2	0.010276	0.000185	0.010377	0.000370	0.278
Commissural height	Group 1	0.102441	0.001271	0.102914	0.002603	0.470
	Group 2	0.102424	0.000664	0.103089	0.001341	0.054

Table 5.4: Static measurements results obtained between male and female within age groups.

small sample size or due to the existence of extreme values. The exception was for the maximum amplitude with a significant difference for males in Group 1 compared to the age Group 2 (Table 5.6). Table 5.4 shows that given the sample size, static features do not provide statistically significant discrimination between gender or age groups. In contrast we have shown that dynamic features are more powerful, given that they are statistically significant even for the small data set.

Our experiments show that temporal facial information — the fusion of onset, apex and offset phases of smile of lip region — provides discriminative information for different age groups and gender differences. While our sample range is relatively small, our results still produce statistically significant results and support previous results [42, 65] on larger data sets but using 2D static and 2D dynamic data respectively that have demonstrated the powerful of dynamic features and have been used in many applications. Our experiment demonstrates that 3D video has good potential for further investigation. 3D data has a distinct advantage in that all measurements are actual physical (3D) distance measures that do not suffer from (2D) projective distortion, yielding more accurate and reliable measures. Further experiments will evaluate this aspect. In summary, the results show that the most significant differences between young people and older

adults are in the maximum acceleration of lip corner movements during smiles for both males and females. The results also show that the old adult people have less amplitude of smile. This conforms with clinical studies that have found that the elasticity of a person's lips decreases with age [72, 74].

Dynamic Features	Groups	Male		Female		(*P value=0.05*)
		Mean	SD	Mean	SD	
Duration	Group 1	0.29083	0.13195	0.36083	0.15776	0.427
	Group 2	0.27416	0.13759	0.33333	0.15700	0.644
Maximum Amplitude	Group 1	0.75227	0.15175	0.89415	0.43618	**0.025**
	Group 2	0.66543	0.11548	0.72926	0.11722	0.614
Mean Amplitude	Group 1	0.72216	0.14594	0.85037	0.42517	**0.028**
	Group 2	0.63976	0.11560	0.69705	0.10695	0.846
Standard deviation	Group 1	0.25546	0.21776	0.34695	0.19738	0.770
	Group 2	0.22103	0.12977	0.26803	0.16529	0.179
Max Speed	Group 1	0.01638	0.017335	0.14634	0.11672	0.090
	Group 2	0.02678	0.05162	0.01024	0.00841	0.062
Mean Speed	Group 1	0.00354	0.00302	0.00418	0.00301	0.883
	Group 2	0.00499	0.00759	0.00301	0.00237	0.095
Max Acceleration	Group 1	0.01705	0.022772	0.00995	0.01012	**0.014**
	Group 2	0.01107	0.01269	0.00724	0.00681	**0.029**
Mean Acceleration	Group 1	0.00035	0.00115	0.00049	0.387	0.387
	Group 2	0.00283	0.00871	0.00005	0.00030	**0.026**

Table 5.5: Descriptive statistics and significance of mean differences of dynamic features between males and females within age groups.

Dynamics features	Male (*P value=0.05*)	Female (*P value=0.05*)
Duration	0.698	0.584
Max Amplitude	**0.049**	0.111
Mean Amplitude	0.055	0.126
Standard deviation	0.547	0.178
Max Speed	0.398	0.180
Mean Speed	0.511	0.186
Max Acceleration	0.211	0.327
Mean Acceleration	0.505	0.456

Table 5.6: Comparisons of dynamics features *Group 1 vs Group 2* within males and females.

5.4.2 Experiment-2: 3D Euclidean and Geodesic Distance

Our experiments using geodesic and Euclidean distances are conducted on the still frames and are concerned only with gender. Researchers in the past have frequently used geodesic and Euclidean distances as features for 3D facial recognition, 3D facial morphology analysis, and gender identification. Gordon and Moreno et al. [99, 148] noted that the sub-parts of the human face have distinct surface curvature properties. High resolution 3D scans record a massive amount of detailed geometrical information of the human face shape [145]. In [83], Farkas et al. used the 3D Euclidean distance to measure the deviation of the morphological face traits from the normal face; these distances were also used by Aldridge et al. in [7] to delineate syndromes. Studies have shown, however, that the geodesic distance is more appropriate for gender identification and for measuring levels of facial masculinity or femininity [96, 97]. Therefore, this work uses the Euclidean and geodesic distances as features for analysing gender differences in different age groups. To compare their performance with

the proposed method, the dynamic features in Section 5.3.4 (i.e., duration, maximum amplitude, mean amplitude, standard deviation, max speed, mean speed, max acceleration, and mean acceleration for the euclidean distances between landmark) were used to evaluate smiles in different age groups and detect gender differences.For each face, six geodesic distances and six Euclidean distances were calculated between the same landmarks that were utilised to extract the geodesic paths shown in Figure 5.12. The fast marching algorithm and Equation 5.4 were used to produce the geodesic and Euclidean distances respectively. Note that the geodesic distance on triangulated surfaces may also be effectively computed using the fast marching method introduced in [126].In addition, studies suggest that geodesic distances may represent 3D models in a better way compared to 3D Euclidean distances [104, 187]. Gupta et al. [103] argue that algorithms based on geodesic distances are likely to be robust to changes in facial expressions. Our experiments show that the geodesic distance has good potential for further investigation and provides discriminative information for age and gender differences. In contrast to the Euclidean distance, which is more suitable for linear spaces, the geodesic distance has the advantage of being able to capture the intrinsic geometric structure of the data [15, 104]. While our sample range was small and time was needed to perform the pre-processing of the data, further experiments will evaluate this aspect. Figure 5.11 shows the 3D Euclidean and geodesic distance between the left mouth corner and the eye corner in neutral and extreme smile expressions of the same individual. The geodesic distance does not vary while there is a significant change in the Euclidean distance. Given these properties of 3D Euclidean and geodesic distances, it seems appropriate to use them for the facial gender classification of images with variations in expression. We used two feature types, i.e., the 3D Euclidean and the geodesic distances between the biologically relevant landmarks. Given 19 landmarks on a 3D face, we extracted 24 3D Euclidean and 23 geodesic distances between them. The extraction of Euclidean distances is a straight forward task. We define geodesic distance as the length of the curve generated by orthogonal projection of the Euclidean line on the 3D facial surface. This is precisely the reason for normalising the pose of

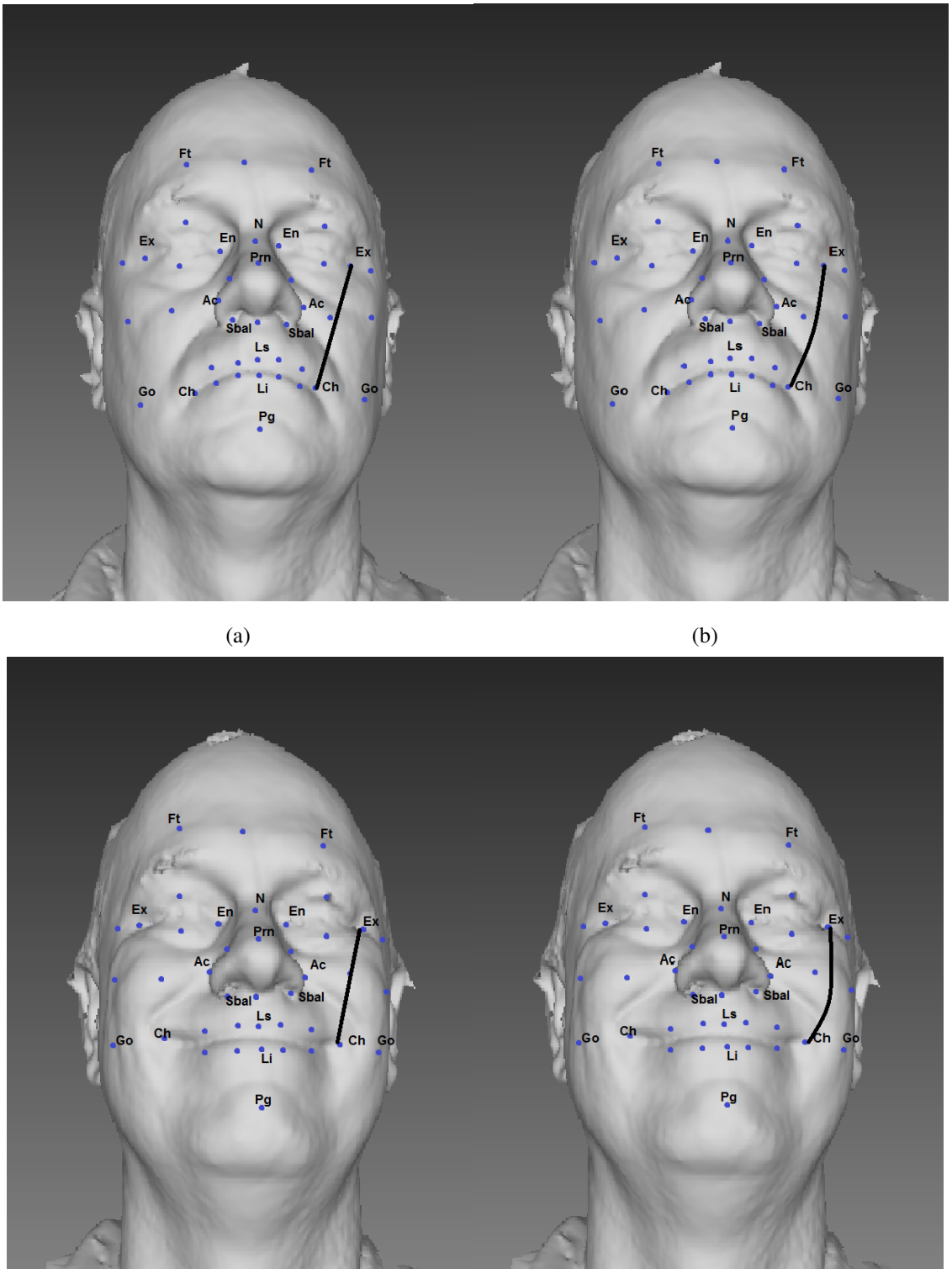


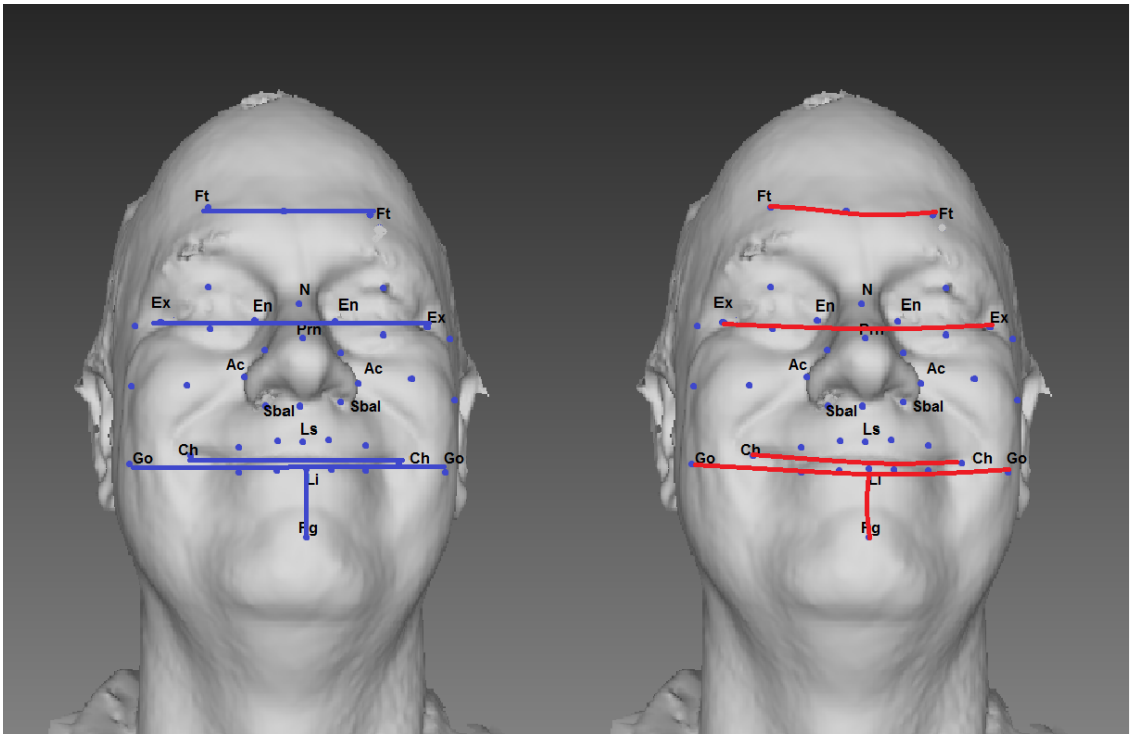
Figure 5.11: Figure shows the 3D Euclidean and geodesic distance between the left mouth corner and left outer eye corner in neutral (a) and (b) and in smile expression (c) and (d). Notice how the geodesic distance is preserved even in the presence of extreme expression.

Landmark	Landmark Definition
(N-Prn)	Nasion and Nose Tip
(Ex-Ex)	Outer Eye Corners
(Li-Pg)	Labrale inferius and Pogonion
(En-En)	Inner Eye Corners
(Go-Go)	Gonion width
(Ft-Ft)	Forehead width
(Ch-Ch)	Mouth Corners

Table 5.7: Definition of the cephalometric significant landmarks [95, 128] that can be used for facial gender classification.

each 3D face, as variations in pose can present a different surface to the viewing angle. Each facial scan was described by a set of 41 landmarks (Figure 5.10). Table 5.7 shows the path definition of the cephalometric significant landmarks that were used for facial gender classification (Tables 5.8 and 5.9).In this experiment, we used 100 subjects (50 females and 50 males); all the facial meshes had a better shape, and the vertices were distributed more evenly. For each face, 7 geodesic distances and 7 Euclidean distances were extracted between the 19 biologically significant landmarks from 6 face regions as shown in Figure 5.10 (b) for gender discrimination. Our proposed algorithm classifies faces correctly as male or female (Table 5.8 and 5.9). It is interesting to note that the outer eye corners (Ex-Ex), mouth corners (Ch-Ch), forehead width (Ft-Ft), nasal bridge length (N-Prn), intercanthal width (En-En) and nasal tip protrusion (sbal-prn-sbal) are selected by the algorithm as the most differentiating features. The results in Table 5.9 show that the geodesic distances are selected as gender discriminating features. These features allow a significant discrimination of facial gender indicating their superiority over 3D Euclidean distances (Table 5.8). Further investigation has shown that although there are clear differences between gender groups, 3D Euclidean distances do not seem

to be the feature of choice for classifying gender (Table 5.8) compared with geodesic distances (Table 5.9). Table 5.9 shows that geodesic distances between biologically significant landmarks yield better gender discrimination. The statistical differences in each table were assessed by applying the MANOVA test as a function of gender groups with consideration of a significance level of 5% (*P value=0.05*)).



(a) (b)

Figure 5.12: 3D Euclidean (a) and geodesic (b) features selected by our experiments to classify facial gender shown on extreme smile expression faces.

Face Part	Male		Female		(*P value=0.05*)
	Mean	SD	Mean	SD	
Ft-Ft	61.9202	27.3218	59.0713	25.5551	0.5357
Ex-Ex	43.7047	26.8847	52.9066	18.8113	0.0502
N-Prn	57.5171	26.0117	55.2887	15.5791	0.6044
Sbal-Prn-Sbal	45.8602	16.1597	48.6633	17.8631	0.4125
En-En	60.9938	22.4661	63.5836	21.9346	0.5611
Ch-Ch	58.6036	34.1149	69.2954	30.7832	0.1031

Table 5.8: Gender discrimination results based on Euclidean distances; significance of means of facial distances determined by MANOVA test between male and female.

5.5 Summary

Generally, a pre-processing stage is required to remove holes and spikes, and to fill missing parts as a result of the capture process. Then, detection of some face landmarks is usually required to separate the face region from the unwanted parts of the obtained face scan. Points of facial features are located in the first frame and tracked during the rest of the smile video using an automatic tracking method, These points are used to calculate the displacement signals of the lip corners. Finally, pose normalisation is also required to minimize overall deviations between the landmark sets. This can be achieved by Procrustes analysis of the landmark sets through translation, rotation, and scaling. Static and dynamic features were extracted separately to evaluate smile in different age groups. Comparison of the results showed that 3D facial dynamics provide more useful information than static features for the characterisation of smiles. Smiles change with age, and their characteristics differ between males and females. We analysed the rich sources of information present in the 3D dynamic

Face Part	Male		Female		(*P value=0.05*)
	Mean	SD	Mean	SD	
Ft-Ft	42.4944	23.1955	33.00564	21.1985	**0.03523**
Ex-Ex	56.6583	22.2045	45.4250	15.8350	**0.0044**
N-Prn	51.5339	15.8017	43.5805	21.5092	**0.03766**
Sbal-Prn-Sbal	44.1256	22.0323	52.3501	20.9414	0.05863
En-En	52.87609	31.1618	63.56847	20.54625	**0.04553**
Ch-Ch	59.0505	23.47065	47.8884	25.4969	**0.02492**

Table 5.9: Gender discrimination results based on geodesic distances; significance of means of facial distances determined by MANOVA test between male and female and statistically significant (*P value=0.05*) are highlighted in bold.

features of smiles to provide more insight into the patterns of smile dynamics. The sources of temporal information that were investigated were the various dynamics of lip movement, which were analysed to extract the descriptive features. Direct Euclidean and geodesic distance measures were extracted between 3D facial landmarks as a local geometric features using the fast marching algorithm. We evaluated the dynamic features of closed mouth smiles of 80 subjects of both genders and all landmark measurements were expressed in pixels.

Chapter 6

Multilevel Principal Components Analysis of Dynamical Smiles

6.1 Introduction

Principal Component Analysis is a key step for building most statistical shape models. Statistical models, including active shape models (ASMs) and active appearance models (AAMs), are commonly used in various applications of computer vision to detect certain shapes or features within images. Nevertheless, in the event that the data set contains naturally-occurring clustering or multilevel data structures, the eigenvalues and eigenvectors determined from the straightforward application of principal components analysis (PCA) are not wholly reflective of the real variability within the collection of images or shapes.

Our 3D video facial scans data has naturally occurring multilevel clustering of the subjects by both age and gender groups (between-group variation); in addition, each subject has their own smile expression (within-group variation). In view of this, a multilevel PCA (mPCA), which is characterised by the simultaneous and autonomous implementation of PCA for the within-group and between-group levels, serves as a viable way to model the structural framework of an image and of 3D data in conjunction with groupings that occur between images. In the context of 3D faces, the within-group level is 'nested' inside the between-group level as shown in Figure 6.1.

Figure 6.1: The "Nested" nature of multilevel data

One of the fundamental benefits of this technique is the possibility of constraining any segmentation, thereby meaning that a model's fit is continuously aligned in a relatively close way to the training set utilised in model creation. A notable process in which mPCA was applied to create ASMs was in the segmentation of the human spine [139]. The spinal column comprises vertebrae linked to one another. Classical statistical models have difficulty dealing with this dependence between items since all points are treated identically whether they belong to the same sub-structure or not. As shown to by the study's findings, mPCA is characterised by considerably more flexibility and by the ability to represent deformations which can not be modelled using conventional statistical models.

In another study, mPCA was used to build ASMs related to the field of dental imaging [85], where proof-of-principle was evaluated for the modelling of basic perioral expressions (approximated to the mouth-lip connection). A Monte Carlo sampling method was applied to create a data set, drawing on a quadratic function $y = cx^2$ (which represented the centre line between the upper and lower lip) in combination with a value of c controlled 'expression'. Various expressions (sad, neutral and happy) were modelled with respect to the between-group level; the modelling of a group of expressions took place correctly, and the modelling of variations in lip width also

took place accurately at the within-group level. One of the notable findings to emerge from the study was that accurate within-group variation and between-group variation of cases were well modelled by mPCA, while the same could not be achieved with conventional PCA.

In addition to this, two clinicians implemented mPCA for the purpose of analysing a dataset, which included landmark points situated on panoramic mandibular radiographs, thereby establishing two distinct groups of landmark points for the group of images. The within-group level of mPCA was employed to model changes in cortical bone shape [9], while variance between the groups of images produced by the experts was examined at the between-group level. As expected, eigenvalues demonstrated that variance owing to modifications in cortical bone shape being more significant than the placement differences resulting from the two independent clinicians evaluations. It is worth noting that the clinicians provided qualitative reports outlining the difficulties associated with situating the point along the boundaries, and it was also found that the initial model of variation for the between-group level was an accurate reflection of this category of change.

In the recent application by Farnell et al. [86], mPCA is utilised for the purpose of studying the landmark points of three-dimensional (3D) laser scans captured of male and female participants from various locations in Europe (specifically, Croatia, Finland, Wales, and England). Following this, eigenvalues and eigenvectors are given, and in turn component scores are derived by fitting the mPCA model to every group of points corresponding to the participants.

Finally, after comparatively examining the outcomes yielded by mPCA and conventional (single-level) PCA, the implications of the key points of variance are examined. The aim of this research was to investigate the impact of naturally-occurring groups within the sample group of Croatian, Finnish, Welsh, and English participants. Following a concise overview of the formal aspects of mPCA, the researchers demonstrated that one of the key advantages of the technique is the flexibility it affords in

modelling variability for multilevel data structures (namely, at the between-group and within-group levels). An investigation of the eigenvalues indicated that each source of variability was consequential, where the modes of variability seemed to be logical. With respect to the within-group mPCA's initial mode, evidence was found to suggest that this was implicated in determining facial width, while regarding the between-group mPCA's initial mode, this appeared to determine nose and/or facial length and shape. The researchers also found evidence to indicate that the preliminary outcomes of mPCA for facial shape data, due to the fact that consistency was observed between males and females for the Croatian, Finnish, Welsh, and English samples, suggest clustering regarding the between-group component scores.

Conventional PCA displayed signs of clustering, but this was ambiguous for the most part. In addition, ias initially expected, point-to-point errors with respect to model fits for PCA and mPCA were reduced while the number of eigenvectors was maintained In view of these findings, the authors reported that where a dataset contains naturally-occurring clustering or multilevel data structures, mPCA is characterised by a higher order of control and adaptability than conventional PCA.

In this thesis, mPCA is applied for the first time to analyse the landmark points of three-dimensional (3D) dynamic smiles of male and female participants in different age groups; details of the data were given in Chapter 3. In the previous chapter, there was a problem in analysing the data due to the grouping that occurred between subjects by age and gender groups and between subjects themselves in smile dynamics, which standard PCA cannot account for.

The new contribution presented in this chapter is the observation that by using mPCA, we present a basic fundamental of mPCA and offer better discrimination between classes and to perform much interpreted. This shows the undoubted potential of the relatively new mPCA approach. This study takes the dynamics of the smile dataset of 80 subjects to explore an application of mPCA to investigate the link between the three levels of the principle components of subjects in level one, the variation between smile

phases in level two, and the variation within smile phases. Our findings demonstrate that the changes in lip shape due to the dynamic change in the lips during the smile is far outstripped by variability in the natural lip shape between subjects, which accounts for 71.97% of the variability of lip shape across our subject set in all phases of the smile. This work, to some extent, follows on from the previous study developed by Farnell in [85, 86], which studied the landmark points of 3D laser scans of the head of subjects in different ethnicities and both gender.

The details of the mathematics that support the mPCA for ASMs techniques are presented in Section 6.2. We present the results of the eigenvectors and eigenvalues, and the component scores are found by projecting the mPCA model to each set of points for each subjects in the dataset and comparing the results of standard (single-level) PCA. The major modes of variation are explored. Further details of the results are presented in Section 6.3.

6.2 Methods

6.2.1 Mathematical Formalism

Following the mathematical notation of Farnell in [85, 86], in this section, we present the formalism for mPCA. The ASM method's use of PCA has been widely documented in the literature [9, 53, 54, 55]. Applying PCA for the covariance matrix and the eigenvalues and eigenvectors is discussed in detail in Chapter 4. Landmark points are represented by a vector, Z_i, and the k^{th} element of this vector is given by Z_{ik}. The total number of such points is n, and the mean shape vector (averaged over all N subjects) is given by $\overline{Z}$. The covariance matrix is found by evaluating

$$C_{k_1 k_2} = \frac{1}{N-1} \sum_{i=1}^{N} (Z_{ik_1} - \overline{Z}_{ik_1})(Z_{ik_2} - \overline{Z}_{ik_2}) \tag{6.1}$$

where k_1 and k_2 indicate elements of the covariance matrix. We find the eigenvalues λ_l

and eigenvectors u_l of this matrix, and we rank all of the eigenvalues Λ_l in descending order. We choose the largest m eigenvalues of largest magnitude to be retained in the model. Any new shape is given by

$$Z = \overline{Z} + \sum_{i=1}^{m} a_i u_i \tag{6.2}$$

The eigenvectors u_l are orthonormal, and so we can determine the coefficients, a_l, for a fit of the model to a new shape vector, z, readily by using

$$a_l = u_l.(Z - \overline{Z}) \tag{6.3}$$

Constraints may be placed on these $a-$coefficients, such as $|a_l| \leq 3\sqrt{\lambda_l}$, which ensures that the subsequent model fits a new shape vector and never 'strays too far' from the normal cases in the training set.

The formalism for mPCA is more complicated, and further details are given in [85]. In mPCA, we observe here that the two covariance matrices for the two-level model are formed. First, there is a within-group covariance matrix, which is the covariance evaluated over all subjects within the group and with respect to their local group means or centroids, and this matrix is then averaged for all groups. Second, there is a between-group covariance matrix that is the covariance matrix of the centroid of the groups with respect to a 'grand' mean shape $\overline{Z}$ of the average of these centroids. The rank of this matrix is limited by the number of groups.

Applying PCA for the within-group covariance matrix of the above equation, the l_{th} eigenvalue is denoted by λ_l^w, and its eigenvector is denoted by u_l^w. Independently, PCA is also applied to the between-group covariance matrix given above. The l_{th} eigenvalue is denoted by λ_l^b and its eigenvector is denoted by u_l^b. All of the eigenvalues are ranked λ^b and λ^w in descending order for the between- and within-group levels separately, and then we retain the m_b and m_w as the largest of such eigenvectors, respectively. The new shape is now given by

$$Z = \overline{Z} + \sum_{l_1=1}^{m_w} a_{l_1}^w u_{l_1}^w + \sum_{l_2=1}^{m_b} a_{l_2}^b u_{l_2}^b \tag{6.4}$$

Constraints may again be placed on these $a-$coefficients, such as $|a_l^b| \leq 3\sqrt{\lambda_l^b}$ and $|a_l^w| \leq 3\sqrt{\lambda_l^w}$, which ensures that the subsequent model fits a new shape vector and never 'strays too far' from the cases in the training set with respect to both the within-group variation and the between-group variation. The covariance matrices are symmetrical and so all the within-group eigenvectors u_l^w are orthogonal to all other within-group eigenvectors (and similarly for the between-group eigenvectors). However, the eigenvectors u_l^w and u_l^b do not necessarily have to be orthogonal with respect to each other, and so an equivalent projection to Equation 6.3 for mPCA becomes problematic. A fit of the model given by Equation 6.4 to a set of candidate points is achieved by minimising the overall (squared) error with respect to the coefficients a_l^w and a_l^b.

6.2.2 4D Acquisition and Preprocessing

The data capture and preprocessing procedure have been described in Chapter 3. Forty-one reliable facial landmarks were automatically tracked for each subject. Each landmark point vector Z was of size 123 $(= 41 \times 3)$. These facial landmarks are shown in 6.1. The numbers of subjects in the two groups were female $(n = 40)$ and male $(n = 40)$ in the first age group $(age < 30)$, and female $(n = 40)$ and male $(n = 40)$ in the second age group $(age > 30)$. Landmark points were scaled by Procrustes transformation so that all sets of points were on broadly the same scale.

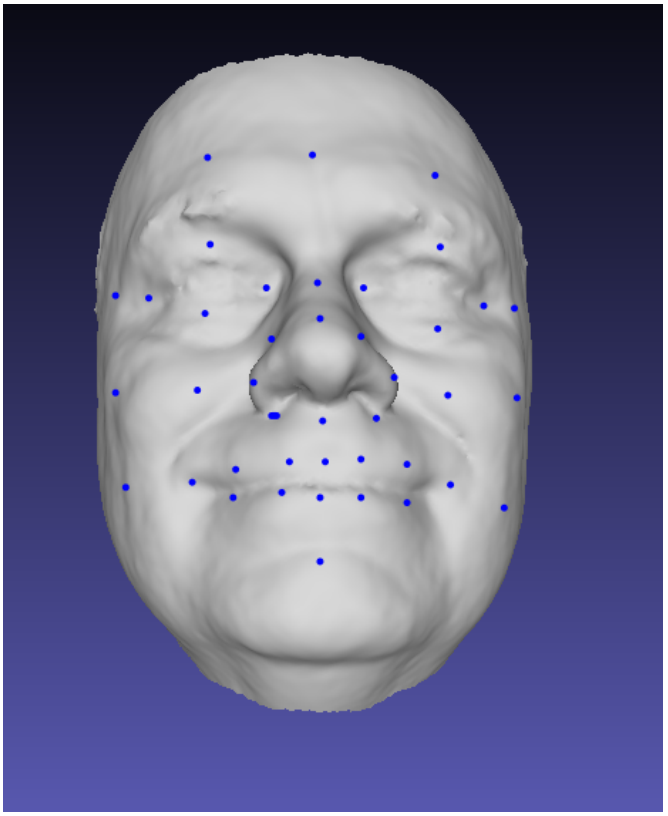


Figure 6.2: Forty-one landmarks points on 3D facial scans of participants

6.3 Results

6.3.1 3 Levels mPCA Model of Dynamical Smiles

In this study, we propose a solution and investigation to build multilevel statistical shape models for the variations of smile dynamics in different age groups and genders. based on multilevel component analysis (mPCA). The concept of mPCA was introduced in [85, 86] as an extension of PCA for multilevel structures.

Our experiment includes our dataset, which consisted of 3D video shape data, where 12 points were placed (and tracked) along the outer boundary of the mouth. After 20 smile subjects were excluded, the smile of 60 subjects of adult staff and students at Cardiff University (31 male and 29 female) were analysed. The number of frames

in the video was between approximately 100 and 250 frames during all phases of a smile for each subject. In these experiment, preprocessing consisted of centring the 3D shapes only. A three-level mPCA model was used for the dataset: level 1, "between subject" due to a natural face shape not attributed to smiling; level 2, "between smile phases" variation due to differences between seven different phases of a smile; and level 3, "within smile phases" variation due to residual differences within the different phases of a smile. Phases of the smile were identified automatically by inspection of the dynamics of smile amplitudes [5] for each subject individually.

We built a hierarchy with three levels. At the highest level of hierarchy (level-1) are subject related factors, such as gender, age, and ethnicity. Situated at the middle level of the hierarchy (level-2) is the variation between smile expressions. Level-2 variables are nested within level-1 groups and are affected by level-1 variables. For example, a smile (level-2) changes with age and differs between males and females, thereby affecting (level-2 variables) in the (level-1 variables). Variables at the lowest level of the hierarchy (level-3) are nested within level-2 groups and also experience the effect of level-2 variables. In our example, variables of smile phases variables, such as onset, apex, and offset (level-3), are situated within (level-2).

Let us consider a three-level mPCA model of dynamical smiles: the idea is to decompose the data into a level-1: the variation between-individual components; level-2: the variation between smile phases; and level-3: the variation within-smile phases. Let us assume a sample with 80 subjects excluded 20 subjects, divided into 2 age groups of 30 subjects each of both genders. The shape of 3D lips was averaged with 12 landmark points over all subjects in each group (age and gender). The number of groups at level 3 was 420 (7 expressions and 60 subjects), level-2 expression level, it contains one group and level-1, subjects level, it contains 60 groups.

Figure 6.3 shows the eigenvalues from the standard single-level PCA and the within-group and between-group eigenvalues from mPCA. The first major mode from the single-level PCA corresponds to between-subject variations. The variations in shape

occur for all phases of a smile (size of mouth, upturned or down-turned mouth). The second major mode occurs during all phases of smile (irrespective of size of mouth, upturned or down-turned mouth). The third major mode from the single-level PCA corresponds to between-subjects variations. Variation within smile phases is negligible. Tables 6.1 and 6.2 show the changes in lip shape due to dynamical changes in the lips

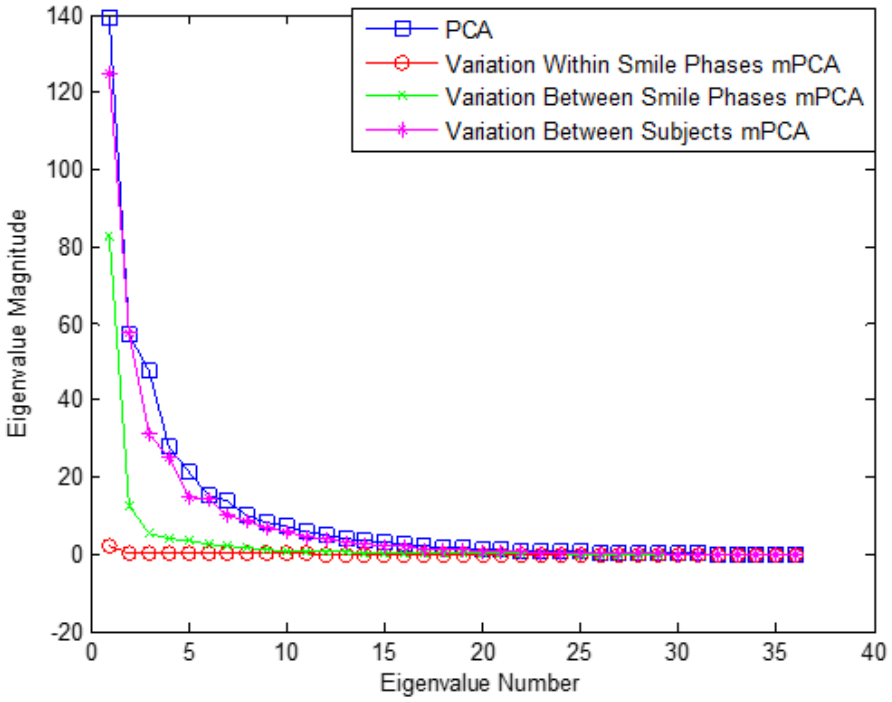


Figure 6.3: Eigenvalues from mPCA level 1 (between-subject variation), level 2 (between-smile variation: facial expression), and level 3 (within-smile variation: smile phases)..

during the smile. These changes in the lips during the smile are far outstripped by variability in the natural lip shape between subjects, which accounts for 71.97% of the variability of lip shape across our subject set in all phases of the smiles. Variation of single-level PCA for eigenvector is computed by

$$j = \frac{100 \times \lambda_i}{\sum_{i=1}^{l} \lambda_i} \tag{6.5}$$

Table 6.1: The Ten-first eigenvalues of the variation of mPCA

	Eigenvalue			
Eigenvalue Number	Single-Level PCA	mPCA:level-1	mPCA:level- 2	mPCA:level-3
1	**139.4**	**124.8**	**82.7**	**2.1**
2	**57.0**	**57.5**	**12.7**	**0.5**
3	**47.5**	**31.3**	**5.4**	**0.4**
4	**28.1**	**25.3**	**4.0**	**0.4**
5	**21.5**	**14.9**	**3.8**	**0.3**
6	**15.1**	**14.2**	**2.7**	**0.3**
7	**13.8**	**10.0**	**2.0**	**0.2**
8	**10.3**	**8.7**	**1.5**	**0.2**
9	**8.4**	**7.1**	**1.2**	**0.2**
10	**7.5**	**6.0**	**1.0**	**0.1**

While the variation of mPCA for eigenvector j at level k is computed by

$$j = \frac{100 \times \lambda_i^k}{\sum_{i_1}^{l_1} \lambda_{i_1}^1 + \sum_{i_2}^{l_2} \lambda_{i_2}^2 + \sum_{i_3}^{l_3} \lambda_{i_3}^3} \tag{6.6}$$

Figure 6.4 indicates the predicted mouth shape from, left, the single-level PCA in the third major mode and, right, the second major mode between-subjects level-1 of the mPCA. These modes are changes in size or fullness of mouth. Figure 6.5 shows the predicted mouth shape: left- single-level PCA second major mode; right- between-subjects level-2 mPCA first major mode. These mode are related to changes in the size of the mouth during a smile irrespective of size of mouth and resting mouth shape, i.e., smiling: mouth becomes wider and taller and the edges of the mouth go slightly upwards, and the corners of the lips pull back strongly in the XY plane.

Table 6.2: The variation of lip shape between subjects

	Variation%			
Eigenvalue Number	Single-Level PCA	mPCA:level-1	mPCA:level- 2	mPCA:level-3
1	**35.9**	**27.37**	**18.1**	**0.5**
2	**14.7**	**12.60**	**2.8**	**0.1**
3	**12.3**	**6.86**	**1.2**	**0.1**
4	**7.2**	**5.54**	**0.9**	**0.1**
5	**5.6**	**3.12**	**0.6**	**0.1**
6	**3.9**	**14.2**	**2.7**	**0.3**
7	**3.6**	**2.19**	**0.4**	**0.0**
8	**2.7**	**1.90**	**0.3**	**0.0**
9	**2.2**	**1.55**	**0.3**	**0.0**
10	**1.9**	**1.31**	**0.2**	**0.0**
Sum	**100**	**71.97**	**26.99**	**1.3**

Figure 6.6 shows the predicted mouth shapes from left the single-level PCA in the first major mode and the first major mode between subjects level-1 of the mPCA at the right. These modes are related to up-shaped or down-shaped mouths irrespective of the smile phase in the XZ plane-some differences in the XY plane.

6.3.2 Component Scores

The scores are used for the visualisation of data or, in conventional classification, clustering or regression methods. When principal component scores are used to visualise high-dimensional data, it is common to display the m first important sample scores in two-dimensional plots (or possibly in 3D). These score plots can be used to com-

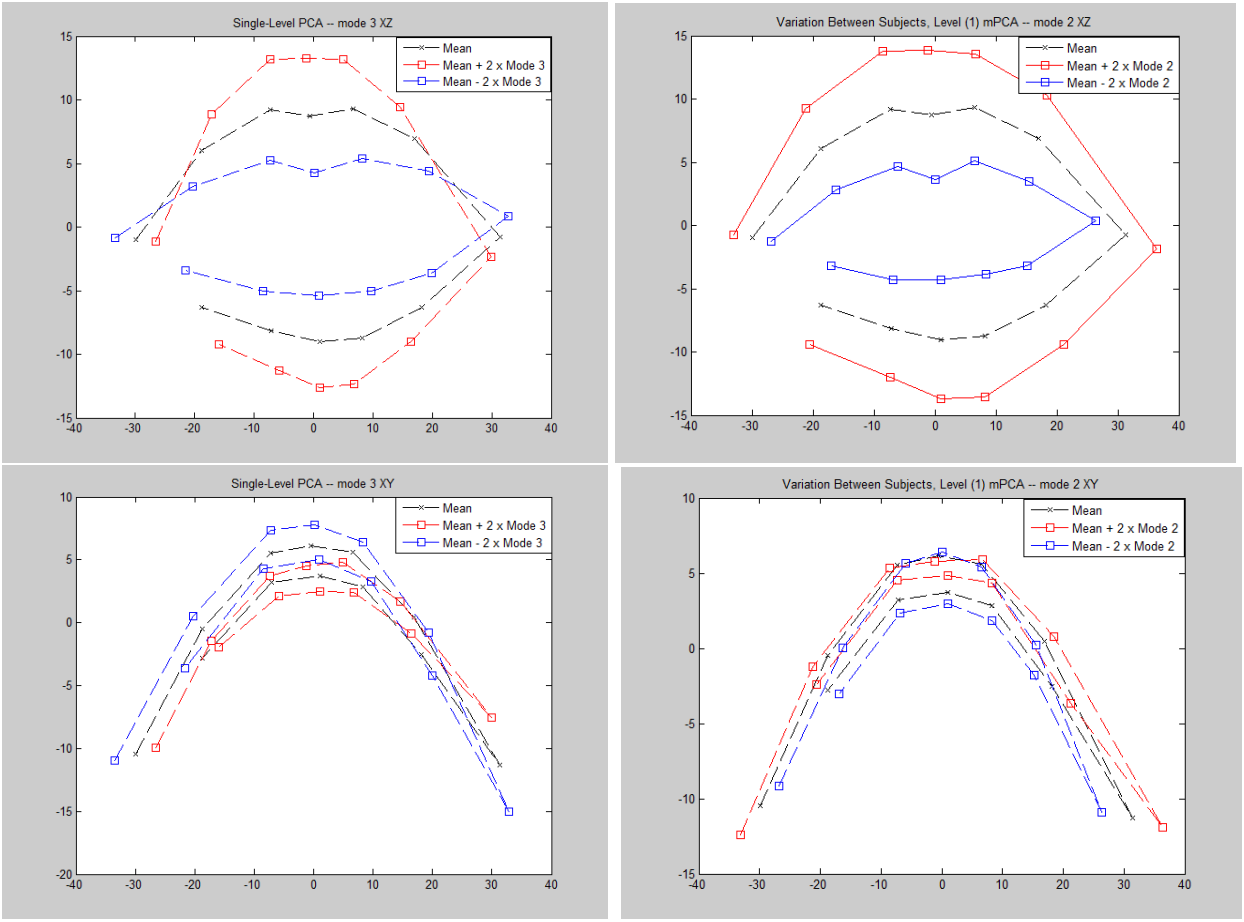


Figure 6.4: The Major modes of variation between subjects of standard PCA and the mPCA in the 3rd and 2nd mode.

pareobservations, detect subgroups, or identify outliers [107]. In our experiment, we used the component scores to detect and visualise the multilevel structure in our data using the mPCA model. Component scores were found by fitting the single level PCA and mPCA models to each set of landmark points for each subject in the data set. Results for single-level PCA, shown in Figures 6.7 and 6.8, indicate some evidence of a difference by gender in the component score for PCA 1 (shape of mouth) for single-level PCA (males in positive part and female in negative part), though any gender difference in the component scores for PCA 2 (smile) is harder to see. Figures 6.9 and 6.10 show some evidence of a difference by gender in the component score for PCA 1 (shape of mouth) for mPCA-level 1 ($mPCA : l_1 = 15$, $l_2 = 3$, and $l_3 = 0$).

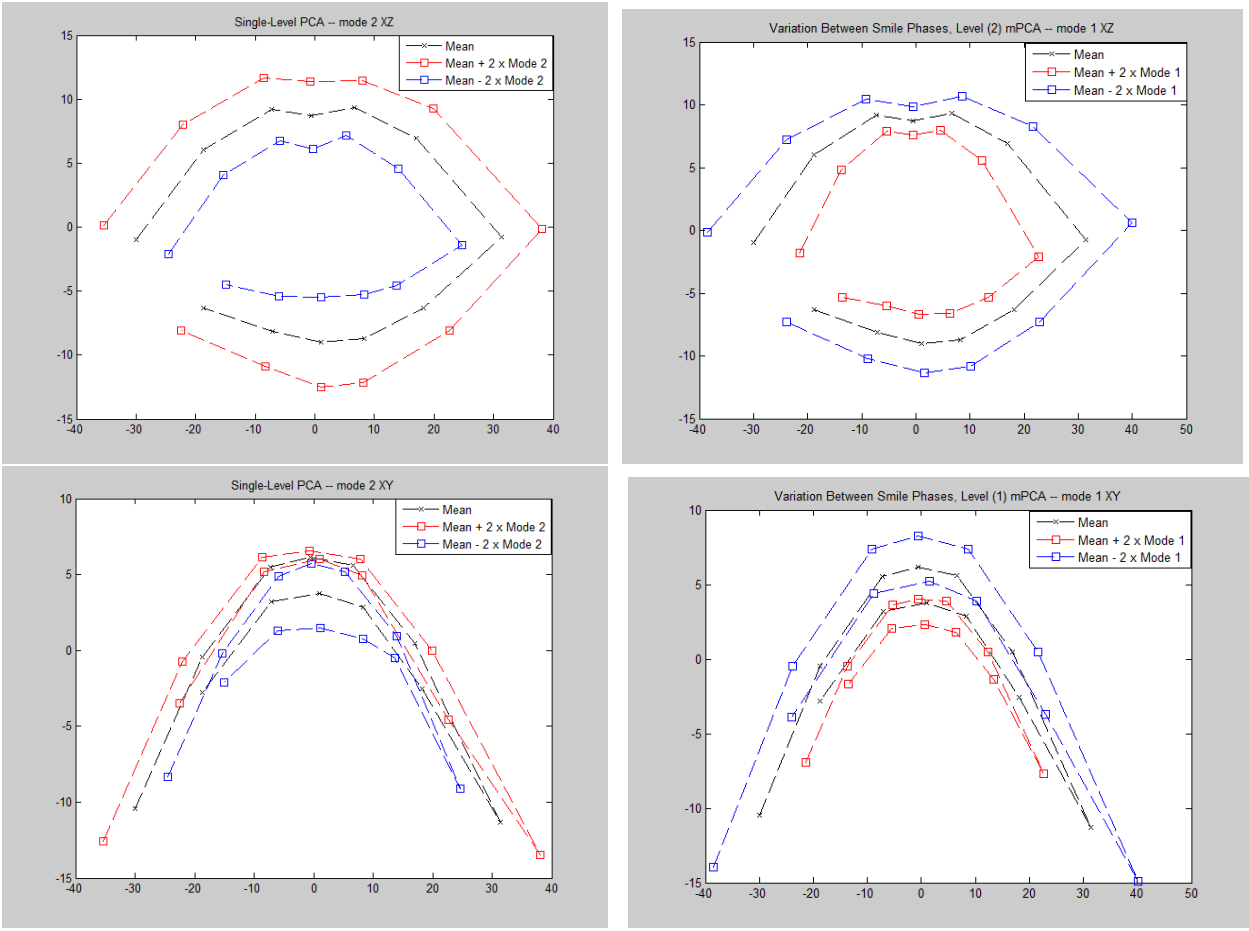


Figure 6.5: The major modes of variation between smile phases of standard PCA and the mPCA in the 1'st and 2nd mode.

No obvious divisions by gender are apparent in Figure 6.11, though the variation in the component score is very large. In general, men and women smile in the same way. However, there is considerable variation by subject in the dynamical smile changes by subject that cannot be excluded from the natural reset mouth shape $(mPCA : l_1 = 15$, $l_2 = 3$, and $l_3 = 0)$.

6.3.3 Distances between centroids

It is possible to infer the distance between group centroids directly from the distances between points in each group. Each group has a well-defined centroid (i.e., average

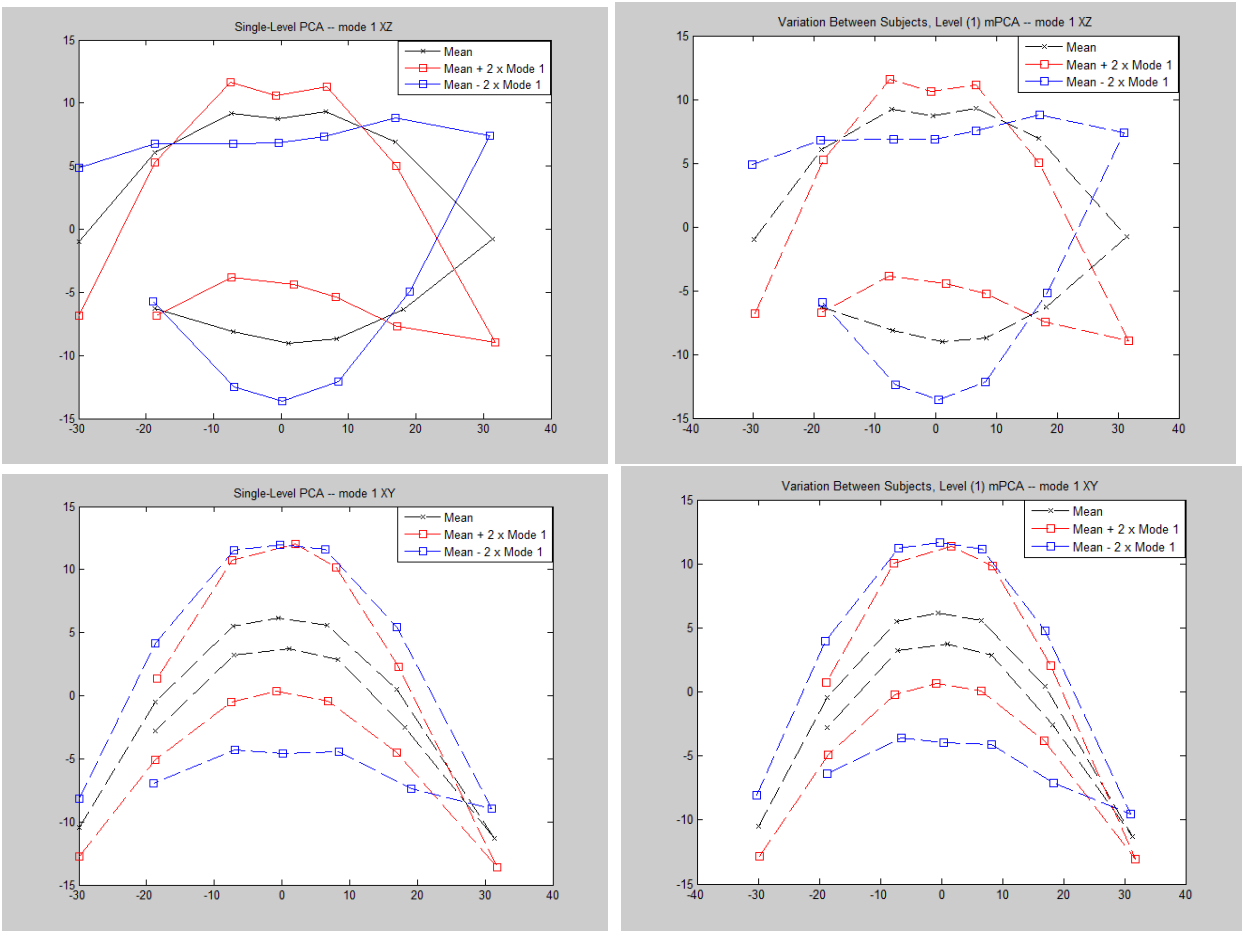


Figure 6.6: The major modes of variation between subjects of standard PCA and the mPCA in the 1'st ans 2nd mode.

across all the points in the group).Representing each group by its centroid, the distance between groups is a distance between centroids. Centroids for the standardised scores for males and females are shown in Table 6.3. Distances between the centroids show some differences in the centroids for subject level-1 mPCA, but not for the between-smile phases level-2 mPCA. Subject level-1 mPCA has the largest distance between the centroids, which is as expected, as this mode relates to mouth size, and so indicates that men have bigger mouths than women. This result supports the study reported by Chetan et al. in [42], where males have more vertical movements whereas females have more horizontal movements during a smile.

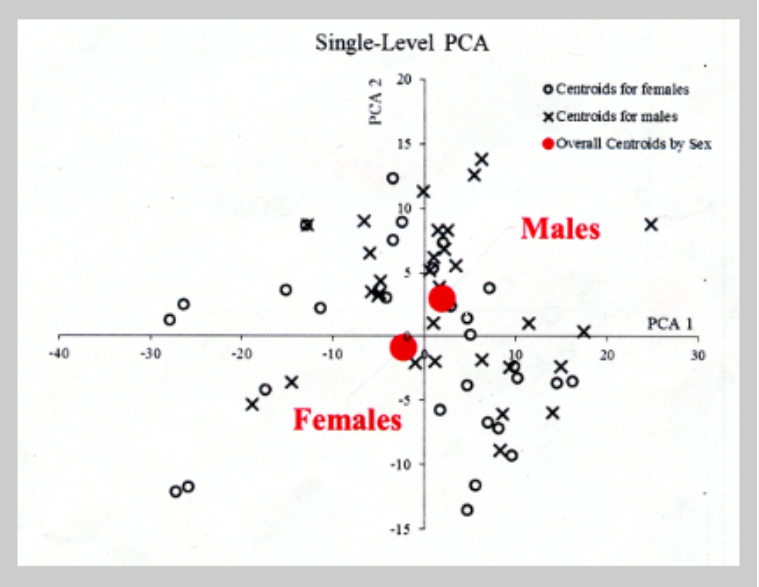


Figure 6.7: Difference by gender in component scores centroids for each subject (Single-Level PCA for PCA 1 and PCA 2) .

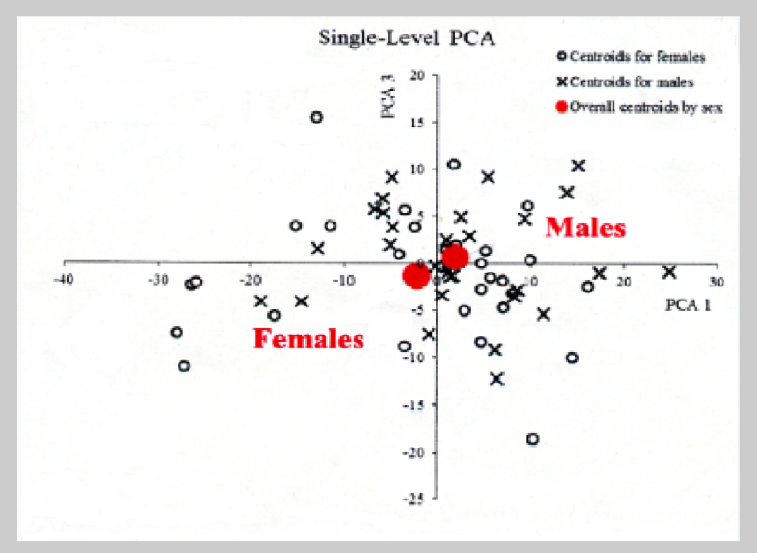


Figure 6.8: Difference by gender in component scores centroids for each subject (Single-Level PCA For PCA 1 and PCA3) .

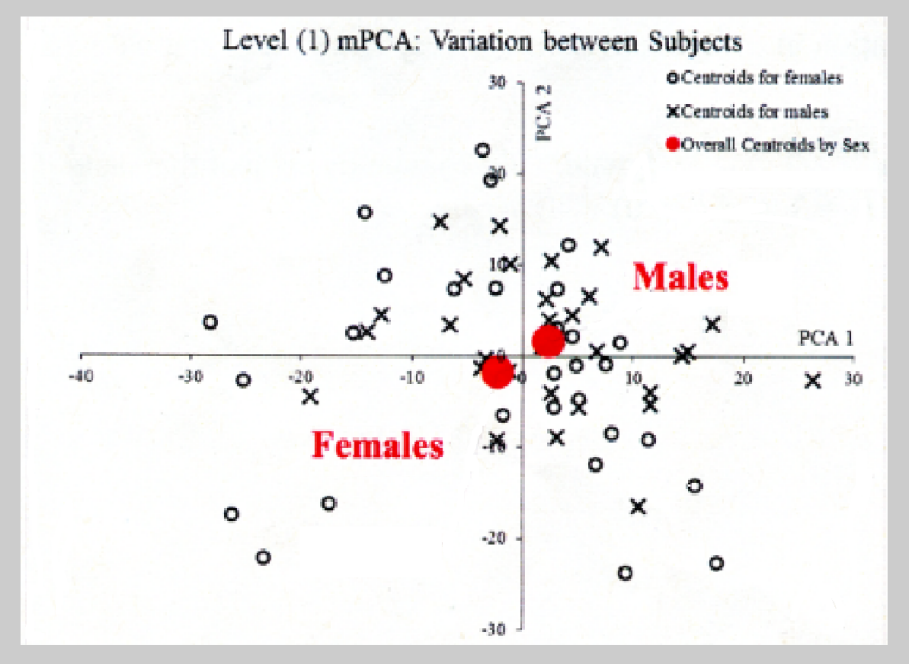


Figure 6.9: mPCA Level (1): Variation between subjects for PCA 1 and PCA 2

	Single-Level PCA			Subject Level (1) mPCA			Between Smile Phases Level (2) mPCA		
	PCA1	PCA2	PCA3	PCA1	PCA2	PCA3	PCA1	PCA2	PCA3
Females	-0.18	-0.13	-0.20	-0.20	-0.26	-0.03	-0.08	0.11	-0.33
males	0.17	0.39	0.09	0.22	0.19	-0.01	-0.19	-0.03	0.60
Distance	0.35	0.52	0.29	0.41	0.45	0.03	0.11	0.14	0.93

Table 6.3: The distances between centroids

6.3.4 Standardisation

Standardisation is achieved by dividing the component scores for a given PCA component by the square root of its eigenvalue. The component scores may finally be standardised to give "standard scores" by dividing the scores for each subject with-respect to each principal component by its standard deviation (i.e., the square root of its eigenvalue: $(a_1^1 \,/\, \sqrt{\lambda_1^1}, a_2^2 \,/\sqrt{\lambda_2^2}, a_3^3 \,/\sqrt{\lambda_3^3})$) such that all scores can be plotted on the same (standardised) scale. The l_{th} eigenvalue at level 1 (gender: male or female)

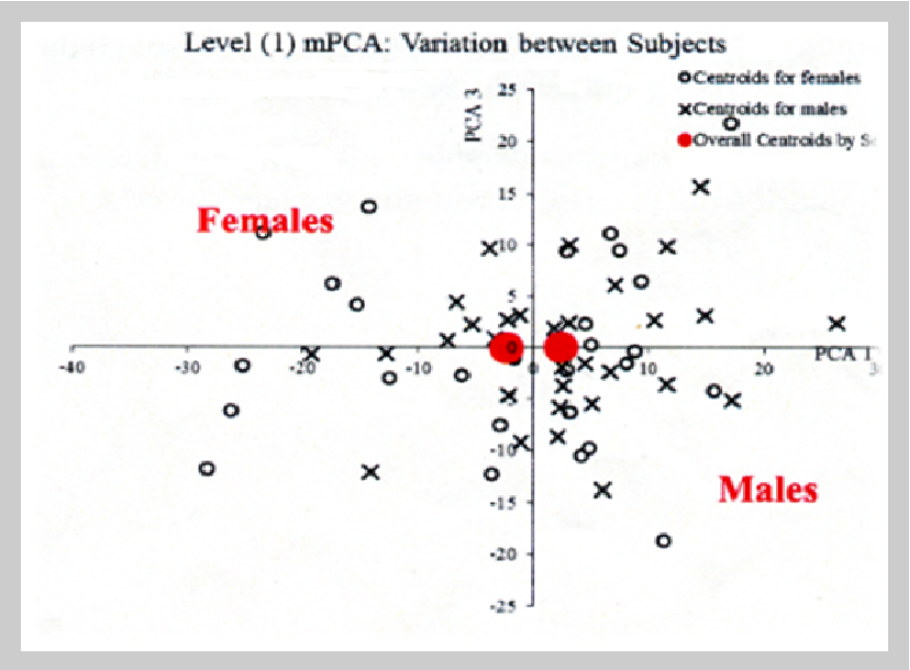


Figure 6.10: mPCA Level (1): Variation between subjects for PCA 1 and PCA 3

is denoted by λ_1^l with associated eigenvector μ_1^l, whereas the l_{th} eigenvalue at level 2 (between-subject variation) is denoted. Figure 6.12 shows that men and women smile in the same way; however, the variation in the component scores is very large. Again there is considerable variation by subject in the manner in which smiles dynamically change by subject, which cannot be excluded from the natural rest mouth shape. A clear cycle is observed in standardised scores at an appropriate level of the mPCA model which reflects these phases in these initial calculations. Figures 6.17 and 6.18 show some evidence of a cycle by phase of smile, especially for the component score PCA 2 for single-level PCA. Figures 6.19 and 6.20 arguably also show some evidence of a 'cycle' or variation by phase of a smile in the component scores in both PCA 1 and PCA 2, which is not what we would expect because this level is intended to capture this source of variation between subjects and not between phases of smile $(mPCA : l_1 = 10, l_2 = 2, \text{and } l_3 = 0)$. This indicates that different people (or different groups of people, e.g., males and females) have different types of smile.

Figure 6.21 shows very strong evidence of a 'cycle' or variation by phase of smile

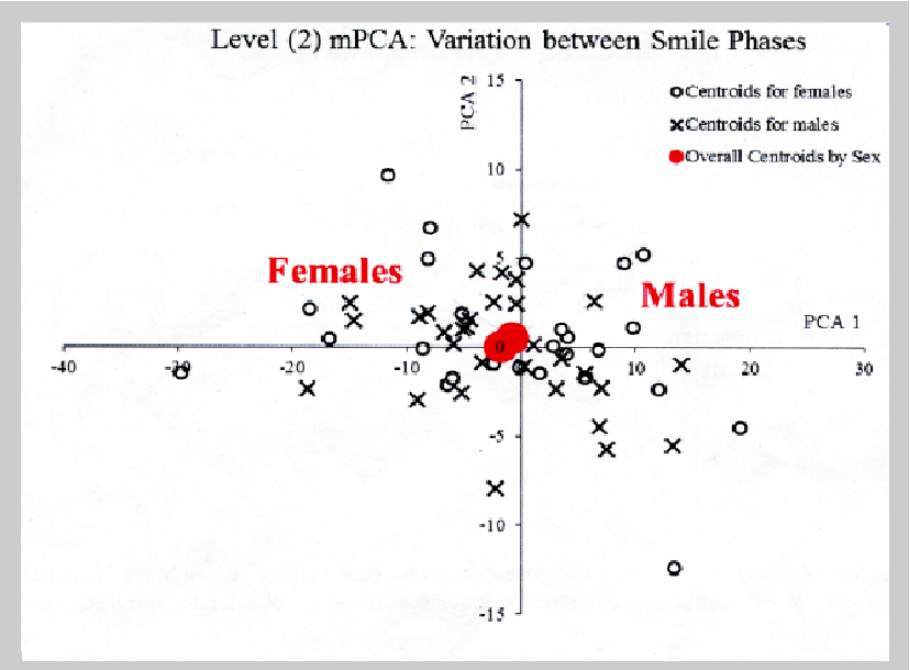


Figure 6.11: mPCA Level (2): Variation between smile phases for PCA 1 and PCA 2.

in the component scores in both PCA 1 and PCA 2, which is what we would expect because this level is intended to capture this source of variation ($mPCA : l_1 = 10$, $l_2 = 0$, and $l_3 = 0$).

In addition, Figure 6.22 shows strong evidence of a 'cycle' or variation by phase of smile in the component scores in both PCA 1 and PCA 2, which is what we would expect because this level is intended to capture this source of variation ($mPCA : l_1 = 15$, $l_2 = 3$, and $l_3 = 0$).

Figures 6.23 and 6.24 also show the same variation by phase of smile in the component scores in both PCA 1 and PCA 2 for mPCA ($mPCA : l_1 = 15$, $l_2 = 3$, and $l_3 = 0$).

Figures 6.25 and 6.26 display strong evidence of a 'cycle' or variation by phase of the smile for the component scores PCA 2 for single-level PCA; indeed, there is clear evidence of a cycle in these scores: rest pre-smile and rest post-smile. This is strong evidence that seven phases of a smile exist (rest pre-smile, onset 1, onset 2, apex, offset 1, offset 2, and rest post-smile).

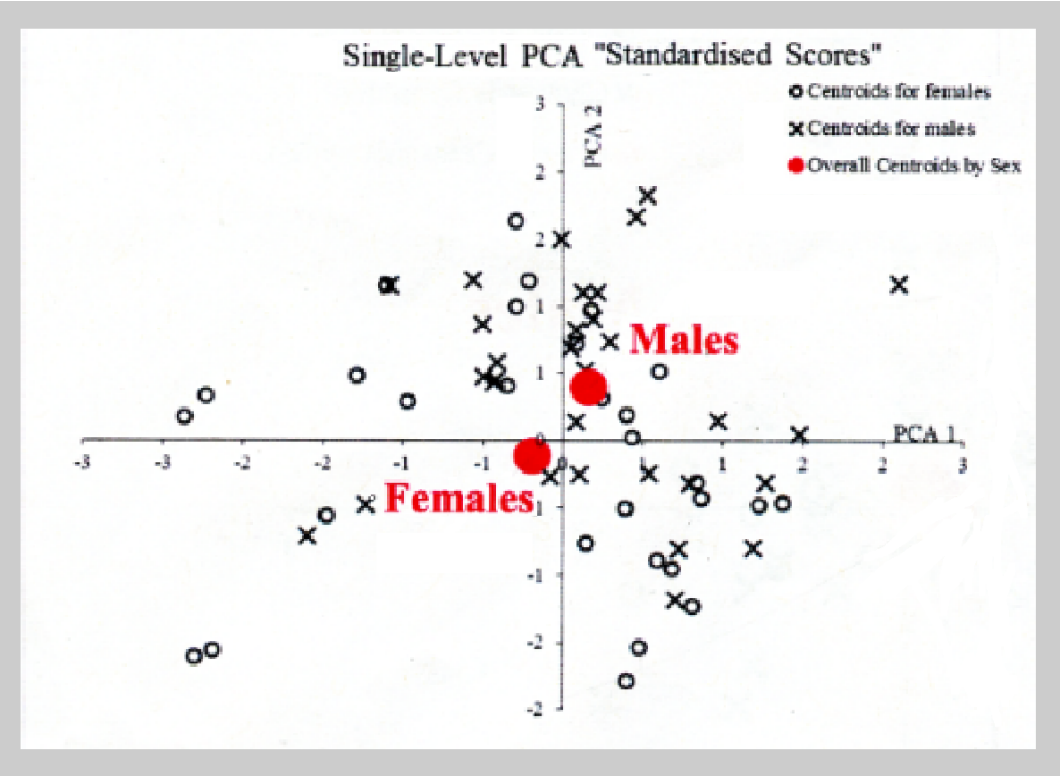


Figure 6.12: Single-Level PCA: Variation between subjects for PCA 1 and PCA 2.

Figure 6.27 again shows strong evidence of a 'cycle' or variation by phase of smile in the component scores in both PCA 1 and PCA 2, which is what we would expect because this level is intended to capture this source of variation ($mPCA : l_1 = 15$, $l_2 = 3$, and $l_3 = 0$).

Figures 6.28 and 6.29 also show some evidence of a 'cycle' or variation by phase of smile for the component scores in both PCA1 and PCA 2. However, now we have standardised the scores, this is much weaker than at level 2. Centroids of standardised scores at level 2 (i.e., between smile phases level) in each of the seven proposed phases of the smile were obtained by averaging these scores over all subjects in each phase. The results are shown in Figures Figure 6.17 to 6.29. We see that there is clear evidence of a seven-phase cycle in these scores: rest pre-smile, onset 1 (acceleration), onset 2 (deceleration), apex, offset 1 (acceleration), offset 2 (deceleration), and rest

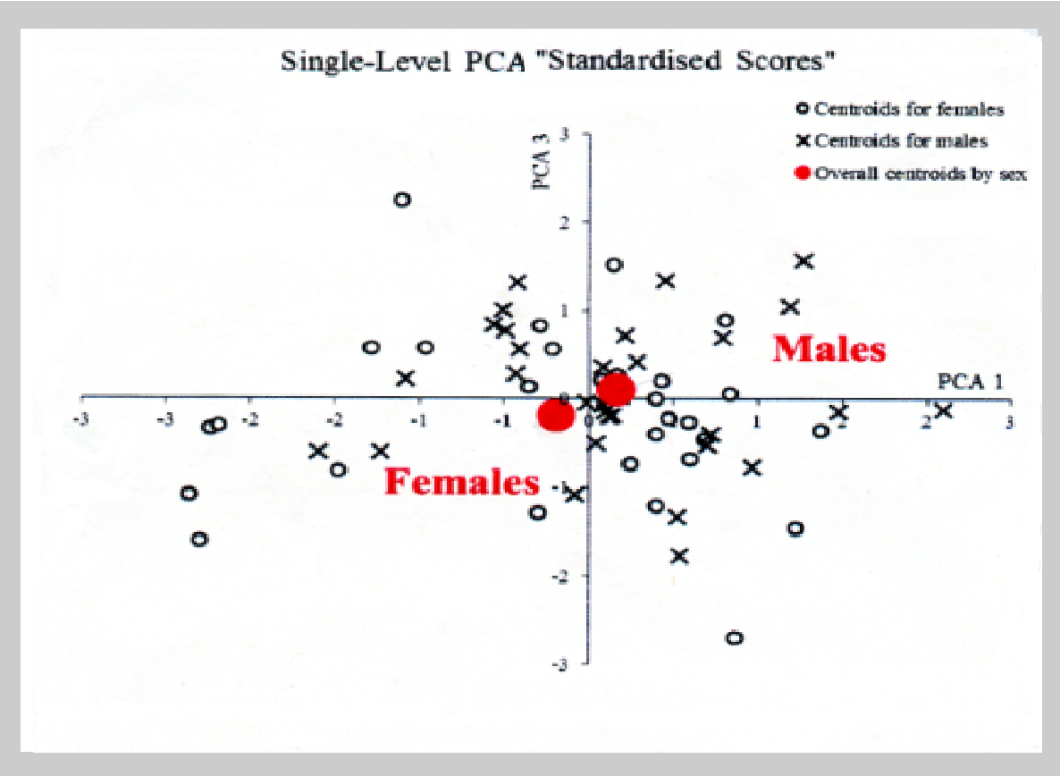


Figure 6.13: Single-Level PCA: Variation Between Subjects for PCA 1 and PCA 3.

post-smile.

6.4 Summary

Multilevel statistical shape models (mPCA) can be used to analyse the effect of naturally occurring groups in a population of individuals for data of smile dynamics. An investigation of eigenvalues indicated that each source of variability was significant in both between and within-group sources of variation. The formalism for mPCA has been described briefly, and we have shown that mPCA allows us to model variations at different levels of structure in the data (between-group and within-group levels). Conventional PCA displayed signs of clustering, but this was ambiguous for the most part. However, the mPCA makes it possible to separate the data in multilevel. The major

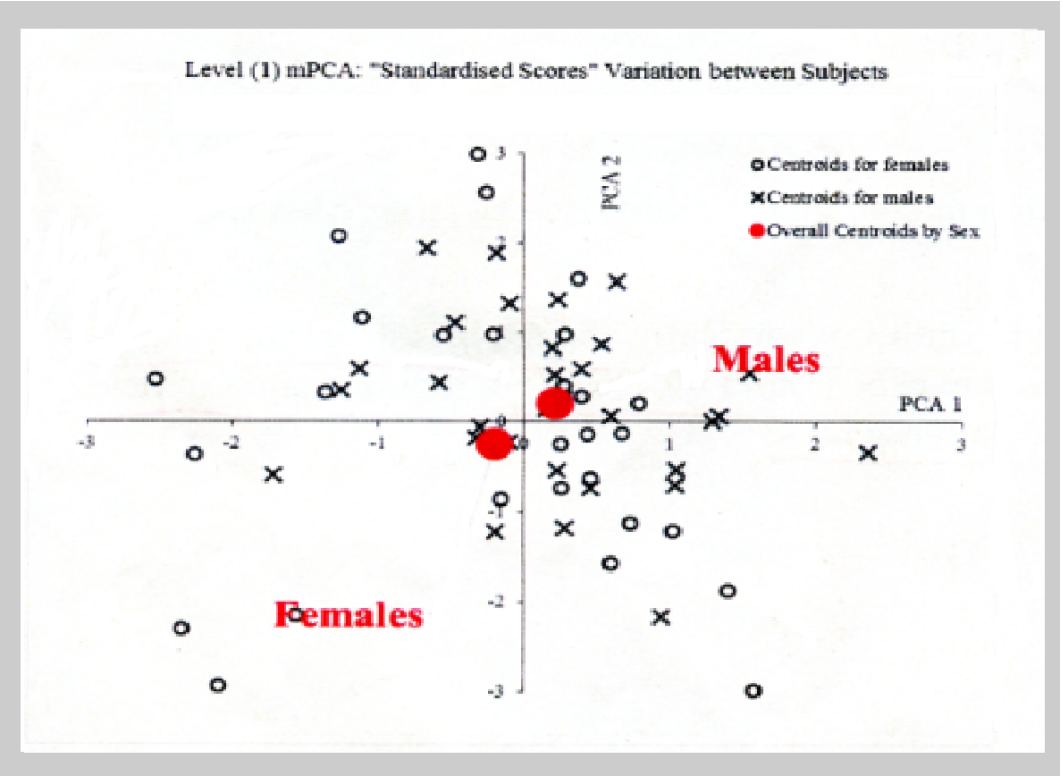


Figure 6.14: mPCA Level (1): Variation between subjects for PCA 1 and PCA 2

modes of variation of the standard PCA and the mPCA in the first three modes are shown in this chapter. We conclude that mPCA correctly decomposes sources of variation due to subjects' variation (shape, gender), facial expression (smile and neutral), and smile phases (onset, apex and offset). This study has therefore been rigorous initial test of the usefulness of mPCA in terms of the modelling of shape.

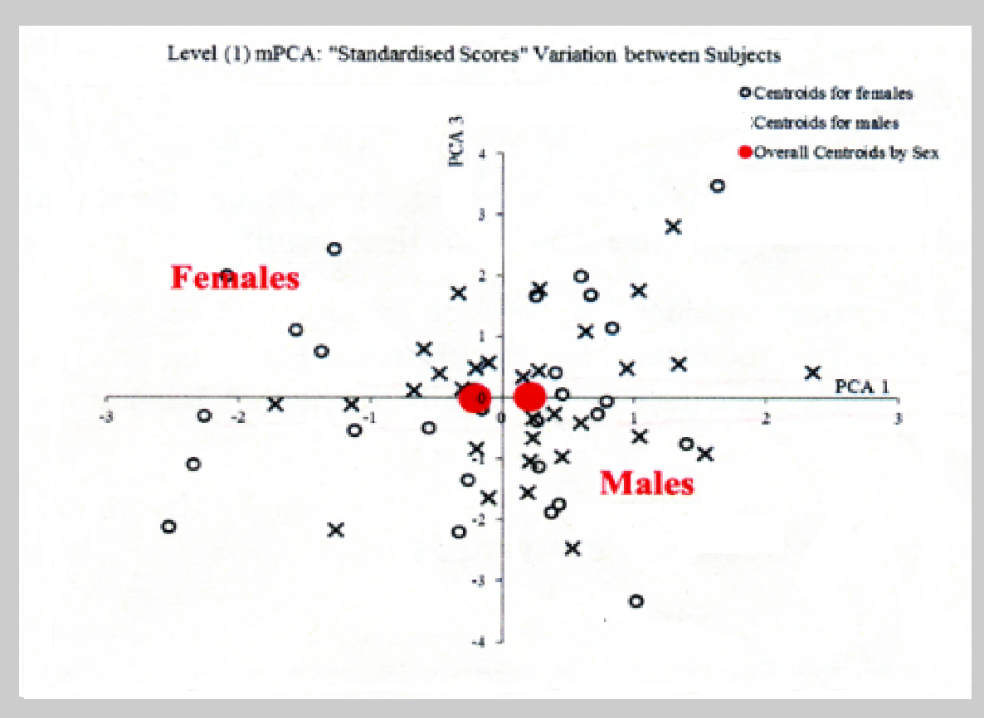


Figure 6.15: mPCA Level (1): Variation between subjects for PCA 1 and PCA 3

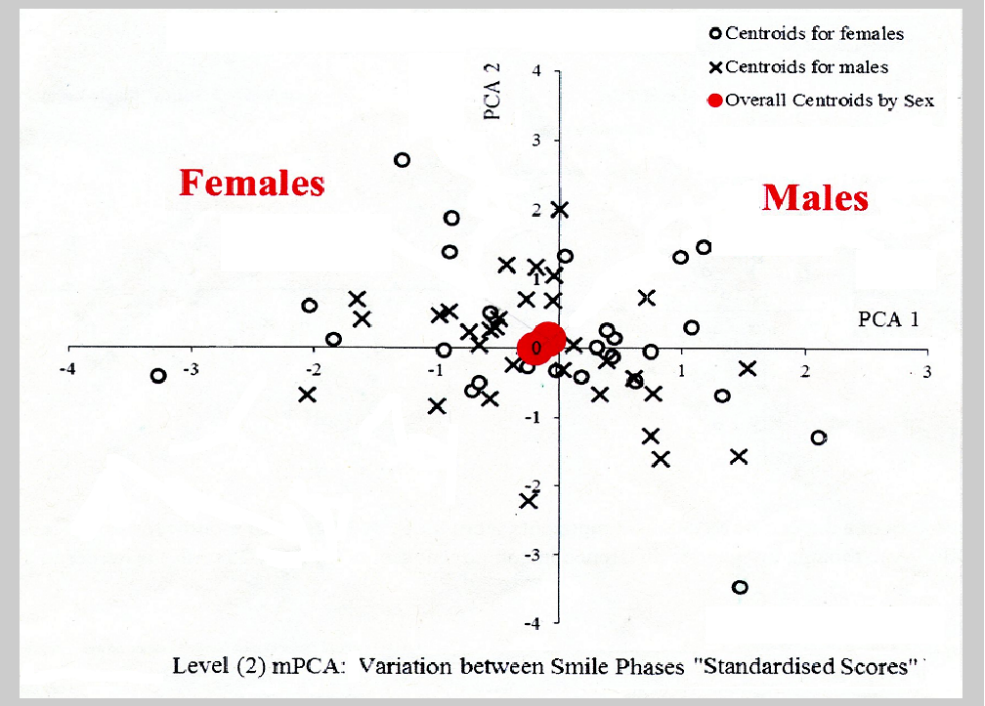


Figure 6.16: mPCA Level (2): Variation between Smile Phases for PCA 1 and PCA 2.

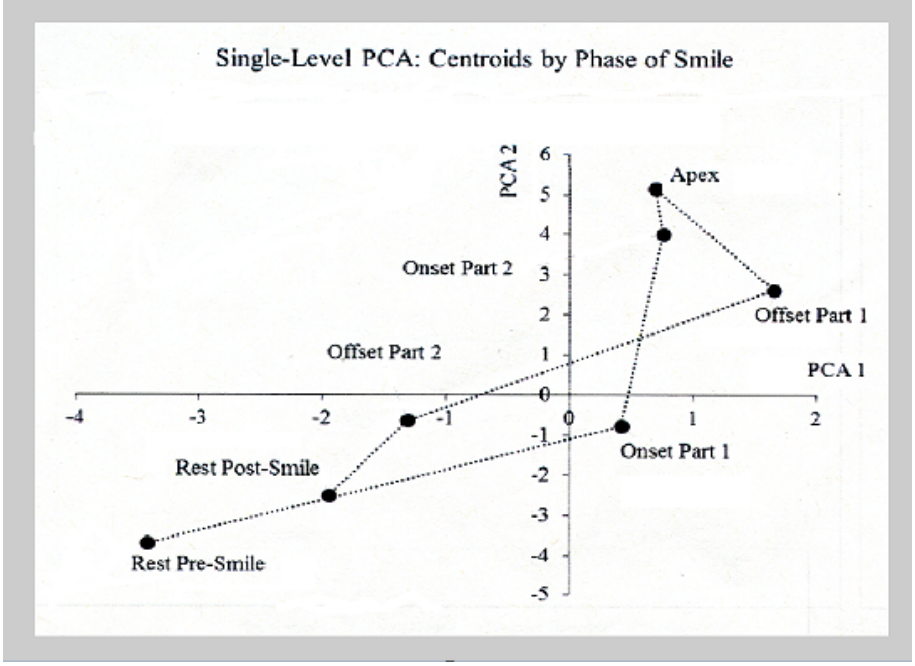


Figure 6.17: Single-level PCA: Centroids by phase of smile for PCA 1 and PCA 2.

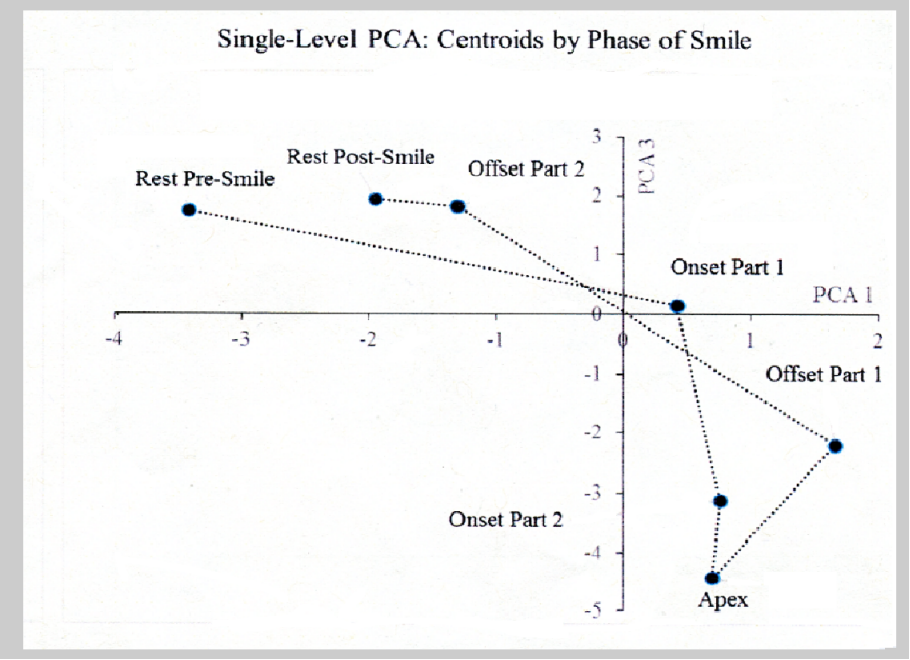


Figure 6.18: Single-level PCA: Centroids by Phase of smile for PCA 1 and PCA 3.

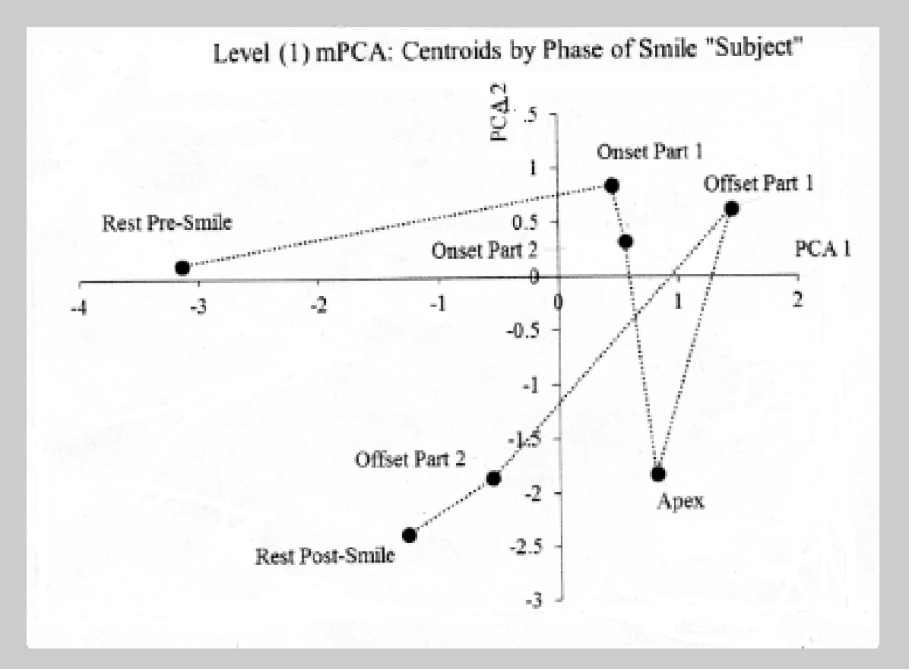


Figure 6.19: mPCA Level (1): Variation between smile phases, centroids by phase of smile.

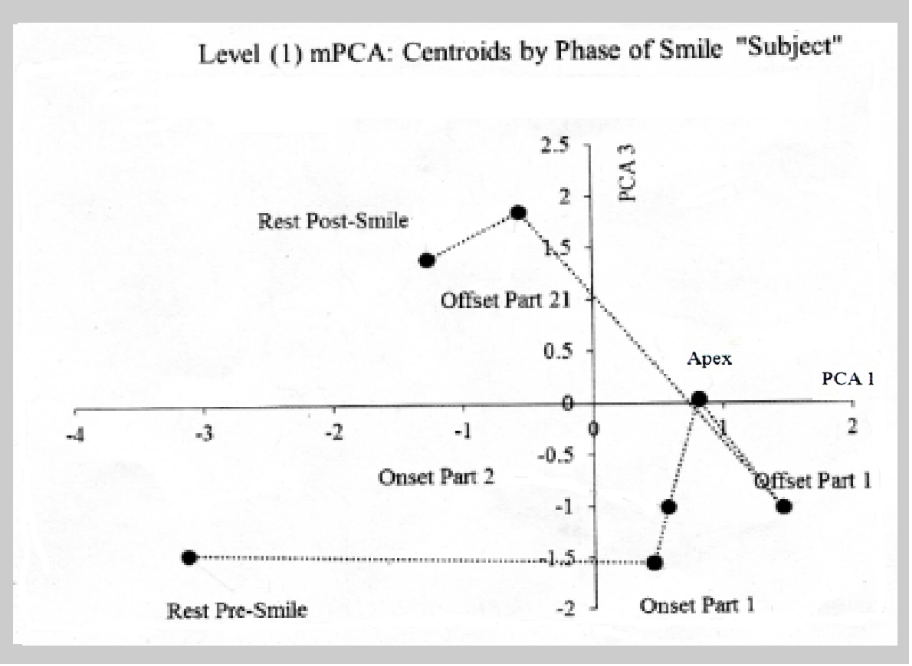


Figure 6.20: mPCA Level (1): Variation between smile phases, centroids by phase of smile.

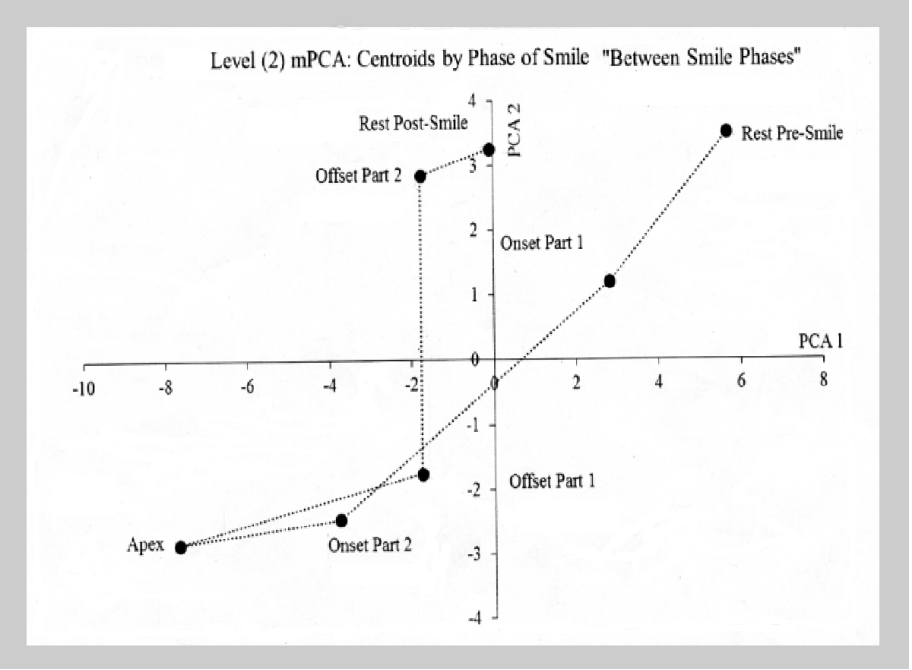


Figure 6.21: mPCA Level (2): Variation between smile phases, centroids by phase of smile.

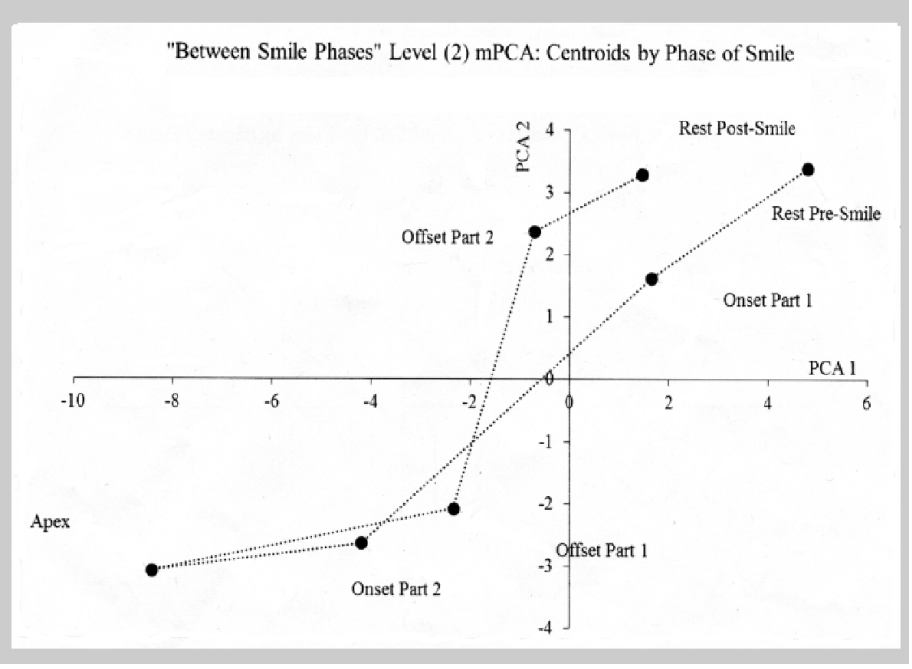


Figure 6.22: mPCA Level (2): Variation between smile phases, centroids by phase of smile.

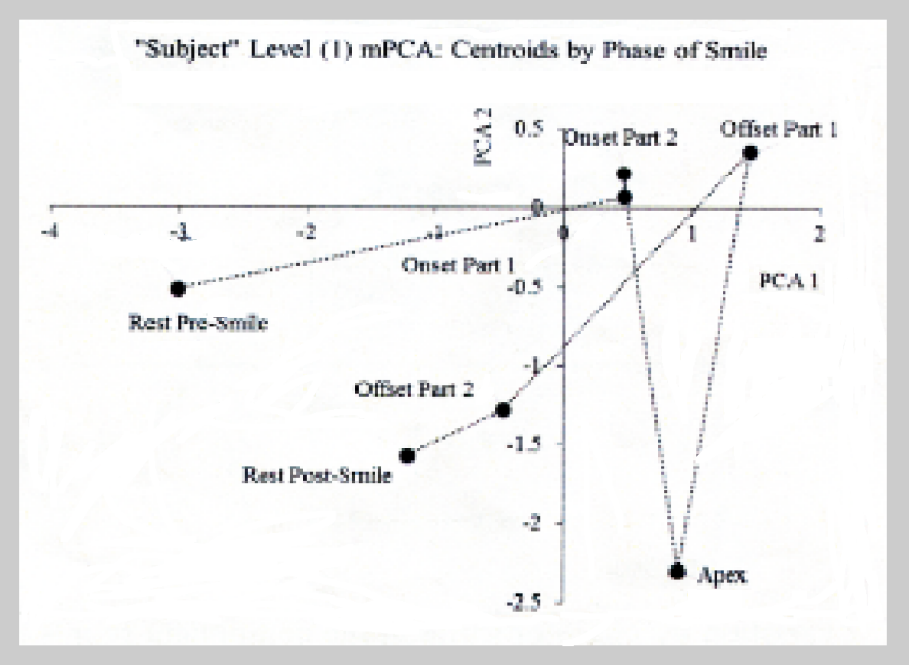


Figure 6.23: mPCA Level (2): Variation between smile phases, centroids by phase of smile.

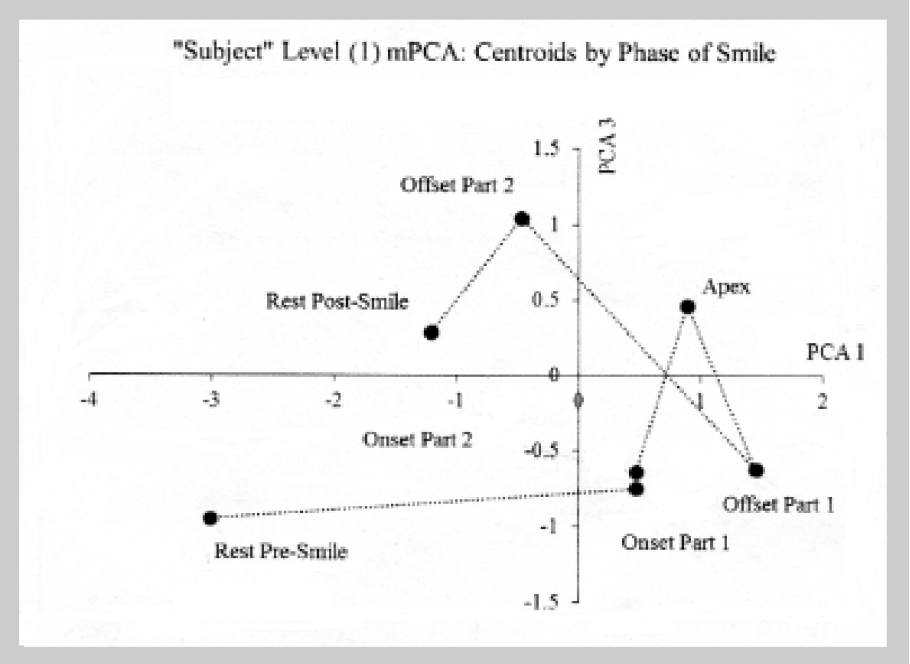


Figure 6.24: mPCA Level (2): Variation between smile phases, centroids by phase of smile.

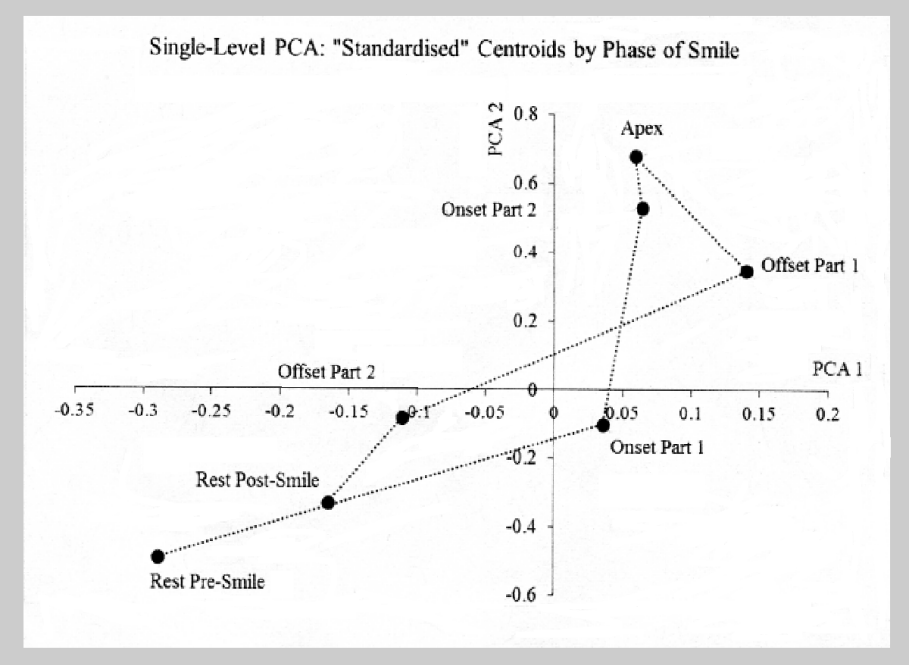


Figure 6.25: mPCA Level (2): Variation between smile phases, centroids by phase of smile.

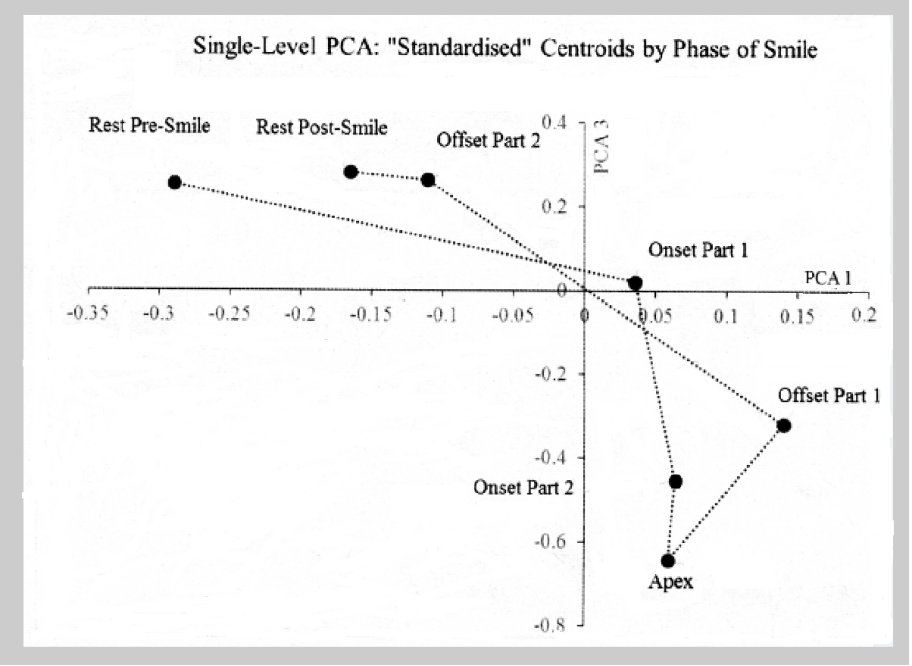


Figure 6.26: mPCA Level (2): Variation between smile phases, centroids by phase of smile.

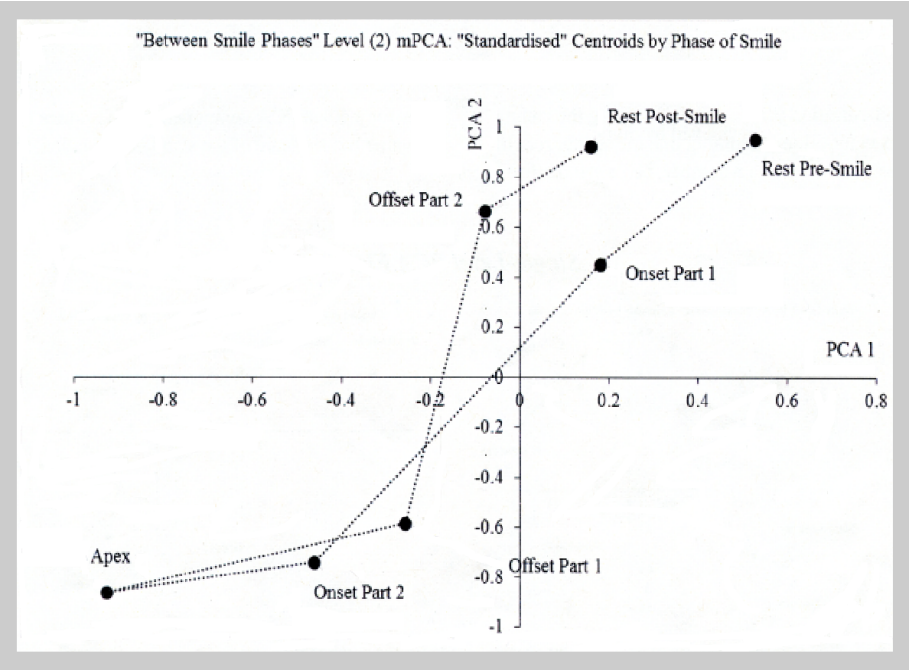


Figure 6.27: Centroids of standardised "scores" from mPCA at level 2 ("between smile phases").

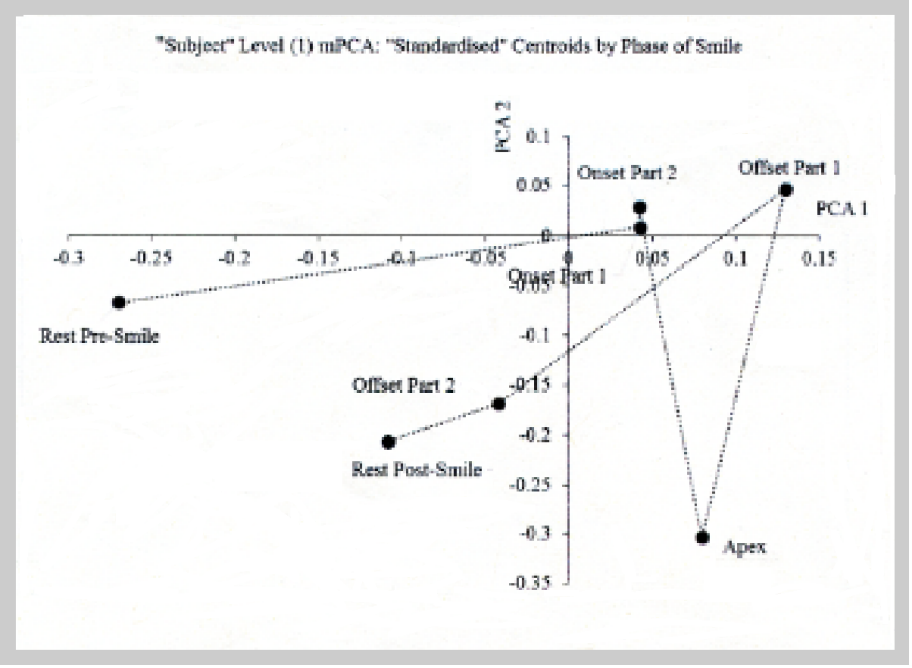


Figure 6.28: Centroids of standardised "scores" from mPCA at level 1 ("between smile phases").

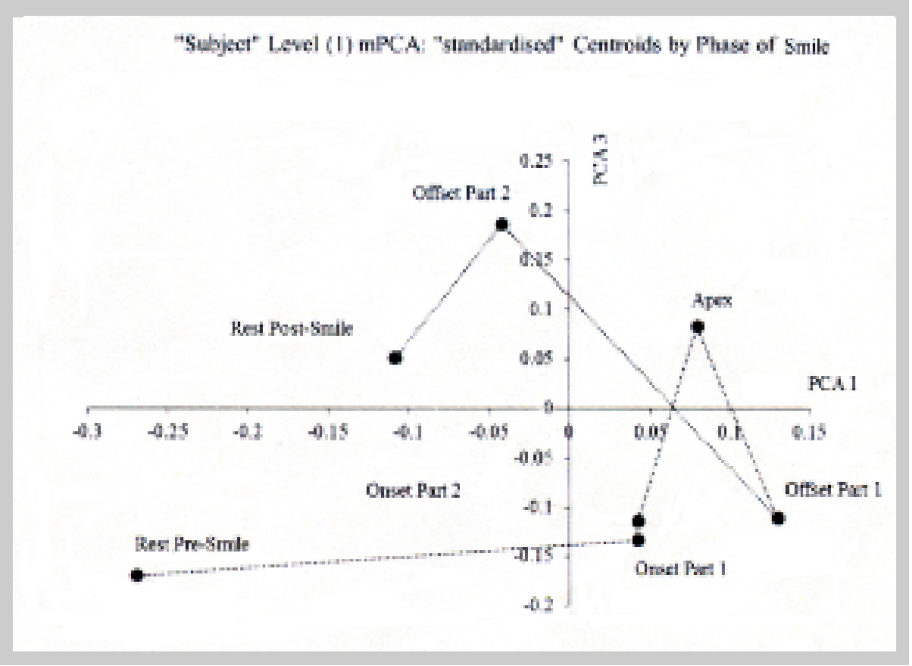


Figure 6.29: mPCA Level (1): Variation between smile phases, centroids by phase of smile.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

Facial ageing is a challenging area of research; every human ages in a different manner under the influence of a wide variety of biological and environmental factors. Facial ageing includes the progress of all factors that lead to changes to the skull, muscles, tissues, cells, and facial wrinkles. Research related to facial motion is a relatively new yet growing area of study. Based on the fact that facial dynamics are a viable biometric feature, and therefore suitable for identity recognition, the aim of this work was to investigate changes in these dynamics due to facial ageing by utilising statistical models to analyse such dynamics, and by using a novel 3D dynamic database collected over several years in an effort to overcome the current limitations associated with 2D data. It thus contributes to the research in a variety of areas, including, but not limited to, security, medical, computer graphics, and forensics applications. Based on this information, the present thesis attempts to analyse the facial motion and to demonstrate that smiles change with age, and their characteristics differ between males and females, through a series of experiments. As opposed to static analysis, research related to facial motion is a relatively growing area of study. The work presented in this thesis takes advantage of a new longitudinal 3D face database that was created with the aim to study facial dynamics more specifically, and to consider individual-specific models with ageing. For these purposes, 3D videos of face scans were collected with larger spans of time between samples of individuals than in traditional ageing data-

bases, including longitudinal spans over nine years. In addition to the scans, the data was collected for individuals, including information such as age and gender, which can be used to study individual-specific ageing factors.

7.1.1 Original Contributions

In this thesis, we have studied the facial dynamics in ageing. The main contributions presented in this thesis can be summarised as follows.

1. In Chapter 3, we introduced the the world's first 4D longitudinal database to include 3D video scans of 316 subjects performing different facial movements over years such as smiles, sentences, and single words. 3D frames were created from a seven BMPs and information was calibrated using 3DMD's MStereo program; in addition, a fully-automatic cleaning process was used for cleaning the meshes and creating a single-view texture map (UTM) from the raw data. This data was used for modelling the facial dynamics in this research to overcome the mesh artefacts which were produced by capture systems. The Pre-processing step was the one of the significant tasks in this research.

    - A tracking and inter-subject registration was performed using the method developed by [202]. This tracking and inter-subject registration approach allowed for the building of statistical models, which was necessary for conducting our research.
    - Moreover, some experiments were also conducted using cubic interpolation techniques, and compared to the linear interpolation method, the latter proved to have the lowest error rates in most cases. Another advantage of the linear interpolation method over the quadratic and cubic methods is that linear interpolation is the simplest method.
    - This research tackled the missing data gaps in the current database by developing a method to predict the length of the sequence in the missing year

and using linear interpolation to estimate the frames in the sequence at any given age.

2. In Chapter 4, different approaches in relation to data description have been considered, including Active Shape Models (ASMs), which provide shape information, and Active Appearance Models (AAMs), which produce texture information. The overall objectives were to extract the dynamic of human faces in the form of dynamic features by projecting the training data into the model and using the dynamic features extracted from the first mode of variation (PC1) to compare the intra-subject dynamics. We presented a performance analysis of subspace-based specific and general models to solve the problem of facial dynamics over age. The experiments were performed using 3D dynamic datasets over time obtained from the longitudinal database. This dataset demonstrate the efficiency on both female and male facial dynamics. In the case of females and age, mPCA achieves greater accuracy compared to PCA.

3. In Chapter 5, smile segmentation is automatically categorised into three phases, namely, onset, apex, and offset. Then, a set of dynamic features were computed from the facial features and the static features for comparison. The aim of this investigation was to measure the effects of age and gender related facial changes using dynamic 3D facial scans. Experiments were carried out on the subjects, who were divided into two groups by age (15-30 years and 31-60 years). Each group was further subdivided by gender. Two-way multivariate analysis of variance (MANOVA) of these features demonstrated that statistically significant age and gender related differences could be detected. In Chapter 5, a new set of 3D geometric features are derived from curvedness measures obtained along the geodesic path between 3D facial anthropometric landmarks. These features describe the face surface shape variation relying on the biological landmarks rather than the simple anthropometric measurements, which provide poor surface information. The impact of these features on face morphology categorisation is

tested compared to anthropometric measurements. We also showed that 3D facial dynamics provide more useful information than static features for the characterisation of smiles.

4. In Chapter 6, we introduced the first time application of the Multilevel Principle Component Analysis (mPCA) method to analyse 3D dynamic smile data. This data has naturally occurring multilevel clustering of the subjects by both age and gender groups (between- group variation). In addition, every subject has their own smile expression (within- group variation), which mPCA uniquely exploits, and is an improvement on more traditional PCA methods. We showed that there is clear evidence that seven phases of a smile indeed exist. The experiments were carried out and tested on 3D meshes of 80 individuals that were taken from the 4D facial ageing database.

The experiments were limited to the dynamics of smiles, as the problem with the data is the time factor, given that pre-processing such a database takes time. Based on the results, it can be concluded that smiles change with age, and their characteristics differ between males and females. We studied the influence of gender and conclude that gender plays a crucial role in facial feature analysis. We analysed the rich sources of information present in the 3D dynamic features of smiles to provide more insight into the patterns of smile dynamics. The sources of temporal information that are investigated are various dynamics of lip movement, which are analysed to extract the descriptive features. We evaluated the dynamic features on closed-mouth smiles of 80 subjects of both genders.

7.2 Future Work

Looking forward, it is clear that this investigation is expandable and can serve as a basis for additional research on facial dynamical analysis as people age.

This data has a potential use as an effective biometric for person identification. There is also evidence that dynamic information is not redundant and may be beneficial for various aspects of face processing, including age, gender, and identity recognition.

In future investigations, it will be possible to use different regions of the face, such as the eyelids and cheeks, to evaluate which dynamic features are more related to age, and it will be possible to apply our methodology to investigate other factors beyond age, such as race and medical conditions[35, 2, 4, 3]. Future research can focus on modelling the effects of ethnicity, gender, age, genetic information, or diseases (e.g., effects perhaps previously hidden in the "5% of variation") on facial shape or appearance using mPCA in terms of modelling of the 3D facial expression shape.

Appendix A: Meshlab CleanMesh Script

<!DOCTYPE FilterScript>

<FilterScript>

CleanMeshScript

Meshlabserver script

This script will Remove non Manifold Vertices and

Edges from the mesh and then close the holes in the mesh.

<filter name="Remove Duplicate Faces"/>

<filter name="Remove Duplicate Faces"/>

<filter name="Select non Manifold Vertices"/>

<filter name="Delete Selected Vertices"/>

<filter name="Select non Manifold Edges "/>

<filter name="Delete Selected Vertices"/>

<filter name="Close Holes">

name="MaxHoleSize" tooltip="The size is expressed as number of edges composing the hole boundary" description="Max size to be closed " type="RichInt" value="30"/>

<Param name="Selected" tooltip="Only the holes with at least one of the boundary faces selected are closed" description="Close holes with selected faces" type="RichBool" value="true"/>

<Param name="NewFaceSelected" tooltip="After closing a hole the faces that have

been created are left selected. Any previous selection is lost. Useful for example for smoothing the newly created holes." description="Select the newly created faces" type="RichBool" value="true"/>
</filter>
</FilterScript>

Appendix B: 3D Imaging System Quick Guide

3D Imaging System - Quick Guide

Turn projectors and panels ON and room lights OFF
"Recording" note on door
Microphone
Face clear from Glasses, Jewellery and Hair (if possible)

Calibration

- Calibrate in mornings and after lunch during scan capturing days and/or before using the system
- *Profile - Calibration.reg*
  - Crosshair may be one dot max from centre
  - Ensure Gen3D Controls is in calibrate mode
  - Snap and Trigger Pulse once for the board positions
    - BMP exports for position 1 and position 2
    - Run move_calib_pos1.bat and move_calib_pos2.bat after each export
      - Run calib.bat to finish calibration

Snapshot Check

- Check before capturing each participant
- *Profile - BMP_Capture.reg*
  - Ensure Gen3D Controls is in normal capture mode
  - Snap and Trigger Pulse once
    - Quick_BMP_Check.py (Streamlined Processing)
    - MStereo_Master.bat (test folder)
      - Open TSB file for check

Capturing

- *Profile - BMP_Capture.reg*
- Run Expression_Capture.py
  - Enter participant ID
    - Trigger Pulses ON and position participant for expression
    - Trigger OFF
    - RECORD
      - Trigger ON **THIS WILL START CAPTURING**
      - Trigger OFF to stop capturing
      - Click on Stop (audio will still record prior to this being pressed)
      - Click associated button on the Expression GUI for each capture

Trigger Pulse ON and OFF before each recording to prevent camera sync problems

Appendix C: Participant Consent Form

CONSENT FORM

Title of Project: **3D Videos to Assess Facial Movement**

Name of Researcher: Hashmat Popat
Clinical Lecturer/Specialist Registrar in Orthodontics
University Dental Hospital
Heath Park
Cardiff
CF14 4XY

Dr David Marshall
Reader
School of Computer Science
Cardiff University
CF24 3AA

Please tick check box

1. I confirm that I have read and understand the information sheet *Version 2, 18th June 2009* for the above study and have had the opportunity to ask questions. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason. ☐
3. I am willing to be contacted for a follow up scan. ☐
4. I agree to take part in the above study and have my videos published on a password protected webpage. ☐
5. I agree to take part in the above study and have my videos published online/in-print. ☐

Name of Participant	Date	Signature
Age in Years of Participant		
Name of Person taking consent (if different from researcher)	Date	Signature
Researcher	Date	Signature

Internal Use: File ID stored on Disc Participant 's Contact School

Bibliography

[1] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12):2037–2041, 2006.

[2] Ala Al Ali, Stephen Richmond, Hashmat Popat, Rebecca Playle, Timothy Pickles, Alexei I Zhurov, David Marshall, Paul L Rosin, John Henderson, and Karen Bonuck. The influence of snoring, mouth breathing and apnoea on facial morphology in late childhood: A three-dimensional study. *BMJ Open*, 5(9):e009027, 2015.

[3] Ala Al Ali, Stephen Richmond, Hashmat Popat, Arshed M Toma, Rebecca Playle, Timothy Pickles, Alexei I Zhurov, David Marshall, Paul L Rosin, and John Henderson. A three-dimensional analysis of the effect of atopy on face shape. *The European Journal of Orthodontics*, 36:506–511, 2014.

[4] Ala Al Ali, Stephen Richmond, Hashmat Popat, Arshed M Toma, Rebecca Playle, Alexei I Zhurov, David Marshall, Paul L Rosin, and John Henderson. The influence of asthma on face shape: A three-dimensional study. *The European Journal of Orthodontics*, 2014.

[5] Khtam Al-Meyah, David Marshall, and Paul L Rosin. 4d analysis of facial ageing using dynamic features. *Procedia Computer Science*, 112:790–799, 2017.

[6] A Midori Albert and Karl Ricanek Jr. The morph database: investigating the effects of adult craniofacial aging on automated face-recognition technology. *Forensic Science Communications*, 10(2), 2008.

[7] Kristina Aldridge, Ian D George, Kimberly K Cole, Jordan R Austin, T Nicole Takahashi, Ye Duan, and Judith H Miles. Facial phenotypes in subgroups of pre-

pubertal boys with autism spectrum disorders are correlated with clinical phenotypes. *Molecular Autism*, 2(1):15, 2011.

[8] Amal Seralkhatem Osman Ali, Vijanth Sagayan, Aamir Malik Saeed, Hassan Ameen, and Azrina Aziz. Age-invariant face recognition system using combined shape and texture features. *IET Biometrics*, 4(2):98–115, 2015.

[9] P Danny Allen, Jim Graham, Damian JJ Farnell, Elizabeth J Harrison, Reinhilde Jacobs, Kety Nicopolou-Karayianni, Christina Lindh, Paul F van der Stelt, Keith Horner, and Hugh Devlin. Detecting reduced bone mineral density from dental radiographs using statistical shape models. *IEEE Transactions on Information Technology in Biomedicine*, 11(6):601–610, 2007.

[10] Thomas R Alley and Katherine A Hildebrandt. Determinants and consequences of facial aesthetics. *Social and applied aspects of perceiving faces*, pages 101–140, 1988.

[11] Fares Alnajar, Caifeng Shan, Theo Gevers, and Jan-Mark Geusebroek. Learning-based encoding with soft assignment for age estimation under unconstrained imaging conditions. *Image and Vision Computing*, 30(12):946–953, 2012.

[12] Zara Ambadar, Jonathan W Schooler, and Jeffrey F Cohn. Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological science*, 16(5):403–410, 2005.

[13] Per Rønsholt Andresen, Fred L Bookstein, K Couradsen, Bjarne Kjaer Ersboll, Jeffrey L Marsh, and Sven Kreiborg. Surface-bounded growth modeling applied to human mandibles. *IEEE transactions on medical imaging*, 19(11):1053–1063, 2000.

[14] C Anitha, MK Venkatesha, and B Suryanarayana Adiga. A survey on facial expression databases. *International Journal of Engineering Science and Technology*, 2(10):5158–5174, 2010.

[15] Djamila Aouada, Shuo Feng, and Hamid Krim. Statistical analysis of the global geodesic function for 3d object classification. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 1, pages I–645. IEEE, 2007.

[16] Greg Attwood, Gill Dyer, and Gordon Skipworth. *Statistics 1*. Bath, UK: Heinemann Educational., New ed. 2000.

[17] Andrew J Aubrey, David Marshall, Paul L Rosin, Jason Vendeventer, Douglas W Cunningham, and Christian Wallraven. Cardiff conversation database (ccdb): A database of natural dyadic conversations. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 277–282. IEEE, 2013.

[18] Bilal M Ayyub and Richard H McCuen. Numerical methodsfor engineers, 1996.

[19] Werner Bailer. A comparison of distance measures for clustering video sequences. In *Database and Expert Systems Application, 2008. DEXA'08. 19th International Workshop on*, pages 595–599. IEEE, 2008.

[20] Aziz Umit Batur and Monson H Hayes. Adaptive active appearance models. *IEEE Transactions on Image Processing*, 14(11):1707–1721, 2005.

[21] Salah Eddine Bekhouche, Abdelkrim Ouafi, Abdelmalik Taleb-Ahmed, Abdenour Hadid, and Azeddine Benlamoudi. Facial age estimation using bsif and lbp. *arXiv preprint arXiv:1601.01876*, 2016.

[22] Peter N. Belhumeur, João P Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):711–720, 1997.

[23] Shenglan Ben, Jiansheng Chen, and Guangda Su. Piecewise linear aging function for facial age estimation. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 2753–2756. IEEE, 2009.

[24] Lanthao Benedikt. *Using 3D facial motion for biometric identification*. PhD thesis, Cardiff University, 2009.

[25] Lanthao Benedikt, Darren Cosker, Paul L Rosin, and David Marshall. Assessing the uniqueness and permanence of facial actions for use in biometric applications. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 40(3):449–460, 2010.

[26] Lanthao Benedikt, Vedran Kajic, Darren Cosker, Paul L Rosin, and A David Marshall. Facial dynamics in biometric identification. In *BMVC*, pages 1–10, 2008.

[27] Christopher M Bishop. Pattern recognition. *Machine Learning*, 128:1–58, 2006.

[28] Soma Biswas, Gaurav Aggarwal, Narayanan Ramanathan, and Rama Chellappa. A non-generative approach for face recognition across aging. In *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, pages 1–6. IEEE, 2008.

[29] Michael J Black and Allan D Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. In *European conference on computer vision*, pages 329–342. Springer, 1996.

[30] Michael J Black, Yaser Yacoob, Allan D Jepson, and David J Fleet. Learning parameterized models of image motion. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 561–567. IEEE, 1997.

[31] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194. ACM Press/Addison-Wesley Publishing Co., 1999.

[32] Volker Blanz and Thomas Vetter. Face recognition based on fitting a 3d morphable model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1063–1074, 2003.

[33] Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 11(6):567–585, 1989.

[34] Prosenjit Bose, Anil Maheshwari, Chang Shu, and Stefanie Wuhrer. A survey of geodesic paths on 3d surfaces. *Computational Geometry*, 44(9):486–498, 2011.

[35] Marko Bozic, Chung How Kau, Stephen Richmond, Nataša Ihan Hren, Alexei Zhurov, Marjana Udovič, Staša Melink, and Maja Ovsenik. Facial morphology of Slovenian and Welsh white populations using 3-dimensional imaging. *The Angle Orthodontist*, 79(4):640–645, 2009.

[36] Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision*, 64(1):5–30, 2005.

[37] D Michael Burt and David I Perrett. Perception of age in adult caucasian male faces: Computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society of London B: Biological Sciences*, 259(1355):137–143, 1995.

[38] Ya Chang, Marcelo Vieira, Matthew Turk, and Luiz Velho. Automatic 3d facial expression analysis in videos. In *Analysis and Modelling of Faces and Gestures*, pages 293–307. Springer, 2005.

[39] Steven C Chapra and Raymond P Canale. *Solutions Manual to Accompany Numerical Methods for Engineers with Programming and Software Applications*. McGraw-Hill, 1998.

[40] Weiyang Chen, Wei Qian, Gang Wu, Weizhong Chen, Bo Xian, Xingwei Chen, Yaqiang Cao, Christopher D Green, Fanghong Zhao, Kun Tang, et al. Three-dimensional human facial morphologies as robust aging markers. *Cell Research*, 25(5):574–587, 2015.

[41] Yi-Lei Chen and Chiou-Ting Hsu. Subspace learning for facial age estimation via pairwise age ranking. *IEEE Transactions on Information Forensics and Security*, 8(12):2164–2176, 2013.

[42] Patil Chetan, Pradeep Tandon, Gulshan K Singh, Amit Nagar, Veerendra Prasad, and Vinay K Chugh. Dynamics of a smile in different age groups. *The Angle Orthodontist*, 83(1):90–96, 2012.

[43] Sung Eun Choi, Youn Joo Lee, Sung Joo Lee, Kang Ryoung Park, and Jaihie Kim. A comparative study of local feature extraction for age estimation. In *Control Automation Robotics & Vision (ICARCV), 2010 11th International Conference on*, pages 1280–1284. IEEE, 2010.

[44] Sung Eun Choi, Youn Joo Lee, Sung Joo Lee, Kang Ryoung Park, and Jaihie Kim. Age estimation using a hierarchical classifier based on global and local facial features. *Pattern Recognition*, 44(6):1262–1281, 2011.

[45] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. Meshlab: an open-source mesh processing tool. In *Eurographics Italian Chapter Conference*, volume 2008, pages 129–136, 2008.

[46] Josephine Clark Weeden, Carroll-Ann Trotman, and Julian J Faraway. Three dimensional analysis of facial movement in normal adults: Influence of sex and facial shape. *The Angle Orthodontist*, 71(2):132–140, 2001.

[47] William S Cleveland. Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*, 35(1):54–54, 1981.

[48] Jeffrey F Cohn, Lawrence Ian Reed, Tsuyoshi Moriyama, Jing Xiao, Karen Schmidt, and Zara Ambadar. Multimodal coordination of facial action, head rotation, and eye motion during spontaneous smiles. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 129–135. IEEE, 2004.

[49] Jeffrey F Cohn and Karen L Schmidt. The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, 2(02):121–132, 2004.

[50] Tim Cootes. Model-based methods in analysis of biomedical images. *Image Processing and Analysis*, pages 223–248, 2000.

[51] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor. Active appearance models. In *European conference on computer vision*, pages 484–498. Springer, 1998.

[52] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685, 2001.

[53] Timothy F Cootes, Andrew Hill, Christopher J Taylor, and Jane Haslam. Use of active shape models for locating structures in medical images. *Image and vision computing*, 12(6):355–365, 1994.

[54] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.

[55] Timothy F Cootes, Cristopher J Taylor, et al. Statistical models of appearance for computer vision, 2004.

[56] Luiza Nayara Almeida Lyra Correia, Silvia Augusta Braga Riis, Ana Claudia de Castro Ferreira Conti, Leopoldino Capelozza Filho, and Renata Rodrigues Almeida-Pedrin. Age-related changes in the Brazilian womanâs smile. *Brazilian Oral Research*, 30(1), 2016.

[57] Darren Cosker, Eva Krumhuber, and Adrian Hilton. A facs valid 3d dynamic action unit database with applications to 3d dynamic morphable facial modeling. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2296–2303. IEEE, 2011.

[58] John P Cunningham and Zoubin Ghahramani. Linear dimensionality reduction: survey, insights, and generalizations. *Journal of Machine Learning Research*, 16(1):2859–2900, 2015.

[59] David Dean, Mark G Hans, Fred L Bookstein, and Krishna Subramanyan. Three-dimensional bolton–brush growth study landmark data: ontogeny and sexual dimorphism of the bolton standards cohort. *The Cleft palate-craniofacial journal*, 37(2):145–156, 2000.

[60] Douglas Decarlo and Dimitris Metaxas. Optical flow constraints on deformable models with applications to face tracking. *International Journal of Computer Vision*, 38(2):99–127, 2000.

[61] Kenneth A Deffenbacher, Thomas Vetter, John Johanson, and Alice J O'Toole. Facial aging, attractiveness, and distinctiveness. *Perception*, 27(10):1233–1243, 1998.

[62] Mohammad Mahdi Dehshibi and Azam Bastanfard. A new algorithm for age recognition from facial images. *Signal Processing*, 90(8):2431–2444, 2010.

[63] Shyam Desai, Madhur Upadhyay, and Ravindra Nanda. Dynamic smile analysis: Changes with age. *American Journal of Orthodontics and Dentofacial Orthopedics*, 136(3):310–e1, 2009.

[64] Hamdi Dibeklioglu, Albert Ali Salah, and Theo Gevers. Like father, like son: Facial expression dynamics for kinship verification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1497–1504, 2013.

[65] Hamdi Dibeklioğlu, Fares Alnajar, Albert Ali Salah, and Theo Gevers. Combining facial dynamics with appearance for age estimation. *IEEE Transactions on Image Processing*, 24(6):1928–1943, 2015.

[66] Hamdi Dibeklioğlu, Theo Gevers, Albert Ali Salah, and Roberto Valenti. A smile can reveal your age: Enabling facial dynamics in age estimation. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 209–218. ACM, 2012.

[67] Hamdi Dibeklioğlu, Albert Ali Salah, and Theo Gevers. Are you really smiling at me? Spontaneous versus posed enjoyment smiles. In *European Conference on Computer Vision*, pages 525–538. Springer, 2012.

[68] Hamdi Dibeklioğlu, Albert Ali Salah, and Theo Gevers. Age estimation using facial expression dynamics. In *Signal Processing and Communications Applications Conference (SIU), 2013 21st*, pages 1–4. IEEE, 2013.

[69] Hamdi Dibeklioğlu, Albert Ali Salah, and Theo Gevers. Recognition of genuine smiles. *IEEE Transactions on Multimedia*, 17(3):279–294, 2015.

[70] Hamdi Dibeklioğlu, Albert Ali Salah, and Furkan Gürpınar. Measurement of facial dynamics for soft biometrics. In *Face and Facial Expression Recognition from Real World Videos*, pages 69–84. Springer, 2015.

[71] Hamdi Dibeklioğlu, Roberto Valenti, Albert Ali Salah, and Theo Gevers. Eyes do not lie: Spontaneous versus posed smiles. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 703–706. ACM, 2010.

[72] Steven T Dickens, David M Sarver, and William R Proffit. Changes in frontal soft tissue dimensions of the lower face by age and gender. *World Journal of Orthodontics*, 3(4), 2002.

[73] Manfredo P Do Carmo. *Differential Geometry of Curves and Surfaces: Revised and Updated Second Edition*. Courier Dover Publications, 2016.

[74] Jin-Keun Dong, Tai-Ho Jin, Hye-Won Cho, and Sang-Chun Oh. The esthetics of the smile: A review of some recent studies. *International Journal of Prosthodontics*, 12(1), 1999.

[75] Stephanie Drummond and Jonas Capelli Jr. Incisor display during speech and smile: Age and gender correlations. *The Angle Orthodontist*, 86(4):631–637, 2015.

[76] Chi Nhan Duong, Kha Gia Quach, Khoa Luu, Hoai Bac Le, and K Ricanek. Fine tuning age estimation with global and local facial features. In *IEEE International Conference on Acoustics, Speech and Signal Processing, Prague, Czech Republic*, pages 2032–2035, 2011.

[77] Mark Jonathan Eaton, Rhys Pullin, JJ Hensman, Karen Margaret Holford, K Worden, and Samuel Lewin Evans. Principal component analysis of acoustic emission signals from landing gear components: an aid to fatigue fracture detection. *Strain*, 47(s1):e588–e594, 2011.

[78] Gareth J Edwards, Christopher J Taylor, and Timothy F Cootes. Learning to identify and track faces in image sequences. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 260–265. IEEE, 1998.

[79] Alon Efrat, Quanfu Fan, and Suresh Venkatasubramanian. Curve matching, time warping, and light fields: New algorithms for computing similarity between curves. *Journal of Mathematical Imaging and Vision*, 27(3):203–216, 2007.

[80] Paul Ekman, Gowen Roper, and Joseph C Hager. Deliberate facial movement. *Child development*, pages 886–891, 1980.

[81] Face, Gesture Recognition Working group, et al. Fg-net aging database, 2000.

[82] Hui Fang, Phil Grant, and Min Chen. Discriminant feature manifold for facial aging estimation. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 593–596. IEEE, 2010.

[83] Leslie G Farkas. *Anthropometry of the Head and Face*. Raven Pr, 1994.

[84] Leslie G Farkas, Tania A Hreczko, John C Kolar, and Ian R Munro. Vertical and horizontal proportions of the face in young adult north american caucasians: revision of neoclassical canons. *Plastic and reconstructive surgery*, 75(3):328–337, 1985.

[85] Damian JJ Farnell, Hashmat Popat, and Stephen Richmond. Multilevel principal component analysis (mpca) in shape analysis: A feasibility study in medical and dental imaging. *Computer methods and programs in biomedicine*, 129:149–159, 2016.

[86] DJJ Farnell, Jennifer Galloway, Alexei Zhurov, Stephen Richmond, Pertti Perttiniemi, and V Katic. Initial results of multilevel principal components analysis of facial shape. In *Annual Conference on Medical Image Understanding and Analysis*, pages 674–685. Springer, 2017.

[87] Yun Fu, Guodong Guo, and Thomas S Huang. Age synthesis and estimation via faces: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 32(11):1955–1976, 2010.

[88] Yun Fu and Thomas S Huang. Human age estimation with regression on discriminative aging manifold. *IEEE Transactions on Multimedia*, 10(4):578–584, 2008.

[89] Yun Fu, Ye Xu, and Thomas S Huang. Estimating human age by manifold analysis of face pictures and regression on aging features. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 1383–1386. IEEE, 2007.

[90] Yun Fu and Nanning Zheng. M-face: An appearance-based photorealistic model for multiple facial attributes rendering. *IEEE Transactions on Circuits and Systems for Video technology*, 16(7):830–842, 2006.

[91] David S Fung. Methods for the estimation of missing values in time series. 2006.

[92] Xin Geng, Kate Smith-Miles, and Zhi-Hua Zhou. Facial age estimation by nonlinear aging pattern subspace. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 721–724. ACM, 2008.

[93] Xin Geng, Zhi-Hua Zhou, and Kate Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 29(12):2234–2240, 2007.

[94] Xin Geng, Zhi-Hua Zhou, Yu Zhang, Gang Li, and Honghua Dai. Learning from facial aging patterns for automatic age estimation. In *Proceedings of the 14th ACM international conference on Multimedia*, pages 307–316. ACM, 2006.

[95] Robert M George. *Facial geometry: graphic facial analysis for forensic artists*. Charles C Thomas Publisher, 2007.

[96] Syed Zulqarnain Gilani and Ajmal Mian. Perceptual differences between men and women: A 3d facial morphometric perspective. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 2413–2418. IEEE, 2014.

[97] Syed Zulqarnain Gilani, Kathleen Rooney, Faisal Shafait, Mark Walters, and Ajmal Mian. Geometric facial gender scoring: objectivity of perception. *PloS one*, 9(6):e99483, 2014.

[98] Mario Gonzalez-Ulloa and Eduardo Stevens Flores. Senility of the face-basic study to understand its causes and effects. *Plastic and Reconstructive Surgery*, 36(2):239–246, 1965.

[99] Gaile G Gordon. Face recognition based on depth and curvature features. In *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 808–810. IEEE, 1992.

[100] Lewis D Griffin. The intrinsic geometry of the cerebrał cortex. *J. theor. Biol*, 166:273, 1994.

[101] Guodong Guo, Yun Fu, Charles R Dyer, and Thomas S Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Transactions on Image Processing*, 17(7):1178–1188, 2008.

[102] Guodong Guo, Guowang Mu, Yun Fu, and Thomas S Huang. Human age estimation using bio-inspired features. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 112–119. IEEE, 2009.

[103] Shalini Gupta, Mia K Markey, and Alan C Bovik. Anthropometric 3d face recognition. *International journal of computer vision*, 90(3):331–349, 2010.

[104] A Ben Hamza and Hamid Krim. Geodesic matching of triangulated surfaces. *IEEE transactions on image processing*, 15(8):2249–2258, 2006.

[105] Sebastian Handrich and Ayoub Al-Hamadi. A robust method for human pose estimation based on geodesic distance features. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, pages 906–911. IEEE, 2013.

[106] Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: a review. *Medical image analysis*, 13(4):543–563, 2009.

[107] Kristoffer H Hellton and Magne Thoresen. When and why are principal component scores a good tool for visualizing high-dimensional data? *Scandinavian Journal of Statistics*, 44(3):581–597, 2017.

[108] Leonard R Herrmann. Laplacian-isoparametric grid generation scheme. *Journal of the Engineering Mechanics Division*, 102(5):749–907, 1976.

[109] https://uk.mathworks.com/help/curvefit/interpolation methods.html.

[110] https://uk.mathworks.com/help/images/ref/imhistmatch.html.

[111] https://uk.mathworks.com/matlabcentral/fileexchange/26710-smooth-triangulated mesh?focused=5150406tab=function.

[112] https://uk.mathworks.com/matlabcentral/fileexchange/6110-toolbox-fast marching.

[113] https://www.norpix.com/products/streampix/streampix.

[114] https://www.python.org.

[115] http://www.blender.org.

[116] Tim J Hutton, BR Buxton, and Peter Hammond. Dense surface point distribution models of the human face. In *Mathematical Methods in Biomedical Image Analysis, 2001. MMBIA 2001. IEEE Workshop on*, pages 153–160. IEEE, 2001.

[117] Fumitada Itakura. Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 23(1):67–72, 1975.

[118] Andrei C Jalba, Michael HF Wilkinson, Jos BTM Roerdink, Micha M Bayer, and Stephen Juggins. Automatic diatom identification using contour analysis by morphological curvature scale spaces. *Machine Vision and Applications*, 16(4):217–228, 2005.

[119] Ranjan Jana, Debaleena Datta, and Rituparna Saha. Age estimation from face image using wrinkle features. *Procedia Computer Science*, 46:1754–1761, 2015.

[120] Michael J Jones and Tomaso Poggio. Multidimensional morphable models: A framework for representing and matching object classes. *International Journal of Computer Vision*, 29(2):107–131, 1998.

[121] Thouis R Jones, Frédo Durand, and Mathieu Desbrun. Non-iterative, feature-preserving mesh smoothing. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 943–949. ACM, 2003.

[122] Heikki Junninen, Harri Niska, Kari Tuppurainen, Juhani Ruuskanen, and Mikko Kolehmainen. Methods for imputation of missing values in air quality data sets. *Atmospheric Environment*, 38(18):2895–2907, 2004.

[123] Eamonn Keogh and Chotirat Ann Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge and information systems*, 7(3):358–386, 2005.

[124] Eamonn J Keogh and Michael J Pazzani. Derivative dynamic time warping. In *Proceedings of the 2001 SIAM International Conference on Data Mining*, pages 1–11. SIAM, 2001.

[125] Ron Kimmel, Arnon Amir, and Alfred M. Bruckstein. Finding shortest paths on surfaces using level sets propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(6):635–640, 1995.

[126] Ron Kimmel and James A Sethian. Computing geodesic paths on manifolds. *Proceedings of the national academy of Sciences*, 95(15):8431–8435, 1998.

[127] Donaldson K.J. Fundamentals of fixed prosthodontics. *British Dental Journal*, 213(8):427–427, 2012.

[128] Salter EM. Kolar JC. Craniofacial anthropometry: Practical measurement of head and face for clinical, surgical, and research use. *1st ed. Springfield: Charles C Thomas Publisher*, 1997.

[129] Eva Krumhuber, Antony SR Manstead, Darren Cosker, Dave Marshall, and Paul L Rosin. Effects of dynamic attributes of smiles in human and synthetic faces: A simulated job interview setting. *Journal of Nonverbal Behavior*, 33(1):1–15, 2009.

[130] Eva Krumhuber, Antony SR Manstead, Darren Cosker, Dave Marshall, Paul L Rosin, and Arvid Kappas. Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, 7(4):730–735, 2007.

[131] YH KWON and N DA VITORIA LOBO. Age classification from facial images. *Computer vision and image understanding*, 74(1):1–21, 1999.

[132] Young Ho Kwon and Niels da Vitoria Lobo. Age classification from facial images. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 762–767. IEEE, 1994.

[133] Dehe Lai, Yewang Chen, Xiangyu Luo, Jixiang Du, and Tian Wang. Age estimation with dynamic age range. *Multimedia Tools and Applications*, 76(5):6551–6573, 2017.

[134] Andreas Lanitis. Comparative evaluation of automatic age-progression methodologies. *EURASIP Journal on Advances in Signal Processing*, 2008:101, 2008.

[135] Andreas Lanitis. A survey of the effects of aging on biometric identity verification. *International Journal of Biometrics*, 2(1):34–52, 2009.

[136] Andreas Lanitis, Chrisina Draganova, and Chris Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1):621–628, 2004.

[137] Andreas Lanitis, Christopher J Taylor, and Timothy F Cootes. Modeling the process of ageing in face images. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 131–136. IEEE, 1999.

[138] Andreas Lanitis, Christopher J. Taylor, and Timothy F Cootes. Toward automatic simulation of aging effects on face images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):442–455, 2002.

[139] Fabian Lecron, Jonathan Boisvert, Mohammed Benjelloun, Hubert Labelle, and Saïd Mahmoudi. Multilevel statistical shape models: a new framework for modeling hierarchical structures. In *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pages 1284–1287. IEEE, 2012.

[140] Marjorie Lee. Physical and structural age changes in human skin. *The Anatomical Record*, 129(4):473–493, 1957.

[141] Zhifeng Li, Unsang Park, and Anil K Jain. A discriminative model for age invariant face recognition. *Information Forensics and Security, IEEE Transactions on*, 6(3):1028–1037, 2011.

[142] Haibin Ling, Stefano Soatto, Narayanan Ramanathan, and David W Jacobs. A study of face recognition as people age. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[143] Raymond Liu and Duncan F Gillies. An estimate of mutual information that permits closed-form optimisation. *Entropy*, 15(5):1690–1704, 2013.

[144] Carlos López-Otín, Maria A Blasco, Linda Partridge, Manuel Serrano, and Guido Kroemer. The hallmarks of aging. *Cell*, 153(6):1194–1217, 2013.

[145] Yan Luximon, Roger Ball, and Lorraine Justice. The 3d chinese head and face modeling. *Computer-Aided Design*, 44(1):40–47, 2012.

[146] Aleix M Martínez and Avinash C Kak. Pca versus lda. *IEEE transactions on pattern analysis and machine intelligence*, 23(2):228–233, 2001.

[147] Baback Moghaddam and Alex Pentland. Probabilistic visual learning for object detection. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 786–793. IEEE, 1995.

[148] Ana Belen Moreno, Angel Sánchez, José Fco Vélez, and Fco Javier Díaz. Face recognition using 3d surface-extracted descriptors. In *Irish Machine Vision and Image Processing Conference*, volume 2, 2003.

[149] RJ Morris, JT Kent, KV Mardia, M Fidrich, RG Aykroyd, and A Linney. Analysing growth in faces. In *Proc. Conf. Imaging Science, Systems and Technology 1999, Las Vegas*, pages 404–410, 1999.

[150] Meinard Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

[151] Mario E Munich and Pietro Perona. Continuous dynamic time warping for translation-invariant curve alignment with applications to signature verification.

In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 108–115. IEEE, 1999.

[152] Hiroshi Murase and Shree K Nayar. Visual learning and recognition of 3-d objects from appearance. *International journal of computer vision*, 14(1):5–24, 1995.

[153] MM Naushad Ali and M Abdullah-Al-Wadud. Image enhancement using a modified histogram equalization. *Computer Applications for Web, Human Computer Interaction, Signal and Image Processing, and Pattern Recognition*, pages 17–24, 2012.

[154] SURYAKANT NIRMAL. An approach of secure face recognition using linear discriminant analysis in network. *International Research Journal of Engineering and Technology (IRJET)*, Volume: 03, Issue: 06.

[155] PAUL O'HIGGINS and Nicholas Jones. Facial growth in cercocebus torquatus: an application of three-dimensional geometric morphometric techniques to the study of morphological variation. *The Journal of Anatomy*, 193(2):251–272, 1998.

[156] Nuria Oliver, Barbara Rosario, and Alex Pentland. A bayesian computer vision system for modeling human interaction. In *ICVS*, pages 255–272. Springer, 1999.

[157] FELIPE GIOACHINO OPERTI. *INTERPOLATION STRATEGY BASED ON DYNAMIC TIME WARPING*. PhD thesis, FEDERAL DO CEAR'A University, 2015.

[158] Emma Otta. Sex differences over age groups in self-posed smiling in photographs. *Psychological Reports*, 83(3):907–913, 1998.

[159] Maja Pantic. Machine analysis of facial behaviour: Naturalistic and dynamic behaviour. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1535):3505–3513, 2009.

[160] George Papandreou and Petros Maragos. Adaptive and constrained algorithms for inverse compositional active appearance model fitting. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[161] Unsang Park, Yiying Tong, and Anil K Jain. Face recognition with temporal invariance: A 3d aging model. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–7. IEEE, 2008.

[162] Frederick I Parke. Computer generated animation of faces. In *Proceedings of the ACM annual conference-Volume 1*, pages 451–457. ACM, 1972.

[163] Eric Patterson, Amrutha Sethuram, Midori Albert, Karl Ricanek, and Michael King. Aspects of age variation in facial morphology affecting biometrics. In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pages 1–6. IEEE, 2007.

[164] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[165] Gabriel Peyré, Mickael Péchaud, Renaud Keriven, Laurent D Cohen, et al. Geodesic methods in computer vision and graphics. *Foundations and Trends® in Computer Graphics and Vision*, 5(3–4):197–397, 2010.

[166] Hashmat Popat. *Determining normal and abnormal lip shapes during movement for use as a surgical outcome measure*. PhD thesis, Cardiff University, 2012.

[167] Narayanan Ramanathan and Rama Chellappa. Face verification across age progression. *Image Processing, IEEE Transactions on*, 15(11):3349–3361, 2006.

[168] Narayanan Ramanathan and Rama Chellappa. Modeling age progression in young faces. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 387–394. IEEE, 2006.

[169] Narayanan Ramanathan and Rama Chellappa. Modeling shape and textural variations in aging faces. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–8. IEEE, 2008.

[170] Narayanan Ramanathan, Rama Chellappa, and Soma Biswas. Computational methods for modeling facial aging: A survey. *Journal of Visual Languages & Computing*, 20(3):131–144, 2009.

[171] Narayanan Ramanathan, Rama Chellappa, Soma Biswas, et al. Age progression in human faces: A survey. *Journal of Visual Languages and Computing*, 15:3349–3361, 2009.

[172] Toni M Rath and Raghavan Manmatha. Word image matching using dynamic time warping. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–II. IEEE, 2003.

[173] Dr Karl Ricanek and T Tesafaye. âmorph face database. *URL: http://faceaginggroup. com*, 2006.

[174] Sakshi Sahni and Sakshi Saxena. Survey: Techniques for aging problems in face recognition. *MIT International Journal of Computer Science and Information Technology*, 4(2):82–88, 2014.

[175] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1):43–49, 1978.

[176] Georgia Sandbach, Stefanos Zafeiriou, Maja Pantic, and Lijun Yin. Static and dynamic 3d facial expression recognition: A comprehensive survey. *Image and Vision Computing*, 30(10):683–697, 2012.

[177] David M Sarver and Marc B Ackerman. Dynamic smile visualization and quantification: Part 1. Evolution of the concept and dynamic records for smile capture. *American journal of orthodontics and dentofacial orthopedics*, 124(1):4–12, 2003.

[178] David M Sarver and Marc B Ackerman. Dynamic smile visualization and quantification: Part 2. Pmile analysis and treatment strategies. *American Journal of Orthodontics and Dentofacial Orthopedics*, 124(2):116–127, 2003.

[179] Richard Carlton Schultz. Anthropometric facial proportions in medicine. *JAMA*, 258(9):1245–1245, 1987.

[180] Stan Sclaroff and John Isidoro. Active blobs: region-based, deformable appearance models. *Computer Vision and Image Understanding*, 89(2-3):197–225, 2003.

[181] Debbie Sell, Anne Harding, and Pamela Grunwell. Gos. sp. ass.'98: an assessment for speech disorders associated with cleft palate and/or velopharyngeal dysfunction (revised). *International journal of language & communication disorders*, 34(1):17–33, 1999.

[182] James A Sethian. A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, 93(4):1591–1595, 1996.

[183] Shylaja S Sharath, Balasubramanya Murthy KN, S Natarajan, et al. Dimensionality reduction techniques for face recognition. In *Reviews, Refinements and New Ideas in Face Recognition*. InTech, 2011.

[184] Kirill Sidorov. *Groupwise non-rigid registration for automatic construction of appearance models of the human craniofacial complex for analysis, synthesis, and simulation*. PhD thesis, Cardiff University (United Kingdom), 2010.

[185] Richa Singh, Mayank Vatsa, Afzel Noore, and Sanjay K Singh. Age transformation for improving face recognition performance. In *International Conference on Pattern Recognition and Machine Intelligence*, pages 576–583. Springer, 2007.

[186] Lindsay I Smith. A tutorial on principal components analysis. *Cornell University, USA*, 51:52, 2002.

[187] Sima Soltanpour, Boubakeur Boufama, and QM Jonathan Wu. A survey of local feature methods for 3d face recognition. *Pattern Recognition*, 72:391–406, 2017.

[188] Olga Sorkine, Daniel Cohen-Or, Yaron Lipman, Marc Alexa, Christian Rössl, and H-P Seidel. Laplacian surface editing. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 175–184. ACM, 2004.

[189] Jinli Suo, Tianfu Wu, Songchun Zhu, Shiguang Shan, Xilin Chen, and Wen Gao. Design sparse features for age estimation using hierarchical face model. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–6. IEEE, 2008.

[190] Jinli Suo, Song-Chun Zhu, Shiguang Shan, and Xilin Chen. A compositional and dynamic model for face aging. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(3):385–401, 2010.

[191] Richard Szeliski et al. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 2(1):1–104, 2007.

[192] Barry-John Theobald. *Visual speech synthesis using shape and appearance models*. PhD thesis, University of East Anglia, Norwich, UK, 2003.

[193] Barry-John Theobald, Sascha Fagel, Gérard Bailly, and Frédéric Elisei. Lips2008: Visual speech synthesis challenge. In *9th Annual Conference of the International Speech Communication Association (Interspeech 2008)*, pages 2310–2313, 2008.

[194] Bernard Tiddeman, Michael Burt, and David Perrett. Prototyping and transforming facial textures for perception research. *IEEE computer graphics and applications*, 21(5):42–50, 2001.

[195] BP Tiddeman, MR Stirrat, and David I Perrett. Towards realism in facial image transformation: Results of a wavelet mrf method. In *Computer Graphics Forum*, volume 24, pages 449–456. Wiley Online Library, 2005.

[196] Arshed Toma. *Characterization of normal facial features and their association with genes*. PhD thesis, Cardiff University, 2014.

[197] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

[198] Kazuya Ueki, Teruhide Hayashida, and Tetsunori Kobayashi. Subspace-based age-group classification using facial images under various lighting conditions. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 6–pp. IEEE, 2006.

[199] Pieter Van der Geld, Paul Oosterveld, and Anne Marie Kuijpers-Jagtman. Age-related changes of the dental aesthetic zone at rest and during spontaneous smiling and speech. *The European Journal of Orthodontics*, 30(4):366–373, 2008.

[200] G Van Rossum. Python tutorial. centrum voor wiskunde en informatica (cwi), amsterdam (1995). *There is no corresponding record for this reference*.

[201] Jason Vandeventer. *4D (3D Dynamic) statistical models of conversational expressions and the synthesis of highly-realistic 4D facial expression sequences*. PhD thesis, Cardiff University, 2015.

[202] Jason Vandeventer, Lukas Gräser, Magdalena Rychlowska, Paul L Rosin, and David Marshall. Towards 4d coupled models of conversational facial expression

interactions. In *Proceedings of the British Machine Vision Conference*, pages 142–1, 2015.

[203] Una Strand Viðarsdóttir, Paul O'higgins, and Chris Stringer. A geometric morphometric study of regional differences in the ontogeny of the modern human facial skeleton. *Journal of Anatomy*, 201(3):211–229, 2002.

[204] Junyan Wang, Yan Shang, Guangda Su, and Xinggang Lin. Age simulation for face recognition. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 3, pages 913–916. IEEE, 2006.

[205] Yongmei Wang, Bradley S Peterson, and Lawrence H Staib. Shape-based 3d surface correspondence using geodesics and local geometry. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 2, pages 644–651. IEEE, 2000.

[206] Frank Wood. Principal component analysis. 2009.

[207] Shuicheng Yan, Ming Liu, and Thomas S Huang. Extracting age information from local spatially flexible patches. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 737–740. IEEE, 2008.

[208] Shuicheng Yan, Xi Zhou, Ming Liu, Mark Hasegawa-Johnson, and Thomas S Huang. Regression from patch-kernel. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[209] Zhiguang Yang and Haizhou Ai. Demographic classification with local binary patterns. In *International Conference on Biometrics*, pages 464–473. Springer, 2007.

[210] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, and Michael Reale. A high-resolution 3d dynamic facial expression database. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference On*, pages 1–6. IEEE, 2008.

[211] Juha Ylioinas, Abdenour Hadid, Xiaopeng Hong, and Matti Pietikäinen. Age estimation using local binary pattern kernel density estimate. In *International Conference on Image Analysis and Processing*, pages 141–150. Springer, 2013.

[212] Hua Yu and Jie Yang. A direct lda algorithm for high-dimensional dataâwith application to face recognition. *Pattern recognition*, 34(10):2067–2070, 2001.

[213] Shipeng Yu, Kai Yu, Volker Tresp, Hans-Peter Kriegel, and Mingrui Wu. Supervised probabilistic principal component analysis. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 464–473. ACM, 2006.

[214] Chengsheng Yuan, Xingming Sun, and Rui Lv. Fingerprint liveness detection based on multi-scale lpq and pca. *China Communications*, 13(7):60–65, 2016.

[215] Lazaros Zafeiriou. *Unsupervised analysis of behaviour dynamics*. PhD thesis, Imperial College London, 2016.

[216] Wenyi Zhao, Rama Chellappa, and P Jonathon Phillips. *Subspace linear discriminant analysis for face recognition*. Citeseer, 1999.

[217] Xiaodan Zhuang, Xi Zhou, Mark Hasegawa-Johnson, and Thomas Huang. Face age estimation using patch-based hidden markov model supervectors. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.