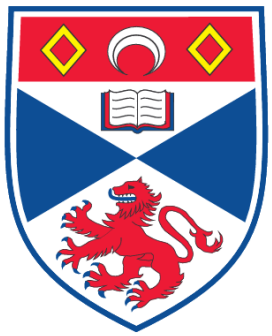
SYNTHESIS OF FACIAL AGEING TRANSFORMS USING THREE-DIMENSIONAL MORPHABLE MODELS

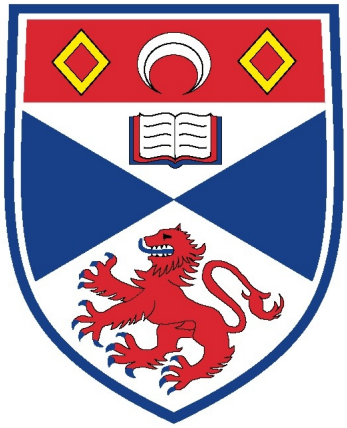
David Hunter

A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews



2009

Synthesis of facial ageing transforms using three-dimensional Morphable Models



A thesis to be submitted to the

UNIVERSITY OF ST ANDREWS

for the degree of

DOCTOR OF PHILOSOPHY

by

David Hunter

School of Computer Science

University of St Andrews

December 2008

Abstract

The ability to synthesise the effects of ageing in human faces has numerous uses from aiding the search for missing people to improving recognition algorithms and aiding surgical planning.

The principal contribution of this thesis is a novel method for synthesising the visual effects of facial ageing using a training set of three-dimensional scans to train a statistical ageing model. This data-base is constructed by fitting a statistical Face Model known as a Morphable Model to a set of two dimensional photographs of a set of subjects at different age points in their lives. We verify the effectiveness of this algorithm with both quantitative and psychological evaluation. Most ageing research has concentrated on building models using two-dimensional images. This has two major shortcomings, firstly some of the information related to shape change may be lost by the projection to two-dimensions; secondly the algorithms are very sensitive to even slight variations in pose and lighting. By using standard face-fitting methods to fit a statistical face model to the image we overcome these problems by reconstructing the lost shape information, and can use a model of physical rotations and light transfer to overcome the issues of pose and rotation. We show that the three-dimensional models captured by face-fitting offer an effective method of synthesising facial ageing.

The second contribution is a new algorithm for ageing a face model based on Projection to Latent Structures also known as Partial Least Squares. This method attempts to separate the training set into a set of basis vectors that best explains the shape and colour changes related to ageing from those factors within the training set that are unrelated to ageing. We show that this method is more accurate than other linear techniques at producing a face model that resembles the individual at the target age and of producing a face image of the correct perceived age.

The third contribution is a careful evaluation of three well known ageing methods. We use both quantitative evaluation to determine the accuracy of the ageing method, and perceptual evaluation to determine how well the model performs in terms of perceived age increase and also identity retention. We show that linear methods more accurately capture ageing and identity information if they are trained using an individualised model, and that ageing is more accurately captured if PLS is used to train the model.

I, David Hunter, hereby certify that this thesis, which is approximately 34129 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September 2003 and as a candidate for the degree of Doctor of Philosophy in September 2003; the higher study for which this is a record was carried out in the University of St Andrews between 2003 and 2008.

date _____ *signature of candidate* _____

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

date _____ *signature of supervisor* _____

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. We have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to Printed copy and electronic publication of thesis through the University of St Andrews.

date _____ *signature of candidate* _____ *signature of supervisor* _____

Acknowledgement

No thesis can be written entirely single handedly, I am obligated to thank the following individuals.

First and foremost I wish express my gratitude to Dr. Bernard P. Tiddeman for his immense patience in supervising my research, for the invaluable advice and suggestions and for proof reading this thesis.

I am extremely grateful to Prof. D. Perrett of the School of Psychology at the University of St Andrews for providing the photographic data used in this thesis along with his advice on conducting perceptual experiments.

I would also like to thank Dr. Jingying Chen, Meng Yu and Zakariyya Bhayat for their help in gathering three-dimensional scans and putting dots on faces. Dr. Tim Storer for providing the LaTeXclass files with wich this thesis is typeset. Norman, Andy, Jose and Jim who keep the schools computers running, the school's secretaries Gina and Joy and the many other people who keep the university functioning around us.

I also have to thank Richard, Sarah and Jonathan for their encouragement and friendship. Finally my parents whoes support and encouragement has never wavered.

This work was funded by Unilever PLC and an EPSRC CASE award.

Published Research originating from this thesis

David W. Hunter and Bernard P. Tiddeman.
Towards individualized ageing functions for human face images.
In *Theory and Practice of Computer Graphics*, Bangor, United Kingdom, 2007. Eurographics Association.

Bernard P. Tiddeman., David W. Hunter, and Yu Meng.
Fibre centred tensor faces.
In *British Machine Vision Conference*, volume 1, pages 449–458, 2007.

David W. Hunter and Bernard P. Tiddeman.
Visual ageing of human faces in three dimensions using morphable models and projection to latent structures.
In *VISAPP 2009: Proceedings of the Third International Conference on Computer Vision Theory and Applications, Lisboa, Portugal, February 05-08, 2009*, 2009.
to appear.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Accurate prediction of how a person's appearance will vary with age has a variety of applications, such as helping in the search for missing persons, planning cosmetic surgery, as well as applications in the film industry and other visual arts. In this work we improve upon current two-dimensional methods by using a technique to fit a statistical face model, known as a Three Dimensional Morphable Model (3DMM) [15], to photographs of human faces. This aims to eliminate the problems associated with pose and lighting as well as approximate the three-dimensional shape of the subject's face. We use this data to investigate various multi-variate statistical methods which, we believe, will provide improved ageing functions by finding correlations between appearance and the way in which an individual ages.

Our method makes use of two databases for its calculations, one a set of 3D scans of individuals of a variety of ages, and the other a set of 2D images of individuals at multiple age points in their lives. The 3D scans are used to create a statistical model, containing the principal components or eigenfaces [85] of the scans. This model is used both to create 3D models from the 2D images and as a coordinate space with which to train the ageing function.

Two dimensional face models, by definition, store no information about the shape of the face in the depth plane, i.e. along an imaginary axis that points into the image. This results in a number of shortcomings in using these models for face analysis. The models are highly vulnerable to changes caused by rotations, perspective effects, or changes in the lighting conditions around the face being studied. As these effects are not related to ageing

it is important to eliminate them before attempting to train an ageing model, to avoid any spurious correlations. As an example, if most or all of the images taken of individuals in one age range were taken face on and most of the images of another age range were at an angle, the naïve method would consider the changes in the image related to rotation to be the strongest correlates to ageing. Previous researchers have attempted to deal with the problem of pose in two-dimensions either by using a standardised image sets, or by using a two-dimensional linear transform to 'de-rotate' images. Standardised image sets where the pose and lighting of the subject can be controlled, are not always available and even small rotations can affect the results, so a method that eliminates the effects of rotations is preferable.

Lighting effects cause similar problems, although lighting can be described in a linear fashion in two-dimensions, either as a low frequency approximation [65] or as a point light source in image template alignment [74], these methods both rely on the absence of rotations and shadowing. Image normalisation can remove the effects of ambient lighting but are still prone to the effects of more directional lighting effects, such as diffuse lighting specular highlights and even area lighting. As a result, lighting effects have been found by some authors [71] to creep into ageing functions even when the images have been normalised.

A two-dimensional model can capture the shading changes related to three-dimensional shape change, provided the lighting is constant, however the lighting sources in our image-set are not constant and exhibit changes in lighting angle, composition and spread.

Using a three-dimensional model to describe the face can deal with these problems by synthesis. The effects of rotation, perspective changes, and lighting transfer can be described using physical modelling. As a result these effects can be used as independent parameters in the description of the fitted face model, and normalised in the age-model to remove their effects.

Another shortcoming of two-dimensional images is the loss of information in the parts of the image occluded, either by rotations causing self occlusions in the face or by other objects.

1.1 System Overview

Figure 1.1 shows an overview of the face ageing system developed in this thesis. The system takes as input a two-dimensional image of a previously unseen individual. A new image is synthesized, based on the input and an ageing function, such that it looks like the same person aged by a specified amount. The system makes use of two databases for its calculations; one a set of three-dimensional scans of individuals of a variety of ages, sexes and lifestyles, and the other a set of two-dimensional images of individuals at multiple age points in their lives. The three-dimensional scans are used to create a generative statistical face model, containing the principal components or eigenfaces of the scans. This model is used both to create three-dimensional models from the two-dimensional images and as a coordinate space within which the ageing function is trained.

Three-dimensional Face Scans

A set of three-dimensional models of individuals' faces is required in order to build a representation of the space of human face shapes and colours. The models were produced by scanning 106 individuals of varying ages from 2 to 60 using a stereoscopic capturing system produced by 3DMD [1]. The models produced by the scanner consist of of a three-dimensional triangle mesh and an image texture captured using flash photography. The mesh consists of a set of vertices and a list that indexes the vertices to form a set of triangles. This is a very flexible format that allows virtually any surface to be approximated as a set of small triangles. The image texture is overlaid on the surface, using texture coordinates defined on each triangle vertex to guide positioning, to describe the colour of the face. The meshes produced by this format are not generally in any meaningful correspondence. That is the vertex corresponding to a feature in one mesh, e.g. tip of the nose, will be in a different position in another mesh. In order to make use of these meshes a meaningful one-to-one mapping has to be found between meshes in the dataset. Statistical methods such as Principal Components Analysis can then be applied to the shape and colour of the faces in order to create a description that approximately spans the space of human faces. The process of generating the mappings is outlined in chapter 3 and the statistical model explained in section 3.3.

Three-dimensional Models from Two-dimensional Images

3D scanners have been developed only recently, so collections of 3D scans of individuals at different ages are rare and incomplete. Waiting for individuals to age in order to rescan them is beyond the time frame of this project. Photography on the other hand has been in existence for over a century and photographs of individuals at different ages are relatively easily obtained. The proposed solution is to fit the 3D statistical model to the photographs and so obtain a 3D model of each face. A technique for obtaining these models has been developed by Blanz and Vetter [16] and has been successfully used in the field of face identification [15]. It has recently been applied by Scherbaum et al. to face ageing, this work was carried out concurrently to this thesis [72]. Scherbaum's system differs from ours as they fit Morphable Models to parameterise three-dimensional scans rather than two-dimensional images. Park et al. also during the course of this thesis used a similar face-fitting method on two-dimensional images to train an ageing model [87], however they used a simple linear ageing method that did not attempt to take individual ageing patterns into account, or attempt to separate ageing related changes from other changes in the training-set. This system still offers the same advantages over 2D image analysis. Variations in pose can be accounted for using rotations and lighting effects, which can distort 2D image analysis, can help shape the 3D model.

Our 2D image dataset consists of 346 images of 43 different individuals taken at various age ranges. Infants between 0 and 1.5 years old, toddlers between 2 and 6 years old, mid child from 6 to 9, late child from 9 to 13, teenagers from 13 to 18 and students from 18 to 23. The images were gathered from images submitted by students of St. Andrews University and vary in quality, pose, lighting and completeness.

1.2 Thesis Contribution

In this thesis we make 4 main contributions;

- A complete system for ageing two-dimensional facial images using three-dimensional Morphable Models.
- A perceptual evaluation of how well the face fitting method retains the identity of the

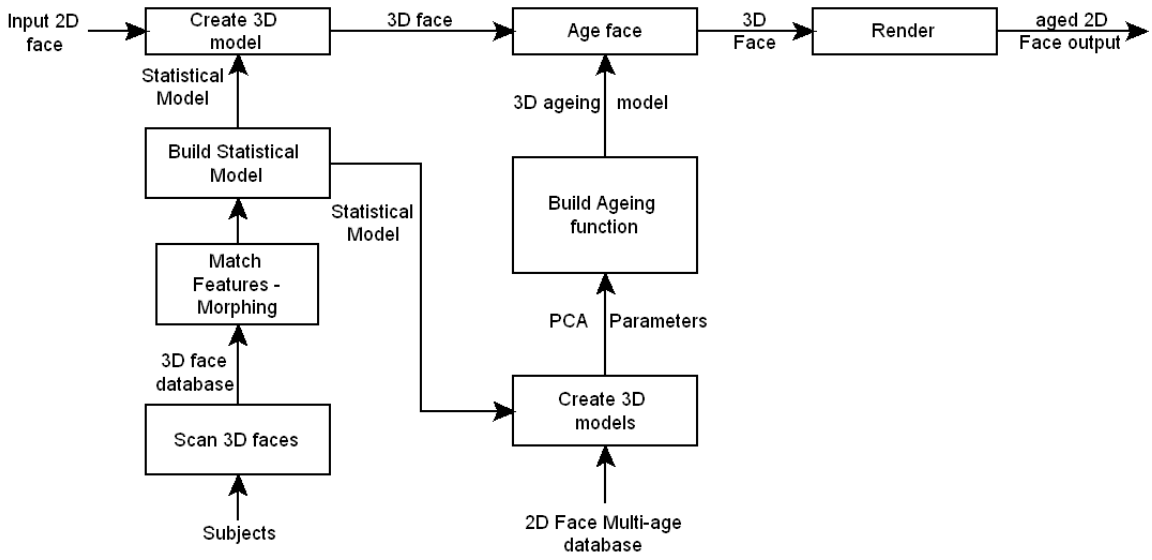


Figure 1.1: System Overview

individual in the image.

- A new statistical ageing method based on Projection to Latent Structures (PLS).
- A quantitative and perceptual evaluation of PLS based ageing and two other well known statistical ageing methods.

1.3 Thesis Outline

Throughout this thesis I argue that by extending the modelling of facial ageing into three-dimensions and fitting three-dimensional models to face images we can improve the accuracy and efficacy of facial ageing methods. An overview of the current state of the art in visual ageing techniques as well as a historical overview of their development is provided in chapter 2. We describe the construction of the statistical face model in chapter 3 as well as its use in synthesising a new face image. Chapter 4 outlines the development of face-fitting methods as well as a description of the techniques. We implement some of the commonly used face fitting methods and argue that in the interests of accuracy a number of reference points need to be specified on each two-dimensional image to guide the fitting process. This is used to build a data-set of three-dimensional face models from two-dimensional images. In chapter 5, we detail two common linear techniques of statisti-

cal modelling of facial ageing; Ageing using prototypical images, and individualised linear ageing, we also introduce a novel ageing method that uses projection to latent structures to remove factors unrelated to ageing from the training set. Experiments to evaluate the accuracy and effectiveness of the ageing methods are described in chapter 6 and the results are presented.

Chapter 2

Literature Review

In the previous chapter we provided an overview of the full overview of both this thesis and the approach we will use to age a given face image. In this chapter we will describe the current state of the art in visual ageing of human face images, as well as providing an historical overview of its development. We will be concentrating particularly on computer modelling of facial ageing and the production of aged face images.

Previous research into ageing a face image has concentrated on transforming a two-dimensional image. At their core, these methods work by applying a shape and colour change to the input image often based on a statistical model. Early methods such as cardioidal strain, were non-statistical and relied on the similarity between mathematical functions and large scale biological changes [59, 60, 50, 88]. More resent researchers have used statistical modelling methods to derive a model from a set of training images [43, 71]. The primary variations in these methods have been the functions and methods used to train the model.

Much of the previous research can be broken down into two major categories; ageing simulation and age estimation. Ageing simulation is the process of synthesising a face image such that it resembles an input face aged a specified number of years. Age estimation is the reverse of age simulation, using computer models to estimate the age of a person based on their physical appearance. Although this research concentrates on ageing simulation many of the ideas and principles behind age-estimation are still relevant. Also, some methods of automated age estimation have been used to perform ageing simulation, where image parameters are altered in order to match the recognised age to the desired age [43].

2.1 Early Methods

One of the earliest recorded techniques for face synthesis was invented by Galton in 1878. His method involved using multiple exposures of a single photographic plate to produce a composite image of a group of individuals. Alignment of individuals in the photographs proved difficult as faces come in a variety of different proportions. The resulting photographs were blurred but features common throughout the group could be perceived [30]. Thompson [82] suggested the use of coordinate transforms for altering the shape of biological organisms. Notably he showed that linear and non-linear transforms could be used to alter the profile of one species such that it approximates the profile of a different but related species.

2.2 Cardioidal Strain

Cardioidal strain has been used by a number of researchers [59, 60, 50, 88] Cardioidal strain approximates shape changes caused by bone growth, making the head smaller, elongating the chin and raising the position of the nose and eyes. It was found by Pittenger and Shaw [59] to positively affect how humans perceived the age of an outline of the face. Similarly Mark and Todd [50] found that applying cardioidal strain to a three-dimensional model of a 15 year old females head also positively affected perceived age. However as reported by Bruce et al. [88] many of the observers did not see the faces transformed to be younger as younger. Cardioidal strain has proved effective at simulating the large scale shape changes caused by ageing, but is less suited to modelling smaller local changes, which do affect how ageing is perceived by a viewer [18]. Ramanathan et al. [66] used a modified version of the cardioidal strain whereby the parameters of the modified cardioidal strain were adjusted such that the shape changes produced corresponded to the changes in a ratio of a number of anthropometric measurements taken at key feature points on the face. These measurements were taken at a number of different age ranges and prototypes generated at 2,5,8,12,15 and 18 years. The cardioidal strain model was fitted to these set of prototypes. They evaluated the results using a recognition experiment based on eigenfaces [86] and found that correct identification rates improved with their method with 58% accuracy as opposed to 44% without, using 109 test images. Although this method can be adapted for individualised shape transforms it is not applicable to colour transforms.

2.3 Overview of Statistical Representation of Faces

Kirby and Sirovich [53] and Turk and Pentland [86] modelled the space of human faces using the Karhunen-Loève theorem to build a set of basis vectors on a set of face images. The set of faces were centred by subtracting the average from each face image, and a set of eigenvectors, known as *eigenfaces* computed from the matrix of covariants. The face images could be approximately reconstructed from a weighted linear combination of a lower dimensional subset of eigenfaces, and these coefficients then used for identification. The representation suffered from blurring effects as the features of the face were not in full alignment, meaning the edges of the same feature (e.g edge of the face) are not likely to be in the same sample area in multiple images. The result is a low frequency approximation. The algorithm was based on the intensity of the image samples only and so shape, view and illumination changes were only implicitly modelled and could not be separated from colour based variances. A more meaningful statistical face model can be constructed by bringing the faces into correspondence so that each pixel sample in the model corresponds to the same position on each face.

In order to bring the face images into correspondence numerous authors have defined landmarks on the image e.g. [23, 70] to bring the images into a course correspondence. These methods relied on manual placement of these points. Craw and Cameron were the first to align a set of face images using a point model. They warped the face images to match a common reference set of points in order to find a dense correspondence between faces in the set. Principle Components Analysis was then performed on the set of aligned faces to build a parameterised face model [23]. By applying PCA to the set of landmark points Cootes et al. produced a Point Distribution Model /citeCootes92trainingmodels. A series of similar PCA based face descriptors, for example, Active Blobs, (Scharloff et al. [73]) and the similar Active Appearance Models (AAMs)(Cootes et al. [19]) describe the face using both shape and colour, as two separate components of the model. The shape is defined as a non-rigid two-dimensional triangle mesh, the space of face shapes is found using Principal Components Analysis. To model the colours of human faces, the faces are warped to average face shape and PCA performed with the features of the face mostly falling in corresponding samples. As before a new face is constructed as a linear combination of shape and colour bases plus the mean. In order to find the parameters of an unseen face, an iterative method is used that minimises the *squared pixel difference* between the target image and the AAM.

The three-dimensional morphable model [16] introduced by Blanz and Vetter is an extension of the idea of using PCA to model variations in face shape and colour from two dimensional face images to three-dimensional models. As two-dimensional images are representations of a three-dimensional object they suffer from problems associated with pose and illumination. The projection of a simple rotation on to a two-dimensional image plane results in a warp that is neither linear nor injective. Linearity is a precondition for effective modelling with a linear basis such as PCA. The mappings are not injective (one-to-one) because points in three-dimensional space can appear and disappear as they become occluded and un-occluded. Illumination also poses a problem, although it has been shown empirically [65] that variations in lighting (including shadowing) can be modelled using a low-dimensional linear basis, this assumes no variation in the shape of the object, it also excludes shadowing. When either pose or shape is altered the relationship between illumination and image intensity becomes non-linear. Using a three-dimensional model these effects can be modelled physically. Like previous methods the morphable model describes the shape and colour of the face separately as a weighted linear combination of basis vectors constructed from PCA plus the mean. Their implementation differed from AAMs in a number of respects; the shape is a mesh of three-dimensional points instead of two-dimensional points, the mesh contains many more vertices than an AAM and thus provides a dense representation of the face shape, the colour components are only defined on the vertex points and linearly interpolated between them. This was justified on the basis of the dense representation of the shape.

2.4 Statistical Methods for Age Transformation

Rather than develop a model of ageing independently, many researchers have used a set of training-data usually in the form of images, although some researchers have used three-dimensional scanning equipment [36] and Morphable Models [72, 87].

Benson and Perrett used image blending along with a warping function to create an average face image [13]. Their method involved delineating 208 key features (eyes, ears, chin etc.) on a set of standardised photographs by hand. A shape average could then be computed by averaging the positions of the feature points. The face was coloured using a per pixel average of pixels at corresponding positions on the face. The correspondences were calculated by warping the face image into the average position using a triangulated linear warp,

with the warp offsets defined as the shift between a feature point position on the face and the feature point's position in the average. Rowland and Perrett [70] extended the method to perform facial transforms. The shape and colour differences between the averages of 20 young faces (males between 25 and 29) and 20 older faces (also males between 50 and 54 years) were used to create a simple transform. The differences were added to a target face using image warping to produce the appearance of ageing. They noted that both shape and colour changes separately produce an increase in perceived age, although the age difference produced by the combined shape and colour transform was significantly less than the 25 year age gap. They postulated that this was caused by the algorithm blurring out textural detail such as wrinkles. Importantly they showed that the transform maintains the identity of the person, thus the resulting image not only looks older but look like the *same* person older.

Burt and Perrett further investigated the process of ageing using these facial composites and transform algorithms [18]. They collected face images of 147 Caucasian males between 20 and 62 and divided the images into 7 sets each spanning 5 years. An average for each group was calculated along with a population average made by combining the groups. They found that the perceived age of the composite average of each group was consistent with the average perceived age of the individuals that made up the group, but noted that raters tended to underestimate the age of the composite images. This underestimation was greater in the older age groups than in the younger age groups. They concluded that the warping and blending process retained most of the age related information and suggested that the underestimation was due to a loss of textural detail in the blending process. In the same paper they described two different ageing transforms, one based on colour caricatures and another based on the vector difference between the oldest and the youngest groups. Colour caricatures were created by doubling the colour difference (in rgb space) between the average of the 50-54 age group and the population average. In the second transform they calculated the difference, in the shape and colour, between the oldest and the youngest age group. The shape and colour differences were then superimposed onto a target image. Experimental evaluation showed that both techniques produced a significant increase in the perceived age, although significantly less than the age difference between the original groups used to train the transform.

Many methods in modern ageing research stems from the work of Lanitis et al. [43]. Ageing functions where generated by fitting polynomial curves through a set of faces pa-

rameterised using PCA. Their technique involved parameterising a set of two-dimensional face images using PCA, in a similar manner to AAMs [19], and then calculating ageing paths through the parameterised space. They delineated key features (eyes, ears, chin etc.) on a set of photographs. A shape average could then be computed by averaging the positions of the feature points. Intensity information was also sampled from within the facial region. The feature points were concatenated into a single shape-vector. The intensity information of the shape-normalised faces were also concatenated into a single colour-vector. Principal Component Analysis was performed on the covariance matrix of the shape-vector deviations were used to find the main axis of variations from the mean, leading to a compact parametric description of the shape of each face. PCA was also performed on the colour-vectors. This method gave them a set of low-dimensional parameters that can be used to both describe a set of faces and also by manipulating the parameters to describe new faces.

Given a set of parameterised faces of a set of individuals at various age points they where able to generate a series of age functions through the PCA face space that describe ageing. Using a genetic algorithm they were able to find polynomial curves, of degree 1,2 and 3, that related the parameters of the face model to the age of the face. This they called a *global ageing function* as it assumed that all faces age in the same manner. These functions were used to estimate the ages of face images. They compared the accuracy of the age estimation produced by the polynomials to the known age of the individual. They found that both the quadratic and cubic polynomials offered a significant improvement over the linear, degree one, age functions. However the improvement offered by the cubic polynomial over the quadratic was slight, and so they chose the quadratic polynomial, as it was the simpler of the two.

In general individuals age differently, as such a global ageing function is inappropriate. A key insight of their paper was that people of similar appearance age in a similar manner. As such, examining the relationship between the parameters of facial appearance and the parameters of the ageing path for a particular person could generate ageing functions tailored to an unseen individual. An ageing path for a specific individual was generated by fitting a quadric polynomial curve to the facial parameters from face images of the same person at different age points in that person's life. They found that the facial appearance parameters and ageing function parameters had a correlation coefficient of 0.55 suggesting that faces with similar facial appearance age in a similar manner. Using this they where able to gener-

ate individualised ageing function for an unseen individual as a weighted sum of the ageing functions for similar individuals in the dataset. The similarity between two faces was estimated using the probability distribution generated from the construction of the PCA model. They also gathered lifestyle information about the individuals in the dataset, information such as gender, socio-economic factors, weather exposure etc, by asking those volunteering facial images to fill in questionnaires. The lifestyle information was vectorised and scaled such that the total variance in lifestyle information equalled the total variance of the facial parameters. In this way a new ageing function can be generated by weighting the ageing functions in the dataset by the combined appearance-lifestyle probabilities. This produced a higher correlation co-efficient of 0.72 suggesting that lifestyle has a significant impact on the visual effects of ageing, thus they were able to confirm known results from (biology, medicine references). By comparing the estimated ages of the face images to the known ages using a leave-one-out method they were able to show that individualised age models produce a more accurate estimation than global ageing functions. This was the case for both appearance based weighting and combined lifestyle appearance weighting. However this method relies on the existence of similar faces in the training set, otherwise the age function tends towards the global age function, as Lanitis et al. showed by attempting to estimate the ages of faces from a different ethnic group than that used to train the age model.

Their work also covered the area of synthesising facial ageing, generating the aged face images using the inverse of the polynomial functions used in age estimation. The results of the age synthesis where evaluated both quantitatively and perceptually. The parameters of the aged faces were compared to the parameters of a face image of the same individual at the target age using the Mahalanobis distance. The rendered face image were also shown to a set of human raters, who were asked to judge whether the synthesised image looked older than the original un-aged image, and whether the rendered image was more similar to the target individual than the original. They concluded from both the quantitative and perceptual results that both global and individual ageing functions produced suitably aged individuals, but that the individualised method was the superior method.

Scandrett et al. [71] investigated ageing functions using combinations of ageing trajectories. Like Lanitis et al. [43] they used a face model parameterised using PCA, and aged the model through this PCA space. In order to eliminate variations caused by pose and expression, the horizontal and vertical rotations as well as the amount of smiling were weighted

subjectively by human observers and then defined in the face space as the sum of score-weighted face parameters. Weighted multiples of these vectors were then used to alter the pose and expression so that they became uniform. This method approximates the rotation of a three-dimensional object on a two-dimensional plane by a linear method, this is a reasonable approximation provided the angles between the face pose and the normalized pose are small. In the event that the angle is large, this approximation becomes less accurate. Even in small rotations parts of the face that are occluded become visible, the textures are unknown and must be approximated, Scandrett at al. achieved this by reflecting the normalized image about its vertical axis. Like other 2D ageing methodologies Scandrett found that lighting variations reduced clustering of face-parameters of around the same age and thus the quality of aged textures. This effect was particularly pronounced with trajectories derived from an individual's history using fewer face samples resulting in less smoothing of errors.

Each of the trajectories were designed to extract different factors that effect ageing, such as personal history, sex, how parents aged etc. The trajectories where defined as the sum of the face parameters centred on the group mean weighted by the mean shifted age of each face. Face images were aged by altering their parameters in the direction of one or more combined ageing trajectories until the target age is reached.

As males and female are known to age in different ways, separate ageing trajectories were produced for male and female in each group. An input face was then compared to these ageing trajectories to determine the comparative influence of the male and female trajectories using a ratio of distances from the face to each trajectory.

It is often the case that multiple images of an individual are available covering a range of ages, all of which pre-date the 'start' age of an ageing function. The 'start' age being the age of the most recent image and therefore the closest to the target age. Scandrett et al. Used these images to construct what they called an 'historical' ageing trajectory, using the age weighted average of the images in the same manner as they had for other groups of images. This historical trajectory could then be combined with the ageing trajectory for the age group of the starting image from the training set, to produce an aged face image. They weighted the trajectories according to a maximum-likelihood metric driving it to be a typical member of the set of trajectories between the source and target age groups, and driving the target of the ageing trajectory to be a typical member of the set of faces at the target age.

The results were analysed using the root-mean-squared error, both on the shape vertices and per pixel, between the resulting face image and a known ground truth image of the individual at the target age. The faces having been converted to grey-scale and normalized to have a mean intensity of zero and standard deviation of one, in order to remove some of the effects of lighting on the results. They found that in general the root-mean squared shape and texture errors were lower when compared to a the ground truth image than with other images in the target age set, and concluded that the ageing methods both aged the individuals appropriately and retained identity through the transform. They found that the most accurate method of ageing varied between individuals and so could not conclude which method had the best performance.

Scherbaum el al. [72] fitted a three-dimensional morphable model to database of laser scanned cylindrical depth-maps. They used a database of 200 adult images and 238 teenagers. The later group ranged in age from 96 months to 191 months. In order to improve the resolution of the face texture map, they reconstructed the textures from three photographs taken at three separate angles. They used the parameters of the model and the age of the subject to train a Support Vector Regression model. The S.V.R. formed a mapping from the high dimensional parameter space of the model to the $\Re$ space of the subjects age. This was used to estimate the age of the subject once the parameters of the morphable model have been found. A new face model could be synthesized from a given set of parameters by 'stepping' through the curved SVR space using a fourth order Runge-Kutta algorithm, using the parameters and an estimated age as the starting point.

They didn't use multiple time-space images of the same individual in building the model, their claim to individualization is the observation that, based on the mean angles between the support vector gradients, the SVR produced different ageing trajectories for different individuals and could therefore be said to be individualized. While this is true, the variation is derived from a large number of single 'snapshots,' i.e. it describes the variations within a population. It may not necessarily capture the variations due to ageing in a particular individual.

An alternative direction based on dynamic Markov models was developed by Suo et al. [78] They used a *Grammatical Model* [94] to describe a set of faces as a hierarchical set of face components, (eyes, nose, skin patches etc.), with an individual face defined as a particular choice of components from the set. An input face was aged in a probabilistic manor using a dynamic Markov chain to select the most likely set of face components at a

target age given the current set.

Park et al. [87] performed a similar experiment to ours fitting a three-dimensional Morphable Model to a set of delineated faces using point data. Ageing was performed by calculating a set of weights between an input face and exemplar faces in the same age group. These weights are then used to build an aged face as a weighted sum of the corresponding faces at the target age. The results were compared using Cumulative Match Characteristic curves to other ageing methods and reported similar results to other methods. They observed that shape modelling in three-dimensions gave improved performance in pose and lighting compensation. Their method differs from ours in that they only fit to the delineated point data rather, whereas our method used both point and pixel information as detailed in the following chapters 4.5.1.

2.5 Age Estimation

Age Classification is the conceptual opposite of ageing synthesis. the age of the face is estimated from an image rather than synthesizing a change resulting from age.

Kwon et al. [41] used the ratio between facial features, the nose, eyes and mouth, as well as wrinkle analysis. If wrinkles were found and ratios indicated an adult face, the image was marked as a senior adult. An image with no wrinkles and a baby-like ratio between features was marked as a baby. Otherwise the image was marked as an adult. This idea was expanded upon by Horng et al [89], who used a three phase method; feature location, extraction and classification. Two geometric features; the ratio distances between eyes and nose and between nose mouth, detected using a Sobel edge detector, and three wrinkle regions were detected. A Sobel edge detector was used to classify wrinkle density, with density defined as $\frac{edges}{area}$. The age was classified to one of four age groups using back propagation neural networks. Kalamani and Balasubramanie used a fuzzy neural net to account for uncertainty in the classification model. Images were classified according to a *degree of inclusion* [39].

Lanitis et al [42] compared four classifiers for age estimation; quadratic ageing curve, Mahalanobis distance (i.e. probability that the input face belongs to a particular group), back propagation neural network (using multilayer perceptrons), and Kohonen self organising

maps. They also introduced three new types of classifier based on the training method. Firstly a classifier they called *age-specific* where the faces were grouped into strata according to age prior to training, where the classifier was only expected to place the input face into the relevant strata. Secondly a classifier they called *appearance-specific* that grouped images according to observation [43] of the relationship between appearance and ageing patterns, divided the individuals into groups of faces that appeared similar or aged in a similar manor. Thirdly a combination of the two. The methods were evaluated and compared using two-fold cross validation with the mean average error in years, between the classification result and the known ground truth. The new classifiers improved accuracy, and offered greater improvements when combined. They used perceptual evaluation of the training images with 20 human raters to gauge the accuracy of human age perception. The raters were shown the whole image, including details such as hair line. This is known to affect how humans rate an individual's age. Human raters out performed the computers albeit on a much reduced number of test images.

A number of authors have used Support Vector Regression [77] for age estimation [31] and synthesis in three-dimensions [72]. Gandhi [31] used Support Vector Regression a modification of SVMs to perform age estimation using a training set of normalised faces images. The images were first compensated for illumination using the Retinex algorithm [37] to perform dynamic range compression and a histogram equalization algorithm to bring the images to the same intensity range. The images were delineated and compensated for pose using an affine transform. Images where the face did not have a neutral expression were rejected as this would affect the formation of wrinkles. A Support Vector Regression machine was trained on the pixel intensities of 818 images ranging from 15 to 99, using a variety of different bases, polynomial, radial, and sigmoid . They found that a polynomial basis of degree 3 produced the most accurate age estimation for an unseen image producing an average of 9.31 years absolute error and a squared correlation coefficient of 0.69, when validated using 4-fold cross-validation. Lanitis [42] used Support Vector Machines to derive a non-Gaussian similarity and age metric, with a hyperplane separating faces in an age or identity group from other faces in the set, and a scalar between 0 and 1 indicating the degree of dissimilarity between an input face and the set. A ageing trajectory was found that maximized the similarity metrics between the target age-group and the groups of individuals, using a sequential quadratic programming method. The identity of an unseen individual was maintained throughout the process by maximizing the sum the of differences between the similarity metrics before and after age progression.

2.6 Fine detail synthesis

Many of the statistical methods used lost textural detail such as wrinkles, a few researchers developed methods that attempted to create appropriate textural detail in aged images. Tiddeman et al. used a wavelet transform [83] and Markov Models [84] , Hussein used Bidirectional Reflectance Distribution Functions [35] and Gandhi used Gaussian filters [31]. These methods work by attempting to replace or adjust the high-frequency components of the image to match the high frequency components of a prototype at the target age.

Hussein [35] synthesised wrinkles by attempting to align the surface normals of two faces, an older and a younger using the relationship between pixel intensity and surface orientation. Under the assumption that the two surfaces shown in the image are co-incident and under the same lighting conditions, surface details such as wrinkles would become the primary changes in intensity. They used the ratio of the two images smoothed with a Gaussian filter multiplied with one of the images so that the fine detail of the other was applied to it. Their method suffered from two main drawbacks, firstly it could not be used under varying lighting techniques, secondly the age was defined from only one image and thus would not in general produce a convincing ageing result for an arbitrary individual.

Gandhi [31] used an Image Based Surface Detail Transfer [47] procedure to map the high-frequency information from an older prototype to a younger, and visa-versa using a Gaussian convolution as a low pass filter. The idea here being to take the high-frequency details of the input image and replace them with the target's. The Gaussian convolution producing two images, one the smoothed original containing the low-frequency large scale detail and the other, the result of applying a standard boost filter, containing the high-frequency fine scale details. An aged image was synthesised by combining the high-frequency of a prototype with the low-frequency of the image. Varying the width of the kernel would vary the size of details captured and thus the perceived age of the person. The prototypes at each age were created by averaging all the images in an age group. Smoothing problems were avoided by combining the high-frequency parts of the training images with the combined average to retain fine detail.

Tiddeman et al. used a Gabor wavelet function to detect edges in the image and decompose it into a pyramid of images containing edge information at varying spacial scales [83].

The edge magnitudes were then smoothed with a B-spline filter to give a measure of edge strength about a particular point in each sub-band. Prototypes at each age were generated using the technique of Benson and Perret [13] and the wavelets were then amplified locally to match the mean of the set. The values of the input wavelet images were modified to more closely match those of the target prototype. These were tested perceptually and found to reduce the gap between the perceived age of the image and the intended age. They then extended their method using Markov Random Fields [24]. An individual was aged using the prototyping method of Burt and Perret described above [18]. Detail was added to the resulting face by decomposing the image into a wavelet pyramid and scanning across the sub-bands using the MRF model to choose wavelet coefficients that match the cumulative probability of the input values. They found that human raters found the resulting image more closely matched the target age of the older group than either the Wavelet method on its own or the prototyping method, it also succeeded in the rejuvenating test where wavelets failed. They also found that humans rated the images more realistic than those generated using Wavelets alone [84].

2.7 Summary

In this chapter previous work in the area of facial age estimation and ageing simulation has been reviewed. In the course of this work we identified several desirable properties for an improved face ageing method, most of which have been included in previous methods, but have not previously been combined in a single implementation, these include:

- The use of 3D models to properly model (and allow removal of) the effects of lighting and out of plane rotations.
- The use of training data that includes within subject age variation to include a degree of individuality in the ageing model.
- The use of modern machine learning and statistical tools for learning and applying the ageing changes.

In the following chapters we will use these observations to build a age synthesis algorithm. As explained in chapter 1 three-dimensional scanning equipment are a relatively recent

invention and sets of three-dimension models of the same individual at multiple age points are not available. As an individualised ageing model has been identified as a significant improvement over a global method, we will describe a face fitting method that can be used to extract three-dimensional information from a set of two dimensional images. The face fitting method is described in detail in chapter 4. In order to develop ageing algorithms using modern statistical tools, and to model the faces in three-dimensions, we will use a Three-dimensional Morphable Model [16] to describe the faces. This model is also used to guide the face-fitting algorithm. In the next chapter we will describe how to construct a Three-dimensional Morphable Model from a set of three-dimensional face scans.

Chapter 3

Constructing a Three-Dimensional Morphable Model

In the previous chapter we provided a detailed overview of current methods in synthesising ageing in human face images. We also identified key properties of these algorithms that are desirable in an improved ageing model. In particular the use of a three-dimensional statistical model to describe the set of human face models. In this chapter, we describe a statistical face model that can be used to parameterise an input face. We also describe how this model can be used to render an image of a synthesised face model under a given set of pose and lighting conditions. Finally, we describe how the textural properties of a face can be reconstructed from partial data.

3.1 Literature Review

The face models produced by the three-dimensional capture system we are using are in the form of a triangular mesh, defined as a set of points (vertices) and a set of edges linking these vertices to form triangles. However each scanned face model is independently produced and as such has an irregular structure. The meshes all have differing edge topologies, that is different edge structures linking vertices in the mesh. Also each point on the surface of a particular mesh has no predefined matching point on the surface of any of the other face models produced by the scanner. In order to build a statistical model these sets of face

models must be brought into a meaningful correspondence across subjects. 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 more serious, requiring detection and interpolation.

A number of algorithms have been developed in the area of registration, tailored to tackle specific problems, such as point alignment, line and edge registration, and surface registration. We wish to look specifically at the area of registering a set of three-dimensional triangular meshes of irregular edge topology such that we can generate of set of meshes are of corresponding edge topology but with varying surface shapes. Our meshes contain holes both around the edges of the mesh and internally that need to be identified and filled in a meaningful manner.

We define the face model as a tuple containing a shape description and a texture-map. The shape is described using a triangle mesh, $\mathcal{M} = (\mathcal{V}, \mathcal{E})$. $\mathcal{V}$ is a set of vertices $\mathbf{v}_i \in \mathbb{R}^3$,$\mathbf{t}_i \in \mathbb{T}^2$,$i = 1, \ldots, n$, where $\mathbf{v}_i$ describes the position of the i^{th} vertex in three-dimensional space and $\mathbf{t}_i$ describes the location in texture space $\mathbb{T}^2$, i.e. $\mathbf{t}_i = (u_i, v_i), u_i, v_i \in [0, 1]$ that holds the i^{th} vertices' colour. $\mathcal{E}$ is a set of edges connecting the vertices, $\mathcal{V}$. We have a set of three-dimensional meshes $\mathcal{M}_j, j = 1, \ldots, m$ that we wish to use to build a statistical face model. We need a method that can construct a set of meshes, $\mathcal{M}'_i = (\mathcal{V}'_i, \mathcal{E})$, which describe surfaces as close as possible to the shape of their corresponding mesh $\mathcal{M}$ but have a common edge topology, $\mathcal{E}$.

The three-dimensional models used by Blanz and Vetter in their original paper on three-dimensional Morphable Models, were built from laser scan data and described a three-dimensional structure using a cylindrical depth-map. As a result they were able to perform alignment using a regularised optical flow method [16]. Given a set of three-dimensional scans $I(h, \phi)$, with vertical component h and rotation component ϕ, optical flow computes a field $(\delta h(h, \phi), \delta\phi(h, \phi))$ such that $||I_1(h, \phi) - I_2(h, \phi)||^2$ is minimised. Optical flow offers poor performance where the scans present few features, the flow vectors in these area were smoothed. The shape and colour of the mesh were obtained using the one-to-one correspondence provided by $(\delta h, \delta\phi)$.

Iterative Closest Point alignment

The Iterative Closest Point alignment (ICP) method can be used to match multiple scans of human bodies to a common template [4, 5]. A set of correspondences between the vertices of the template mesh and the surface of the template mesh is found by locating the nearest point to each vertex on the target mesh. Obviously this assumes that the meshes are already in close proximity. A deformation field is then found that matches the displacements of each set of correspondences. The template mesh is updated with this deformation field and a new set of closest point correspondences generated. These steps are repeated iteratively until the meshes are sufficiently aligned. Not all the possible correspondences between template and surface are valid, and so a regulatory term is typically added. Besl and M^{c}Kay defined the field globally [14], Feldmar and Ayache [26] defined the affine transforms locally over a spherical region. Allan et al. [4] and Amberg et al. [5] defined the field as an affine transform per vertex, this transform is not sufficiently constrained by a single correspondence and so used a regularising term to constrain the result. The closest point is generally found either by searching along the normal, [3, 33] or by finding the closest point in any direction [14, 26, 4, 5]. A search along the normal has an advantage over the closest point in that the space of searches matches the curve of the surface, however on surfaces exhibiting rapid changes in direction the search can potentially cross before finding a match. A regularisation term ensures a smooth deformation field between the two surfaces, Feldmar and Ayache [26] used the two principal curvatures of the surfaces to drive the matching towards similar features. Allen et al [4] used the sum of the Frobenius norm between affine transforms defined for adjacent vertices on the template mesh as part of the minimisation to weight the fitting towards smoothly varying deformations fields. Amberg et al. modified this metric to allow a weighting between the rotational and skew parts of the deformation at each vertex [5].

Not all the points on the template will be matched to points on the surface of the target mesh, in most cases this is due to holes in the target mesh. Early ICP algorithms assumed that the target mesh was complete. Kähler et al. used the template to define the surface of missing parts of the target mesh, warping the surface to match the area surrounding the hole in that target mesh using Radial Basis functions, [38]. Allen et al. also used the template mesh to define the area of the hole, however they used the smoothing term over the deformation field to drive the template to an approximation of the missing surface [4].

Remeshing

ICP algorithms assume that even if the vertices and topology of meshes are not in any sort of correspondence, the surfaces of the meshes are already closely aligned. If, however the surfaces are not in close correspondence ICP can produce some spurious results. Methods based on completely reconstructing meshes can find correspondences between surfaces that are not already in close proximity.

Instead of fitting a template mesh to a set of meshes, a new mesh can be constructed from the input meshes in a consistent manner so that the resulting meshes are in one-to-one correspondence. This method is known as remeshing. The method relies on the fact that a well defined triangle mesh forms a surface, a mapping can then be generated to map this surface from three-dimensional space to a two-dimensional space, a new mesh can then be generated by sampling at regular intervals within this two dimensional space and mapping back into the original three-dimensional space. The method of generating this mapping is known as parameterisation and was first described by Tutte [79]. Here I outline the method described by Floater [27] using mean-value coordinates [28].

Given a triangular mesh $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ we desire to create a one-to-one mapping $\mathbf{u} : \Re^3 \rightarrow \Re^2$ from the 3D mesh surface to a plane. As the surface is three-dimensional, the resulting mapping will distort the mesh, a mapping that best preserves the angles and distances between the vertices when mapped onto the two-dimensional space is desirable.

The position of a vertex in the two-dimensional plane is defined as a weighted sum of the positions of its neighbours in the edge graph $\mathcal{E}$. The weights are derived from the three-dimensional structure of the mesh, and it's these weights, defined along the edges, $\mathcal{E}$, that help preserve the relative positions and angle of the vertices when mapping to a two-dimensional plane. A vertex $\mathbf{v}_i \in \mathbb{R}^3$ and its corresponding vertex on the plane, $\mathbf{u}_i \in \mathbb{R}^2$ form the mapping on the vertex. Between the vertices the mapping is defined using barycentric coordinates. The weighted sum definition of a point only applies to internal vertices where the neighbouring triangles form a complete circle around the point, this is not the case for boundaries and so they have to be defined separately. $\mathbf{u}_i$ and $\mathbf{v}_i$ are redefined to include only the internal vertices and $\mathbf{b}_i \in \mathbb{R}^2$ defined to be the position of the boundary vertices in the two dimensional plane. These are usually fixed by projecting them onto a unit circle and thus are known before the start of the parameterisation. Floater

defined the position on the plane of the i^{th} vertex as,

$$\mathbf{u}_i - \sum_{j=1}^{J} \lambda_{i,j}\mathbf{u}_j = \sum_{k=1}^{K} \lambda_{i,k}\mathbf{b}_k,\ i = 1, \ldots, n \tag{3.1}$$

where J and K are the number of internal and external vertices respectively. Here $\lambda_{i,j}$ defines the weight along the edge i, j in $\mathcal{E}$. If edge (i, j) is not in $\mathcal{E}$ then $\lambda_{i,j} = 0$. The weight is defined in such a way as to preserve the structure of the mesh, minimise stretching and preserve the angles between vertices.

$$\lambda_{i,j} = \frac{w_j}{\sum_{k\in\Omega} w_k},\ w_j = \frac{tan(\alpha_j/2) + tan(\alpha_{j-1}/2)}{||\mathbf{v}_i - \mathbf{v}_j||} \tag{3.2}$$

where $k \in \Omega$ is the one-ring of neighbouring vertices to i, that is, all vertices directly connected to the i^{th} vertex by and an edge. The middle component of the equation is a normalisation term ensuring that $\sum_{j=1}^{k} \lambda_{i,j} = 1$ for all i. The final part of the equation defines an un-normalised weight in terms of the angles about the vertex i along both sides of the edge (i, j) and the length of the edge (i, j), see figure 3.1.

The two-dimensional positions $\mathbf{u}_i$ are found by solving the linear equation defined by equation (3.1).

A more detailed survey of parameterisation techniques for three-dimensional meshes can be found in [29].

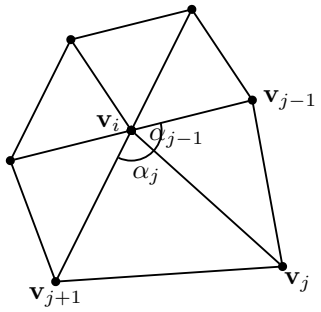


Figure 3.1: The triangles of the one-ring around vertex $\mathbf{v}_i$, illustrating the construction of the weight $\lambda_{i,j}$ along the edge (i, j). The weight is computed as a function of angles α_j, α_{j-1} about the edge and the inverse of its length.

Praun et al. used this flatting method to create a set of consistent parameterisation between a set of face meshes in order to produce an average face model. They chopped the meshes

into a set of triangular patches using a common high level topology. Each patch was separately parameterised by fixing the boundary vertices to a triangle in two dimensions and generating a mapping using Floaters parameterisation method (described above) [62].

The method of parameterisation is better able to cope with matching surfaces with poor initial correspondence. However, due to the parameterisation method only applying to internal vertices, it is unable to deal effectively with holes in the mesh.

3.2 Constructing a Mesh Correspondence

Before a statistical model can be built, the meshes must be brought into a one-to-one correspondence. In order to do this we define a template mesh $\mathcal{S} = (\mathcal{V}, \mathcal{E})$ and adapt this mesh to the input meshes, so that all the meshes share the same topological structure, $\mathcal{E}$.

We implemented two methods for building a mesh correspondence. The first is based on the method by Praun et al. [62] using parameterisation and the second is based on ICP which is described below.

3.2.1 Iterative Closest Point alignment using multi-level free-form deformation

We adapt a reference face model to each subject's face as outlined in figure 3.3. The two surfaces, the subject and the reference are brought into alignment by translation to a common mean and removing rotational variance between the two meshes. The translation to the mean is found by making the centre of mass of the point sets equal. The meshes are scaled by normalising the deviation from the mean and the required rotation found using SVD on the cross-covariances between pairs of meshes.

The reference mesh topology we use is just an individual selected on the basis of the quality of the scan (i.e. all major features visible). We adapt this to each subject using a standard two stage warping process. The first stage is a feature based warping method, in which manually placed landmarks are used to drive a multi-level free-form deformation (MFFD), which is a hierarchy of B-spline interpolating functions with progressively finer resolution

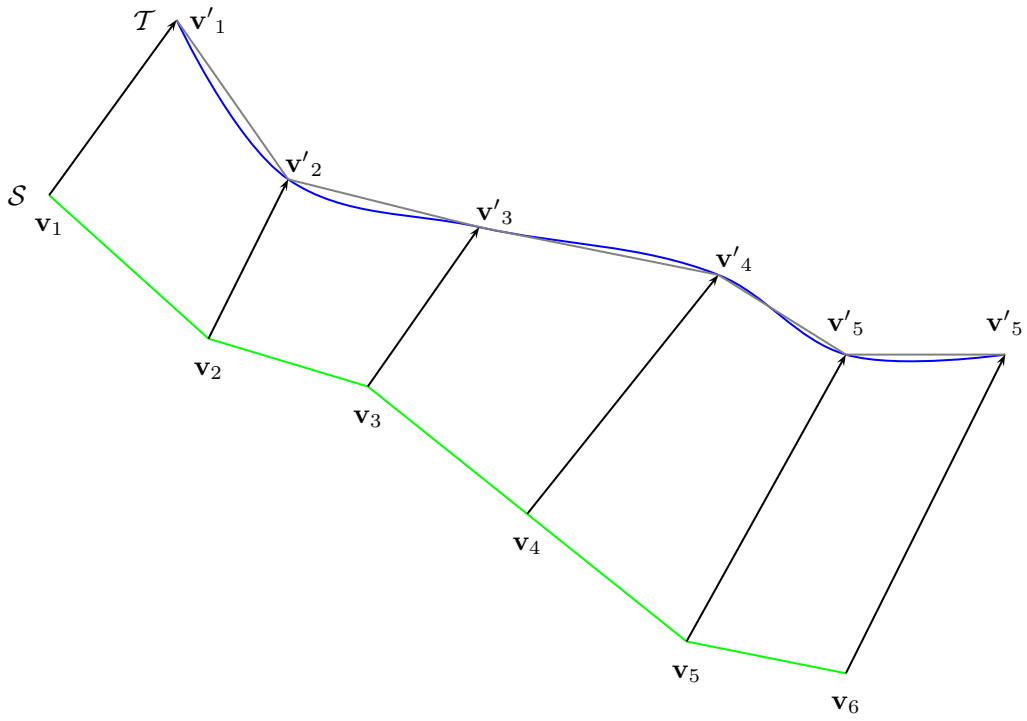


Figure 3.2: Surface $\mathcal{S}$ is matched to surface $\mathcal{T}$ using Iterative Closest Point alignment. Each vertex $\mathbf{v}_i$ is matched to a vertex $\mathbf{v'}_i$ by searching along the surface normal.

[45, 46]. We have implemented the MFFD warping using a space and search efficient octree data structure. Once the face meshes are in approximate alignment correspondences are found using a standard ray-tracing algorithm. Rays are traced out of the reference mesh from each vertex in both directions along the surface normal and the first intersection (within a maximum radius) with the target face is found using an octree ray-tracing method (see figure 3.2). Not all vertices will find a target, and so these displacements are interpolated (again using MFFDs) across the reference mesh. This brings the reference mesh into good alignment with the subject.

3.2.2 Surface alignment using Parameterisation

As with our ICP implementation we delineated a set of points on each of a set of scanned faces. The parameterisation method does not require a predefined template mesh, this is defined procedurally. For each mesh we computed a mapping from the three-dimensional surface of the mesh to a two dimensional plane, $w : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ using Floaters' parameterisation [28]. For each internal vertex of the mesh, $\mathbf{v}_i \in \mathbb{R}^3$, a corresponding position

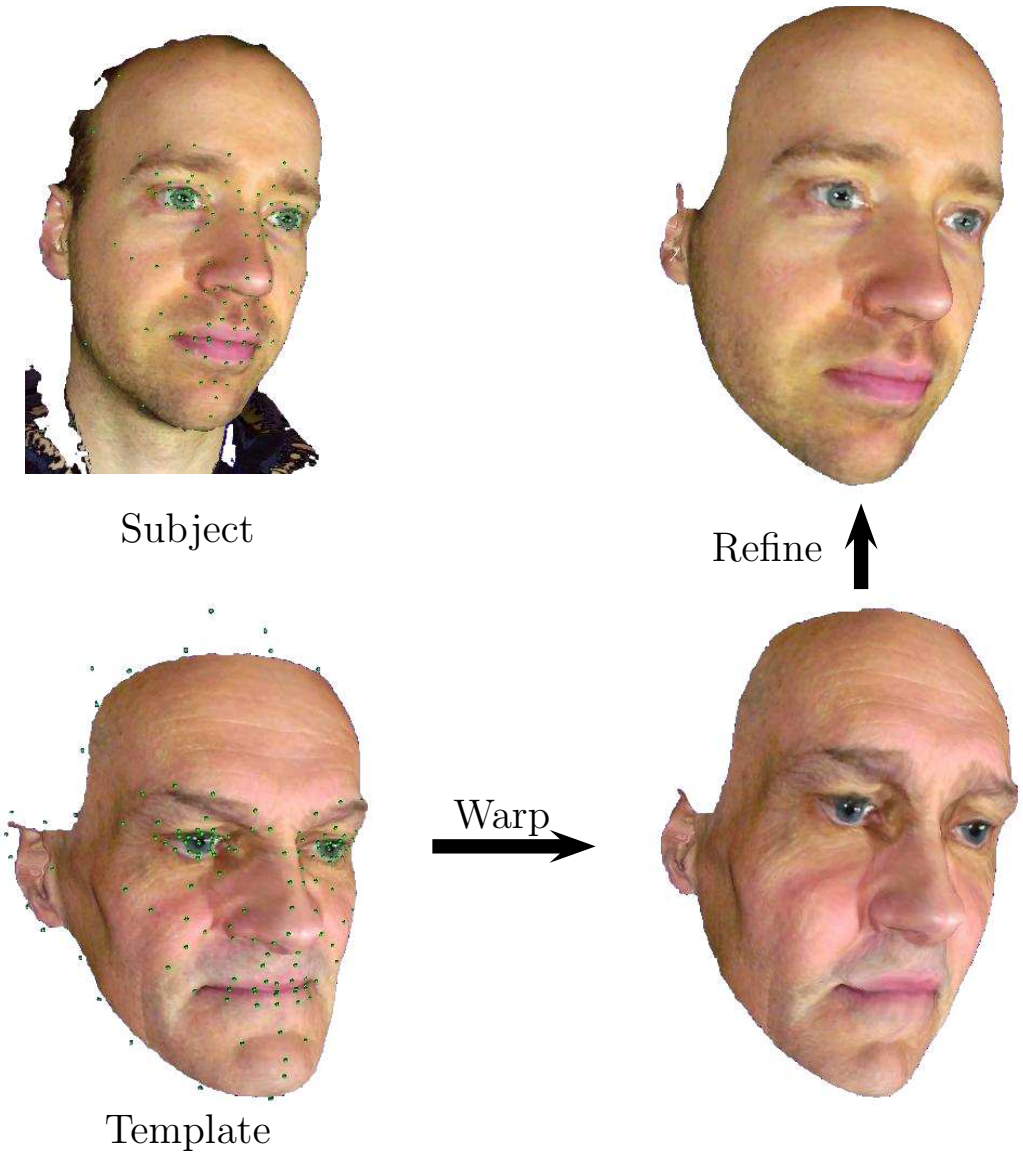
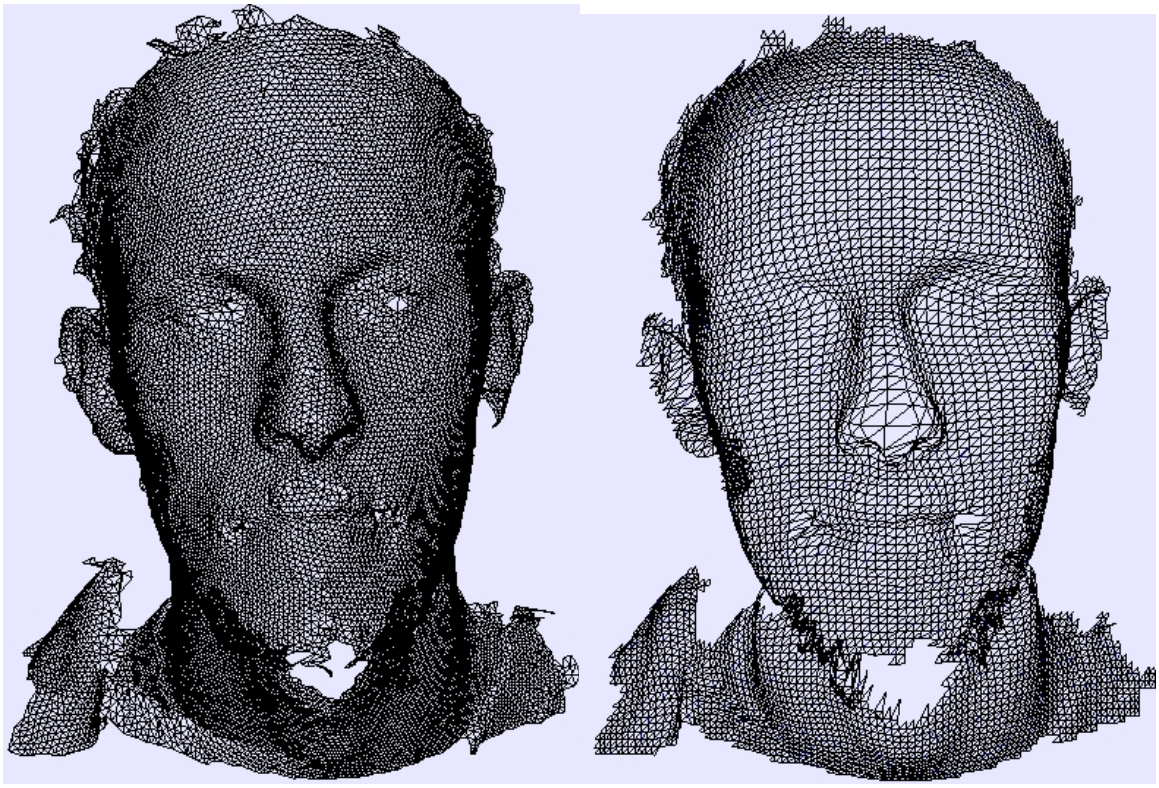


Figure 3.3: An outline of the Iterative Closest Point Alignment algorithm. First both meshes are delineated by hand. A template mesh is then warped to the rough shape of the subject mesh using the delineated points. The warp is refined by tracing along the surface normals of the subject model and matching at the point of intersection with the template mesh. Finally the texture is reconstructed by applying an inverse warp to the texture-map.

$\mathbf{u}_i \in \mathbb{R}^2$ is found on the plane, such that the mesh is 'flattened' on to the surface of the plane in a manner that best preserves the angles between edges of the mesh and the lengths of edges (see section 3.1). Each position within a triangle on the surface of the mesh can be mapped onto the plane by finding its associated triangle in the flattened mesh and determining its position in this mesh using the method of barycentres. This produces a set of flattened meshes. These meshes can be brought into correspondence using the delineated points, for each delineated point on the surface of the mesh the corresponding point on the plane as found using the mapping, w. This produces a set of templates in two-dimensions for each input face. An average of these templates was found using Generalised Procrustes alignment [25] and warping between each template and the average generated using Thin Plate Splines (TPS). The application of this warp to flattened meshes brings them into correspondence on the plane. A new mesh can be generated by regularly sampling along a predefined grid on the plane in the area of the warped mesh. Using the TPS warp on the vertices of the mesh is possible but is not easily defined within the triangles of the mesh, instead we use an inverse warp on the sample points of the regular grid this allows us to sample from the triangles as if they were warped. For each sample point in the regular grid the triangle of the flattened mesh containing it is located and its corresponding point on the surface of the three-dimensional mesh found by inverting the mapping, w. If no triangle is found containing the sampling point the sample point is marked as missing. A new mesh is constructed by triangulating the located sample points in a 'chess-board' pattern with two triangles being formed in each quad.

3.2.3 Results

Figures 3.4a and 3.4b provide a 'before and after' view of the remeshing process. The two methods produced similar results in terms of fitting. The ICP method relied on the quality of the original mesh, and produced some spurious fitting around areas of high curvature due to the closest point search finding poor correspondences. The parameterisation method was vulnerable to stretching in mapping the surface to two-dimensions, this results in under-sampling. As stretching often occurs in areas of high curvature, this means that the under-sampling will occur most frequency in areas of most interest. The parameterisation method also lacked a hole-filling method. This deficiency was the main reason the ICP method was the chosen method for producing the correspondences between meshes.



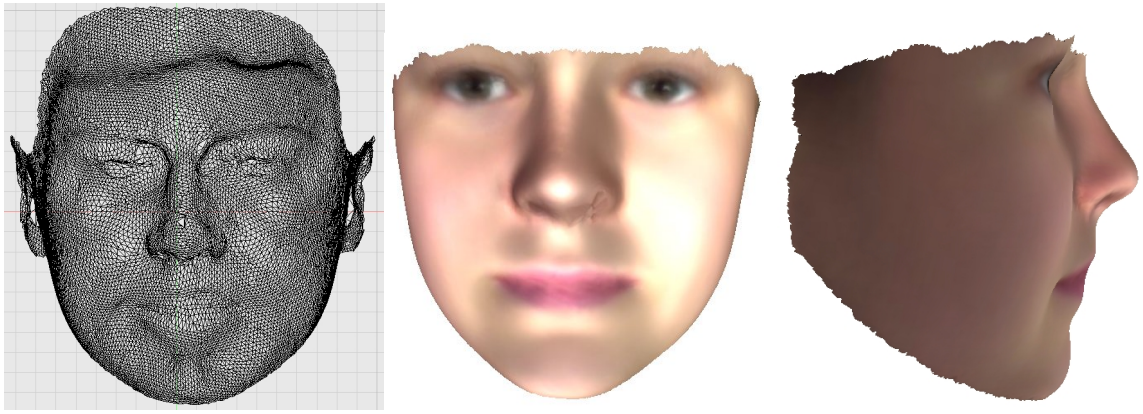
(a) Irregular Mesh (b) Regular Mesh

Figure 3.4: 'Before and after' view of a reconstructed mesh. The left hand image shows a 'wire-frame' view of a mesh as produced by the scanner. The irregular structure of the mesh is clearly visible as are the holes. The right hand image shows a reconstructed mesh, the regular grid pattern is visible as are the un-patched holes in the mesh.

The average of a set of 106 face models produced by the three-dimensional scanner was produced by aligning these meshes to a 'reference mesh' using the ICP fitting method. These aligned models were used to build a face model as described in the next section. The mean of the face models can be found in figure 3.5.

3.3 Principal Component Analysis

When attempting to discern a three-dimensional object from a two-dimensional image we need to reduce the space of possible models to the set of models that are faces, and to find in a maximum likelihood sense the face that most accurately describes the face shown in the image. To do this we need a descriptor that best spans the space of possible face surfaces



(a) Average Mesh (b) Front View (c) Profile View

Figure 3.5: Average of a set of face meshes constructed using Iterative Closest Point alignment. Left is a wire-frame image of the full average mesh. Centre and right is are a face-on and profile rendering of the mean shape and colour of the set of face meshes. The centre and right meshes have been clipped to the facial area.

and arranges these according to their likelihood.

Principal Component Analysis (PCA) is a dimensionality reduction and decorrelation technique. It finds a set of orthogonal basis vectors that span the data such that the covariance matrix is diagonal. In other words the data is linearly decorrelated. The first vector of the basis, the first principle component, is parallel to the direction of greatest variance in the data. The second vector is parallel to the direction of the next greatest variance and so on. Here we describe explicitly the method we used to calculate PCA on the shape data, the same method is also applied to texture data. This is a standard method [63]. In order to build a statistical model, the face shapes were first brought into a one-to-one correspondence between each face (see section 3.2). The three-dimensional positions of the vertices of the mesh were then concatenated to form a shape-vector,

$$\mathbf{s} = (X_1, Y_1, Z_1, X_2, Y_2, Z_2, \cdots, X_n, Y_n, Z_n)^T, \tag{3.3}$$

where n is the number of vertex points in the face models. With a set of m exemplar shape-vectors $\mathbf{s}_i$ where $i \in 1 \ldots m$ was centred by subtracting the mean $\hat{\mathbf{s}}$ and arranged in a matrix D, the eigenvectors and eigenvalues of its covariance matrix C were computed using the Jacobi Method [63] on C.

The face shapes are generally unaligned and require registration in order to calculate a meaningful average. The mean, $\hat{\mathbf{s}}$ is defined as that shape which minimises the residuals

from between the mean and the set of shapes under a Euclidean similarity (scale, translation and rotation) transformation. Here we define the matrix $R \in SO(3)$ to be the rotation matrix, a vector describing translations, $\mathbf{t} \in \Re^3$, and a scale $\kappa \in \Re$. Given two sets of vertex points $\mathbf{p}$ and $\mathbf{q}$ we align $\mathbf{q}$ with $\mathbf{p}$ by,

$$\mathbf{p} = \kappa \mathbf{q} R + 1_m \mathbf{t}^T + E \tag{3.4}$$

where E is an $m \times 3$ error matrix and 1_m is a $1 \times m$ vector of 1s. In two dimensions this can be solved by converting the two-dimensional points into a set of points in complex space,(x+y1) and finding the mean as the principal eigenvector of the complex covariance matrix, this is known as the full Procrustes algorithm [25]. In three dimensions it was solved in an iterative manner. First we calculated the translation, $\mathbf{t}$ by making the centre of mass of the point sets equal.

$$\hat{\mathbf{p}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{p}_i. \tag{3.5}$$

where $\mathbf{p}_i$ is the i^{th} point in the vertex point set $\mathbf{p}$. Similarly for $\mathbf{q}$. The scale factor was chosen such that the l^2-norm of $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$ are both equal to 1. The mean centred and unit scaled points sets are denoted $\dot{\mathbf{p}}$ and $\dot{\mathbf{q}}$. The rotation matrix R was found using SVD on the cross-covariance matrix,

$$\dot{\mathbf{p}}^T \dot{\mathbf{q}} = V \Lambda U^T \tag{3.6}$$

V and U form a set of orthonormal basis vectors in $SO(3)$. R was updated as $R = UV^T$ with $\kappa = trace(\Lambda)$

The covariance matrix of residuals from the shape mean, C, can be calculated using the outer product of the matrix D with itself,

$$C = \frac{1}{m} D D^T \tag{3.7}$$

However, this matrix is huge, having a width and height of the length of the vector $\mathbf{s}$. We used a more practical alternative making use of the fact that the non-zero eigenvalues of $\frac{1}{m} D^T D$ are identical to the non-zero eigenvalues of $\frac{1}{m} D D^T$, using the inner product, $\frac{1}{m} D^T D$, which has a width and height of the number of samples, m. The eigenvectors and eigenvalues of $\frac{1}{m} D^T D$ can be calculated using the Jacobi Method [63]. We then used these to calculate the eigenvectors and eigenvalues of the covariance matrix C. If we denote the i^{th} eigenvector of C as $\mathbf{x}_i$ and the i^{th} eigenvector of $\frac{1}{m} D^T D$ as $\mathbf{y}_i$, then $\mathbf{x}_i = D^T \mathbf{y}_i$. That is the eigenvectors were found using a weighted sum of the residual vectors using the eigenvectors of $\frac{1}{m} D^T D$ as weights.

The eigenvectors of the covariance matrix, C, are rearranged in descending order of their associated eigenvalues. These rearranged eigenvectors form the columns of the matrix U, and the eigenvalues form the vector $\mathbf{d}$. We denote the i^{th} column of the matrix U as $\mathbf{u}_i$. Here the vectors $\mathbf{u}_i$ are simply the reordered eigenvectors denoted $\mathbf{y}_i$ above. These eigenvectors form a linear basis that exactly spans the space of the faces used to build the PCA model. The eigenvectors $\mathbf{d}$ are related to the variance of the distribution as $\sigma_i^2 = \frac{d_i}{m}$ where d_i is the i^{th} eigenvalue in $\mathbf{d}$ and σ_i is the standard deviation in the direction of the corresponding eigenvector $\mathbf{s}_i$.

A significant benefit of PCA is that we can perform a dimensionality reduction on the face-space. We selected a subset of vectors of U to form a basis from which each exemplar face can be approximately reconstructed and new faces synthesised. As the columns of U had already been reordered we could form a basis in l dimensions simply by taking the first l column vectors in U. We denote this basis $S_{l\times n}^{(l)} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, y_l]$. This optimally minimises the residual error under the $l^2 - norm$. We wished to find the minimal value of l such that an acceptable amount of the variance in the data D is explained by $S^{(l)}$. If we define the cumulative energy as $g_i = \frac{\sum_{i=1}^{l} d_i}{\sum_{i=1}^{m} d_i}$, i.e. the cumulative sum of the singular values up to l weighted by the total sum. l is therefore the minimum value such that $g_l \geq \epsilon$, where ϵ is the amount of variance the model is required to explain . The first l vectors are known as the principal components.
We were therefore able to define a new shape descriptor as a linear combination of principal components:

$$\mathbf{s} = \hat{\mathbf{s}} + \sum_{i=1}^{l} \alpha_i \mathbf{y}_i \tag{3.8}$$

The new descriptor spans the space of the shape exemplars up to an accuracy of g_l. The probability distribution of the shape descriptor is:

$$p(\mathbf{s}) \sim e^{-\frac{1}{2}\sum_i \frac{\alpha_i^2}{\sigma_{\mathbf{s},i}}} \tag{3.9}$$

The range of fitting is thus constrained in a maximum-likelihood sense to those most likely to correspond to the identity of the face shown in a particular image.

A colour descriptor was defined similarly. Before PCA can be performed the colour-maps have to be brought into correspondence. On two-dimensional images, this can be achieved by warping the images such that the shape matches an average prototype using Thin-Plate-Splines [17]. For each three-dimensional face model in the data set, the texture-space

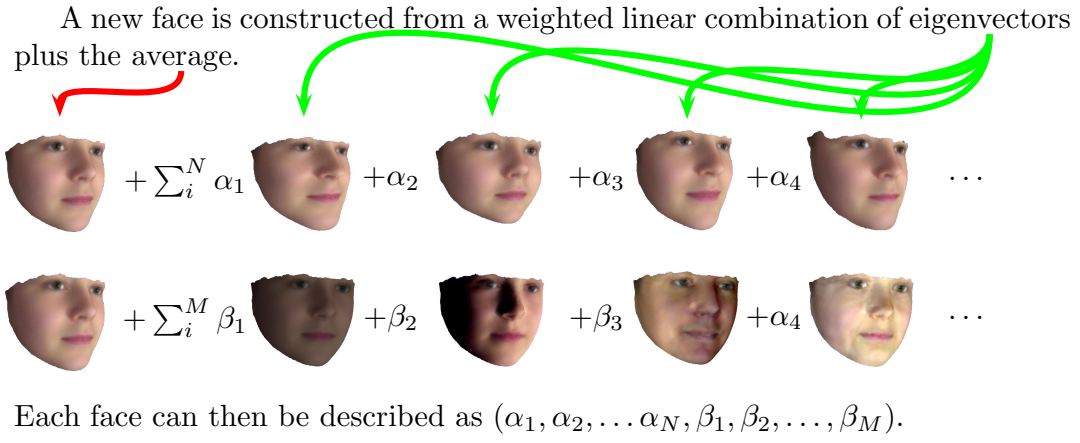


Figure 3.6: Constructing a Morphable Model.

was defined as a two-dimensional image mapped onto the surface of the model using a set of discrete texture coordinates defined for each vertex and linear interpolation to define the mapping within the triangles of the mesh. The colour-maps for each of the face models were registered and the discrete texture coordinates computed during the morphing process.

The pixel values of each input image were concatenated as,

$$\mathbf{t} = (R_1, G_1, B_1, R_2, B_2, G_2, \cdots, R_o, G_o, B_o)^T \tag{3.10}$$

and PCA performed, by the same method as for the shape components, to produce,

$$\mathbf{t} = \hat{\mathbf{t}} + \sum_{i=1}^{l} \beta_i \mathbf{t}^i \tag{3.11}$$

where $\mathbf{t}^i$ is the i^{th} eigenvector of a covariance matrix of the registered colour-maps and β is a set of colour components.

As before the space of the colour model is reduced to the first l principal component vectors. A new face colour can be constructed as a linear combination of the principal components weighted with a set of parameters β. The probability distribution of the new colour descriptor is,

$$p(\mathbf{t}) \sim e^{-\frac{1}{2}\sum_i \frac{\beta_i^2}{\sigma_{\mathbf{t},i}}} \tag{3.12}$$

A new complete face model can be described using the shape parameters α and the colour parameters β combined (see figure 3.3). The combined probability distribution of both the

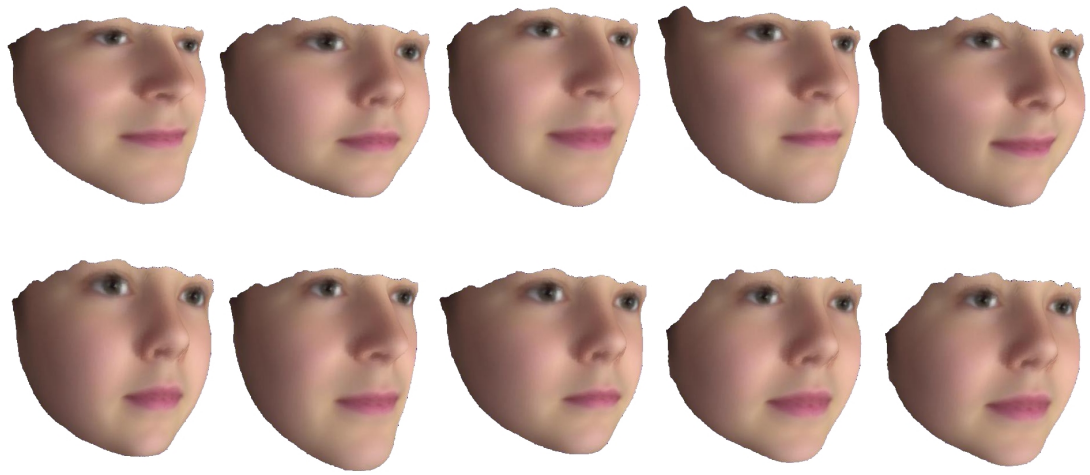


Figure 3.7: Examples of the shape changes associated with the first five Principal Components. Each of the images shows a model rendered with the specified component increased by an arbitrary amount in the top row and decreased by same amount in the bottom row.

shape and colour is defined as,

$$p(\mathbf{s},\mathbf{t}) = p(\mathbf{s}).p(\mathbf{t}) \sim e^{-\frac{1}{2}\sum_i \frac{\alpha_i^2}{\sigma_{\mathbf{s},i}}}.e^{-\frac{1}{2}\sum_i \frac{\beta_i^2}{\sigma_{\mathbf{t},i}}} \tag{3.13}$$

Figures 3.7 and 3.8 show the shape and colour changes associated with the first five Principal Components of the face-set. The first few eigenvectors of the colour principal components clearly contain significant lighting information, this is due to the variable lighting condition under which the faces were scanned and the texture maps produced. This means that some of the model parameters may contain unwanted lighting information in later stages, such as fitting or ageing.

3.4 Summary

In the chapter we have described a method for constructing a three-dimensional Morphable Model. First a set of meshes have to be brought into a one-to-one correspondence using a surface alignment method. We provide an overview of current surface alignment methods and identify desirable features of the alignment method, in particular,

Figure 3.8: Examples of the colour changes associated with the first five Principal Components. Each of the images shows a model rendered with the specified component increased by an arbitrary amount in the top row and decreased by same amount in the bottom row.

- The ability to generate a set of meaningful one-to-one correspondences.
- A technique to fill holes in the mesh.

We also described the method we used to align a set of face models. Second, we described how to construct a statistical face model using Principal Components Analysis. This PCA model provides a probabilistic model of the space of human faces that can be used to guide the process of fitting a three-dimensional face-model to a two-dimensional image. The PCA is also used to describe a face model as a set of lower-dimensional parameters and used to train a statistical ageing model. The Morphable Model forms at the core of the method of synthesising face ageing described in this thesis. It is used to describe a set of faces using a lower-dimensional set of parameters, ageing functions can be constructed through this face space and new faces synthesised from them. The Morphable Model is also used to guide the process of fitting a three-dimensional face-model to an image in order to extract three-dimensional information about the face from that image. The face-fitting process with be described in the next chapter.

Chapter 4

Fitting a Three Dimensional Morphable Model to an Image

In our literature review of current methods for synthesising face ageing, chapter 2, we identified an individualised ageing model as superior in accuracy to a global ageing model. This requires a set of three-dimensional face models of the same individual at multiple age points. At the time of writing such data-set were unavailable, due to the relative novelty of three-dimensional scanning equipment. A suitable set of three-dimensional face models can be generated by extracting three-dimensional information from a set of two-dimensional image. Two-dimensional image-sets being much more readily available. This extraction is achieved by fitting a three-dimensional Morphable Model to the images, by minimising the error between the rendered Morphable Model and the target image.

In this chapter, we outline the algorithms that have been developed for fitting a three-dimensional Morphable Model to a two-dimensional image. From these, we select a suitable algorithm, describe our implementation, and present some results of our implementation. For the face-fitting method to be useful in developing ageing models it must be able to extract an accurate three-dimensional representation of the individual depicted in the image. Towards the end of this chapter we will describe an experiment to evaluate how well the three-dimensional models represent the individual in the image by determining if a set of human observers can recognize the individual from their rendered model.

The face models extracted by this process are used to train a statistical ageing model in

order to synthesis the aged appearance of a face.

We begin by describing some of the most commonly used algorithms for fitting Active Appearance Models (AAM), the two-dimensional analog of three-dimensional Morphable Models (3DMM) and discuss their application to fitting 3DMMs. Then we outline current methods devised specifically for fitting 3DMMS to a two-dimensional image.

4.0.1 Feature Extraction

A description of the shape and colour of a human face in an image can be extracted by aligning the statistical face model with the image in such a way that the error between the rendered face model and the image is minimised. This is normally achieved by varying the parameters of the face descriptor in such a way that a cost function is minimised. Although the parameters of the Morphable Model are defined as a linear combination of basis vectors, the fitting operation is not linear. In general, the intensity of a pixel is not related to its position in the image plane, as a result there does not exist a linear relation between shape changes in the face model and the resulting change in the image intensity of a particular sample point. Other sources of non-linearity result from the three-dimensional nature of the face model, rotations in particular result in occlusions and in changes in relative position of two parts of the surface on the two-dimensional image plain. Illumination changes can be modelled in a linear fashion provided the pose and shape are fixed. However, if the shape or pose are altered the angle of the surface relative to the light sources is also altered resulting in non-linear changes in lighting intensity. Changes in the distribution of shadows, including partial occlusions, due to changes in the surface or rotation relative to light sources, can have dramatic effects on image intensity. Changes in the shape of the model will also results in changes in the texture space as the texture will be stretched or compressed as the surface area is altered by the shape changes.

4.0.2 Alignment based methods

Alignment based methods attempt to fit a rendered model to an image by minimizing the error between a rendered image of the face and an input image. The rendered face image is adjusted such that it matches as closely as possible the input image, thus minimising

the error. In these systems the face is described using an Appearance Model, or Morphable Model, to describe both the shape and colour of the face. Details related to the pose, lighting are separated from the Appearance Model using physical modelling of these attributes, with a set of adjustable parameters so that the simulated physical attributes can be made to match those depicted in the image. The parameters of both the Appearance Model and the physical model are iteratively adjusted such that the error function is reduced towards a global minimum that represents the best match between the rendered face model and the input image. The error function used is normally the l^2-norm or *sum of squared pixel error*. Although variants such as weighted l^2-norm exist, which are potentially more robust to noise and occlusions in the image. Normalized cross-correlation can also be used, in this case the error function is maximized. The error-functions are normally minimised by finding a relationship between the changes in pixel intensity brought about by varying the models physical and appearance parameters and the differences in intensity between the rendered and input images. However this does not in general result is a linear relationship. Many changes in both physical (e.g. a translation) and shape parameters of the model do not result in a linear-change in the intensity values of the pixels. This problem is compounded in three-dimensions as a transform that is linear in three-dimensions is not necessarily linear when projected onto a two-dimensional plane, this is the case with rotations. Alterations of parameters such as rotation, position of lighting and some shape changes introduce changes to the face's silhouette, or distribution of shadows, that can have a marked effect on pixel intensity value while representing a small change in the offending parameter. Finally some of the face can be occluded by a non-face object resulting in pixel values that are unrelated to the face model. It is in tackling these problems that much of the variety between various fitting methods is produced.

By separating the key elements of alignment based fitting we can get an overview of the various directions researchers have taken in tackling the problem of fitting a face model to an image.

1. The sophistication of the rendering model: The more accurately the rendering model can synthesise human face images in variety of physical conditions, the more accurately it can match the pixel values in a particular image, given the correct parameters. Rendering Models range from having a simple point-source and ambient lighting model [16], use of a 9-D Spherical Harmonic basis for lighting [95] to detection and modelling of specular highlights [67].

2. The error function: The error function describes the difference between the rendered face model and a input image in a sensible manner. The ideal error function both ignores irrelevant features, e.g. occlusion, shadows etc. and weights the fitting towards features relevant to the face image. It should also be continuously differentiable so that a gradient descent method can be used to find the global minimum. The correctly match face must minimise this function. Most fitting methods, in both two and three-dimensions, use a squared pixel difference metric. Patterson et al. [58] evaluated the l^2-norm together with the Mutual Information, in terms of the individual and joint entropy of the rendered and input images, and a correlation ratio between the images.

3. Outlier removal: Many methods make use of, or are derived from, the l^2-norm of pixel differences between the input and rendered image. This has the disadvantage of being highly vulnerable to outliers. Outliers in the pixel difference can be an indication of occlusion of the face by a foreign object, or the presence of facial blemish, e.g. a mole, or facial hair that is not captured by the Appearance Model. One method of removing outliers involves weighting the l^2-norm according to a measure of confidence that the sample point is part of the face image and not part of the background. Romdhani et al. [69] use a Talwar function to remove occluded parts of the face model. Romdhani also used a *Lorentzian* estimator to remove outliers [67]. De Smet et al [76] modelled a visibility map as a binary Markov Random Field to take advantage of the spatial coherence of outliers. In two dimensions, Theobald et al. [81] compared the effectiveness of a number of different methods of detecting outliers on AAMs and concluded that, where known, a weighted probability function based on the standard deviation of residuals resulting from fitting AAMs unoccluded image at each pixel was superior to the other methods tested. But in the case where this data is unavailable, the Talwar or Cauchy functions out-perform the others. The choice of functions investigated was limited to those that required minimal re-computation of the Hessian matrix.

4. The method of minimization (or maximization): The algorithm to minimize the chosen error function(s). Most of the error functions used in face fitting suffer from local minima, a varitey of methods to avoid these have been developed. Blanz and Vetter used a stochastic gradient descent method [16], Romdhani and Vetter [69] adapted the inverse compositional alignment method of Baker and Matthews [10] to 3DMMs, this method is based on Gauss-Newton gradient descent. These methods

rely on the computation of the derivatives of the cost function in the image-space or some suitable proxy. This is possible using l^2-norm based function, but non-trivial in other cases. They also suffer from local-minima problems when the data is noisy or the error function is non-convex. Moghaddam used a *downhill Simplex* method to minimize a cost function based on aligning silhouettes [54]. Learning methods such as the regression technique of Cootes et al. [20] have also been used.

5. The method of combining different sources of error: Some of the algorithms use multiple error functions or separately use multiple filters on the images (both input and rendered). These are applied to the model separately, resulting in an iterative multi-pass algorithm [95, 76] or combined using a Bayesian metric to create a single parameter update each iteration [67].

4.1 Literature Review

The face analysis methods of Benson and Perrett [13] and the manipulation techniques of Rowland and Perrett [70], relied on manual placement of points on individual face features for each face and made no attempt to automate the delineation process. With large scale databases or automated systems this method is somewhat unsatisfactory.

Early methods to automate face feature detection, such as Active Shape Models [80, 20], fitted statistical point models to images based on contours in the image intensity. Models were created by delineating points on a set of example faces and performing PCA on the set of points. When fitting to an image, the data points were driven towards strong edge features in the image and constrained to lie within a hyper-ellipsoid defined by mapping the perturbed points into the PCA space and setting a maximum value for the parameters in PCA space. These methods were limited to describing objects with clear and distinct edges. Should the edges of the model be indistinct or close enough that points are perturbed towards an incorrect edge, the model will invariably fail to fit correctly. They extended the method by building local grey-level models in which image intensity was sampled in a line along the normal at each profile point. They were normalised and the grey-level model created by building a covariance matrix of the normalised samples. The points were fitted by minimising the Mahalanobis distance between the samples and mean grey-level [22]. As part of the grey-level model extended along the normal, beyond the area of interest the

method was vulnerable to the random nature of the intensities of the background, around the model.

Descriptors that can use the whole area of interest, e.g. the face, for fitting, tend to be more robust as they don't depend on edge contours and use more information about the area between points to help guide the fitting. Active Blobs (Sclaroff et al. [74]) and Active Appearance Models (Cootes et al. [19]) use triangulated meshes in two dimensions to describe the target shape and colour. Sclaroff described a mesh using PCA to describe the deformations, including translation, rotation, scale etc. The eigenvalues of the covariance matrix used in the PCA were used to define a stiffness for their associated eigenvectors and thus constrain the model. The error between the image and the target was minimised using Levenberg-Marquardt gradient descent. To avoid the influence of outliers the l^2-norm normally associated with Least Squares minimisation was replaced with a Lorentzian influence function. Active Appearance Models use PCA to describe the face model, rather than deformations as in Active Blobs and were introduced by Cootes et al. [19]. The term Active Appearance Model applies to both the face model and the fitting method, to avoid confusion we will use the term Appearance Model (AM) to describe the face model alone.

4.1.1 Active Appearance Models

Active shape models describe the space of two-dimensional face images using a point distribution model [20]. They consist of a statistical model of point distribution and an individual statistical appearance model for a patch about each point. Each point is fitted individually with the shape model being used to guide the model towards the correct features. This only models the shape of the face and not its colour, it is also highly vulnerable to changes in pose and illumination. Active Appearance Models also by Cootes et al. [19] model both the shape and colour of the space of face images. Each face consists of a two-dimensional mesh with the vertices placed on key feature points and a colour map. The AAM is defined as a linear combination of basis vectors of the shape points plus the average positions of the points. A new face shape s can be constructed using a set of parameters p as,

$$\mathbf{s}(\mathbf{p}) = \hat{\mathbf{s}} + \sum_{i}^{N} p_i \mathbf{s}_i \text{ ; where } \mathbf{p} = \{p_1, p_2, \ldots, p_N\} \tag{4.1}$$

where $\hat{\mathbf{s}}$ is the average of a set of shapes. The basis vectors $\mathbf{s}_i$ from a matrix S and are found using Principal Component Analysis (see section 3.3). The colour map is defined similarly, the faces are first warped such that the shape points align with the average, and a set of basis vectors calculated using PCA on the vectorized pixels from the aligned exemplar face images. Shape and colour can be further combined by a subsequent application of PCA to the combined shape and colour parameters. This model is similar to 3DMMs in many respects since they are both built from separate PCA models of vectorized shape and colour components, but they differ in a number of important respects. Firstly 3DMMs are typically more dense, requiring more vertices to describe the three-dimensional shape of the model, lighting can be explicitly realized on a 3DMM through simulation of physical light transport, whereas in AAMs it normally has to be removed before hand or statistically modelled. Vetter's original implementation only defined colour on the vertex points, whereas AAMs and our implementation of the 3DMM defines colour between vertex points using a texture map.

3DMMs and their two-dimensional cousin AAMs provide an holistic model of the entire face. ASMs on the other hand break the face up into a small number of individual points to be fitted independently, using the ASM as a constraint [20, 44]. The number of sample points on a 3DMM is significantly larger. A compromise between the flexibility of the point distribution based model with the improved surface description of the 3DMM is found by segmenting the face into regions, e.g. nose, eyes, mouth etc. and creating separate shape and colour models for each region. The boundaries between regions are subsequently smoothed to provide a neat join. The surface problem also applies to 3DMMs as nothing in the description prevents unrealisable faces caused by intersecting surfaces. Provided the parameters are kept to within a few standard deviations of the mean, this should not be a problem.

4.1.2 Fitting an Active Appearance Model to an image

In the original Active Appearance Model paper Cootes et al. assumed a constant linear relationship between the error in the images and the error in the parameter updates [19]. They defined a difference vector as the difference between the rendered AAM with the concatenated shape and colour parameters $\mathbf{p} = \{\alpha, \beta\}$ and the face image,

$$\delta\mathcal{I}(\mathbf{x}) = \mathcal{I}(AAM(\mathbf{x}; \mathbf{p})) - \mathcal{T}(\mathbf{x}) \tag{4.2}$$

where $\mathbf{x}$ is a two-dimensional sample point, $\mathcal{T}(\mathbf{x})$ is the intensity of the sample at position $\mathbf{x}$ on the target image and $\mathcal{I}(AAM(\mathbf{x};\mathbf{p}))$ is the rendered appearance model with parameters $\mathbf{p}$ with the sample point $\mathbf{x}$ warped to a new location $AAM(\mathbf{x};\mathbf{p})$ by the shape changes imposed on the appearance model by the parameters $\mathbf{p}$. A parameter update is found by finding a change in $\mathbf{p}$ that minimises the square magnitude of the distance vector $||\delta\mathcal{I}||^2$. Under the assumption that the starting point and the current parameters were close.This parameter update is calculated from the relationship between the image error $\delta\mathcal{I}$ and the parameter error $\delta\mathbf{p}$,

$$\delta\mathbf{p} = A\delta\mathcal{I} \tag{4.3}$$

The update matrix A was calculated by linear regression between known perturbations of the parameters, $\mathbf{p}$, and the resulting errors between the rendering of the unperturbed appearance model and the rendering of the perturbed model. Cootes et al. also suggested using an estimate of the Jacobian, $\mathbf{J} = \frac{\partial\mathbf{x}}{\partial\delta\mathbf{p}}$, to perform conjugate gradient descent. This algorithm assumes that the matrix A or its equivalent is constant, when in fact it varies with respect to the parameters $\mathbf{p}$ [51]. In common with other two-dimensional alignment methods it is also poor at fitting to faces that are rotated from the angle of the training image. A Multi-view AAM trained on facial images from multiple poses was proposed by Cootes et al. as a solution to this problem [21].

4.1.3 The Kanade Lucas Tomasi Algorithm

The Kanade Lucas Tomasi (KLT) algorithm [49] is one of the most popular image-alignment and motion tracking techniques in computer vision. Originally developed as an algorithm for point feature tracking it has been extended to cover numerous applications ranging from optical-flow and tracking to Active Appearance Model (AAM) fitting. There are numerous incarnations of this algorithm, their primary common element is the solution of a cost function by Gauss-Newton optimisation. They differ in terms of whether they use an additive or compositional update; the method of descent, Gauss-Newton, Newton-steepest descent or Levenberg-Marquardt and in the cost function minimised. Finally whether the rendered image is fitted to the target image, or vice-versa can have important implications on the efficiency of the algorithm. These algorithms are reviewed in detail by Baker and Matthews in [9, 8, 6, 7, 11]. Here we focus on their application to face-fitting, focusing on two algorithms; the forward additive algorithm and the inverse-compositional algorithm.

Both using Levenberg-Marquardt gradient descent using the l^2-norm as the error function. The algorithm was extended for fitting appearance models by Matthews and Baker [51].

The aim of the KLT algorithm is to minimise the l^2-norm of the difference in pixel intensity between the target image, $\mathcal{T}$, and a warped input image, $\mathcal{I}$, by varying the parameters of the warp. The flexibility of the method is that assuming the parameters are independent we can apply this method directly to rendered images of the Morphable Model, provided that we are able to differentiate the model with respect to each parameter of the model. We use the notation of Baker and Matthews and define an image warp $W(\mathbf{x};\mathbf{p})$, parameterised with parameters $\mathbf{p}$. $W(\mathbf{x},\mathbf{p})$ maps a pixel at location $\mathbf{x}$ to another (sub-pixel) location. Here the mapping defines a general warp, e.g. an affine warp. In the next sections, we will outline how this method was extended for use with AAMs and its applicability to 3DMMs.

The sum of squared errors in given by,

$$\chi^2 = \sum_{\mathbf{x}\in\Omega} (\mathcal{I}(W(\mathbf{x};\mathbf{p}+\delta\mathbf{p})) - \mathcal{T}(\mathbf{x}))^2 \tag{4.4}$$

where Ω is the subset of all samples in the image. The equation is solved in a iterative manner with the parameters being updated in an additive manner. Lucas et al. [49] assumed that there exists a linear relationship between the target image and the warped input image.

$$\mathcal{T}(\mathbf{x}) \approx \mathcal{I}(W(\mathbf{x};\mathbf{p})) + \nabla\mathcal{I}(W(\mathbf{x};\mathbf{p}))\frac{\partial W(\mathbf{x};\mathbf{p})}{\partial \mathbf{p}} \tag{4.5}$$

where $\nabla\mathcal{I}(W(\mathbf{x};\mathbf{p})) = (\frac{\partial\mathcal{I}}{\partial x}\frac{\partial\mathcal{I}}{\partial y})$ is the gradient of image $\mathcal{I}$ evaluated at $W(\mathbf{x};\mathbf{p})$ and,

$$\frac{\partial W(\mathbf{x};\mathbf{p})}{\partial \mathbf{p}} = \left(\frac{\partial W(\mathbf{x};\mathbf{p})}{\partial p_0}, \frac{\partial W(\mathbf{x};\mathbf{p})}{\partial p_1}, \ldots, \frac{\partial W(\mathbf{x};\mathbf{p})}{\partial p_N}\right) \tag{4.6}$$

is the Jacobian of the warp with respect to each of the its parameters. This assumption only holds if $\mathcal{T}$ and $\mathcal{I}(W(\mathbf{x};\mathbf{p})$ are close. Applying a first order Taylor series expansion to equation (4.4) and differentiating w.r.t. to the parameters $\mathbf{p}$ gives an update for the parameters $\mathbf{p}$ at each iteration as,

$$\delta\mathbf{p} = H^{-1}\sum_{\mathbf{x}\in\Omega}\left[\nabla I\frac{\partial W}{\partial \mathbf{p}}\right]^T [T(\mathbf{x}) - I(W(\mathbf{x};\mathbf{p}))] \tag{4.7}$$

where the Hessian H is defined as:

$$H = \sum_{x\in\Omega}\left[\nabla I\frac{\partial W}{\partial \mathbf{p}}\right]^T \left[\nabla I\frac{\partial W}{\partial \mathbf{p}}\right] \tag{4.8}$$

The parameters of the warp, $\mathbf{p}$ are found by iteratively solving equation (4.7) and updating $\mathbf{p}$ as $\mathbf{p} \leftarrow \mathbf{p} + \delta\mathbf{p}$.

The only requirement on the set of warps is that they are differentiable w.r.t. $\mathbf{p}$ [9]. As a result, this algorithm can be adapted directly for use with AAMs or 3DMM (see section).

4.1.4 Inverse KLT algorithm

In the general KLT algorithm there is a relationship between the derivatives of the parameters and the parameters themselves, the Jacobian has to be recomputed at each iteration resulting in longer computation times. The Inverse composition algorithm reversed the role of the rendered image and the target, and updated the parameters by composing the warp, as a result the derivatives are always evaluated at $\mathcal{W}(\mathbf{x}; 0)$ and are thus constant [10].

Here the roles of the the input image and the target are reversed in the error function and instead of an additive update step the warp $\mathcal{W}(\mathbf{x}; \delta p)$ is inverted and composed with the current warp.

$$\chi^2 = \sum_{\mathbf{x}\in\Omega} [W(\mathbf{x}; \delta\mathbf{p}) - \mathcal{I}(W(\mathbf{x}; \mathbf{p}))]^2 \tag{4.9}$$

$$\mathcal{W}(\mathbf{x}; \mathbf{p}) \leftarrow \mathcal{W}(\mathbf{x}; \mathbf{p}) \circ \mathcal{W}(\mathbf{x}; \delta\mathbf{p})^{-1} \tag{4.10}$$

as with the KLT algorithm, equation (4.9) is minimised by expansion with a first order Taylor series expansion and derivation w.r.t. to $\mathbf{p}$.

$$\chi^2 = \sum_{\mathbf{x}\in\Omega} \left[\mathcal{T}(W(\mathbf{x}; 0)) + \nabla\mathcal{T}(W(\mathbf{x}; 0)) \frac{\partial W(\mathbf{x}; 0)}{\partial \mathbf{p}} \delta\mathbf{p} - \mathcal{I}(W(\mathbf{x}; \mathbf{p})) \right]^2 \tag{4.11}$$

This gives the solution to the least-squares problem as,

$$\delta\mathbf{p} = H^{-1} \sum_{\mathbf{x}} \left[\nabla\mathcal{T}(; 0) \frac{\partial W(; 0)}{\partial \mathbf{p}} \right] [\mathcal{I}(W(; \mathbf{p})) - \mathcal{T}] \tag{4.12}$$

$$H = \sum_{\mathbf{x}} \left[\nabla\mathcal{T} \frac{\partial W(; 0)}{\partial \mathbf{p}} \right]^T \left[\nabla\mathcal{T} \frac{\partial W(; 0)}{\partial \mathbf{p}} \right] \tag{4.13}$$

since the Jacobian and Hessian are both evaluated at $\mathbf{p} = 0$ they don't depend on $\mathbf{p}$ and don't require updating between iterations.

This relies on the template $\mathcal{T}$ not changing, and the alignment consisting only of a change in shape that can be described using a warp. Both 3DMMs and AAM consist of both

shape and colour models and so changes in intensity resulting from updating the colour parameters have to be taken into account. It also places restrictions on the warp, it requires that W be differentiable, invertible, associative, exhibit closure (i.e. a warp composed with another produces a warp) and contain the identity element (in this case $W(\mathbf{x};0)=\mathbf{x}$), i.e. it must form a group.

4.1.5 Projecting Out Appearance Variation

Matthews and Baker [51] modified the Inverse Compositional Kanade Lucas Tomasi image alignment algorithm [9] to fitting two dimensional AAMs. The KLT image alignment algorithm has similarities to the AAM fitting method of Cootes et al. [19] and the Morphable Model fitting method of Blanz and Vetter [16], in that it attempts to minimise the l^2-norm of the image difference between the target image and the rendered face model using a gradient descent technique. In order to modify the inverse compositional KLT algorithm to fitting AAMs they used the project-out method of Hager and Belhumeur [32], to separate texture changes in the model from shape changes. Their method involved defining the error function as follows,

$$||\hat{\mathbf{t}}(W(\mathbf{x},\delta\alpha))+\sum_i \beta_i \mathbf{t}_i(W(\mathbf{x},\delta\alpha))-\mathcal{I}(W(\mathbf{x};\alpha))||^2 \tag{4.14}$$

Where $\hat{\mathbf{t}}$ is the average texture from the AAM model and the vectors $\mathbf{t}_i$ are the eigenvectors describing the colour variations. α,β are the shape and colour parameters of the AAM respectively. Here $W(\mathbf{x};\alpha)$ denotes the displacement of the sample point $\mathbf{x}$ by an AMM shape transform with parameters α. Note that the update of the shape parameters, $\Delta\alpha$ appears in the first term rather than the second. We solve the equation by differentiating w.r.t the parameters α and β and ρ simultaneously. We follow their notation and denote the subspace spanned by $\mathbf{t}_i$ by $span(\mathbf{t}_i)$ and its orthogonal complement by $span(\mathbf{t}_i)^\perp$ give the error function,

$$||\hat{\mathbf{t}}(\mathbf{x})+\sum_i \beta_i \mathbf{t}_i(\mathbf{x})-\mathcal{I}(W(\mathbf{x};\alpha))||_{span(\mathbf{t}_i)^\perp}+||\hat{\mathbf{t}}(\mathbf{x})+\sum_i \beta_i \mathbf{t}_i(\mathbf{x})-\mathcal{I}(W(\mathbf{x};\alpha))||^2_{span(\mathbf{t}_i)} \tag{4.15}$$

where $||.||^2_L$ is the L^2-norm of the vector projected in the linear subspace L. Since the inner product of any vector in the subspace $span(\mathbf{t}_i)$ and any vector in the subspace $span(\mathbf{t}_i)^\perp$ is zero, components in the subspace of $\mathbf{t}_i$ can be ignored. The first term therefore does not depend on β_i. The minimum of the second term evaluates to zero for any α or ρ.

The two terms can be minimised separately, first by finding α and ρ using the first term and then find β setting α and ρ as constants. As β_i are orthogonal from the definition of the PCA model the second term has a simple linear solution;

$$\beta_i = \sum_{\mathbf{x}} \mathbf{t}_i.[\mathcal{I}(W(\mathbf{x};\alpha,\rho)) - \hat{\mathbf{t}}]. \tag{4.16}$$

As the shape displacement of the first term is calculated in the subspace $span(\mathbf{t}_i)^{\perp}$, the gradients have to be projected into $span(\mathbf{t}_i)^{\perp}$.

$$J_i = \nabla\hat{\mathbf{t}}\frac{\partial W}{\partial \alpha} - \sum_i \left[\sum_{\mathbf{x}} \mathbf{t}_i.\nabla\hat{\mathbf{t}}(\mathbf{x})\frac{\partial W}{\partial \alpha}\right]\mathbf{t}_i(\mathbf{x}) \tag{4.17}$$

The first term of (4.15) is computed using a Taylor series expansion,

$$\sum_{\mathbf{x}} = \left[\mathcal{I}(W(\mathbf{x};\alpha)) - \hat{\mathbf{t}}(W(\mathbf{x};0)) + \nabla\hat{\mathbf{t}}\frac{\partial W}{\partial \alpha}\delta\alpha\right] \tag{4.18}$$

Assuming that $W(\mathbf{x};0) = \mathbf{x}$ this is solved as a least squares problem,

$$\delta\alpha = H^{-1}\sum_{\mathbf{x}} J^T[\mathcal{I}(W(\mathbf{x};\alpha)) - \hat{\mathbf{t}}] \tag{4.19}$$

where $H = J^T J$ is the Hessian and J the projected-out derivatives of α. The efficiency of this algorithm is due principally to the Jacobian no longer being dependent on α, it can be evaluated at the start of the algorithm and considered constant thereafter. The separation of the texture terms also introduces efficiency as these can be evaluated once after the shape parameters have been found. However the project out method assumes that the derivatives of the shape parameters are not affected by changes in the colour parameters, this assumption does not hold in general. As an example if the subject has a beard this will provide edges that can be fitted to that don't exist on a face without facial hair, or a different style of facial hair. As a result this method is only suitable for a fitting of a single subject [64].

The shape parameters are updated by composition with the current shape warp, $W(\mathbf{x};\alpha) \leftarrow W(\mathbf{x};\alpha) \circ W(\mathbf{x};\Delta\alpha)^{-1}$. To compose the AAMs shape transformation with an inverted shape transformation they used a per-triangle piece-wise affine warp based on barycentric coordinates within the triangle, they then used a first-order approximation that allowed them to define $W(\mathbf{x};\Delta\alpha)^{-1} = W(\mathbf{x};-\Delta\alpha)$. As pointed out by Romdhani [69] these displacements actually occur at two different points, $\mathbf{x}$ for the warp W and $W(\mathbf{x};\Delta\alpha)$ for the warp W^{-1}, for displacements within a triangle this approximation holds as the

warp is linearly interpolated within the triangle containing $\mathbf{x}$, however for points outside the triangle this approximation introduces errors. The method also assumes that the colour components $\mathbf{t}_i$ do not vary with any of the parameters, however the colour of the face model and the distribution of the intensities does vary with changes in lighting. The Jacobian of the shape update is also derived for two-dimensional warps, it has to be updated due to changes caused by rotations.

4.2 Fitting a Morphable Model

Blanz and Vetter [16] used a stochastic gradient descent technique to minimise the l^2-norm between the rendered image of the Morphable Model and a target image. We can describe their method by defining $\mathcal{M}(\mathbf{x};\mathbf{p})$ to be the intensity of $\mathcal{M}$ sampled at point $\mathbf{x} \in \Re^2$ and $\mathcal{I}(\mathbf{x})$ to be the intensity of the image $\mathcal{I}$ sampled at point $\mathbf{x}$. All parameters necessary for rendering the 3DMM; position, rotation, scale, perspective, shape and colour, are concatenated in $\mathbf{p}$. The l^2-norm can be defined as,

$$\chi^2 = \sum_{\mathbf{x}\in\Omega} (\mathcal{M}(\mathbf{x};\mathbf{p}) - \mathcal{I}(\mathbf{x}))^2 \tag{4.20}$$

Using the probability density function from the PCA face model they defined the probability of observing an input as:

$$E = \frac{\chi^2}{\sigma_G^2} + \sum_i \frac{p_i}{\sigma_i} \tag{4.21}$$

where χ^2 is the sum of squared pixel error (4.20), σ_G^2 is the standard deviation of a Gaussian model to describe noise in the image and σ_i is the standard deviation of the i^{th} parameter in face model. For shape and colour these are defined from the PCA model (equation (3.9) in section 3.3), for the translation, rotation, scale and lighting parameters they are manually chosen such that they cover a 'reasonable' range of values for the particular parameter. This equation can be used as a cost function and the parameters updated using the derivatives of the cost function,

$$p_i \mapsto p_i - \lambda_i . \frac{\partial E}{\partial p_i} \tag{4.22}$$

where λ_i is a step-size operator. The partial derivatives $\frac{\partial E}{\partial p_i}$ can be calculated from the equation (3.9) and (4.22) using the chain-rule.

To improve the speed of the fitting they used a subset of 40 triangles and used a single

sample point in the centre of each triangle. Shadowing effects and hidden surfaces were recalculated every 1000 steps.

Romdhani et al. [68] used optical flow to generate a set of vectors describing the displacement between the rendered and input images. Shape changes could then be calculated by solving a linear equation between the optical-flow vectors and the shape change derivatives. Translation, rotation, scale and lighting are non-linear and so were solved using Levenberg-Marquardt minimisation [63]. By inverting the lighting model the texture model could be updated linearly.

4.2.1 Extending the Inverse KLT algorithm to three-dimensional Morphable Models

Romdhani et al. [69] extended Matthews and Baker's method (see section 4.1.4) for use with Morphable Models. Unlike Matthews and Baker they made a conceptual distinction between the image and the reference frame that allowed them to avoid using the linear approximation to the inverse shape transform and apply a direct inversion based on searching a triangle list. For a point in the image frame, say $\mathbf{x}$ they found the triangle in the Morphable Model projected by $M(;\alpha,\rho)$ and determined its relative position within it. The position of the inverse shape projection of $\mathbf{x}$ was defined as the same relative position in the corresponding triangle of the Morphable Model in the reference frame. This definition avoided the problem of approximating a transform that crossed triangular boundaries and the cost of performing a search through the triangles of the Morphable Model. The resulting shape updates were composed using a series of projection and inverse projection transforms using the current parameters α^c, the derivative parameters α^d (the shape parameters on which the derivatives were calculated, usually the starting point of the search) and the updates $\Delta\alpha$,

$$M^*(\mathbf{u_i}) = M(\mathbf{u};\alpha^c) \circ M^{-1}(\mathbf{x};\alpha^d) \circ M(\mathbf{u}_i, \alpha + \Delta\alpha) \tag{4.23}$$

Rather than update the three-dimensional model each iteration with a new α^c they found that it was preferable to defer this step, replacing $M(\mathbf{u};\alpha^c)$ with a set of points in two dimensions generated from the remaining two projections. This avoids a time consuming step and they claimed is more accurate in general as M^* does not lie within the span of the Morphable Model. However this method also presents a disadvantage as it can introduce errors, resulting in malformed face models. They found empirically that the project out

method resulted in errors in the shape Jacobian that affected fitting, they therefore found the texture parameters of the model simultaneously with the shape models. In common with its two dimensional counter-part Romdhani et al's inverse composition algorithm was vulnerable to changes in pose, however unlike two-dimensional AAMs it did not need to be retrained to deal with slight changes in pose. Initialising the algorithm to capture the derivatives at as close an angle as possible to the correct face pose was necessary.

Romdhani et al [69] described only how to project discrete points, it was not specified whether the derivatives of the image $\mathcal{I}$, $\nabla\mathcal{I}$ were being correctly scaled in projections and inverse projection. Although it is simple from the definition to define a within triangle linear approximation that can warp the image derivative from the image frame to the reference frame. This is necessary as each iteration required $\nabla\mathcal{I}$ to be warped by two different projection, $M^{-1}(;\alpha^d) \circ M(;\alpha^d)$ and $M(;\alpha^c)$. The first is a null projection in the frame of the derivatives the second is in the frame of the current model shape. The second projection can cause scaling and rotation effects in the direction of the image derivative.

4.2.2 Feature Alignment

Due to the lack of accuracy in fitting the Morphable Model using image intensity alone some authors have attempted to guide the fitting by isolating a set of features in the image and guiding the fitting towards them. Blanz and Vetter [15] fitted points manually to the input image, both to initialise the model and to guide the fitting. These points, such as the centre of the eyes, mouth nose, ears (if visible), were delineated by hand on both the set of images and on the Three-Dimensional Morphable Model.

Romdhani [67] used a *canny edge detector* to find a set of two dimensional edge points in an image. Edge points on the three-dimensional model are brought into correspondence with the edge points in the image by finding the point on the image closest to it using the *Chamfer Distance Transform*. The distance between these two points is minimized using the probability of the edge map given a set of view and model parameters as a prior.

Two Dimensional Methods

A method of extracting three-dimensional models from two-dimensional AAMs was explored by Xiao et al. [93]. They observed that the space of possible 3DMMs under a weak perspective projection can be spanned using a AAM of at most $6n$ dimensions where n is the number of dimensions of the 3DMM. This combined 2D-3D model also spans a large number of parameters that do not realize a legitimate face model. They attempted to amortize this problem by maintaining two simultaneous models, one 2D and one 3D and applying a cost function to the AAM fitting algorithm to constrain the search space to realizable models [52].

4.2.3 Shape from Shading

Shape from shading is the process of recovering a three-dimensional surface shape from the intensity of an image. The surface normal is estimated based on solving partial differential equations linking the image intensity to the reflectance map based on the assumption that the surface is Lambertian. In this case the illumination I is related to the reflectance as $\hat{\mathbf{l}}.\hat{\mathbf{n}}$ where $\hat{\mathbf{l}}$ is the normalized light vector and $\hat{\mathbf{n}}$ is the surface normal. The set of surface normals are constrained to fall on the illumination cone defined by Lambert's Law $I - \mathbf{l}.\mathbf{n} = 0$. The surfaces are regularized by applying a constraint such are curvature or gradient consistency in order to generate more reasonable surfaces [61, 91]

Wu et al. [92] used shape from shading to perform gender classification. They did not attempt to reconstruct the three-dimensional surface, instead performing Independent Component Analysis directly on the needle-map. As the normal vector exists only on the surface of a sphere of radius 1 i.e. $\mathbf{n} \in S^2$ statistical methods on $\Re^3$ fail to produce an accurate result. The authors used Principal Geodesic Analysis to find the mean and principal axes on the surface of the sphere.

Current Shape from shading formulations rely on specific lighting and camera set-ups, e.g. a distant light source or a light source at the optical centre of the image. Due to the varied nature of lighting conditions in our images this made shape-from-shading unsuitable for our purposes.

4.2.4 Error Functions

Alignment based fitting functions attempt to minimize the error between a rendered image of the 3DMM and the input image of the face. The choice of these functions is therefore appropriate for consideration. One of the simplest and most common choice of function is *Ordinary Least Squares* or l^2*-norm* . We define $\mathcal{I}$ and $\mathcal{R}$ as as the input and rendered image respectively. An individual sample at a point $\mathbf{x}$ in the image $\mathcal{I}$ is defined as $\mathcal{I}(\mathbf{x})$. Then the l^2-norm is:

$$\chi^2 = \sum_{\mathbf{x}} (\mathcal{I}(\mathbf{x}) - \mathcal{R}(\mathbf{x}))^2 \tag{4.24}$$

Combined with a gradient descent method, 4.24 will minimize the squared residual between the two images. Due to the squared term this function has a tendency to bias towards outliers, in the context of face-fitting this has the advantage that it drives the fit towards prominent features as these exhibit the largest pixel value differences, for example the boundary between the white of the eye and eye-lids. If the distribution of outliers is skewed, estimates are biased. As an example, the presence of glasses induces outliers that are not part of the model and thus distorts the fitted result. The presence of outliers can also slow the algorithm down as more iterations are required for convergence Alternatively it is possible to use a function weighed by a confidence value. The value is determined by the likelihood a particular samples value belongs to the relevant distribution.

$$\chi^2 = \sum_{\mathbf{x}} \lambda_{\mathbf{x}} E(\mathbf{x})^2 \tag{4.25}$$

Where $E(\mathbf{x}) \equiv (\mathcal{I}(\mathbf{x}) - \mathcal{R}(\mathbf{x}))$. This function integrates poorly with the Inverse-composite KLT algorithm as it requires the Hessian matrix to be recomputed each iteration.

Talwar Function

Numerous variations on the function to calculate the weights $\lambda_{\mathbf{x}}$ for each sample point $\mathbf{s}$ exist, they call attempt to reduce the influence of outliers on the fitting by tapering the functions for higher values of $(\mathcal{I}(\mathbf{x}) - \mathcal{R}(\mathbf{x}))$. One of the simplest is the *Talwar* function [34], that simply clips the function at a predetermined error.

$$\lambda_{\mathbf{x}}(E(\mathbf{x})) = \begin{cases} 1 & E(\mathbf{x}) \leq c \\ 0 & E(\mathbf{x}) > c \end{cases} \tag{4.26}$$

where c is predefined constant, we defined this using the median of the function $E(\mathbf{x})$. This function has been used by some authors [69, 76] to remove hidden-surfaces from three-dimensional models.

Cauchy Function

The Cauchy function uses the assumption that the samples from the two images $\mathcal{I}$ and $\mathcal{R}$ are independently identically normally distributed, i.e. $\mathcal{I}(\mathbf{x}), \mathcal{R}(\mathbf{x}) \in N(\mu, \sigma)$, this assumption is reasonable as we are attempting to match the two images. The distribution of the error function $E(\mathbf{x}) \equiv (\mathcal{I}(\mathbf{x}) - \mathcal{R}(\mathbf{x}))$ is normal with a mean of zero and an unknown standard deviation. The confidence that a sample $E(\mathbf{x})$ is within the distribution of the image residuals and not an outlier caused by occlusion can be determined by a t-test with one-degree of freedom. This special case of the t-distribution is the Cauchy distribution. Cauchy weights can be defined as,

$$\lambda_{\mathbf{x}}(E(\mathbf{x})) = \frac{1}{1 + (\frac{E(\mathbf{x})}{c})^2} \tag{4.27}$$

again c is a constant derived from the median of $E(\mathbf{x})$. This function has been used by Theobald [81] on AAMs and Romdhani [67] on 3DMMs.

4.3 Rendering

When attempting to fit a Morphable Model to a two dimensional image, it is important that the model is rendered as accurately as possible. The effects of pose and illumination must be correctly modelled. The pose is modelled by projecting the model onto a two-dimensional surface via a linear transform. The lighting is more complicated and requires a mathematical model of both direct and indirect illumination on the surface.

The 3DMM's shape is projected onto a two-dimensional surface using a linear transform, describing the view and perspective transforms. The view matrix combines a three-dimensional rotation, scaling and translation to map the shape vectors $\mathbf{s}$ from an object centred position to one relative to the view point.

$$V = R_\gamma R_\theta R\phi.S + t \tag{4.28}$$

where

$$R_\gamma = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} R_\theta = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} R_\phi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\theta \\ 0 & \sin\phi & -\cos\theta \end{bmatrix}$$

$$S = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix}, t = t_x, t_y, t_z$$

The angles θ, ϕ and γ are rotations around the principal axes vertical, horizontal and depth respectively. The concatenated parameters $\theta, \phi, \gamma, s_x, s_y, s_z, t_x, t_y, t_z$ are denoted as ρ. The shape of the Morphable Model is projected onto the two dimensional image plane by first building the Morphable Model using the shape parameters α using the view-perspective matrix V.

$$s^{2D} = V[\bar{\mathbf{s}} + \alpha^T s] \tag{4.29}$$

4.3.1 Inverse Shape Projection

This projection can be used to define a transform $M(\mathbf{u}; \alpha, \rho)$ that maps a point $\mathbf{u}$ on the face model surface to a point on the two dimensional surface. This mapping is defined for each vertex in the face model, and for points in the triangles of mesh using linear interpolation. The mapping is thus continuous in $\mathbf{u}$ and injective provided occlusions are ignored. The inverse mapping is not injective as the face only covers part of the image, it is also not surjective as occlusions will result in points in the texture-space $\mathbf{u}$ not having a point on the image plane that maps to them. It is possible however to compute an inverse projection $M^{-1}(\mathbf{x}; \alpha, \rho)$ that maps from the image plane $\mathbf{x}$ to the texture space $\mathbf{u}$ as outlined by [69]. First the Morphable Model is projected onto the image plane using $M(; \alpha, \rho)$, then the triangle containing the point $\mathbf{x}$ is found and the point $\mathbf{u}$ is the relative position of that point in the equivalent triangle in the texture-space. This inversion function is only defined for parts of the image plane covered by the triangles of the projected Morphable Model. Within the area covered by the Morphable Model in both the image plane and the texture-space, $M = M^{-1}M^{-1}$ and $M(\mathbf{x}; 0) = \mathbf{x}$ is the null transform.

Indirect Illumination

Experimental work by numerous authors, including [65] have shown that the 'space of images produced by an object under a wide range of lighting conditions lies near a low dimensional linear subspace in the space of all possible images.' Ramamoorthi [65] used principal component analysis of face images under multiple lighting conditions but constrained pose to show that the first five eigenvectors accounted for 95% of image variance. Basri and Jacobs [12] and Ramamoorthi [65] independently showed that under conditions with no shadows cast on the object by external sources, the space of illumination functions can be accurately approximated using an expansion of spherical harmonic functions. Let Y_n^m be a spherical harmonic function of degree n and order m where $n = 0, 1, 2, \ldots$ and $-n \leq m \leq n$ then:

$$Y_n^m(\theta, \phi) = \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}} P_n^{|m|}(\cos\theta) e^{im\theta} \tag{4.30}$$

where P_n^m are the *associated Legendre function*, defined as

$$P_n^m(z) = \frac{(1-z^2)^{\frac{m}{2}}}{2^n n!}\frac{d^{n+m}}{dz^{n+m}}(z^2-1)^n \tag{4.31}$$

These equations form an orthonormal basis, thus any piecewise continuous function f can be described as a linear combination of an infinite series of harmonics. In the case where f is known, it can be evaluated at some point $u \in S^2$ on the sphere by:

$$f(\mathbf{u}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} f_n^m Y_n^m(\mathbf{u}) \tag{4.32}$$

f_n^m is scalar value of

$$f_n^m = \int_{S^2} f(\mathbf{u}) Y_n^{m*}(\mathbf{u}) d\mathbf{u} \tag{4.33}$$

where $*$ denotes the complex conjugate of Y_n^m
The function f can be approximated to a desired degree using a finite series of components of order N. Thus:

$$f(\mathbf{u}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} f_n^m Y_n^m(\mathbf{u}) \approx \sum_{n=0}^{N}\sum_{m=-n}^{n} f_n^m Y_n^m(\mathbf{u}) \tag{4.34}$$

We describe the illumination function of a face model using $f(\mathbf{u})$, defining $\mathbf{u}$ as the surface normal, $\hat{\mathbf{n}} = (\theta, \phi)$ of a point of the face model. $\mathbf{u}$ is therefore a point on the unit sphere. The illumination function then becomes a finite series of order N (and length N^2)

spherical harmonic functions. Which can then be parameterised as a set of N^2 parameters using f_n^m. These methods required capturing a large set of images in order to compute the illumination function, whereas we need to carry out the inverse operation, that is compute the illumination function from an image.

Zhang and Samaras [95], solved for the harmonic functions using a least squares approach. For each of the 9 basis functions, delta images $\Delta\mathcal{I}_i$ are generated where a single parameter of $f(\mathbf{u})$ is altered by a small amount δ, thus isolating the changes in intensity caused by varying the illumination function with that parameter.

We denote the combined lighting parameters $f_{r,n}^m, f_{g,n}^m, f_{b,n}^m, n = 0, 1, 2, 3$, concatenated in to a vector τ. The parameters are calculated as the solution to a linear equation.

$$\tau_i = (\Delta\mathcal{I}_i.\Delta\mathcal{I}_j)^{-1}\Delta\mathcal{I}_i.(\mathcal{T} - \mathcal{I})\,\forall i, j \tag{4.35}$$

This linear equation is then solved and the parameters updated.

A significant failing of the 9D illumination model is the inability to handle shadows. The illumination integral can be performed as a dot product of the spherical harmonic basis and the weighting parameters, only in the case of a double product integral. The addition of a shadowing term to the integral turns it into a triple-product integral which is significantly more complicated to solve [56] and also slower. Soft shadowing effects of sufficiently low frequency and large illumination sources can be modelled.

Direct Illumination

The 9D subspace assumes that the light illuminating the subject is distant and diffuse, this is not the case for many of the subjects in the database as they were taken using flash photography. The flash can be modelled as a single point light source. The spherical harmonic basis functions are a poor model of this as the point would exist in a high frequency space, whereas only three orders of the basis are used. The spherical harmonics illumination model also assumes that the surface is Lambertian diffuse, human skin has a noticeable specular highlight. These effects can be modelled using a standard point light model. The point illumination $\mathbf{l} = r_l, g_l, b_l$ is defined as:

$$\mathbf{l} = \mathbf{l}_{amb} + \mathbf{l}_{diff}\hat{\mathbf{p}}_l.\hat{\mathbf{n}} + \mathbf{l}_{spec}(\hat{\mathbf{r}}.\hat{\mathbf{v}})^s \tag{4.36}$$

where $\mathbf{l}_{amb} = r_{amb}, g_{amb}, b_{amb}$ estimated ambient light in the scene, $\mathbf{l}_{diff} = r_{diff}, g_{diff}, b_{diff}$ and $\mathbf{l}_{spec} = r_{spec}, g_{spec}, b_{spec}$ are the estimated ambient and specular components of the surface. The vector $\hat{\mathbf{p}}_l$ is the normalised direction to the light-source from the surface point, $\hat{\mathbf{v}}$ the normalised direction from the surface point to the view point and $\hat{\mathbf{r}}$ is direction to the light source reflected by the surface, that is $\hat{\mathbf{p}}_l$ reflected by the surface normal $\hat{\mathbf{n}}$. $\hat{\mathbf{r}} = \hat{\mathbf{p}}_l - 2\hat{\mathbf{n}}(\hat{\mathbf{n}}.\hat{\mathbf{p}}_l)$.

Colour Transformation

In order to allow for a wide range of tones in our source images, we apply a colour transform [16] to the rendered image. The colour of output pixel, post lighting, is transformed using a scaling metric or gain in red, green and blue ($\mathbf{g} = \{g_r, g_g, g_b\}$) and an offset translation in colour space $\mathbf{o} = \{o_r, o_g, o_b\}$ together with a scalar contrast c applied to each colour channel. The final output colour $\mathbf{c}_{out}$ is computed as:

$$\mathbf{c}_{out} = M.\mathbf{c}_{in} + \mathbf{o} \tag{4.37}$$

where

$$M = \begin{bmatrix} g_r & 0 & 0 \\ 0 & g_b & 0 \\ 0 & 0 & g_g \end{bmatrix} . \left(I + (1-c) \begin{bmatrix} 0.3 & 0.59 & 0.11 \\ 0.3 & 0.59 & 0.11 \\ 0.3 & 0.59 & 0.11 \end{bmatrix} \right) \tag{4.38}$$

The colour transform parameters $g_r, g_g, g_b, o_r, o_g, o_b \& c$ are denoted ρ.

4.3.2 Calculating lighting parameters

The colour transform parameters ρ, the indirect lighting parameters τ and the colour parameters from the Morphable Model β are all orthogonal and linear, both within each set of parameters and between the sets. The point light model however is non-linear due to the movement of the point light source. The parameters of a combined Direct and Indirect illumination model can be calculated in an iterative fashion.

4.4 Colour reconstruction

In theory, the colour of the fitted face could be reconstructed from the colour parameters of the morphable model. This approach has two problems. Firstly this throws away a lot of textural information that can be easily extracted from the image. The second problem is more subtle, the fitting methods assume that the lighting model is independent of the colours of the morphable model. As the lighting conditions during morphable model capture were not totally consistent the morphable model contains lighting information, and is therefore not independent of the lighting model used for fitting.

In order to reconstruct the colours of the face images for ageing analysis, we used a grab-back and lighting-removal technique. The shape of the morphable model is reconstructed with the shape parameters, α, generated from the fitting algorithm. Each vertex v of the morphable model is then projected onto the 2D image plane using the rendering equation 4.28 and the view parameters ρ. The colour values $\{r, g, b\}$ are sampled from the image at this point. This is the grabbed-back part of the technique. The full texture map of the face can be reconstructed by setting the colour of vertices of the face model to the sampled colour and linearly interpolating between vertices. Some of the colours sampled by this method will be inappropriate due to occlusion, where the part of the face being sampled is occluded by another part of the face. In this case the sample will be from the visible part of the face. The case where the face is occluded by an object that is not part of the face is not considered as this does not happen in our dataset. If removal of samples from part of the face occluded by non-face objects, or hair is necessary, they can be removed by statistical methods of outlier removal see 4.2.4. Occluded vertices are discovered by comparison with a depth-map as described above in subsection 4.3.1, and their associated samples from the source image are discarded and set to the colour of the associated image from the fitted Morphable Model. In order to avoid sharp edges between samples from the source image and samples from the Morphable Models reconstructed texture-map, the two are blended using a linear weighting based on the inner-product of the surface normal of the reconstructed face shape and the direction to the view point. Most occluded vertices will be on the far side of the face from the viewer and so their inner product with the view angle will be negative, this creates a smooth blend starting from the profile edge of the face.

If we define $g(\mathbf{v}) = \mathcal{I}(R^{(2)}.\mathbf{v})$ as the sample from the source image at the point where $\mathbf{v}$ is projected onto the image plane. Let $h(\mathbf{v}; \beta) = \hat{\mathbf{c}}_{\mathbf{v}} + \sum_j^m \beta_j \mathbf{c}_{j,\mathbf{v}}$ be the colour of the vertex

$\mathbf{v}$ of the reconstructed face model. The colour of the grabbed back face is computed as:

$$i(\mathbf{v};\beta) = g(\mathbf{v}).(\hat{\mathbf{n}}.\mathbf{d})^{+} + h(\mathbf{v};\beta).(1 - (\hat{\mathbf{n}}.\mathbf{d})^{+}) \tag{4.39}$$

where $\hat{\mathbf{n}}$ is the normal to the surface at the vertex $\mathbf{v}$ and $\mathbf{d}$ is the direction to the view point from vertex $\mathbf{v}$. The plus symbol indicates that only positive values of $\hat{\mathbf{n}}.\mathbf{d}$ are taken i.e. $x^{+} = x$ iff $x > 0$, $x^{+} = 0$ otherwise.

4.4.1 Removing Lighting

The texture-maps grabbed back from the face contain significant amounts of lighting information specific to the environment the image was captured in. In order for an ageing model to model colour changes over time it's necessary to remove this lighting information.

The fitting process also involved an estimate of the lighting conditions, we can use this estimate to remove the lighting from the texture using the lighting equation.

$$\mathbf{I} = \mathbf{tl} \quad \text{where } \mathbf{I},\mathbf{t},\mathbf{l} \in \{r, g, b\} \tag{4.40}$$

where $\mathbf{I}$ is the resulting sample colour, $\mathbf{t}$ is the colour of the surface and $\mathbf{l}$ is the computed lighting intensity at the sample point given a set of view parameters (ρ) and a set of colour parameters (β). In this case $\mathbf{t}$, the original surface colour is the unknown and is found by division of the grabbed-back colour $\mathbf{I}$.

4.5 Implementation

We implemented a face-fitting algorithm based on an adaptation of the KLT algorithm to three-dimensional Morphable Models, combined with a point alignment method to bring delineated points into correspondence. As the purpose of the fitting is to produce models for training of a statistical ageing model, efficiency is of secondary importance to accuracy.

We start by describing the Inverse Compositional Kanade Lucas Taylor algorithm adapted for 3DMMs [48] (see section 4.1.3). The flexibility of the method is that assuming the parameters are independent we can apply this method directly to rendered images of the Morphable Model, provided that we are able to differentiate the model with respect to each

parameter of the model. We define $\mathcal{M}(;\mathbf{p})$ to be the image resulting from rendering a 3DMM with parameters $\mathbf{p}$. $\mathcal{M}(\mathbf{x};\mathbf{p})$ to be the intensity of $\mathcal{M}$ sampled at point $\mathbf{x} \in \Re^2$ and $\mathcal{I}(\mathbf{x})$ to be the intensity of the image $\mathcal{I}$ sampled at point $\mathbf{x}$. We redefine $W(\mathbf{x};\mathbf{p})$ to be the mapping of a point $\mathbf{x}$ on the two-dimensional surface to another point of the two-dimensional surface caused by the shape change, translation rotation and scaling, from their initial positions, of a Morphable Model with parameters $\mathbf{p}$.

The error metric is given by,

$$\chi^2 = \sum_{\mathbf{x}\subset\Omega} [\mathcal{M}(W(\mathbf{x};\mathbf{p})) - \mathcal{I}(W(\mathbf{x};\mathbf{p}))]^2 \tag{4.41}$$

All parameters necessary for rendering the 3DMM; position, rotation, scale, perspective, shape and colour, are concatenated in $\mathbf{p}$. Ω is the subset of all samples in the image. The equation is solved in a iterative manner. Baker et al. [9] made the same assumption as is used in the AAM model algorithm of Cootes at al. [19], that there exists a linear relationship between the target image and the rendered image. This linear relationship can then be described as a weighted sum of the derivatives of the Morphable Model.

$$\mathcal{M}(\mathbf{x};\mathbf{p}+\delta\mathbf{p}) \approx \mathcal{M}(\mathbf{x};\mathbf{p}) + \delta\mathbf{p}.\nabla M(\mathbf{x};\mathbf{p}) \tag{4.42}$$

where

$$\nabla\mathcal{M}(\mathbf{p}) = \left(\frac{\partial\mathcal{M}}{\partial p_0}, \frac{\partial\mathcal{M}}{\partial p_1}, \ldots, \frac{\partial\mathcal{M}}{\partial p_N}\right) \tag{4.43}$$

is the gradient of the model with respect to each of the model parameters. Taking a first order Taylor series expansion on (4.41) and differentiating w.r.t. to each of the parameters yields;

$$\delta\mathbf{p} = (\nabla^2\mathcal{M}(;0))^{-1}(\nabla\mathcal{M}(;\mathbf{p}).(\mathcal{I} - \mathcal{M}(;\mathbf{p})) \tag{4.44}$$

where

$$\nabla^2\mathcal{M}(\mathbf{p}) = \left\{\frac{\partial\mathcal{M}}{\partial p_i}.\frac{\partial\mathcal{M}}{\partial p_j}\right\} \tag{4.45}$$

is an approximation of the *Hessian* matrix. As the Hessian matrix is evaluated at $\nabla^2 M(;0)$ it is constant over all iterations, and thus can be pre-computed rather than updated each frame. The warp is then updated using a compositional update step (equation (4.10)), however this is undefined for a three-dimensional mesh as the warp cannot be applied to points that are not on the surface of the mesh [9]. We found that a subtractive update step provided a sufficient approximation. Romdhani noted that (4.41) does not necessarily require that $\mathcal{I}$ be warped into $\mathcal{M}(W(\mathbf{x};\mathbf{p}))$ but can be warped into another frame

of reference. He also performed the inversion of the 3DMM warp by finding the triangle that would contain the point in the forward warp and recovering its relative position in the triangle on the reference frame.

In order to define the warp $W(\mathbf{x};\mathbf{p})$ in three-dimensions we have to define if for all points in three-dimensional space, however the warp function on the 3DMM is only defined on the surface of the mesh so W cannot be directly calculated.We attempt to approximate a pseudo warp using a sampling technique. Here $W(\mathbf{x};\mathbf{p})$ defines a set of points on the two-dimensional image produced when a set of sample points on the morphable model surface are projected onto the surface using the current view parameters. Thus changes to position, pose, scale and shape resulting in an updated set of sampling points, in effect warping the image into the same frame of reference as $\mathcal{M}(;0)$.

In this implementation we have deviated slightly from the algorithm as outlined by Baker and Matthews [8], who applied the algorithm to two-dimensional warps, in particular affine transforms rather than to three-dimensional Morphable models. The difference is in the calculation of the derivative images, or steepest descent images $\frac{\partial \mathcal{M}}{\partial p_i}$. A simple per sample linear approximation can be used:

$$\frac{\partial \mathcal{M}(\mathbf{x})}{\partial p_i} = \frac{\mathcal{M}(\mathbf{x};\mathbf{p}+\delta_i\mathbf{p}) - \mathcal{M}(\mathbf{x};\mathbf{p}-\delta_i\mathbf{p})}{2}, \forall \mathbf{x} \in \Omega \tag{4.46}$$

where $\delta_i\mathbf{p}$ is a small change in the value of parameter i and zero elsewhere. They used the chain-rule to combine the image derivative $\nabla\mathcal{I}$ with the parameter update warp (4.6) producing *steepest-descent parameter update* images,

$$\frac{\partial M}{\partial p_i} = \nabla I \frac{\partial M_i}{\partial p_i} \tag{4.47}$$

Rotations and shape changes can cause parts of the model to become occluded or unoccluded over the course of several iteration. We used a depth-map generated each iteration to determine which sample points were occluded and set the value of occluded samples in to be zero. This resulted in their influence in the update being reduced, however it added further inaccuracies to the computation. Finally as observed by Gross et al. [64] the Hessian is only constant if the colour of the face does not change, so colour and lighting updates require recomputation of the Hessian. Because of the cummulative effect of these approximations we found it necessary to recompute the Hessian frequently, generally recomputing it should the shape changes fail to find a suitable update for the parameters.

Figure 4.1 details our implementation of the Inverse *Subtractive* KLT algorithm, using the

Levenburg-Marquardt algorithm [63]. A graphical overview of the core component of the algorithm, the parameter update equation can be seen in figure 4.5.

Algorithm 4.1 Inverse Subtractive KLT algorithm

Require: V_0 is the current view/scale/perspective matrix. Create from the view scale and perspective parameters

Require: $\mathbf{v}$ a set of sample points on the surface of the Morphable Model

Transform sample positions into screen-space $\mathbf{u}_0 = V_{0,p}.\mathbf{v}$

Build delta images $\Delta_j \leftarrow \frac{M(;0+\delta_j)-M(;0-\delta_j)}{||\delta||}$

Build Hessian matrix H where $H_{n,m} = \delta_n.\delta_m$.

repeat

Render current model with current parameters. $\mathcal{R}_i \leftarrow V_i.M(;\mathbf{p}_i)$

Sample from input image. $\mathbf{t}_i = \text{sample}(\mathbf{u}_i, \mathcal{I})$

Sample from rendered image $\mathbf{r}_i = \text{sample}(\mathbf{u}_i, \mathcal{R}_\rangle)$

$\mathbf{d}_i = \mathbf{t}_i - \mathbf{r}_i$

Apply depth mask. Set obscured samples to zero.

$\mathbf{b}_j = \Delta_j.\mathbf{d}_{i,j} + \frac{p_{i,j}}{\sigma_j}$

Solve $H\mathbf{x}_i = \mathbf{b}_i$

if $\mathbf{d}_{i+1}^2 < \mathbf{d}_i^2$ **then**

Update $p_{i+1} \leftarrow \mathbf{p}_i + \mathbf{x}_i$

else

$\lambda_{i+1} = 10\lambda_i$

end if

update $V_i + 1$ from $p_i + 1$

until $\mathbf{d}_{i+1}^2 > \mathbf{d}_i^2$ && $\lambda_{i+1} > 1000.0$

We found that in practice the algorithm performs better if parameters that affect movement like shape, position and angle are separated from parameters that are entirely linear in pixel intensity, i.e. lighting and colour. Thus making the algorithm effectively multi-pass. Typically one iteration of colour was performed before multiple iterations of shape.

4.5.1 Point Alignemt

Due to the lack of accuracy in fitting the Morphable Model using the previous methods, we added a set of delineated points to guide the fitting in the manner of [16]. We define $\mathbf{q}_i \in$

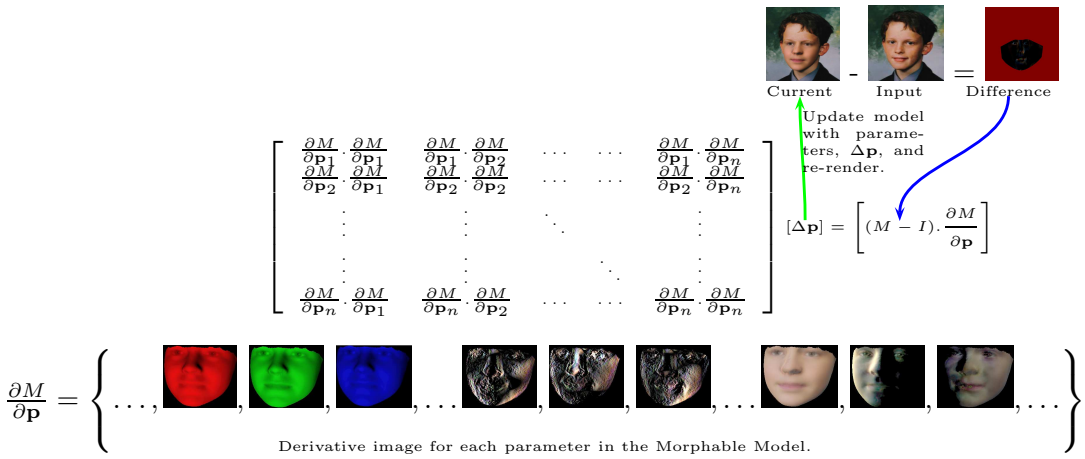


Figure 4.1: Calculating the parameter update for iterative face-fitting. The figures shows graphically the core update step from the face fitting algorithm using equation 4.44. The 'Input' image is the image containing the target face image. The 'Current' image shows the result of superimposing the rendered face model with the current parameter on the 'Input' image. This is an implementation detail to ensure that background detail is ignored as it will be zeroed when the 'Input' is subtracted from the 'Current' image. The 'Difference' image is the result of subtracting the 'Input' image from the 'Current' image and sub-sampling in the area of the rendered face model. Each iteration the 'Current' image is updated using the parameter updates, $\Delta\mathbf{p}$, and re-rendering. The bottom row of images show some of the derivative images, $\frac{\partial M}{\partial \mathbf{p}}$, resulting from making a small change to their corresponding parameter (4.46). The first three show examples of changing the diffuse lighting parameters. The second group of three show examples of the changes resulting from modifying the parameters corresponding to the first three shape eigenvectors. The last three show examples of the changes resulting from modifying the parameters corresponding to the first three colour eigenvectors.

$\Re^2$ as the i^{th} point on the two-dimensional input image and $\mathbf{r}_i \in \Re^3$ as the corresponding point on the surface of the Morphable Model, we further define $v_j \in \mathcal{V}$ as the vertex corresponding to $\mathbf{r}_i$. A new 3DMM can be generated as a subset of the original 3DMM (3.8) as:

$$\mathbf{s}' = \bar{\mathbf{s}}' + \sum_{i=1}^{l} \mathbf{p}_i \mathbf{s}'^i \tag{4.48}$$

Similar to before $M'(;\mathbf{p})$ describes the Morphable Model subset built with equation (4.48) using the parameters $\mathbf{p}$. In order to fit the 3DMM subset to the points on the image we minimize the l^2-norm between the two-dimensional points $\mathbf{q}$ and the projection onto the image plane of the three-dimensional points $\mathbf{r}$ using the same projection as in the previous algorithms, (4.28). Such that $V.M'(;)$ is the projection onto the plane, where V is the $4x4$ projection matrix. Only the first two coordinates of the vector $V.M'(;)$ are required for projection to a plane. We define $M(\mathbf{r};\mathbf{p}) \in \Re^2$ as the projection of the point (r) onto the image plane. We define the l^2-norm as:

$$\chi^2 = \sum_i (M(\mathbf{r};\mathbf{p}) - \mathbf{q})^2. \tag{4.49}$$

We apply an additive update and minimize it by expansion with a first order Taylor series approximation.

$$\sum_i \left(M(\mathbf{r};\mathbf{p}) + \frac{\partial M(\mathbf{r};\mathbf{p})}{\partial \mathbf{p}} - \mathbf{q} \right)^2 \tag{4.50}$$

As before this can be solved by minimizing the partial derivatives of $\frac{\partial \chi^2}{\partial \mathbf{p}_i}$ using an iterative method based on additive updates of the parameters $\mathbf{p}$. The updates $\Delta\mathbf{p}$ are calculated as,

$$\begin{bmatrix} \frac{\partial M}{\partial \mathbf{p}_1}.\frac{\partial M}{\partial \mathbf{p}_1} & \frac{\partial M}{\partial \mathbf{p}_1}.\frac{\partial M}{\partial \mathbf{p}_2} & \cdots & \frac{\partial M}{\partial \mathbf{p}_1}.\frac{\partial M}{\partial \mathbf{p}_n} \\ \frac{\partial M}{\partial \mathbf{p}_2}.\frac{\partial M}{\partial \mathbf{p}_1} & \frac{\partial M}{\partial \mathbf{p}_2}.\frac{\partial M}{\partial \mathbf{p}_2} & \cdots & \frac{\partial M}{\partial \mathbf{p}_2}.\frac{\partial M}{\partial \mathbf{p}_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial M}{\partial \mathbf{p}_n}.\frac{\partial M}{\partial \mathbf{p}_1} & \frac{\partial M}{\partial \mathbf{p}_n}.\frac{\partial M}{\partial \mathbf{p}_2} & \cdots & \frac{\partial M}{\partial \mathbf{p}_n}.\frac{\partial M}{\partial \mathbf{p}_n} \end{bmatrix} [\Delta\mathbf{p}] = \left[M(\mathbf{r};\mathbf{p}) - \mathbf{q}.\frac{\partial M}{\partial \mathbf{p}} \right] \tag{4.51}$$

The similarity of the forms for this equation with the update step of the K.L.T. algorithm allows them to be combined.

$$(\alpha H_k + \beta H_p)\Delta\mathbf{p} = \alpha \left[\mathcal{M}(x;\mathbf{p}) - \mathcal{I}(x) \right] + \beta \left[M(\mathbf{r};\mathbf{p}) - \mathbf{q}.\frac{\partial M}{\partial \mathbf{p}} \right] \tag{4.52}$$

where H_k is the Hessian matrix from the Inverse K.L.T. algorithm equation (4.44), H_p is the Hessian matrix of the point fitting equation (4.51). α and β are weighting factors

to skew fitting towards one or other fitting type. In our case the weightings reflected the differing number of samples in the two fitting methods, the Inverse K.L.T. algorithm used in the order of ten thousand samples, whereas the point-fitter used 165.

Finally to guide the fitting towards statistically likely face shapes we applied a Gaussian prior to the update step using the probability function from the PCA model used to build the Morphable Model (see equation (3.9)).

$$\left(H + \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sigma_n^2} \end{bmatrix} \right) .[\Delta \mathbf{p}] = \mathbf{b} + \begin{bmatrix} \frac{\mathbf{p}_1 - \mu_1}{\sigma_1} \\ \frac{\mathbf{p}_2 - \mu_2}{\sigma_2} \\ \vdots \\ \frac{\mathbf{p}_n - \mu_n}{\sigma_n} \end{bmatrix} \tag{4.53}$$

where μ_i and σ_i are the i^{th} mean and standard deviation of the i^{th} component on the Morphable Model, or an appropriate choice of constants for the view and lighting components. $\mathbf{b} = \alpha \left[\mathcal{M}(x; \mathbf{p}) - \mathcal{I}(x) \right] + \beta \left[M(\mathbf{r}; \mathbf{p}) - \mathbf{q}.\frac{\partial M}{\partial \mathbf{p}} \right]$ from equation (4.52). Such priors have be used by previous authors to guide fitting [16, 10].

Results of fitting using the combined KLT based pixel error method and the point fitting method can be see in figure 4.2.

GPU Implementation

Due to the simple, repetitive nature of parts of the algorithm is ideally suited to *Single Instruction Multiple Data* (SIMD) parallel computation, such as that provided by a sufficiently powerful (shader version 3.0) Graphics Processing Unit (GPU) found in many modern PCs. Each pixel in a texture can be thought of as a cell in a two dimensional grid containing four floating point values. The pixel shader unit (PSU) executes simple shader programs (also known as fragment programs in OpenGL) outputting the result to a single pixel in the grid. The PSU can output to multiple grids but only one cell per grid. Input to the PSU can be supplied by setting 'shader constants' i.e input variables that are the same each time the shader program is called, and from reading data from the cells in the form of a texture lookup. Data can also be supplied from the vertex shader but this is limited in scope. There is no ability to store information between calls to the shader program.

This architecture allows programs to be run that are unrelated to the original purpose of

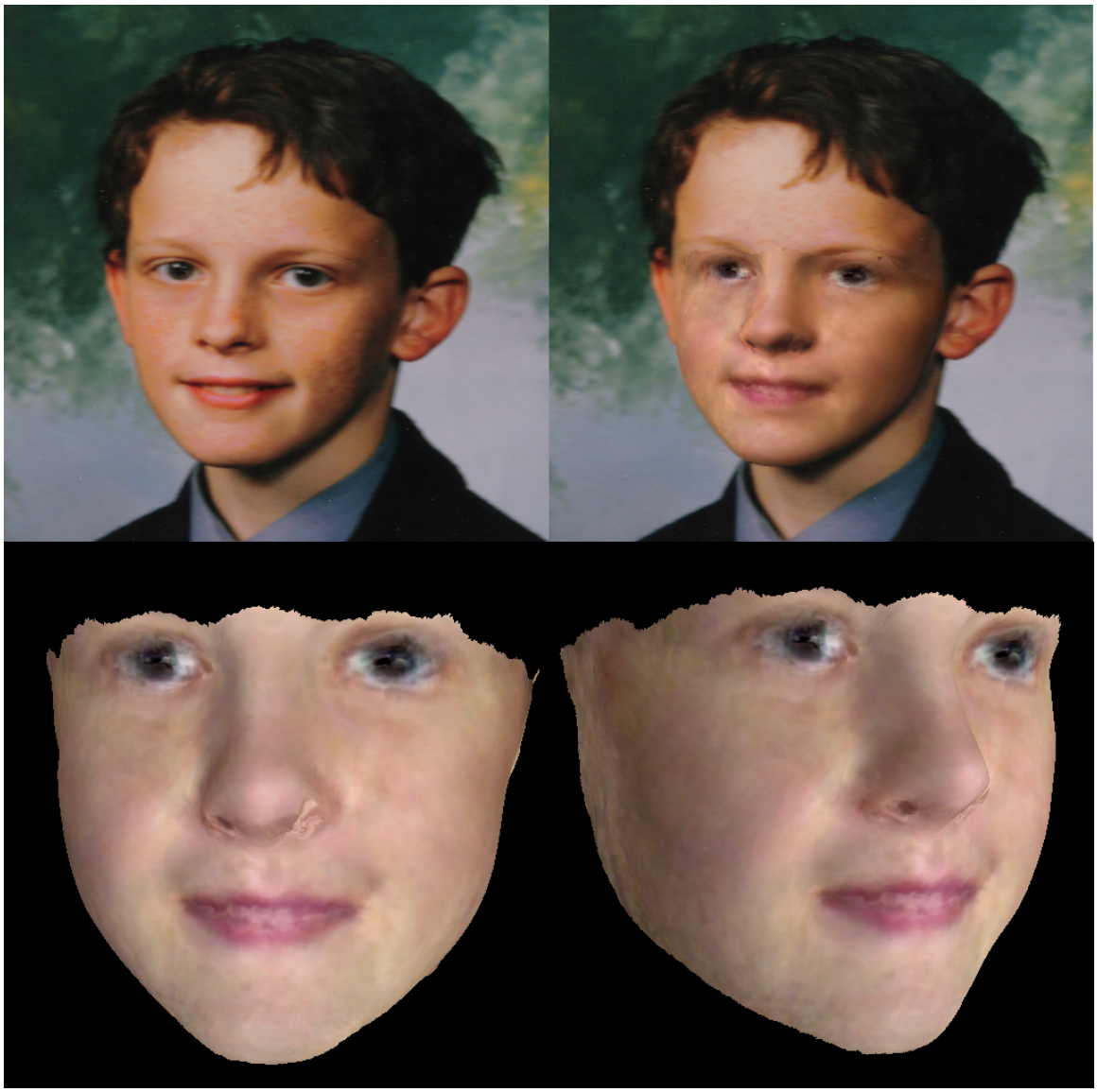


Figure 4.2: An example of a Three-dimensional Morphable Model fitted to a face image. The image on the top left is the original photograph. The image on the top right shows the results of the fitting, rendered in-situ. The images on the bottom row are of the same model rendered under neutral lighting conditions from different angles.

rendering geometry. Several basic algorithm skeletons have been adapted for use on the GPU. The *map* skeleton is implemented simply by calling a shader program once for each cell in a grid. The *scan* and *reduce* skeletons are more complicated as the shader programs cannot store state between calls. The skeletons are implemented using an algorithm known as 'ping-ponging' which produces the results in $log_2(O(n))$ time. A recent algorithm to perform the scan operation in $O(n)$ was presented by Sengupta et al. [75]. A more detailed overview can be found in [57].

4.5.2 The software

The face-fitting software was the single largest part of the project, taking two and a half years to complete and containing 22000 lines of code. It was written in C++ using the Microsoft DirectX Application Programming Interface (API).

A key component of the software was the ability to carry out many of the repetitive and memory intensive operations on the Graphics Processing Unit of a standard PC graphics card. As an example the single most common and computationally expensive part the of the face fitting algorithms are the dot products of two large sets of image samples. These dot-produces are required to build the Hessian matrix and to project the difference images onto the Jacobian (see section 4.5). Dot-products are simply a sequence of additions and multiplications over a large memory space. Attempts to implement the same algorithms on the CPU proved to be too slow, especially with a database of over 200 images to fit faces to. GPU computation however was observed to offer an improvement of as much as a thousand fold over a na ive CPU implementation. At the time of development APIs for performing computation on a Graphics Processing Unit were in their infancy and were often incompatible with our hardware, so it was decided to implement our own version in DirectX. Our GPU API consisted of a series of 'wrapper' functions that handled the process of the formating and sending the information to the graphics card, and and retrieving it after computation. It also contained functions to setup and call GPU functions on the data once it had been correctly formated and transferred. All the functions used with the GPU code were written in HLSL and complied using the DirectX API.

The face fitting code consisted of two components; one to build and render a Morphable Model, the other to fit it to a given image. As the Morphable Model has to be rebuilt at least once per iteration of the face-fitting algorithm, it was also implemented on the GPU to improve efficiency. The majority of the development time was spent writing code to perform various methods of fitting and to tune the fitting parameters associated with the methods in order to get the best performance out of each method attempted. The order of construction and refinement was roughly the order in which the algorithms are outlined in the review section of this chapter. We developed three-dimensional implementations of the Kanade Lucas Taylor, algorithm (outline in section 4.1.3), an approximation to the inverse KLT method described by Matthews and Baker, also converted to three-dimensions. Finally we refined these methods by guiding the fitting using a large hand delineated point template

in a similar manner to other researchers, such Blanz and Vetter [16], see section 4.5.1. Each stage was implemented in order to achieve either greater computational efficiency, or more importantly, greater accuracy in the fitted-face. Along the way we also investigated applying various filters to the images to extract details such as edges or to eliminate outliers. However we found that when using edge filters face details in the Morpable Model were frequently matched to the wrong edges in the image, e.g. the chin the matched to the bottom lip. Outlier removal was found to make the Hessian progressively less accurate as each outlier removed meant that fewer samples were being used to compare with the Jacobian than were being used to construct the Hessian. If the Hessian was recomputed each iteration a small improvement in accuracy could be gained at the expense of speed.

4.6 Fitting Accuracy

The method of fitting a 3DMM to an image as outlined in this chapter is a fundamental part of the overall ageing methods implemented. It is important that the fitting algorithm is able to capture enough information about the target face for the face to still be recognisable to an observer. To test this we ran an experiment using human raters, presenting them with a series of 2AFC tests asking them to choose between the original face image and a distractor image of another individual of the same age group. The raters were presented with a stimulus image of a rendered face model, the result of fitting the three-dimensional morpable model to a face image. The original image along with a distractor was presented and the users asked to choose between them. The images were presented on a uniform black background with all the images cropped to remove hair (except facial hair), the image order was randomised. The experiment contained three tests with three different sets of stimuli, 1) The fitted face models of the mid-child age group in the same pose and lighting conditions as the original image, 2) The fitted face models of the mid-child age group in standardised pose and lighting, 3) The fitted models of the student age group in standardised pose and lighting.

Three images were presented, one in the top row, two in the bottom. The top row image was a stimulus image from one of the three tests. The bottom row contained the original image and a distractor in random order. 11 subjects took part in the trails, they where presented with three stimuli for each of 33 subjects. Figure 4.3 shows an example of the test displayed to each rater.

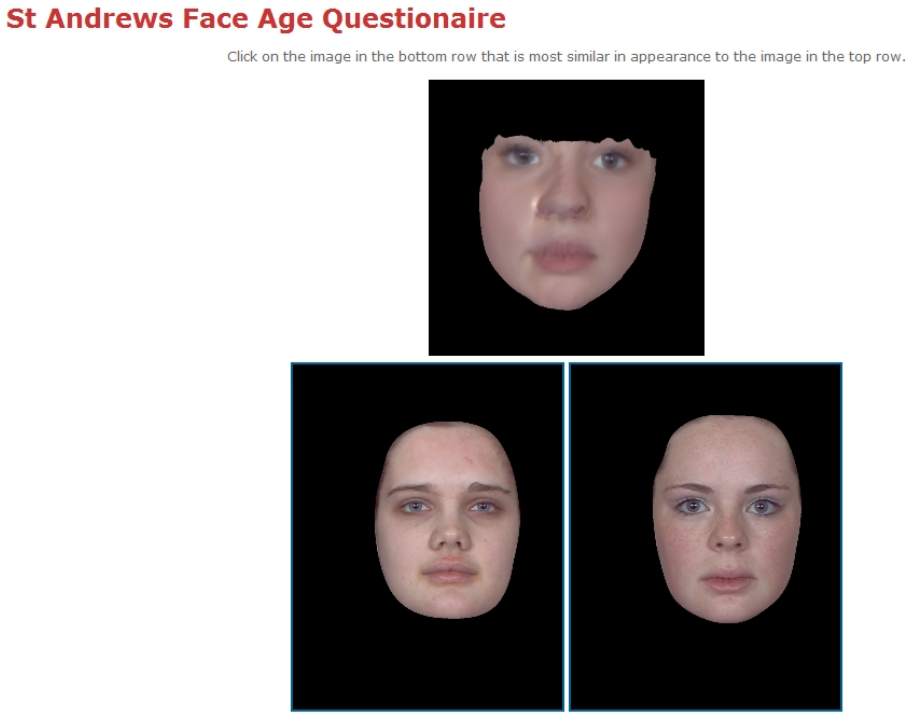


Figure 4.3: An example stimulus shown to a user in the web based experiment. The top image shows a rendered face model. The bottom two images show clipped images of faces from the training set. The user is ask to select which of the two bottom images most closely resembles the rendered face image. The one on the left was used to align the face model shown in the top image.

The results of the experiments were analysed in the same manner as the Identity Retention experiment (section 6.2.1), with the proportion of correct responses and the d' for each test appearing in table 4.1. From this table we can see that raters were mostly able to identify the correct individual in cases of matching pose and lighting, with more than one standard deviation separating the means of True Positives ($p(H)$) and False Negatives ($p(F)$). Of course in these cases the pose and lighting itself could be the cue that raters were using to correctly identify the individual, although they were given no indication that there was a relationship between pose or lighting and the subject, also the tests were mixed together so raters would have seen plenty of tests where the pose and lighting did not match either response image. The accuracy decreases for the second test with standardised pose and lighting, so it safe to conclude that raters were making some use of these variations in making their choice. The proportion correct is still above 0.5, however the difference of means (d') is less than one standard deviation. The raters were able to discriminate between the correct individual and the distractor using the fitted face model. Raters were less successful at the third test with the student images, with the Proportion Correct near to 0.5 and d' of 0.21 standard deviations. This is likely due to the greater uniformity between images in the student age group as they were taken under standardised conditions. The child group displays significantly greater variation across the set including expression variation that is not considered in the experiment and lighting information that is not completely removed. All three have a $\chi^2 > 6.6349$ and therefore show a statistically significant difference from random chance.

It should be noted that the task given to the raters was hard one. With the image area croped to show only the face, many cues potentially used to identify individuals, such as hairstyles, clothes, and basis physical attributes such has height, were missing. In this context the relatively low recognition scores are to be expected.

Table 4.1: The Proportion Correct and d' for identification of fitted face models. 1) Mid-Child matching pose and lighting, 2) Mid-child standardised pose and lighting, 3) Student standardised pose and lighting.

Test	Proportion Correct	d'	chi^2
1	0.869	1.588	612.02
2	0.729	0.861	234.86
3	0.559	0.209	39.439

4.7 Summary

In this chapter we have described current research into fitting a three-dimensional Morphable Model to a two-dimensional image together with relevant research into fitting two-dimensional face models to an image. Although much of the previous research focused on minimising differences in intensity between the model and the target image, we found, from our own experience, that the fitting is more accurate and generally more likely to resemble the target face if points are manually delineated on the image. The process by which we extracted three-dimensional face models from our image is described in detail along with a detailed description of the rendering and lighting models along with a process to eliminate the effects of lighting in a photograph. The results were evaluated using human raters to determine how well the fitted models retained the identity of the person in the image a key requirement of any ageing method.

- Human raters were able to correctly identify the individual depicted by a rendered face model when presented with the original two-dimensional image of that person with a reasonable degree of accuracy.
- This ability was reduced but not lost when the pose and lighting of the model did not match that of the face in the image.

In this chapter we have described how to extract a face-model from a two-dimensional image with a reasonable degree of accuracy and efficiency. The models produced by the face-fitting method are in the form of parameterised Morphable Models. In the next chapter we will describe how ageing functions can be derived from a set of these face-model parameters.

Chapter 5

Synthesising Facial Ageing

So far in this thesis we have described the construction of three-Dimensional Morphable Models (see chapter /refch:3DMM)from a set of scanned models of human faces. We have described how the 3DMM is used to fit a rendered Morphable Model to an image of a human face in order to extract three dimensional information about the face (chapter /refch:Fitting). These extracted three-dimensional face models form the training set for the ageing algorithms used in this thesis. In chapter /refch:LiteratureReview we provided an overview of the state-of-the-art, at the time of writing, in visual facial ageing synthesis and well as a chronology of developments in this field. In this chapter we describe in detail two well known statistical ageing methods; ageing using prototypes [18] and ageing using individualised linear functions in PCA space [43]. We also introduce a new method based on Projection to Latent Structures (PLS). These ageing methods will then be evaluated in detail in the next chapter.

5.1 Ageing using Three Dimensional Morphable Models

In order to generate ageing functions, we take a set of models of a individuals at various points in their lives and fit an 'ageing path' through the face space. We have a dataset of photographs of individuals at various age points between infancy and early adulthood donated by student volunteers. These images contain significant variations in pose, expression and lighting conditions. In order to study the shape changes associated with ageing

in these age groups we extracted the shape information using the Point and Sample fitting method described in section 4.5.1. Colour information for each face was grabbed-back from the image, colour and intensity due to lighting removed and principal Components Analysis preformed to create a new set of colour parameters as described in section 4.4.

The resulting faces are therefore described as a set of combined shape and colour parameters $\mathbf{p}$, and can be reconstructed as a weighted sum of linear shape and colour bases. The colour basis used are a reduced set of basis vectors from the PCA of the set of estimated facial texture maps after lighting is removed, see section 4.4. The shape basis vectors are the same basis vectors used to fit the Morphable Models to the images. The faces are reconstructed as,

$$f(\mathbf{p} = \{\alpha, \beta\}) = \begin{cases} \hat{\mathbf{s}} + \sum_i^n \alpha_i \mathbf{s}_i & \text{where } \alpha_i,\ i = 1 \ldots n \text{ are the shape parameters} \\ \hat{\mathbf{c}} + \sum_j^m \beta_j \mathbf{c}_j & \text{where } \beta_i,\ i = 1 \ldots m \text{ are the colour parameters} \end{cases} \tag{5.1}$$

A new face can be constructed by varying the parameters $\mathbf{p}$, these parameters approximately span the space of human faces. A path through this space can be defined and if this path describes the shape and colour changes caused by ageing then it is known as an *ageing path*.

This chapter is devoted to methods for discovering appropriate ageing paths and to verifying their effectiveness.

Stratification

The images in the ageing set had been taken at regular intervals spanning from infancy to early adult-hood. The subjects had been asked to supply images from a number of key ages, as a result the dataset lent itself to stratification. The set is divided into five strata according to the age of the individual when the image was taken (Table 5.1). The strata show varying numbers of subjects, this is because the dataset does not contain images for every subject in every age-range. Algorithms and methods that require that each individual has an image in each of the included age ranges exclude those individuals for whom data is incomplete. This results in a reduced dataset of 35 individuals.

Table 5.1: Ageing dataset stratification

Name	Age Range	Number of subjects	Mean Age	Standard deviation
Infant	2-5	51	3.12	0.82
Mid Child	5-8	50	6.54	0.85
Late Child	8-12	49	10.7	0.94
Teenager	12-17	47	14.24	1.18
Student	17-23	47	20.02	1.69

5.2 Ageing using Prototypes

Prototypes have been used by Burt and Perrett [18] to age two-dimensional face images. Using a set of delineated faces, a prototypical face for each age strata was produced by averaging the delineated points of all the faces within the stratum and applying a triangulated linear warp. The per-pixel colours of the warped faces were averaged for all the faces within the age-group. The differences in positions of the face points in the two prototypes are used to drive a triangulated linear warp. The per-pixel differences in colour between the two prototypes defines a colour transform. This warp and transform can then be applied to an input face to age it such that its apparent age is brought closer to the age of the target group. We used the three-dimensional analogue of this technique as the first method to age three-dimensional face models.

As we are using a 3DMM rather than a set of delineated points, we will be manipulating a vectorised set of shape and colour parameters $\mathbf{p}$. As these are derived from the shape and colour components of a PCA model the parameters are linearly-independent and form a basis in $\Re^n$ where n is the number of combined shape and colour components.

Average Prototypes

Prototype face models were created for each age-stratum by averaging the parameters over all faces in the stratum.

$$\hat{\mathbf{f}}_{\mathbf{s}} = \sum_{i}^{m} \mathbf{p}_i \tag{5.2}$$

where $\hat{\mathbf{f}}_\mathbf{s}$ are the parameters of the averaged face model of all the faces in the stratum s. Here $\mathbf{p}_i$ is the vector of parameters for the i^{th} face model in the stratum. m is the number of faces in the stratum.

A linear transform is created between two strata by creating a vector between the average of each stratum and dividing it by the age difference. Thus, to generate a transform from stratum j to stratum k we take,

$$\mathbf{t} = \frac{\hat{\mathbf{f}}_\mathbf{k} - \hat{\mathbf{f}}_\mathbf{j}}{\hat{a}_k - \hat{a}_j} \tag{5.3}$$

where $\hat{a}_j$ and $\hat{a}_k$ are the average ages of the individuals within strata j and k respectively. An input $\mathbf{f}_{input}$ in stratum j is aged towards the age group of stratum k by moving it in the direction of the vector t, multiplying t by the desired number of years. See figure 5.2.

$$\mathbf{f}' = \mathbf{f}_{in} + (a_t - a_s)\mathbf{t} \tag{5.4}$$

where $\mathbf{f}'$ is the set of model parameters at the target age, a_s and a_t are ages of the input face and the target age respectively. Clearly this transform is most valid if the target age is within the range of years of the target stratum k.

5.3 Individualized Linear Transform

It is well known that faces do not age in an identical manner [43], and therefore have their own unique or almost unique ageing paths that depends on numerous factors such as environment, lifestyle and individual genetic make-up. These ageing paths nevertheless exhibit similarities, implying an underlying relationship that can be exploited in generating ageing paths. One such relationship is between the appearance of the input face and the ageing path for that individuals. Lanitis et al. [43] found a correlation coefficient of 0.55 between the Mahalanobis distance between two face models and the Euclidean distance between the parameters of the ageing functions. This suggests that we might be able to use the initial appearance to guide a more personalised ageing transform.

In order to construct an individualized linear ageing transform for an unseen individual, we first identify the strata of the starting and ending age. For each individual in the dataset a linear ageing path is defined as a vector from one sample face in the starting stratum to another in the target stratum containing the end age. If no suitable pair of sample faces can

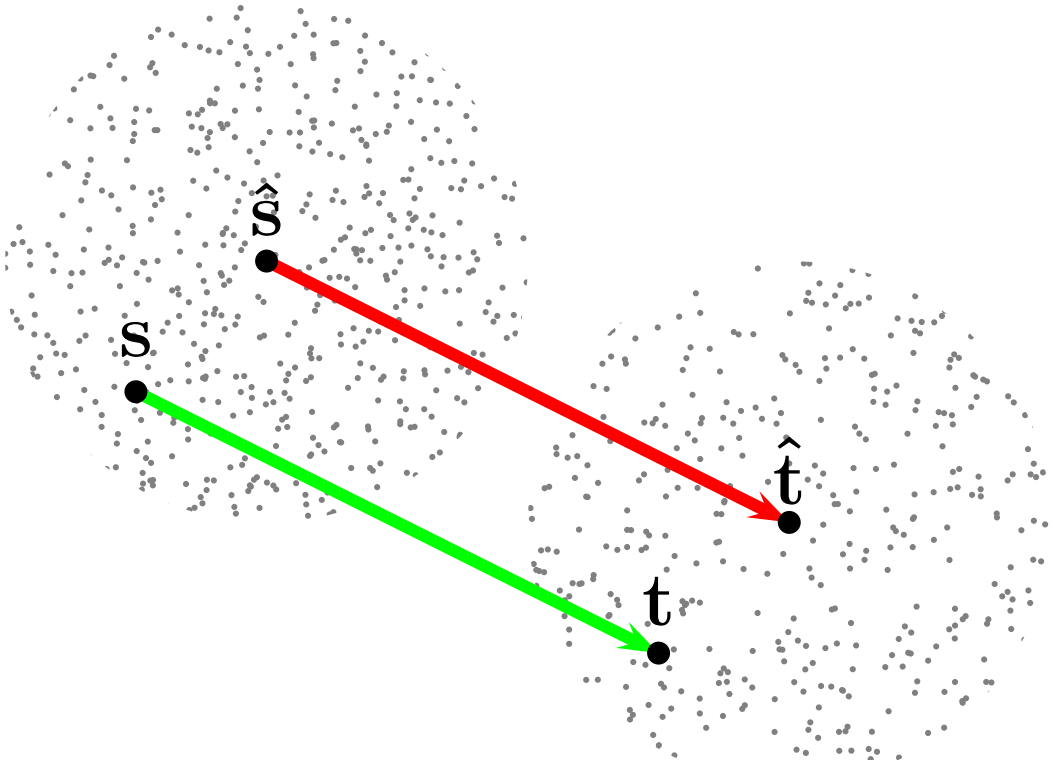


Figure 5.1: Ageing using Prototypes. $\hat{\mathbf{s}}$ and $\hat{\mathbf{t}}$ are the averages of the start and end strata respectively. The ageing trajectory is a vector from the point $\hat{\mathbf{s}}$ to the point $\hat{\mathbf{t}}$, a new face model, represented by $\mathbf{s}$, can be aged by moving it through the PCA space in the direction $\overrightarrow{\hat{\mathbf{s}}\hat{\mathbf{t}}}$ to produce the target face, $\mathbf{t}$.

be found the individual is excluded from the dataset. We denote s,e as the start and end ages of the transform respectively, and $\mathbf{p}_i$ and $\mathbf{q}_i$ as the parameters of the face models of the i^{th} individual taken from the starting and end strata respectively. We define a single linear ageing function such that the j^{th} parameter of the face model of the individual i at time t is,

$$\mathbf{f}(\mathbf{t})_j = t.\mathbf{a}_{i,j} + \mathbf{b}_{i,j} \tag{5.5}$$

where $\mathbf{a}$ and $\mathbf{b}$ are sets of weights and $\mathbf{a}_{i,j}$ and $\mathbf{b}_{i,j}$ are the j^{th} weights for the i^{th} individual in the training set. The vector $\mathbf{a}$ defines the gradient of the path in $\Re^n$ and the vector $\mathbf{b}$ defines the parameters of the face at time $t = 0$. These are defined as,

$$\mathbf{a} = \frac{\mathbf{q} - \mathbf{p}}{e - s}$$
$$\mathbf{b} = \mathbf{p} - s\mathbf{a}$$

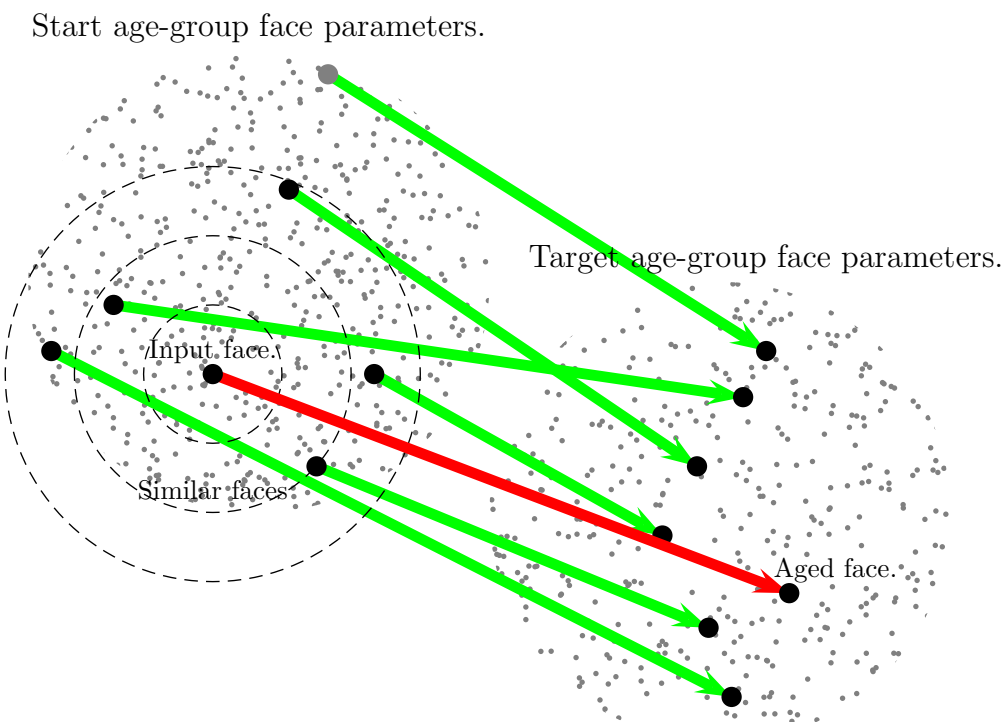


Figure 5.2: Ageing using an Individualized Linear Transform. Each face in the start age-group is matched to a face of the same individual in the target age-group, and an ageing vector calculated between them (green-lines). A new ageing vector for an unseen face (red) as calculated as a sum of the ageing vectors for individuals in the training set weighted by their similarity to the input face.

These functions can be parameterised using $\mathbf{a}_i$ and $\mathbf{b}_i$ to describe the ageing function f_i for the i^{th} individual. A new ageing path for an unseen individual can be created using a linear weighted sum of the parameters of the ageing functions for each individual in the training set.

$$\mathbf{f}' = \sum_{i}^{n} \rho_i \mathbf{f}_i, \qquad \sum_{i} \rho_i = 1 \tag{5.6}$$

where ρ_i are a set of weights relating the unseen individual to the ageing path of the i^{th} individual in the dataset. The ρ_i's sum to one, so that the function does not add a scaling factor to the ageing path.

As in [43] we use a form of regularised interpolation to combine ageing paths from a number of individuals, based on the similarity of the input face model and the model in the training set associated from the ageing path. The weighting ρ is defined using the probability distribution of the PCA space of the face model. Given two face models with parameters embedded in a Gaussian PCA space this function is,

$$p(\mathbf{p}_{in}, \mathbf{p}_i) = e^{-\sqrt{\sum_j^n \frac{(p_{in,j} - p_{i,j})^2}{2\sigma_j^2}}} \tag{5.7}$$

where $\mathbf{p}_{in}$ and $\mathbf{p}_i$ are the parameters of the input and i^{th} face model respectively. $p_{in,j}$ is the j^{th} parameter of the input face model. σ_j^2 is the variance of the PCA space in the j^{th} dimension. This function is closely related to the Mahalanobis distance.

To obtain the face model parameters $\mathbf{p}_i$ for equation (5.7) we use the parameters generated from the function (5.5) with t set to the age of the input individual. This is preferred to using a face model matching the individual in the source age group, for two reasons. Firstly there won't necessarily be a face model for a given individual in the dataset at the age of the input face model, in fact that will generally be the case. Secondly the individual's perceived age may be slightly different from their biological age.

5.4 Partial Least Squares Regression

The data-set of parameters contains a significant amount of information that is not relevant to ageing. Any statistical analysis needs to separate those factors related to ageing from those that are not related either explicitly or implicitly.

Partial Least Squares [90] also known as a Projection to Latent Structures is a statistical technique similar to Principal Components Analysis that describes mean centred data as weighted linear combination of basis vectors. Unlike PCA which finds directions of maximum variance in the input data, PLS attempts to describe a set of dependent variables from a set of predictors. It works by extracting a set of latent vectors that decompose both the dependent and independent matrices in such a manner as to explain as much of their covariance as possible. Using PLS we can produce a set of vectors that approximately spans the space of face models in such a way as to explain much of the variance in the face model set due to ageing.

The PLS decomposition is performed as follows. If we take the parameters of the face models in the data-set $\mathbf{f}_i$ and use them to build the matrix $X = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n]^T$ such that each row contains the parameters of an individual face model. We define $Y = [age_1, age_2, \ldots, age_n]^T$ where age_i the corresponding ages to the i^{th} face. The rows of both X and Y are then mean centred and scaled by the inverse standard deviation $\frac{1}{\sigma}$.

As described by Abdi in [2], we aim to decompose the independent variables, that is the face models, as $X = TP^T$ with $T^TT = I$. T is the *score* matrix and P is the loading matrix. We *estimate* the dependent variables Y, that is the ages in our case, as $\hat{Y} = TBC^T$. The diagonal matrix B hold the regression weights, and C is the weight matrix of the dependent variables. See Adbi [2] for further details on what these mean in practice. The columns of T are the latent vectors, they form an exact decomposition of the face models, X, but only an approximation to the ages, Y. The decomposition is found using an iterative algorithm where a latent vector is found that maximizes the covariance between X and Y and is then subtracted from both. The proportion of variance explained by this vector is found by dividing the sum of squares of the residuals by the the sum of squares of the input matrices X and Y. The algorithm is outlined in 5.1.

The decomposition created by PLS can be truncated in the same manner as PCA. We denote the dimensionally reduced loading matrix as $P^{(l)}$ where l is the number of reduced dimensions. A new set of PCA face model parameters can be constructed from the PLS face model as,

$$\mathbf{f}(\mathbf{q}) = \mathbf{f} + \sum_{i}^{l} \mathbf{q}_i \mathbf{P}^{(l)}_{.,i} \tag{5.8}$$

where $\mathbf{P}^{(l)}_{.,i}$ is the i^{th} column of the truncated l columned loading matrix and $\mathbf{q}$ the parameters of the PLS face. The new parameters $\mathbf{f}$ can be used to build a face model as before. l is chosen in such a way as to minimize the number of dimensions while maximizing the covariance explained.

Conceptually we wish to divide the face data into two sets, one describing the age related components of the data-set, and one describing non age-related factors, e.g. identity. The age related information is described using the PLS model with separate models for shape and colour. The PLS modelling does not fully span the space of the face model and is truncated to those components that best describe ageing. The residuals that result from computing the PLS components are considered to contain the non age-related information about the face. This way we can separate age and identity.

Algorithm 5.1 PLS regression algorithm

Require: Matrix X is the matrix of Z-score model parameters.

Require: Matrix Y is the matrix of Z-score ages.

SS_x,SS_y are the corresponding total sum of squares.

$i = 0$;

repeat

fill $\mathbf{u}$ with random values

repeat

$\mathbf{w} \propto X^T\mathbf{u}$. Estimate X weights.

$\mathbf{t} \propto X\mathbf{w}$. Estimate X factor scores.

$\mathbf{c} \propto F^T\mathbf{c}$. Estimate Y weights.

$\mathbf{u} = F\mathbf{c}$. Estimate Y scores.

Here $\propto$ means normalize the result.

until $\Delta\mathbf{u} < \epsilon$

$b = \mathbf{t}^T\mathbf{u}$.

$\mathbf{p} = X^T\mathbf{t}$.

$X = X - \mathbf{t}\mathbf{p}^T$.

$\frac{\mathbf{p}\mathbf{t}^T}{SS_x}$ =variance of X explained.

$Y = Y - b\mathbf{t}\mathbf{c}^T$.

$\frac{b\mathbf{t}\mathbf{c}^T}{SS_y}$ =variance of Y explained.

Fill the matrices with the resulting vectors.

$P(:, i) = \mathbf{p}$

until $b < \epsilon$

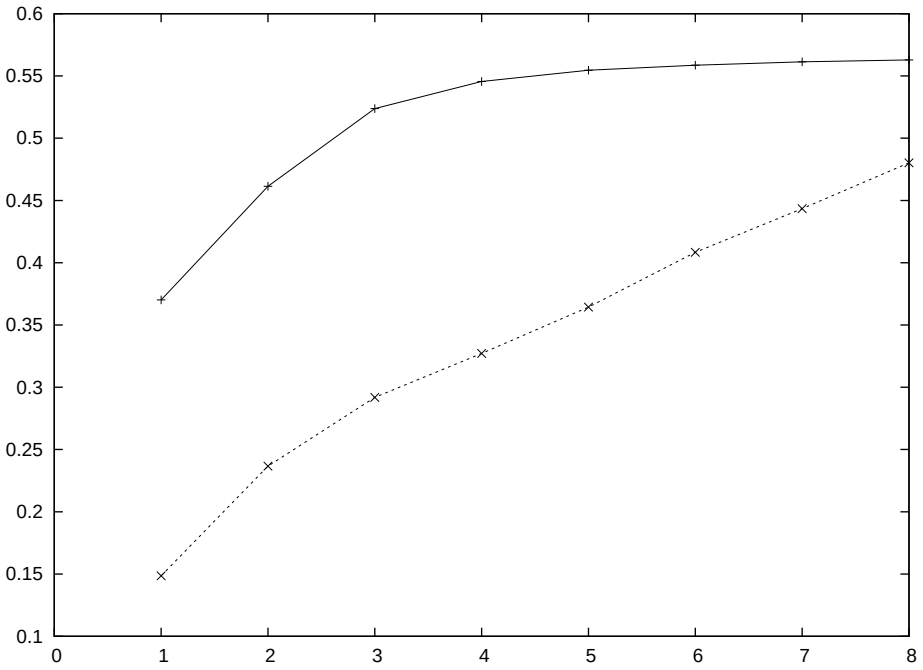


Figure 5.3: The variance of both the dependent and independent variables using Projection to Latent Structures. The percentage of age data variance (Dependent) explained is shown by the solid line (top) and the percentage variance of the model parameters (independent) is shown by the dashed line (bottom).

PLS was performed to find the decomposition that explains the covariance between the face models and their ages. We found that the first 5 principal vectors, $\mathbf{P}^{(5)}$, of the decomposition explained 56.3% of the variance in the face ages and showed little improvement in accuracy thereafter, as shown in figure 5.3.

We now have a model that explains 55.6 of the age variance and 52.8% of the parameter variance in the data using 6 parameters. Taking more than 6 parameters would improve the coverage of the model paramters but not the ageing paramters, this is counter productive as we want only that data that pertains to ageing.

We separate the parameters into two components; the components most related to ageing and a remainder. As the data used to train the PLS model has been converted to Z-scores by centring the data on the mean and scaling by the standard deviation, we must convert

the parameters of the input face $\mathbf{f}$ to Z-scores also. We denote the Z-score converted face as $\bar{\mathbf{f}}$. $\dot{\mathbf{f}} \approx \mathbf{g}P$. Since the loading matrix P is not generally orthogonal in PLS regression $\mathbf{g}$ is computed as,

$$\mathbf{g} = (P^T P)^{-1} P\bar{\mathbf{f}} \tag{5.9}$$

The PCA face model parameters can be recovered from the PLS space as $\bar{\mathbf{f}}' = \mathbf{g}P$ and converted from Z-scores to the original PCA parameter space using $\mathbf{f}' = \bar{\mathbf{f}}'\sigma + \hat{\mathbf{f}}$.

In general the recovered $\bar{\mathbf{f}}' \neq \mathbf{f}$, so we compute the residual $\mathbf{r}$ as $\bar{\mathbf{f}} = \mathbf{g}P + \mathbf{r}$. Ageing is performed using the Individualized Linear ageing model 5.3 on the PLS model parameters instead of the PCA model parameters. After the face is aged the residuals $\mathbf{r}$ are added back in. The overall algorithm is outlined in 5.2.

(5.10)

Algorithm 5.2 PLS ageing algorithm

Train PLS model.
for Face model i in the training set **do**
Convert parameters of model i to Z-scores. $\bar{\mathbf{f}} = \frac{\mathbf{f}-\hat{\mathbf{f}}}{\sigma}$.
Calculate parameters in PLS space using $\mathbf{g}_i = (P^T P)^{-1} P\bar{\mathbf{f}}_\mathbf{i}$, equation (5.9).
end for
Train Individualised Linear ageing model, equation (5.3) on PLS face models $\mathbf{g}$.

Require: Input face model with parameters $\mathbf{p}$
Convert parameters $\mathbf{p}$ to Z-scores. $\bar{\mathbf{f}} = \frac{\mathbf{f}-\hat{\mathbf{f}}}{\sigma}$.
Convert $\mathbf{p}$ from PCA model space to PLS space. $\mathbf{g} = (P^T P)^{-1} P\bar{\mathbf{f}}$, equation (5.9).
Calculate residuals, $\mathbf{r} = \bar{\mathbf{f}} - \mathbf{g}P$.
Age $\mathbf{g}$ using Individual Linear ageing model in PLS space.
Recover PCA parameters from PLS parameters. $\bar{\mathbf{f}}' = \mathbf{g}'P$.
Add residuals $\mathbf{r}$ to $\bar{\mathbf{f}}'$.
Convert from Z-scores to original model-space. $\mathbf{f}' = \bar{\mathbf{f}}'\sigma + \hat{\mathbf{f}}$.

5.5 Summary

In this chapter we have described how ageing functions can be created by training linear paths through a three-dimension PCA based face space. Focusing on three methods for

Figure 5.4: Examples of aged face images. Each row contains a different individual. The columns show, from left to right, the original mid-child face model for each individual, the mid-child face model aged to student age using the Prototyping method, the Individual Linear method and the PLS method. The right-most column shows the original face model for the individual at student age. The final row shows the same individual as the 3^{rd} rotated to a show a $\frac{3}{4}$ profile.

visual facial ageing; ageing using average prototypes, ageing using individualised linear ageing trajectories and a novel method of ageing using Projection to Latent Structures (PLS). In the next chapter we will evaluate, in detail, the ability of these algorithms to age a face model in terms of accuracy and resulting perceived age.

Chapter 6

Results

In the previous chapters we have described how to generate a synthetically aged image from an input image. First a three-dimensional face model is extracted by fitting a Morphable Model to the image, then the face is aged by altering the parameters of the face model using a linear statistical ageing function. Finally the synthesised aged face image is produced by rendering the Morpahble Model with the new parameters from the ageing method. In this chapter we will evaluate, in detail, the effectiveness of the statistical ageing models outlined in the previous chapter; Ageing using Prototypes, Individualised Linear Ageing, and PLS based ageing. We will evaluated these methods, for both statistical similarity of an aged face model to a known face model of the individual at the target age, and the ability of human observers to perceive the face models as being aged. Finally as it is important for most applications of a visual ageing synthesis system for the individual to be recognizable after the ageing process.

Leave One Out Evaluation

Each ageing method was tested using a *Leave One Out* paradigm. All the faces associated with a particular individual were removed from the training-set and the ageing model trained over the remaining face models. For each individual in the training-set the face model in the mid-child age group is aged to the age of that individual's face model in the student age group. For each individual tested we have a face model at the target age which has been removed from the training set. We compare the results of the ageing methods to

this model to determine accuracy, and refer to it as the *ground-truth* model.

6.1 Quantitative Evaluation

Using the Mahalanobis distance as a similarity measure we can quantify the error between the synthetically aged models and a known ground-truth for each individual as the Root Mean Squared Error between the Mahalanobis distances between the shape and colour parameters of the two models. The face models of the individuals in the mid-child and late-child strata were aged to their matching student ages. The aged model and a model of the individual at the target age are then compared using the distance measure to give an error for the ageing model. This is repeated for all the individuals in the training set. The Root Means Squared Error for all the individuals in the set is used to evaluate the model.

$$RMSE = \sqrt{\sum_{i} d(\mathbf{f}(\mathbf{x_i^s}), \mathbf{x}_i^s)^2} \tag{6.1}$$

where $\mathbf{x}_i^s$ and $\mathbf{x}_i^t$ are the parameters of a set of individuals at two different ages, the start age and the target age respectively. $\mathbf{f}(\mathbf{x_i^s})$ are the parameters of the input face $\mathbf{x}_i^t$ at the starting age and d is the Mahalanobis distance function,

$$d(\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{y})) = \sqrt{\sum_{i=1}^{n} \frac{(\mathbf{x}_i - \mathbf{y}_i)^2}{\sigma_i^2}}, \tag{6.2}$$

where σ_i is the standard deviation of the face-space along axis i.

This method is applied to the algorithms detailed above and the results can be found in Table 6.1. The figure shows the Root Mean Squared Error (RMSE) for each of the methods detailed above; the Prototyping method (section 5.2), the Individual Linear method (section 5.3) and the PLS method (section 5.4).

We can see from the root mean squared error (RMSE), that all the methods result in a smaller error than with no-ageing ('null'). Each method offers improvement over the one before. The standard deviations of the errors of the methods are larger than the differences in the RMSE between them, so a statistical test of significance is required to determine the level of confidence that the differences are not the result of random chance. For this we used a *dependent t-test* comparing the means of two of the tests to determine if the difference is statistically significant. The t-test assumes that the distribution of the errors is

Table 6.1: Standard deviation weighted RMSE between shape and colour parameters of aged face model and a known ground-truth model for each individual in the data-set.

Ageing Method	RMSE	Standard Deviation	Count
No ageing	10.64	2.06	70
Prototyping	8.86	1.84	70
Individual Linear	8.69	1.92	70
PLS	7.4	1.4	70

Gaussian, as the errors are an absolute distance from a point, a Gaussian distribution is not appropriate. However the pair-wise differences between two distributions are more reasonably assumed to be Gaussian. We use a dependent t-test, comparing the distribution of the differences between errors by the different ageing methods to a mean of zero. The t-tests showed all the differences to be statistically significant. The dependent t-test between the Prototyping method and the unaged images produced [p=1e-15,t=-7.4654,df=92], between the Individual Linear and Prototyping method [p=1e-15,t=-6.1989,df=92], and between the PLS method and the Individual [p=1e-15, t=-6.8198,df=92]. The probabilities (p) are the probability that the mean difference between each of the paired errors is zero.

6.2 Perceptual Evaluation

The quantitative analysis performed above evaluated the performance of the algorithms compared to a known ground truth using a quantitative distance measure, the Mahalanobis distance. These measures are a linear measure of shape and colour difference and so may miss cues associated with ageing that humans are able to perceive. It is also unable to account for errors introduced by the experiment such as those introduced by the face-fitting method. Humans are also the end-user of an ageing system and so it is worth-while evaluating how effective it is from a perceptual standpoint. We tested the ageing methods performance in two key metrics; ability to make the individual look the correct age and the ability of the algorithm to retain the individual's identity through the transform.

The experiments were run using a web interface, where visitors to the school's 'Face of the Future' web-site were asked to participate.

Ageing Accuracy

In this test one image was shown to the user. The user was asked to estimate the age of the individual displayed in years. The images consisted of renderings of the 3DMM of the mid-child aged faces aged to student age as well as renderings of the 3DMMs of the mid-child and student originals. Renderings of the same individuals aged using the 2D equivalents of the techniques were shown also. The ordering of the images was randomized.

Evaluation of Human accuracy

Although human raters can use many more visual cues in determining age than in most computer vision algorithms and therefore is frequently more accurate, it is still necessary to evaluate how effectively humans can rate the age of the individuals in the data set when only their facial features are shown, i.e. when they have no visible hair etc. For this we showed a group of raters the original photographs from the training set and asked them to estimate the age of each one.

6.2.1 Identity Retention

The specification of the ageing model requires that the algorithm not only makes the face appear the appropriate age, but also retains the identity of the individual through the ageing process. We wish to know whether the aged image is still recognisable as the person being aged. We devised an experiment to determine if the individual being aged can be distinguished from another random individual. Human raters were presented with a *Two Alternative Forced Choice* (2AFC) classification task [55]. Three images were shown to the user in two rows. The top row contained an image of an individual aged from the mid-child age to student age-group by one of the ageing methods. The bottom row contained two photographs of student aged members of the data-set, cropped to show only the faces. One of these matched the identity of the aged individual in the top row, the other, a distractor, did not. Users were asked to pick which of the two images in the bottom row most closely resembled the image in the top row, with the following question, “Click on the image in the bottom row that is most similar in appearance to the image in the top row.” The face models in the aged images were rendered using a standard position, pose and lighting.

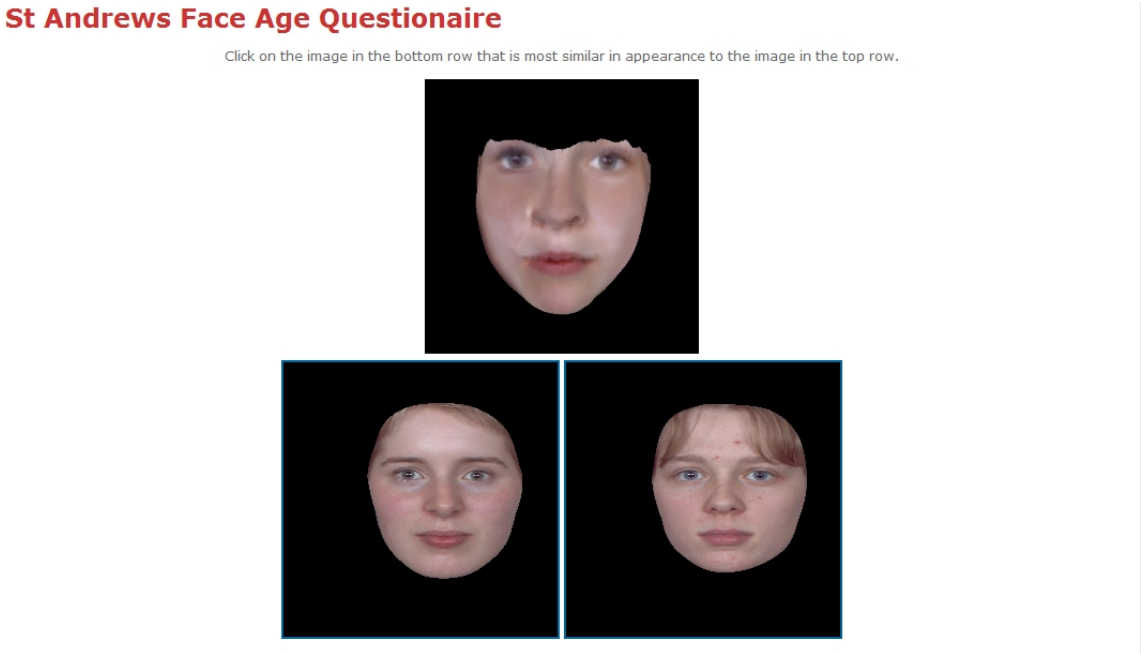


Figure 6.1: An example stimulus shown to a user in the web based experiment. The top image shows an aged face model. The bottom two images show clipped images of faces from the training set. The user is ask to select which of the two bottom images most closely resembles the rendered face image.

The students in the photographs were all in the face forward pose with standardised lighting and a neutral expression. The individual in the distractor image was also of the same or similar age to the matching image. Thus individuals are not distinguishable by variation in these areas. The users are presented with 35 tests for each ageing method, in a random order, with the order of the stimuli on the bottom row also randomised. Figure 6.1 shows an example of the stimulus shown to each rater.

In the following we denote the image containing the correct individual for the i^{th} test as I_i+ and the distractor image as I_i-. The user selects one of the two images, the selected image is denoted S_i+ and the other S_i-.

As the users are presented with two alternatives, one containing the correct individual and one containing the incorrect individual, the user will have a 0.5 probability of selecting the correct individual by random chance. One would hope that in normal circumstances the probability of choosing the correct image $p(H)$, will be greater or equal to 0.5. Evidence for a strong ability to discriminate between the correct and incorrect individual requires a large value of p(H). Using *Bayes' Theorem* we can calculate the post-test probabilities of a

true positive, i.e. correct identification of the individual and a false negative, i.e. incorrectly identify the individual as not appearing in the image, as;

$$p(\mathrm{H}_i) = p(I_i + |S_i+) = \frac{TP_i}{TP_i + FP_i}$$
$$p(\mathrm{F}_i) = p(I_i + |S_i-) = \frac{FP_i}{TP_i + FP_i} \tag{6.3}$$

where the probability that the individual is present in the image, $p(I+)$ and the probability that the individual is not present $p(I-)$ are both equal to 0.5.

The sensitivity index d' is calculated for a 2AFC test as [55],

$$z(p(H)) - z(p(F)) = \sqrt{2}d' \tag{6.4}$$

where $z(.)$ is the the inverse normal cumulative distribution function. d' measures the distance between the means of the true positive and false positive distributions. A value of 0 means that the probability of user selection of the correct individual is equal to the probability of selecting that individual by random chance. Higher values of d' indicate increasing ability of the user to select the correct individual.

Pearson's χ^2 test can be used to determine the probability of the results being obtained by chance. The choices can be modelled as a binomial distribution with mean $np(I_i+)$ and variance $np(I_i-)p(I_i+)$. From the central limit theorem $z = (TP - np(I_i+))$ has an approximately Normal distribution, which when squared has the distribution $z^2 \approx \chi^2_{(1)}$.

$$\chi^2 = \frac{TP - np(I_i+)}{np(I_i+)p(I_i-)} \tag{6.5}$$

From this distribution the confidence that the responses are not random chance can be calculated. The null hypothesis being tested is that the responses are random chance and has a mean of 0.5, this is rejected at the 90% confidence level if $\chi^2 > 2.7055$.

The proportion of correct responses and the d' for each ageing method is shown in table 6.2.

The proportion of correct results is greater than chance on all the models. The task set for raters is a hard one, with raters often failing to recognise the correct individual when shown an image of the original face model (see section 4.6) The Individualised Linear ageing transform offers a slight improvement on the Prototyping transform in matching the target individual. This is agrees with earlier quantitative results showing that the Individualised

Table 6.2: The Proportion of Correct responses d' and χ^2 for each ageing method in the test for retention of identity through the ageing transform.

Ageing Method	Proportion Correct	d'	chi^2
Prototyping	0.542	0.151	8.1002
Individual Linear	0.601	0.363	46.061
PLS	0.526	0.093	3.0675

Linear ageing method is more accurate than the Prototyping method. However users attempting the task with faces aged using the PLS method were less able to correctly identify the target individual. This contradicts the results of the quantitative analysis which showed that PLS was more accurate than both the Prototyping and Individualised Linear methods. The Individual Linear ageing method has a χ^2 of 46.061 with a p-value of $p = 1e - 20$ causing us to reject the null hypothesis and conclude that there is a statistically significant difference between the mean and random chance mean 0.5. The Prototyping method has $\chi^2 = 8.1$ which has a p-value of $p = 0.004$ allowing us to reject the null hypothesis has conclude that there is statistically significance difference between the mean the distribution of random chance. The PLS method has $\chi^2 = 3.0675$ and $p = 0.08$ again allowing us to reject the null hypothesis and conclude a statistically significant difference, but at the lower confidence level than with the other two methods. The high levels of statistical significance of the results is due to the large number of trails $(n = 1122)$, d' measures the statistical sensitivity of users to the tests, a high d' shows a high ability to distinguish between the two individuals and a low d' shows a lower ability to distinguish between them. Unlike the χ^2 the d' statistical shows a relatively low ability to distinguish the correct individual.

6.2.2 Perceived Age

Quantitative measures may miss ageing cues that human raters would be able to detect. In order to measure how well the ageing methods perform at simulating the visual effects of ageing we performed a series of tests where humans were asked to estimate the age of a face image. Each rater was shown a single image of a rendered face model at a time and asked to estimate the age of the face shown (see Figure 6.2. The age was selected from a range between 5 and 30 to the nearest year. The stimuli were a selection of mid-child faces aged to student age by the three-methods; prototyping, individualized linear and PLS, together

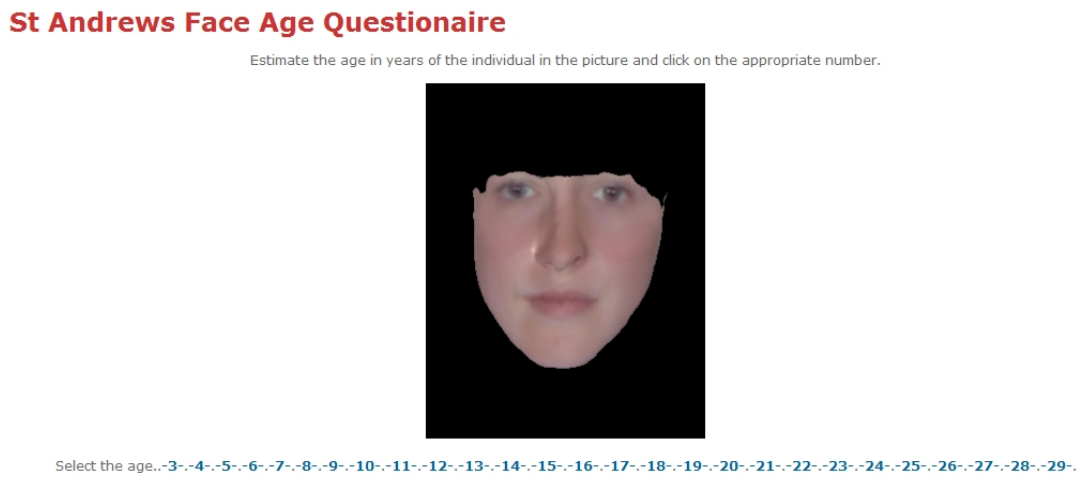


Figure 6.2: An example stimulus shown to a user in the web based experiment. The image is of a rendered face model, the user is asked to select from the choices listed along the bottom of the screen what age the individual shown appears to be.

with the rendered face models of the individuals at the source and target ages. Thus we have five groups for each individual in the face model data-set.

A histogram of the total number of responses from human raters matching each age over all the images presented to all raters showed that the distribution of responses approximated a Normal distribution, but that users had tendency to over select values at the extreme end of the range as can be seen in Figure 6.3. We removed ratings that selected the extreme value 29, as was the largest single response, it would have given the distribution with a mean of 16.269 a mode of 29.

Table 6.3 shows the mean perceived age in years for the face models aged by the different methods as well as the mean ages of the rendered models of the original face models. The mean ages of the original face models, labelled ‘Student’ and ‘Mid Child’ are the mean responses of the human raters to the original face models, resulting from fitting the morphable model to two dimensional images. Thus they are an indicator of the ability of the face models and face fitting to capture ageing information. The mean biological age of the Mid Child group was 6.54, the estimate given by human raters was 12.762, this is substantially higher than the biological age. However it is significantly lower than the average perceived age of the Student age group. Human raters were able to distinguish between the two groups so ageing information was retained by the process. The mean biological age of the Student age group was 20.02, the mean perceived age was 16.728. There is also a

tendency for the raters to underestimate the age of older subjects and overestimate the age of young subjects, this tendency is visible in the histogram of all the age ratings as shown in figure 6.3, with the raters tending towards the age of 15.

The mean perceived age of the face models produced by all three ageing methods was higher than the perceived age of the Mid Child age group, showing that all three-methods successfully aged the faces. Of the three, the most effective, with mean perceived age closest to that of the original face models, was the PLS ageing method, followed by the Prototyping method with the Individual Linear method last. This differs from the the quantitative results (section 6.1) in that the Individual Linear method performed less well than the Prototyping method.

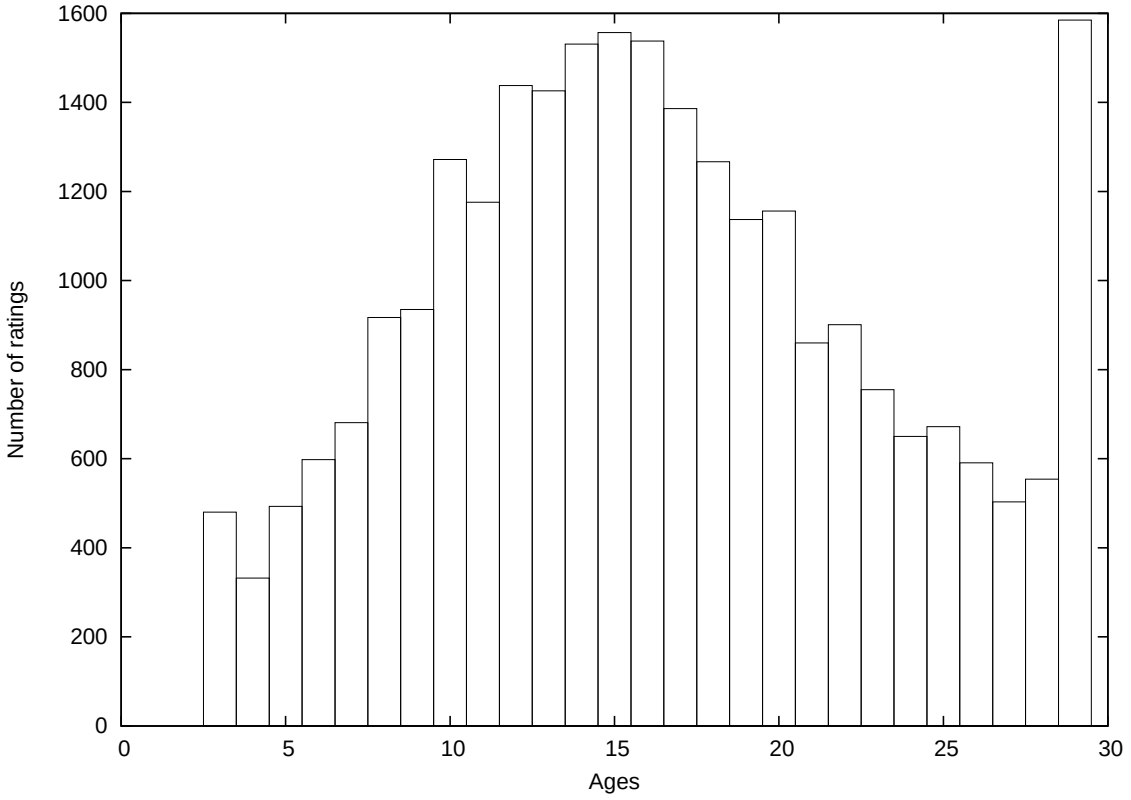


Figure 6.3: Distribution of age responses from human raters for rendered face models. Note large number of responses for extreme values.

The mean perceived ages only provides a guide to the error in as each age strata included a range of ages. To give a more precise indication of the error we also calculated the mean of the differences between the percieved and actual or target age. The mean of the age differences between the perceived age of the rendered face model and the biological age

Table 6.3: Mean (μ) and standard deviation (σ) of the human rated ages for faces ages by each method

Ageing Method	μ	σ	Count
Prototyping	16.289	5.9237	4596
Individual Linear	16.106	5.9848	4646
PLS	16.394	5.8209	4551
Student	16.728	5.2372	5855
Mid Child	12.762	6.0626	4678

of the original face or the target age of an aged face model can be found in Table 6.4. There is a clear difference between the biological age of the face model and the perceived age from the web experiment, with the perceived age of the Student age group being on average 3.3737 years less than the biological age. The same applies to the Mid Child age group with an average difference of 6.2135. As before this shows a tendency to under-rate Student aged faces and over-rate Mid Child aged faces. This over and underestimation is a reflection of the difficulty of the task presented to the raters. The face images have been cropped removing some of the cues normally used to judge age in human faces, such as hair-style. The face is also taken out of context, the user is unable to judge the size of the person depicted and the relative size of the head to the body, thus removing more cues to age.

The mean age difference in years between the aged face models using the three ageing methods and the target age is also shown in table 6.4. As before the Individual Linear method performs the least well with the largest mean difference between the perceived age and the biological age. This is followed by the Prototyping method, with PLS offering the least difference and the best performance.

As before the standard deviation is greater than the difference between the two methods so we used a one tailed independent t-test to determine if the differences were significant. The results of the t-test are shown in figure 6.5. The means of the aged face models by the three methods are significantly lower than the mean of the original student face models, as can be seen from the first row and column of the table, only the PLS method shows a slight (0.05) possibility that means are greater than or equal, but we can still be 95% confident that the perceived age of the PLS aged model is consistanty less then the

perceived age of the original model. The closeness of the mean perceived ages of the three methods reduces the confidence that one method offers an improvement over the others. In the case of the Individual Linear method we can be at least 94.5% confident that the other two methods offer an improvement, 98% in the case of the comparison with the PLS method. The confidence in improvement of the PLS method over the Prototyping method is less at 77.5%. This is consistent with the findings of the mean age differences and shows them to be statisticaly significant, although the confidence of the improvement of the PLS method over the prototyping method is only 77.5%, thus there is a 22.5% probability that the improvement result is chance.

Table 6.4: Mean (μ) and standard deviation (σ) of the error in years between the perceived age and the true or target age in human rated ages for faces aged by each method.

Ageing Method	μ	σ	Count
Prototyping	-3.6614	6.1888	4596
Individual Linear	-3.8674	6.1688	4646
PLS	-3.5643	6.1098	4551
Student	-3.3737	5.4273	5855
Mid Child	6.2135	6.0815	4678

Table 6.5: Probability of accepting null hypothesis of one-tailed independent t-tests between perceived age rating grouped by ageing method. The null hypothesis is that the means of the perceived ages is identical. The alternative is that the mean of the method in the rows is lower than the mean for the method in the columns.

Ageing Method	Original Student	Prototyping	Individual Linear	PLS
Original Student	-	0.99427	1.0	0.95363
Prototyping	0.0057345	-	0.94544	0.22495
Individual Linear	0.0000	0.054561	-	0.017934
PLS	0.046367	0.77505	0.98207	-

6.3 Summary

In this chapter we have compared three statistical ageing methods. Two well known methods, one using average prototypes, one using Individualised Linear trajectories and a third

using Projection to Latent Structure. These methods have been assessed using both quantitative and perceptual evaluation. The results can be summarised as follows,

- All three methods successfully aged the input images as shown by both quantitative and perceptual evaluation.
- Of the three methods assessed the PLS method produced the most accurate ageing when compared to a face model of the same individual at the target age using quantitative evaluation.
- The *PLS* method also produced images with perceived age closest to the target age measured using perceptual evaluation.
- All three methods retained some of the individuality of the input face as shown by the two alternative forced-choice perception results.
- The *Individualised Linear* method proved most capable of retaining the identity of the individual in the input face, being the most frequently recognised by the human raters.
- The individuals aged using the *PLS* method were the least recognisable to the human raters.

At first sight the results for the individualised ageing method seem somewhat anomalous, in the quantitative evaluation it out performs the Prototyping method, but in the perceptual evaluation it doesn't. It is also the most successful at retaining identity through the ageing process. It seem likely that these two observations are in some way correlated, i.e. that more a face model is successfully aged by the ageing process the more the face model loses some of its' identity. This is likely to be the same for any face transform.

Chapter 7

Conclusions and Future Work

7.1 Summary

This thesis investigated the use of 3D Morphable Models fitted to face-images for the training and application of ageing to 2D images. The effectiveness of known top-down fitting methods were investigated and evaluated in terms of their accuracy in capturing face shape and colour. To my knowledge ours is the first attempt to compare an estimated face model to a three-dimensional scan of the same individual. We found that in general the fitting is more accurate if a number of training points are used to align and train the model. In our case we used up to 105 hand delineated points to train each face model. In perceptual analysis of the fitted face models we found that raters were able to correctly identify the identity of the individual depicted by the face model from a two-dimensional image of the subject. However raters were less able determine the correct identity of individuals if the pose and lighting were normalised. Users were also significantly less able to recognise students from their fitted face model. This could mean that raters were detecting features other than identity, residual effects of lighting or the expressions of the subject that were missing in the student face models. However the ability of raters to correctly identify and individual from a relatively small number of shape and colour components suggests that the method is effective at capturing details of the individual from the image. I also present a novel method to perform visual facial ageing in using a Projection to Latent Structures technique to separate those factors relating to ageing from those that do not. Its efficacy in producing aged images of an individual has been demonstrated using perceptual evaluation.

In these experiments it was found that the PLS method successfully produced images that human raters perceived as being older than the starting image of the individual, and also that this age increase was greater than that produced by the Prototyping and Linear ageing techniques. The PLS method was also found to be more accurate in quantitative evaluation. The Mahalanobis distance metric on Morphable Models also serves as a method of identifying a subject [15] thus it can be reasonably concluded that the technique retains identity information as well as capturing ageing effects. The extension to three-dimensions of Lanitis et al's [43] individualised linear ageing method confirmed that their results applied in three-dimensions as well as two, both in terms of the improvement over a linear training method and in finding a correlation between the model parameters and the Mahalanobis distance between individuals.

7.2 Future improvements

The results of the modelling of ageing is dependent on the accuracy of face fitting model, using a more sophisticated face-fitting method, such as that developed by Romdhani [67] would potentially reduce errors caused by the fitting techniques. However Romdhani's method is not applicable to image functions for which a Baysian prior is unavailable, investigating methods, such as edge alignment [40], for which such a prior is not easy calculated offers an alternative that is more applicable to contour generating structures such as shadows. An obvious improvement on this would be the creation of a data-set of three-dimensional scans of the same individual at several time points, if such data-sets were available the process of fitting to an image, which will always be inherently lossy, will not be required. As three-dimensional scanners become cheaper and more common-place such data-sets are more likely to be created.

The data-set used in this thesis was relatively small, containing only 35 different individuals. The statistical methods used to train the ageing models would be more accurate with a larger data-set. The individuals depicted in the original photographs exhibited a wide range of expressions, an investigation of methods for expression normalisation would amortise the affect of expressions on the training set. The three-dimensional data-set used to create the morphable model was also limited both in size and scope. A larger set of three-dimensional models used to build the Morphable Model would result in a more accurate description of the space of human faces, resulting in more accurate face-fitting and in

turn more accurate data with-which to train the ageing model. A wide range of expressions in the Morphable Model's training set would result in a greater ability of the morphable model to describe the varied expression in the faces it is fitted to.

This thesis concentrated on linear methods of training the ageing data. Investigation of non-linear methods such as the set of statistical models based Kernel methods may provide ways of modelling variations such as the 'growth spurts' children go through. Support Vector Regression has been investigated by Scherbaum [72], however other methods such as Kernel PLS may offer a better correlation between the face model parameters and the subject's age, due to its ability to 'factor out' those parts of the higher dimensional space unrelated to ageing. Kernel PLS does suffer from an inability to reduce the number of 'support vectors' and as a result the matrix that maps an input to a higher dimension is the square of the number of training subjects.

However not all future extension need be non-linear. The approach of Scandrett et al. [71] in two-dimensions using exemplar faces of the individual at ages younger than the input age to build an 'historical' ageing trajectory, could be combined with the face-fitting strategy to create three-dimensional face models.

This thesis does not tackle the area of fine detail analysis of ageing, concentrating instead on a holistic approach. Fine details, such as wrinkles significantly affect the perceived age of the subject especially in later adult life. Most methods for fine detail analysis rely on high-frequency analysis using wavelets [83] or Gaussian image pyramid [31]. The holistic approach of this thesis can be considered a low-frequency approach. Clearly both are needed to make the ageing model as accurate as possible. An approach that could combine sources of age-related change from both high detail methods and low-frequency methods would offer further improvements.

Finally the use of three-dimensional models allows extensions and applications not available to two-dimensional ageing techniques. The ability of a three-dimensional model to easily and accurately describe rotation makes method such as this ideally suited to video related applications. For example by updating the face fitting algorithm to track a moving head, and applying the ageing algorithm to the fitted model each frame, an actor can be made to look older or younger than they really are.

7.2.1 Final remarks

Visual facial age simulation has applications in many areas, from assisting in locating missing persons, guiding reconstructive surgery, simulating ageing of actors in films or understanding the effects of skin care products. In this thesis, we have used a mixed two/three dimensional approach in which a generative three-dimensional face model is first built and then adapted to a set of two-dimensional face images in order to capture age related changes. This combines the main advantage of a two-dimensional image set, wide availability, with the ability of a three-dimensional model to describe the shape of human head. The main difficulty with this approach has been producing accurate three-dimensional face models from unstandardised two-dimensional images. However improvements in the field of face-fitting continue to be made and it is not unreasonable to expect that accurate and efficient methods for extracting a three-dimensional face model from single images and video sequences will be developed in the near future. Overall, this method seems to produce plausible images of an individual aged by a specified number of years and is readily extensible to other statistical ageing methods.

Bibliography

[1] 3dmd systems. 3q technologies.

[2] H. Abdi. Partial least square regression (pls regression). In *In N.J. Salkind (Ed.): Encyclopedia of Measurement and Statistics.*, pages 740–744. Thousand Oaks (CA): Sage., 2007.

[3] Brett Allen, Brian Curless, and Zoran Popović. Articulated body deformation from range scan data. *ACM Trans. Graph.*, 21(3):612–619, 2002.

[4] Brett Allen, Brian Curless, and Zoran Popović. The space of human body shapes: reconstruction and parameterization from range scans. In *SIGGRAPH '03: ACM SIGGRAPH 2003 Papers*, pages 587–594, New York, NY, USA, 2003. ACM.

[5] Brian Amberg, Sami Romdhani, and Thomas Vetter. Optimal step nonrigid icp algorithms for surface registration. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1–8, 2007.

[6] Simon Baker, Ralph Gross, and Iain Matthews. Lucas-kanade 20 years on: A unifying framework: Part 3. Technical Report CMU-RI-TR-03-35, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, November 2003.

[7] Simon Baker, Ralph Gross, and Iain Matthews. Lucas-kanade 20 years on: A unifying framework: Part 4. Technical Report CMU-RI-TR-04-14, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, February 2004.

[8] Simon Baker, Ralph Gross, Iain Matthews, and Takahiro Ishikawa. Lucas-kanade 20 years on: A unifying framework: Part 2. Technical Report CMU-RI-TR-03-01, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, February 2003.

[9] Simon Baker and Iain Matthews. Lucas-kanade 20 years on: A unifying framework: Part 1. Technical Report CMU-RI-TR-02-16, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, July 2002.

[10] Simon Baker and Iain Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221 – 255, March 2004.

[11] Simon Baker, Raju Patil, Kong Man Cheung, and Iain Matthews. Lucas-kanade 20 years on: Part 5. Technical Report CMU-RI-TR-04-64, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, November 2004.

[12] Ronen Basri and David Jacobs. Lambertian reflectance and linear subspaces. *7th IEEE International Conference on Computer Vision (ICCV)*, 02:383, 2001.

[13] Perrett D I Benson P J. Extracting prototypical facial images from exemplars. volume 22, pages 257–262, 1993.

[14] Paul J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, 1992.

[15] V. Blanz and S. Romdhani. Face identification across different poses and illuminations with a 3d morphable model. In *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, page 202, Washington, DC, USA, 2002. IEEE Computer Society.

[16] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co.

[17] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(6):567–585, 1989.

[18] D.M. Burt and D.I. Perrett. Perception of age in adult caucasian male faces: computer graphic manipulation of shape and colour information. *Proceedings of Royal Society of London*, B-259:137–143, 1995.

[19] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *Lecture Notes in Computer Science*, 1407:484–, 1998.

[20] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models—their training and application. *Comput. Vis. Image Underst.*, 61(1):38–59, 1995.

[21] T. F. Cootes, K. Walker, and C. J. Taylor. View-based active appearance models. In *FG '00: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, page 227, Washington, DC, USA, 2000. IEEE Computer Society.

[22] T.F. Cootes, C.J. Taylor, A. Lanitis, D.H. Cooper, and J. Graham. Building and using flexible models incorporating grey-level information. pages 242–246, 1993.

[23] I. Craw and P. Cameron. Parameterising images for recognition and reconstruction. Turing Institute Press and Springer Verlag, 1991.

[24] George Robert Cross. *Markov random field texture models*. PhD thesis, East Lansing, MI, USA, 1980.

[25] Ian Dryden. General shape and registration analysis. In *In*. Chapman and Hall, 1997.

[26] Jacques Feldmar and Nicholas Ayache. Rigid, affine and locally affine registration of free-form surfaces. *IJCV*, 18(18):99–119, 1996.

[27] Michael S. Floater. Parametrization and smooth approximation of surface triangulations. *Comput. Aided Geom. Des.*, 14(3):231–250, 1997.

[28] Michael S. Floater. Mean value coordinates. *Comput. Aided Geom. Des.*, 20(1):19–27, 2003.

[29] Michael S. Floater and Kai Hormann. Surface parameterization: a tutorial and survey. In N. A. Dodgson, M. S. Floater, and M. A. Sabin, editors, *Advances in multiresolution for geometric modelling*, pages 157–186. Springer Verlag, 2005.

[30] Francis Galton. *Composite portraits made by combining those of many different persons into a single figure*, volume 8. 1879.

[31] Maulin R. Gandhi, Maulin R. G, and Mrs Rajesh N. G. A method for automatic synthesis of aged human facial images. Technical report, Masters thesis, McGill University,2004. 1, 2004.

[32] Gregory D. Hager and Peter N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:1025–1039, 1998.

[33] Adrian Hilton, Jonathan Starck, and Gordon Collins. From 3d shape capture to animated models. *3D Data Processing Visualization and Transmission, International Symposium on*, 0:246, 2002.

[34] Melvin J. Hinich and Prem P. Talwar. A simple method for robust regression. *Journal of the American Statistical Association*, 70(349):113–119, 1975.

[35] Hussein Karam Hussein. Towards realistic facial modeling and re-rendering of human skin aging animation. In *SMI '02: Proceedings of the Shape Modeling International 2002 (SMI'02)*, page 205, Washington, DC, USA, 2002. IEEE Computer Society.

[36] Tim J. Hutton, Bernard F. Buxton, Peter Hammond, and Henry W. W. Potts. Estimating average growth trajectories in shape-space using kernel smoothing. *IEEE Trans. Med. Imaging*, 22(6):747–753, 2003.

[37] D.J. Jobson, Z.U. Rahman, and G.A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. 6(7):965–976, July 1997.

[38] Kolja Kähler, Jörg Haber, Hitoshi Yamauchi, and Hans-Peter Seidel. Head shop: generating animated head models with anatomical structure. In *SCA '02: Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 55–63, New York, NY, USA, 2002. ACM.

[39] D. Kalamani and P. Balasubramanie. Age classification using fuzzy lattice neural network. In *ISDA '06: Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA'06)*, pages 225–230, Washington, DC, USA, 2006. IEEE Computer Society.

[40] M. Keller, R. Knothe, and T. Vetter. 3d reconstruction of human faces from occluding contours. In *Model-based Imaging, Rendering, Image Analysis and Graphical special Effects*, pages 261–273, 2007.

[41] Young H. Kwon and Niels da Vitoria Lobo. Age classification from facial images. *Computer Vision and Image Understanding: CVIU*, 74(1):1–21, 1999.

[42] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Trans. Systems, Man and Cybernetics*, 34(1):621–628, February 2004.

[43] A. Lanitis, C. J. Taylor, and T. F. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):442–455, 2002.

[44] Andreas Lanitis, Chris J. Taylor, and Timothy F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):743–756, 1997.

[45] Seung-Yong Lee, Kyung-Yong Chwa, and Sung Yong Shin. Image metamorphosis using snakes and free-form deformations. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 439–448, New York, NY, USA, 1995. ACM Press.

[46] Seungyong Lee, George Wolberg, and Sung Yong Shin. Scattered data interpolation with multilevel b-splines. *IEEE Transactions on Visualization and Computer Graphics*, 3(3):228–244, 1997.

[47] Zicheng Liu, Zhengyou Zhang, and Ying Shan. Image-based surface detail transfer. *IEEE Computer Graphics and Applications*, 24(3):30–35, 2004.

[48] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. pages 674–679, 1981.

[49] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. pages 121–130, April 1981.

[50] L.S. Mark and J.T. Todd. The perception of growth in three dimensions. *Perception and Psychophysics*, 33(2):193–196, 1983.

[51] Iain Matthews and Simon Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135 – 164, November 2004. In Press.

[52] Iain Matthews, Jing Xiao, and Simon Baker. 2d vs. 3d deformable face models: Representational power, construction, and real-time fitting. *Int. J. Comput. Vision*, 75(1):93–113, 2007.

[53] M.Kirby and L. Sirovich. Application of karhunen-loève procedure for characterization of human faces. *Transactions on PAMI*, 12(1):103–108, 1990.

[54] Baback Moghaddam, Jinho Lee, Hanspeter Pfister, and Raghu Machiraju. Model-based 3d face capture with shape-from-silhouettes. In *In IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, page pages, 2003.

[55] C. Douglas Creelman Neil A. Macmillan. *Detection Theory: A User's Guide*.

[56] Ren Ng, Ravi Ramamoorthi, and Pat Hanrahan. Triple product wavelet integrals for all-frequency relighting. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*, pages 477–487, New York, NY, USA, 2004. ACM.

[57] John D. Owens, Mike Houston, David Luebke, Simon Green, John E. Stone, and James C. Phillips. Gpu computing. *Proceedings of the IEEE*, 96(5), May 2008.

[58] James Paterson and Andrew Fitzgibbon. 3d head tracking using non-linear optimization. In *In Proc. BMVC., 609–618*, pages 609–618, 2003.

[59] J.B. Pittenger and R.E. Shaw. Aging faces as viscol-elastic events: Implications for a theory of nonrigid shape perception. *J. Experimental Psychology: Human Perception and Performance*, 1(4):374–382, 1975.

[60] J.B. Pittenger, R.E. Shaw, and L.S. Mark. Perceptual information for the age level of faces as a higher order invariant of growth. *J. Experimental Psychology: Human Perception and Performance*, 5(3):478–493, 1975.

[61] Emmanuel Prados and Oliver Faugeras. *Shape from Shading*. Springer, 2005.

[62] Emil Praun, Wim Sweldens, and Peter Schr oder. Consistent mesh parameterizations. In *Proceedings of ACM SIGGRAPH 2001*, pages 179–184, August 2001.

[63] William Press, Saul Teukolsky, William Vetterling, and Brian Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, 2nd edition, 1992.

[64] Iain Matthews Ralph Gross and Simon Baker. *Generic vs. person specific active appearance models*, 23(1):1080–1093, November 2005.

[65] Ravi Ramamoorthi. Analytic pca construction for theoretical analysis of lighting variability in images of a lambertian object. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(10):1322–1333, 2002.

[66] N. Ramanathan and R. Chellappa. Modeling age progression in young faces. In *Computer Vision and Pattern Recognition Conference (CVPR '06)*, pages 387–394, 2006.

[67] Sami Romdhani. *Face Image Analysis using a Multiple Feature Fitting Strategy*. PhD thesis, University of Basel, 2005.

[68] Sami Romdhani, Volker Blanz, and Thomas Vetter. Face identification by fitting a 3d morphable model using linear shape and texture error functions. In *Computer Vision – ECCV'02*, volume 4, pages 3–19, Copenhagen, Denmark, 2002.

[69] Sami Romdhani and Thomas Vetter. Efficient, robust and accurate fitting of a 3d morphable model. In *9th IEEE International Conference on Computer Vision (ICCV)*, pages 59–66, 2003.

[70] Duncan A. Rowland and David I. Perrett. Manipulating facial appearance through shape and color. *IEEE Computer Graphics and Applications*, 15(5):70–76, 1995.

[71] Catherine M. Scandrett, Christopher J. Solomon, and Stuart J. Gibson. A person-specific, rigorous aging model of the human face. *Pattern Recogn. Lett.*, 27(15):1776–1787, 2006.

[72] Kristina Scherbaum, Martin Sunkel, Hans-Peter Seidel, and Volker Blanz. Prediction of individual non-linear aging trajectories of faces. In *The European Association for Computer Graphics, 28th Annual Conference, EUROGRAPHICS 2007*, volume 26 of *Computer Graphics Forum*, pages 285–294, Prague, Czech Republic, 2007. The European Association for Computer Graphics, Blackwell.

[73] Stan Sclaroff and John Isidoro. Active blobs. Technical Report 1997-008, 5, 1997.

[74] Stan Sclaroff and John Isidoro. Active blobs. In *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*, page 1146, Washington, DC, USA, 1998. IEEE Computer Society.

[75] Shubhabrata Sengupta, Mark Harris, Yao Zhang, and John D. Owens. Scan primitives for gpu computing. In *GH '07: Proceedings of the 22nd ACM SIGGRAPH/EUROGRAPHICS symposium on Graphics hardware*, pages 97–106, Aire-la-Ville, Switzerland, Switzerland, 2007. Eurographics Association.

[76] Michael De Smet, Rik Fransens, and Luc Van Gool. A generalized em approach for 3d model based face recognition under occlusions. In *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1423–1430, Washington, DC, USA, 2006. IEEE Computer Society.

[77] Alex J. Smola and Bernhard Sch Olkopf. A tutorial on support vector regression. Technical report, Statistics and Computing, 1998.

[78] J. Suo, F. Min, S.C. Zhu, S.G. Shan, and X.L. Chen. A multi-resolution dynamic model for face aging simulation. In *IEEE Computer Vision and Pattern Recognition or CVPR*, pages 1–8, 2007.

[79] Tutte W. T. How to draw a graph. In *Proceedings of the London Mathematical Society*, volume 13, pages 743–768, 1963.

[80] C. J. Taylor. Active shape models - 'smart snakes. In *In British Machine Vision Conference*, pages 266–275. Springer-Verlag, 1992.

[81] Barry-John Theobald, Iain Matthews, and Simon Baker. Evaluating error functions for robust active appearance models. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 149 – 154, April 2006.

[82] D'Arcy Wentworth Thompson. *On Growth and Form: The Complete Revised Edition*. Dover, 1992.

[83] Bernard Tiddeman, Michael Burt, and David I. Perrett. Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications*, 21(5):42–50, 2001.

[84] Bernard Tiddeman, Michael Stirrat, and David I. Perrett. Towards realism in facial image transformation: Results of a wavelet mrf method. *Comput. Graph. Forum*, 24(3):449–456, 2005.

[85] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, pages 586–591, 1991.

[86] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1):71–86, 1991.

[87] Y. Tong U. Park and A. K. Jain. Face recognition with temporal invariance: A 3d aging model, September 2008.

[88] T. Doyle V. Bruce, M. Burton. Further experiment on the perception of growth in three dimensions. *Perception and Psychophysics*, 46(6):528–536, 1989.

[89] Cheng-Ping Lee Wen-Bing Horng and Chun-Wen Chen. Classification of age groups based on facial features. In *Tamkang Journal of Science and Engineering*, volume 4, page 183192, 2001.

[90] H. Wold. Estimation of principal components and related models by iterative least squares. *Multivariate Analysis*, pages 391–420.

[91] Philip L. Worthington and Edwin R. Hancock. New constraints on data-closeness and needle map consistency for shape-from-shading. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(12):1250–1267, 1999.

[92] J. Wu, W.A.P. Smith, and E.R. Hancock. Gender classification using shape from shading. In *British Machine Vision Conference*, 2007.

[93] Jing Xiao, Simon Baker, Iain Matthews, and Takeo Kanade. Real-time combined 2d+3d active appearance models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 535 – 542, June 2004.

[94] Zijian Xu, Hong Chen, and Song-Chun Zhu. A high resolution grammatical model for face representation and sketching. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, pages 470–477, Washington, DC, USA, 2005. IEEE Computer Society.

[95] Lei Zhang and Dimitris Samaras. Pose invariant face recognition under arbitrary unknown lighting using spherical harmonics. In *ECCV Workshop BioAW*, pages 10–23, 2004.

Appendix A

Appendix

A.1 Mathematical Notation

$\mathbf{v}$	**Bold** lower case letters denote vectors.
$\mathbf{v}_i$	Denotes the i^{th} element of the vector $\mathbf{v}$
$\mathbf{v} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$	Defines an n-length vector with each element filled with $\mathbf{v}_1$ to $\mathbf{v}_n$ in order.
A	Matrices are denoted by **bold** upper case letters.
$A_{n \times m}$	Denotes an m column and n row matrix.
A^T	Denotes the transpose of the matrix A
$A_{i,j}$	Denotes the element in the j^{th} column along the i^{th} row.
$A_{.,j}$	Denotes the vector formed from the j^{th} row.
$A_{i,.}$	Denotes the column vector formed from the i^{th} column.
$[\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$	Defines a matrix with column vectors $\mathbf{v}_1$ to $\mathbf{v}_n$ in order.
I	Identity matrix. Dimensions determined from context.
$< \mathbf{a}, \mathbf{b} >$	Inner product of the vectors $\mathbf{a}$ and $\mathbf{b}$.

$\mathcal{I}$	Denotes an image.
$\mathcal{I}(\mathbf{x})$	r, g, b sample at position $\mathbf{x} = x, y$ in image $\mathcal{I}$.
$\mathcal{M}(;\mathbf{p})$	Morphable Model built using parameters $\mathbf{p}$.
$\mathcal{M}(\mathbf{x};\mathbf{p})$	sample at position $\mathbf{x}$ on image produced by $\mathcal{M}(;\mathbf{p})$.
$W(\mathbf{x};\mathbf{p})$	Denotes a warp defined at the point $\mathbf{x}$ that warps $\mathbf{x}$ to another point parameterised by $\mathbf{p}$.