A Statistically Rigorous Approach to the Aging of the Human Face
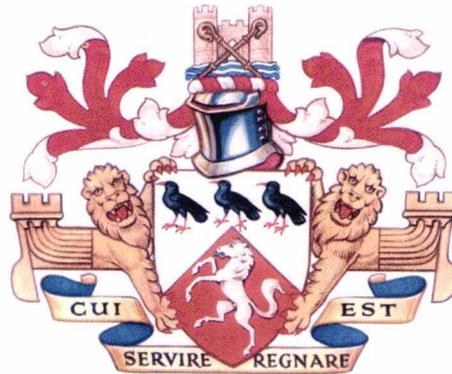
Catherine Mary Scandrett

FORENSIC IMAGING GROUP

SCHOOL OF PHYSICAL SCIENCES

UNIVERSITY OF KENT, U.K

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

SUPERVISED BY: DR. CHRISTOPHER J. SOLOMON

2007

Preface

This thesis has been prepared in the Forensic Imaging Group in the School of Physical Sciences at the University of Kent in partial fulfillment of the requirements for the Degree of Doctor of Philosophy, Ph.D., in Physics. Unless otherwise stated and referenced in the thesis, all thesis work was carried out by myself.

Catherine Mary Scandrett   i

Abstract

The ability to accurately age an image of the human face in an automatic and rigorous fashion has widespread potential applications. This thesis is concerned with the development and testing of a new approach to computerised age progression based on a statistical learning procedure.

The thesis begins with an overview of existing methodologies for age progression and outlines the need for improved procedures. After a review of the underpinning mathematical techniques, the theoretical basis of the new age-progression methodology is then presented. In this new approach, age progression is achieved through the calculation of optimised trajectories within a model space constructed from a principal component analysis of the shape and texture of a training sample of images. The statistical framework proposed extends naturally to include both generic and person-specific influences on the changes in facial appearance as aging progresses. Specific, physiological developmental periods, facial appearance at a previous age and the tendency to resemble close relatives are all incorporated into the model.

The methodology is then computationally implemented and tested. Quantitative and perceptual tests both confirm the essential validity and accuracy of the techniques. This new methodology demonstrates that near photographic-quality, age-progressed images may be obtained based on rigorous scientific principles and considerably more quickly than is currently possible using forensic artistry. It is concluded that the algorithms may, in the future, be used to augment or even replace the existing artistic methodology.

Publication List

Publications emanating from the work presented in, or related to, this thesis.

Papers published in journals and refereed conference proceedings

- C.M. Scandrett (née Hill), C.J. Solomon and S.J. Gibson. A person-specific, rigorous aging model of the human face. *Pattern Recognition Letters: Vision for Crime Detection and Prevention*, 2006, 27(15): 1776-1787.

- C.M. Scandrett (née Hill), C.J. Solomon and S.J. Gibson. Towards a Semi-automatic Method for the Statistically Rigorous Aging of the Human Face. *IEE Proceedings Vision, Image & Signal Processing*, 2006, 153(5): 639-649.

- C.M. Hill, C.J. Solomon and S.J. Gibson. Towards a Semi-automatic Method for the Statistically Rigorous Aging of the Human Face. *Proceedings of the IEE International Conference on Visual Information Engineering: Convergence in Graphics and Vision*, 2005, pp 9-15, Glasgow, UK.

- C.M. Hill, C.J. Solomon and S.J. Gibson. Aging the Human Face — a Statistically Rigorous Approach. *Proceedings of the IEE International Symposium on Imaging for Crime Detection and Prevention*, 2005, pp 89-94, London, UK.

- C.M. Hill, C.J. Solomon and S.J. Gibson. Plausible Aging of the Human Face using a Statistical Model. *Proceedings of the 7th IASTED International*

Conference on Computer Graphics and Imaging, Editor M.H. Hamza, 2004, pp 56-59, Kauai, Hawaii, USA.

Practical Applications List

Practical, real-world applications of the work presented in this thesis are listed below and are examined in more detail in Appendix H.

- Age-progression of the celebrity faces of Richard Hammond and Charlotte Church, commissioned by the BBC for use in a television programme.

- Age-progression of the face of the presenter for the Australian television programme, "Beyond Tomorrow". This was produced for Beyond International Ltd. production company.

- Age-progression of a missing child for comparison with a forensic artist's age-progressed image. This was performed for forensic artist Cathy Charsley of the Metropolitan Police.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank the following individuals:

- My supervisor, Dr. Chris Solomon, for his guidance, advice and support throughout my Ph.D.

- Dr. Andreas Lanitis and the FG-NET consortium for permitting me to use their face database in my work.

- My colleague, Dr. Matthew Maylin, for his help in producing the on-line perception tests for the analysis of my results.

- Dr. Stuart Gibson, Dr. Alvaro Pallares-Bejarano and Dr. Matthew Maylin for work-related discussions and help and for their friendship throughout my time at the University of Kent.

- Fr. Peter Geldard, Fr. Sean O'Connor and the members of the University of Kent Catholic Society, past and present, for their prayers, guidance and companionship.

- My parents and my brother, my step-parents, parents-in-law and brother-in-law for their love, support and encouragement.

- I would especially like to thank my wonderful husband, Craig, for his unconditional love, endless patience and boundless faith in me. I could not have done this without you.

Chapter 1

Background

1.1 Introduction

Forensic age-progression of the face is a discipline currently employed by law enforcement agencies during investigations into missing persons, often missing children. It is used to age an image of a child to predict their probable current facial appearance and thereby assist in their recognition and recovery. An age-progressed image may also provide the link necessary to establish new information on adult cases, for example, in "cold case" murder enquiries or in instances of criminal abscondment. However, child age-progresions are typically utilised more than adult age-progressions because, over a relatively short period of time, the face and skull of a child are undergoing such rapid growth and development that a subject may be unrecognisable if not seen for even a few months [1]. Once adulthood is reached, the changes to the face with time are predominantly textural (for example, the development of wrinkles) rather than structural.

The Police and other law enforcement agencies, as well as missing persons' charities, currently rely on the skills of highly trained forensic artists in order to perform

age-progression. According to the National Missing Persons Helpline [2] in the UK, approximately 6-12 age-progressions are carried out by forensic artists per year, a relatively low number compared to the demand from families. This represents approximately 0.35 per million, only about 1% of the estimated number of very long-term missing children[1] (see Appendix A).

Although the results obtained from age-progressions carried out by forensic artists are visually appealing and plausible, there is no assurance that they are scientifically accurate as such approaches have never been subject to systematic testing. In addition, the high level of skill required means that the production of age-progressed images is time-consuming, typically requiring twenty-four hours[2] to complete [2, 4]. Furthermore, only limited success may be achieved if the images are of insufficient quality [5, 6, 7] or if the faces are of unconstrained pose. To effectively address these problems and allow a non-specialist to produce accurate, rapid and scientifically principled results, a new approach is required.

The work in this thesis outlines such an approach using a rigorous methodology. Statistical learning methods are used to construct an abstract model space through which aging directions may be calculated from an appropriate sample of training data. A primary motive for this approach is an expected increase in speed of production which would permit a greater number of age-progressions to be performed and hence aid the recovery of more missing subjects than is currently possible. Moreover, scientifically principled age-progression might reasonably be expected to be more accurate than subjective, artistic methods [8], since the rigorous modelling of facial growth might be expected to give more objective results than those of

---

[1]A long-term missing child is one who has been missing for more than one year. A very long-term missing child is classed as one who has been missing for more than five years [3].

[2]In extreme cases, up to seven days may be required [4].

artistic interpretation.

It is hoped that the work introduced herein will eventually be used as a practical tool by the Police and missing persons bureaux. Ultimately, the objective of this, and of future work, is to produce a semi-automatic, integrated system for facial aging in which a user could present to the system a subject image, stipulate the required degree of aging and produce a near-photographic quality aged result quickly and effortlessly.

Automated age-progression methods could also find a number of commercial applications (see Appendix B). Possible examples include the animation and film industries to produce aging effects for characters, as a novelty package to give parents an insight into how their children may appear in the future and, used in reverse, automated age-regression might be used in cosmetic surgery planning to give a patient some estimate of their post-operative appearance.

Owing to the wide range of potential applications, not least in the search for missing persons, the work presented in this thesis is of potential interest and relevance across many fields and disciplines.

1.2 Thesis Overview

The remainder of this thesis is organised as follows. This chapter describes the background to the work in terms of the physiology of aging and outlines the currently used artistic approaches to age-progression. In addition, scientific methodologies in the literature are explored. In chapter 2, the mathematical techniques which underpin the novel aging algorithms are discussed. Chapter 3 introduces the theoretical basis

for the three new aging algorithms. A simple approach is first presented, based upon the average growth and development trends of a peer group sample. To incorporate the non-linear nature of aging, a related approach is discussed, constructed upon age ranges of known developmental growth periods. Finally, the most sophisticated technique is outlined, which allows the incorporation of familial influences as well as the appearance of a subject at previous ages to predict the future appearance. In chapter 4, the algorithms are implemented computationally and representative results displayed and discussed. A quantitative comparison is then made between an aged subject and its target. Chapter 5 presents a qualitative assessment of the results in terms of human perception. Comparison is also made between the results and those obtained from forensic artistry. Chapter 6 provides a summary of the thesis and appropriate conclusions are drawn. Finally, future work, based on the work described in this thesis, is discussed.

1.3 Chapter Aims

The aim of this part of the chapter is to discuss the current artistic and scientific approaches to age-progression such that the key strengths and weaknesses of these methodologies are identified. Firstly, the physiology of aging is briefly presented, the main stages in facial development are outlined and the distinctions between male and female facial features are highlighted. This is to provide an anthropological justification and context for the assumptions and procedures of the novel aging approaches outlined in later chapters. Artistic techniques for age-progression are subsequently discussed to demonstrate the most dominant and popular approaches currently in use. Finally, the more rigorous and scientific developments in this field are introduced, beginning with the initial work on cardioidal strain [9, 10] and concluding by examining the work of Lanitis et al. [11].

1.4 The Physiology of Growth & Development

According to accepted models in anthropology [12], the human growth and development process can effectively be summarised by six stages:

- **Infancy** — ∼0-3 years [12, 13]. Before the age of about three years, all children have the same basic rounded face shape. The face appears relatively short and wide due to the large brain, which is developmentally precocious compared to the face[3]. The face of the infant grows out from under the brain and facial structures grow proportionally for a longer length of time the further they are located from the brain case (for example, the mandible grows very slowly compared to the eyes, which, like the brain, are developmentally precocious and appear large and widely-set on the infant face). The growth of the facial features during the period of infancy is determined by other areas of physiological development. For instance, the infant has a short nose (with a low nasal bridge) because of the diminutive lung size. The nasal region is a keystone of facial architecture, a key part upon which other surrounding parts are dependent for placement and stability [14]. Since the nose is small, the forehead is bulbous and upright and the cheekbones are prominent. The mandible is small and underdeveloped relative to the rest of the face and the ramus is short due to the slow, transient development of the masticatory system. The emergence of the teeth has a large effect on the face as the infancy stage progresses, leading to the development of the chin (which appears incompletely formed in young infants [14, 15]).

- **Childhood** — ∼3-7 years, characterised by a constant growth rate. By the age of 7, a child's face has progressed into one of the three types — euryprosopic

---

[3]The growth of the cranium is more advanced from infancy to mid-childhood and is closely related to the development of the nervous system. Facial growth is, in contrast, independent of brain case growth and is related to the development of the pulmonary and masticatory systems.

(round), leptoprosopic (long) or dinaric (combination) [14]. During this period, as the body continues to grow at a constant rate, the lung size increases and therefore the nose increases in size. Accordingly, the forehead becomes gradually more sloping as age increases. In addition, the chin develops further and the jaw begins to catch up in size.

- **Juvenile** — $\sim$7-10 years. During this stage, the rate of growth declines to the lowest level since birth (apart from adulthood). As the teeth begin to be replaced by the emergence of the adult dentition, the chin develops further, the ramus of the mandible becomes enlarged, the masticatory muscles expand and the gonial angles become more flared. These factors lead to the whole lower face taking on a more U-shaped appearance [16].

- **Adolescence** — $\sim$10-20 years, where there is an acceleration in the growth of almost all skeletal tissue. The cranium, which grows very slowly after early childhood, experiences an acceleration in the growth of both head length and head breadth. Full maturity in head length, breadth and circumference occurs between approximately 10 and 14 years of age in females and between 13 and 15 years of age in males, [13]. Final head height in both sexes reaches adult size at about 13 years of age, [17]. In the face, the growth of most dimensions accelerates to reach maximum velocity a few months after the peak velocity in stature. Hence, by considering the chronology and velocity of the changes in stature, analogies may be drawn to the rate at which the face is also developing. Craniofacially, throughout adolescence, the greatest change that occurs is in the mandible, which has until this stage, lagged behind the rest of the face in its development. It is hence growing faster than any other facial component and this may explain why it responds the most at adolescence. In this particular facial component, approximately 25% of the total growth in the height of the

ramus occurs between the ages of 12 and 20 years. The mandible ramus has the largest growth spurt of all the facial measurements but the body of the mandible also accelerates both in antero-posterior length (from front to back of the skull) and in depth from the lower incisor teeth to the point of the chin. Therefore, as a result of the spurt, the jaw becomes considerably longer in relation to the rest of the front part of the face. It also becomes thicker and more projecting in appearance [17]. The growth of the maxilla, directly above the upper incisor teeth, also increases slightly in its forward motion so that, in total, the prognathism (projection of the jaws, a jutting forward of the facial skeleton and jaws) of both the upper and lower jaw increases, with the lower jaw more affected than the upper. In addition to these changes, the profile of the face becomes straighter with the chin more pointed and the incisors of both jaws more upright. There is also acceleration in nose growth during the adolescent growth spurt for the majority of children, particularly in the antero-posterior direction. This is in contrast to the growth of the cranial base (base of the skull) where only 6-7% of the total growth occurs during this adolescent period, [18, 19].

The maturation of such facial components is considerably greater in males than in females after the age of approximately 13 years [14, 20] (female facial development begins to slow markedly after about 13 years of age), indicating sexual dimorphism in the face as adolescence proceeds [14, 21]. The width of the cranial cavity increases more in boys than in girls between the ages of 9 and 14 years and they also have a greater increase in soft tissue width [18]. For instance, in males, the soft tissues of the nose as well as the nasal bone also experience a growth spurt, causing the point of the nose to move further forwards and downwards in relation to the rest of the face. This effect

varies greatly on a individual basis however. In association with the skeletal changes that occur in adolescent males, there is also a subtler change in musculature and subcutaneous tissue distribution. Males develop less subcutaneous fat than females, giving the male a more angular and leaner face than its female counterpart. The development of subcutaneous fat in females in contrast causes a rounding and softening of the contours of the face and body [22]. These changes may be caused by the effects of hormones on the facial tissue [19]. Figure 1.1 [19] gives an example of the change in masculine facial appearance during adolescence.

a)   b)   c)   d)
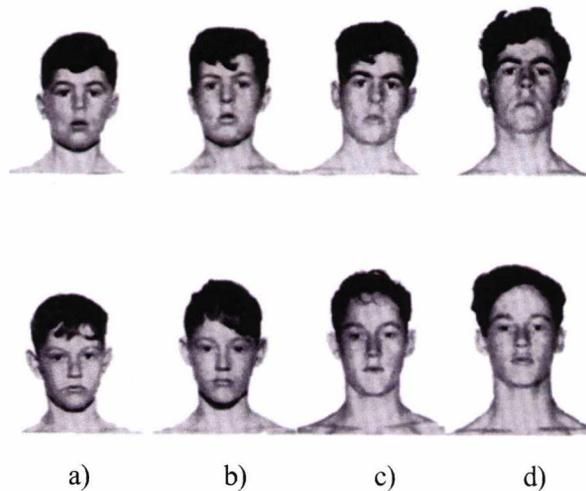
Figure 1.1: **Changes in male facial appearance during adolescence**. The mandible becomes longer, thicker and more projecting and there is an increase in nose growth. The face also becomes leaner and more angular. Upper subject: a) Age 11 years b) Age 12.5 years, c) Age 13.25 years, d) Age 14.5 years, Lower subject: a) Age 12 years b) Age 13.75 years, c) Age 14.75 years, d) Age 15.5 years. Image courtesy of [19].

- **Adult & Old Age** — ~20-70 years. The adult stage represents a period of homoeostasis with old age characterised by physiological decline [21]. As aging progresses, creases may appear across the forehead [5], the eyelids may

droop, lines may appear around the eyes and mouth [14, 21] and the lips may become thinner [23, 24]. Lifestyle effects such as UV exposure, excessive alcohol consumption and smoking all serve to accelerate these processes. The process of aging in adults is, however, highly individual in nature and none of the aforementioned age-related changes are definite.

Figure 1.2 [3] shows an artist's impression of aging from infancy to old age.
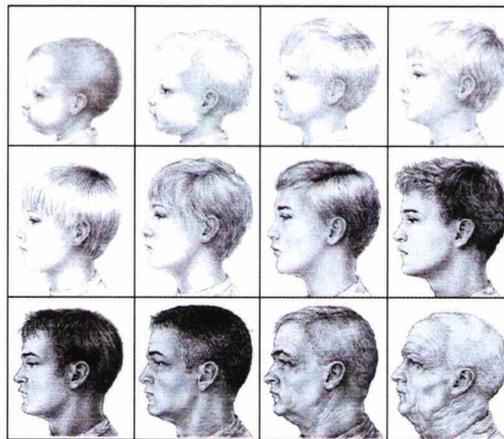
Figure 1.2: **An artistic impression of aging from infancy to old age**. There is rapid development in infancy followed by an approximately constant growth rate in early childhood. The juvenile period between 7 and 10 years is characterised by a decline in the growth rate and this is followed by adolescence, where there is an acceleration in the growth of almost all skeletal tissue. Adulthood represents a period of homoeostasis and, finally, old age is characterised by a drooping of the eyelids, wrinkling around the eyes and mouth and a thinning of the lips. Image courtesy of [3].

1.5   Comparing Male and Female Facial Features

The main differences between the facial features of a male and those of a female are the following:

The male body is overall of larger dimension than that of the female. As a result of this, there is a difference in lung size between the sexes and this means that males and females differ in the size and configuration of their noses. Consequently, there are differences in other structures of the face since the airway is a fundamentally key aspect of facial development, the nasal part of the face strongly influencing an individual's facial form. Owing to the larger lung size, the male usually has a proportionally larger nose than his female counterpart. Additionally, the male nose also tends to be more protrusive, longer and wider. This leads to the male having a more protrusive, sloping forehead and pronounced eyebrow ridges with the female correspondingly characterised by a bulbous and upright forehead. Additionally, the male appears to have more deep-set eyes with the female eyes appearing more forward on the face. The male mandible is also larger than that of the female and the angle of the jaw is more acute. The amalgamation of all of the facial features render the female face flatter and proportionally wider in appearance than the male face. In addition, the female face is usually smaller overall than the male face, with a weaker mandible [14][4]. These physiological distinctions between the sexes must be appropriately treated by any successful age-progression algorithm.

---

[4]It is, however, possible for a male face to appear more round and hence embody a more feminine appearance. Likewise, a female face may take on more masculine qualities.

1.6 Artistic Techniques for Forensic Aging of the Face

Computerised age enhancement was first carried out in the 1980s by artist Nancy Burson, [25, 26]. Her pioneering approach was based upon knowledge of normal, structural bone development with age. To achieve age-progression by a given number of years, these average, predictable changes were appropriately applied to the facial features. The soft tissue was subsequently aged by adjusting the pixel values to reflect the well-documented, average alterations that occur in the skin over time. This technique led to the production of the first age-progressed images of missing children. Around this time, medical illustrators Scott Barrows and Lewis Sadler used craniofacial growth data to produce age-progressed images [5]. Their work emphasised the use of quantifiable growth data for people of different age ranges and ancestral backgrounds. Using more than forty anatomical landmarks on the face, they reviewed the photographs of thousands of children and developed a database of facial measurements. They then generalised from this sample to predict the likely changes in facial measurements for a subject child to be aged. Later studies involved digitised images, whereby the database of training examples was queried to discover the change in facial measurements between the current age of a subject and a target age.

Building on this early work, artists at the National Center for Missing and Exploited Children (NCMEC) in the United States currently perform age-progression by combining artistic skill with the use of computer software packages [6]. In their approach, incorporation of the hereditary traits of aging is considered to be central and population influences are only employed in the absence of familial data. A current photograph of the subject child is required, along with photographs of the

biological parents or older siblings at approximately the same age as the target for the progression. The goal is to age the child's face in accordance with the tendency for an individual to resemble the appearance of immediate family members. To simulate normal craniofacial growth, the face of the child is appropriately "stretched" using the familial images to estimate the extent required[5] — that is, the overall size of the face is increased and the size and shape of the features are adjusted. Subsequently, the intensity values of equivalent pixels from the child and parental images are averaged, in essence permitting the artist to "borrow" the more mature features of the reference face to complete the age-progression of the child. A limitation of this approach is that sufficiently high quality familial images are often unavailable. To complete the final, age-progressed image, many subtle changes must often be made artistically, for example, naturally formed creases or wrinkles must be preserved to retain the individual's unique appearance. Figure 1.3 shows an example of age-progression performed by artists at the NCMEC.

a)   b)   c)
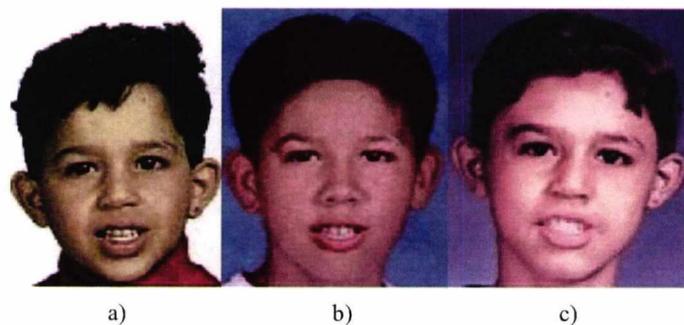
Figure 1.3: **Artistic age-progressions performed by NCMEC specialists** — a) Original subject, 6 years old, b) Subject artistically age-progressed to 11 years old, c) Subject artistically age-progressed to 14 years old. Images courtesy of http://www.lynnpolice.org/jesus_delacruz.htm.

In conjunction with the NCMEC, the FACES Laboratory at Louisiana State

---

[5]If no family pictures are available, suitable reference images are used.

University also perform age-progressions, [4]. A similar technique is used to that at the NCMEC, with parental and sibling images playing a vital role in addition to information about medical conditions that could affect the subject's appearance. The technique also explicitly incorporates quantifiable growth data to predict the structural changes typical of specific developmental periods [27] and the face is reconstructed to reflect this growth. The resultant image is then "merged" with the photograph of the relative — a weighted average is calculated using the intensity values of equivalent pixels to experiment subjectively with different combinations of features. Good quality, full frontal facial photographs are a pre-requisite for the procedure and each age-progression may take up to seven days to complete. The technique used is considered to be effective, since it is claimed that the resultant age-progressed images play an important role in the recovery of one in every seven children reported missing to the NCMEC [4].

The National Missing Persons Helpline (NMPH) in the UK has staff trained specifically for child age-progressions[6] [2, 28]. A similar technique is effected to that of the NCMEC with additional knowledge of the underlying bone structure and dentition used to augment the approach. An example of a child age-progression performed by the NMPH is given in figure 1.4.

The NMPH also performs age-progression on adult faces. Since the age-related changes in an adult are less structural and more dependent on environmental and lifestyle factors, the age-progression involves more uncertainty with respect to factors such as skin texture, hairstyle and weight. Hence, to effect a realistic age-progression, it is useful for an artist to have access to information about the subject's lifestyle habits, such as smoking or alcohol consumption [27] in addition to photographs

---

[6]To warrant an age-progression, the child must be at least two years old at the time of disappearance and must have been missing for a period of at least two years.

Figure 1.4: **Artistic age-progressions of children performed by NMPH specialists** — a) Original subject, 3 years old, b) Subject artistically age-progressed to 7 years old, c) Original subject, 2 years old, d) Subject artistically age-progressed to 7 years old. Images courtesy of http://www.missingpersons.org/unidentified_age.asp.

of family members at approximately the target age. An example of adult age-progression is shown in figure 1.5.

Figure 1.5: **An artistic age-progression of an adult performed by NMPH specialists** — a) Original subject, 18 years old, b) Subject artistically age-progressed to 34 years old. Images courtesy of http://www.missingpersons.org/unidentified_age.asp.

In addition to the work of agencies, freelance forensic artists also perform age-progression. They apply knowledge of structural bone development as well as information from parental images to give an estimate of current appearance, [5, 29, 30]. Figure 1.6 [5] depicts an example of such a child age-progression performed by artist Karen T. Taylor. Some artists also perform age-progression and regression for more

commercial applications such as family portraits [31].
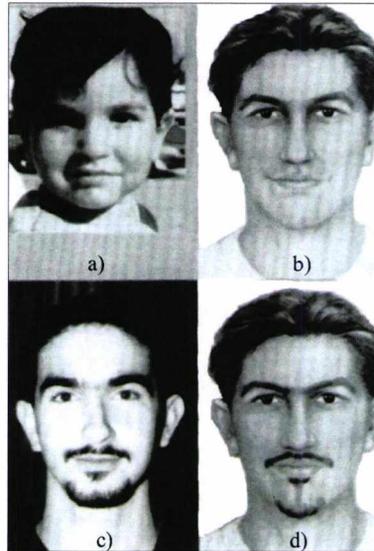
Figure 1.6: **An example of age progression by freelance artist, Karen T. Taylor** [5] — a) Original subject at 2.5 years of age, b) Subject progressed to 18 years of age using knowledge of structural bone development in addition to artistic skills, c) Actual appearance of the subject when located at age 18, d) Age-progressed image equivalent to image b) but with facial hair added. Images courtesy of [5].

Knowledge of photogrammetry and anthropometry is also used by some experts to perform age progression. Police artist Bette Clark [32] and surgeon Leslie G. Farkas [13] employ the following method. Firstly, anthropometric measurements on the image of the subject face are used to enlarge it to life-size to obtain objective facial data at the time of disappearance. Particular anthropometric measurements are extracted from the enlarged image and compared to the equivalent measurements from the current age population norms (calculated from a representative sample of children of the subject's age, ethnicity and sex). To age the face, the measurements are adjusted to reflect an equivalent position relative to the population norms for the target age. The proportions of the face are suitably retained to preserve subject

identity — however, if necessary, the measurements may be modified such that the face incorporates physical characteristics associated with those of family members. To complete the facial aging, artistic interpretation is required for the soft tissues of the skin and features. Figure 1.7 [13, 32] depicts an example of an age progression performed in this fashion.
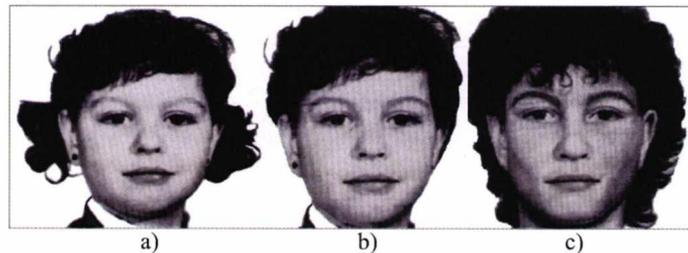
Figure 1.7: **An example of age progression by Police artist, Bette Clark** [13, 32] — a) Original subject at 5 years of age, b) Subject progressed to 8 years of age using knowledge of photogrammetry and anthropometry, c) Subject progressed to 18 years of age. Images courtesy of [13, 32].

The subjective approaches outlined here generally give visually pleasing and realistic results. However, the need for dedicated software to perform both age-progression and regression is widely accepted by forensic artists, [2, 33]. The high demand from families coupled with the length of time needed to perform each progression or regression means that the industry would benefit immensely from such a tool. Commercial systems already available include the APRIL software [34], which utilises learned information from a large training set to age an individual in accordance with physiology and lifestyle habits. Although the motivation for its use is educational, potential forensic applications also exist. However, to date, these have not been implemented. The FACE age software, developed by Vision Control International, has attempted to provide another solution [35]. To develop the software, information about typical skull growth in children was used to calculate a

normal pattern of growth[7]. Firstly, stable "landmarks" on the face were identified. These landmarks refer to positions on the face that remain approximately constant throughout the growth process, for example, landmarks around the orbital regions. Changes in skull growth were then computed relative to these stable positions over time. A skull growth algorithm was then produced to stretch the face of the subject according to these landmarks to produce the anatomically correct facial shape, on a sliding scale between 1 and 18 years, down to an increment of months. At the current time, however, the software is unable to automatically process soft tissue growth. Hence, the soft tissue information must be modified artistically using a graphics package to approximate growth and development in these features. An example of an age-progression performed using the FACE age software is given in figure 1.8 [35].

Figure 1.8: **An example of age progression using the FACE age software** [35]. The original image (left) depicts a subject who was 8 years old at the time of her criminal abduction. The other two images show her age-progressed appearance at age 21, with two different hair styles. Figure courtesy of [35].

---

[7]This information was obtained from dental records and longitudinal photographic studies of male and female Caucasian subjects. Using the acquired data, the information was divided into the general facial shapes of round, medium and long such that normal growth could be modelled for different facial types.

There is hence still a need for software that can age both the shape and texture of a face in an objective and rigorous manner. As observed by Feik and Glover [36]:

> *"At present, prediction of the way in which a particular child's face will age still relies largely on subjective artistic impressions. However, as our anatomical knowledge and understanding improve, the use of scientific and computer modelling techniques should lead to improved results in predicting the growth of faces."*

1.7 Robust Techniques for Facial Aging

In addition to the approaches to age-progression discussed in section 1.6, several attempts have also been made to address the problem using more robust, scientific means. In this section, the basis for several of the more important techniques is outlined.

1.7.1 Geometric Transformations - Cardioidal Strain

Some of the earliest work on the rigorous modelling of growth was pioneered by D'Arcy Thompson. He observed that a condition of *strain*, a result of stress on a body, is a direct stimulus to growth itself (Wolff's Law) [37]. In developing Thompson's ideas, the first attempts to model 2D craniofacial growth using geometric transformations were performed by Pittenger and Shaw [10, 38]. The most successful transform for modelling growth was found to be *cardioidal strain* [9, 39] and is given in polar coordinates by:

$$\theta' = \theta$$
$$R' = R(1 - k\cos\theta) \tag{1.1}$$

where θ' and R' are the post-transform coordinates, θ and R are the pre-transform coordinates and k is a free parameter that may be adjusted as required (see figure 1.9).

Subsequently, Todd, Mark, Shaw and Pittenger [40] sought to further explore the notion of Wolff's Law and the use of cardioidal strain for modelling facial growth. To this end, studies were performed to investigate the effects of gravity on a growing

head [9]. In this hydrostatic analysis, the head was treated as a fluid-filled, spherical water tank and was assumed to grow in accordance with the exerted pressure, the analysis allowing the so-called *"revised" cardioidal strain transformation* to be derived. This may be written as:

$$\theta' = \theta$$

$$R' = R(1 + k(1 - \cos\theta)) \tag{1.2}$$

where the variables have the same meanings as in equation (1.1). Figure 1.9 [17] shows the effect of applying varying amounts of the revised cardioidal strain transformation to the profile of an infant. The transformed profiles appear older as successive amounts of the transform are applied.
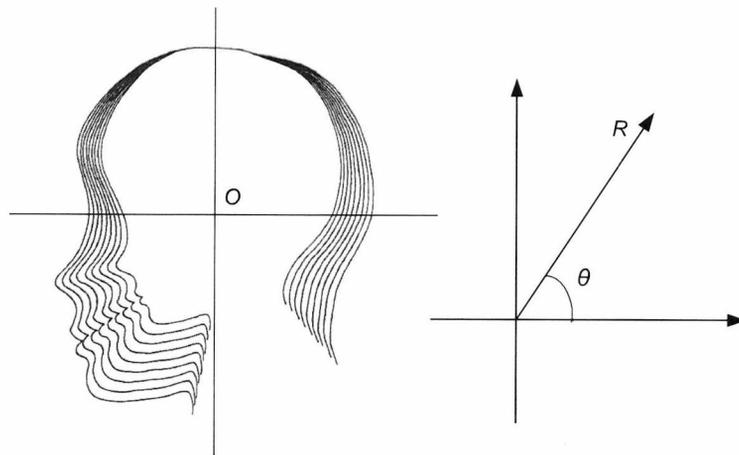
Figure 1.9: **The successive application of the revised cardioidal strain transformation to the profile of a child.** The transformation is initially applied to the profile of an infant (innermost profile). Successive application of the transformation produces profiles which appear older until the outermost profile has the appearance of an adult. The origin for the transformations is depicted by O. The figure has been adapted from [17].

To evaluate its use further, Pittenger and Shaw [10] performed comparison experiments between cardioidal strain and an alternative transformation called affine shear[8]. Applied to facial profiles, they found that cardioidal strain had a more powerful and consistent effect than affine shear on estimates of age. This finding was reinforced by Mark and Todd [39] who discovered that observers were more sensitive to small growth differences produced by cardioidal strain than to comparable differences produced by affine shear when applied to 2D profiles. Mark et al. [41] subsequently found that simulated facial growth by cardioidal strain was judged as more similar to actual growth, compared to other transformations. Hence, cardioidal strain appears to provide a naturally salient depiction of growth. However, exact comparisons cannot be made as certain craniofacial changes are not captured by growth transformations[9].

Later studies performed on digital images and three-dimensional busts [42] also demonstrated the ability of the cardioidal strain transformation to model growth and aging on more complete representations of the human face. However, in similar work using 3D faces, Bruce et al. [43] found that many observers could not accurately judge the relative ages of the original and transformed faces. This implies that cardioidal strain may not be the most appropriate transformation to embody facial aging. In addition, Lanitis et al. [11] compared cardioidal strain with their derived aging algorithm. They found that the transformed shapes were more similar to the target age prototypes for the aging algorithm than for cardioidal strain. Fi-

---

[8]Affine shear may be described in Cartesian coordinates by:

$$Y' = Y$$
$$X' = X + Y \tan \theta$$

where Y' and X' represent the coordinates after the transform has been applied, Y and X represent the coordinates prior to the transform application and θ is the angle of rotation.

[9]For instance, the development of the frontal sinus (a bump above the bridge of the nose that enlarges considerably after puberty) and changes in the size and shape of the nose.

nally, Ramanathan and Chellappa [44] recently demonstrated that cardioidal strain provides a perceptually acceptable prediction of facial growth for small age changes but a poor prediction for larger changes.

In summary, cardioidal strain as a means of modelling growth has never been subjected to quantitative assessment but has only been evaluated in a perceptual sense. Furthermore, the transformation is restricted to describing growth in shape only. As soft tissue changes cannot be captured, cardioidal strain is necessarily limited as a method to accurately model age-progression.

1.7.2   Aging using Facial Composites (Burt & Perrett, 1995)

To further the advancements in systematic facial aging begun by the work on cardioidal strain, Burt and Perrett [45] employed facial composites to investigate other visual cues to age, such as textural changes in the skin. Using an appropriate training sample, facial shape was modelled by a set of landmark points placed around the facial outline and features (see section 2.3). To model the texture, pixel intensities were extracted from each face subsequent to the adjustment of the shape to a consistent configuration (see section 2.3.2). Two facial composites (prototypes) were then produced in both shape and texture to investigate the notion that changes in facial appearance with age can largely be characterised by the average differences between young and old faces. These prototypes were calculated as follows:

$$\bar{\mathbf{s}}_{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{s}_{\mathbf{y}i}$$

$$\bar{\mathbf{t}}_{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{t}_{\mathbf{y}i} \tag{1.3}$$

where $\bar{\mathbf{s}}_\mathbf{y}$ and $\bar{\mathbf{t}}_\mathbf{y}$ represent the shape and texture prototypes respectively for the young age group (25-29 years), $\mathbf{s}_{\mathbf{y}_i}$ and $\mathbf{t}_{\mathbf{y}_i}$ represent the ith shape and texture vectors respectively of the young group and N is the number of young examples. Similarly,

$$\bar{\mathbf{s}}_\mathbf{o} = \frac{1}{M} \sum_{i=1}^{M} \mathbf{s}_{\mathbf{o}i}$$

$$\bar{\mathbf{t}}_\mathbf{o} = \frac{1}{M} \sum_{i=1}^{M} \mathbf{t}_{\mathbf{o}i} \tag{1.4}$$

where $\bar{\mathbf{s}}_\mathbf{o}$ and $\bar{\mathbf{t}}_\mathbf{o}$ are the equivalent prototypes for the older age group (50-54 years) and M is the number of old examples. To calculate the shape and textural changes with age, the difference in coordinate values was computed between corresponding points on the old and young prototypes and added to the appropriate shape or texture vector of subject faces. That is:

$$\mathbf{s}_{\mathbf{y}_j}{}' = \mathbf{s}_{\mathbf{y}_j} + (\bar{\mathbf{s}}_\mathbf{o} - \bar{\mathbf{s}}_\mathbf{y})$$

$$\mathbf{t}_{\mathbf{y}_j}{}' = \mathbf{t}_{\mathbf{y}_j} + (\bar{\mathbf{t}}_\mathbf{o} - \bar{\mathbf{t}}_\mathbf{y}) \tag{1.5}$$

where the left hand side represents the aged shape and texture of the jth young subject. The age-progression was performed by adding the vector difference between the older and younger prototype faces to the subject face. Figure 1.10 [45] depicts these transforms in shape and texture (displayed in grey-scale, as in [45]) on an example face.

In addition to this work, Burt and Perrett also explored the effect of caricaturing on age perception. A texture prototype was produced for a population of example faces between 20 and 54 years. This was then mapped (see section 2.3.2)
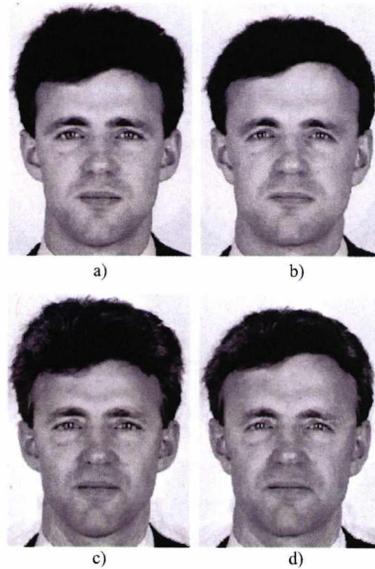
Figure 1.10: **Age transformation of facial shape and texture (Burt and Perrett, 1995)** — a) Original subject, b) Subject transformed using shape difference calculated from the 25-29 years and the 50-54 years prototype, c) Subject transformed using colour (RGB) difference calculated from the 25-29 years and the 50-54 years prototype, d) Subject transformed according to both shape and colour calculations. Images courtesy of [45].

into the shape of the 50-54 prototype and the difference in RGB values between corresponding pixels calculated. This may be expressed as follows:

$$\delta = \overline{\mathbf{t}}_{\mathbf{o}} - \overline{\mathbf{t}}_{\mathbf{b}} \qquad (1.6)$$

where $\overline{\mathbf{t}}_{\mathbf{b}}$ is the texture prototype of example faces between 20 and 54 years, mapped to shape $\overline{\mathbf{s}}_{\mathbf{o}}$. Subsequently, to produce a caricature of $\overline{\mathbf{t}}_{\mathbf{o}}$, the following equation was used:

$$\overline{\mathbf{t}}_{\mathbf{o}}{}' = \overline{\mathbf{t}}_{\mathbf{o}} + 2\delta \qquad (1.7)$$

where $\overline{\mathbf{t}}_{\mathbf{o}}{}'$ represents the caricature. The result of this procedure is given in figure

1.11 [45]. It was found that observers perceived the caricatured image as older than the original image. It was also discovered that transforming both the shape and the texture of subject faces, as described initially, significantly increased the perceived age compared to shape and texture transformations alone. It was concluded that texture contained considerable information about the age of faces.

a) b)
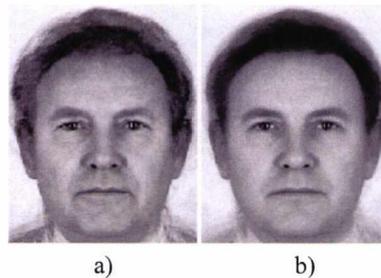
Figure 1.11: **Image caricaturing in texture to investigate cues to facial aging** — a) Result of image caricaturing - the RGB difference between the population prototype and the 50-54 years prototype is exaggerated and added onto the colour values of the latter, b) The 50-54 years prototype. The result is that face a) is perceived as older than face b). Images courtesy of [45].

Burt and Perrett's approach has several limitations. Firstly, they used only full-frontal, neutral expression, high resolution images in their analysis. The modelling is therefore not robust to any perturbations in head orientation or expression. Additionally, this technique is only effective if a sufficient number of images can be used to form well-defined prototypes. Finally, this approach uses simply the average changes in a population to effect aging. To more accurately model age-progression, individual influences such as those of genetics or lifestyle are required.

1.7.3 Aging using Caricatures (O'Toole et al., 1997)

O'Toole et al [46] also performed work on caricaturing the human face, applying a standard caricaturing algorithm to 3D laser scans of human heads. They found

that the process of caricaturing seemed to produce an increase in the apparent age of the individual. Results of the caricaturing and its effects are displayed in figure 1.12 [46]. Further work by O'Toole et al. [47] on shape-normalised and texture-normalised faces[10] revealed that these altered faces appeared younger than their original counterparts, indicating that psychologically meaningful attributes such as age can be successfully modelled in the appropriate face space.
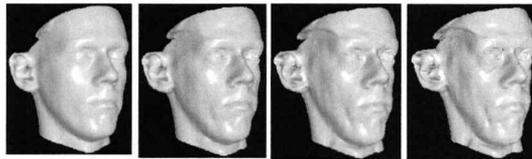
Figure 1.12: **Caricaturing of 3D laser scans of human heads to investigate perceptual cues to aging**. A 3D laser scan of a 27 year old male face (far left) is shown, with caricaturing at increased distances from the mean. The faces are perceived as increasing in age. Figure courtesy of [46].

The main limitation of the work by O'Toole et al. is that the aging effect discovered was incidental upon application of the caricaturing algorithm. There is no guarantee that the age-progressions are accurate or reliable and specific amounts of aging in years cannot be explicitly applied.

1.7.4 Estimating Growth Trajectories (Hutton et al., 2003)

Hutton et al. [48] have investigated computerised aging from a medical perspective. Their rationale was to produce a model of normal facial growth for the study of syndromes or pathologies that cause abnormal growth. Using three-dimensional scans of human faces, an aging trajectory for the average individual was determined through a high-dimensional shape-space, calculated from dense surface point distri-

---

[10]A shape-normalised face refers to a morphing of the texture maps from individual faces onto the average head shape. Conversely, a texture-normalised face is created by morphing the average texture onto the shape of each individual face.

bution models [49]. To calculate the average age trajectories (with male and female faces treated separately), kernel smoothing[11] was used to compute an average face for any given age. The aging trajectory through the space was then defined to be the path that joins these averages. A face may hence be synthesized at any age along this trajectory, as shown in figure 1.13 [48].
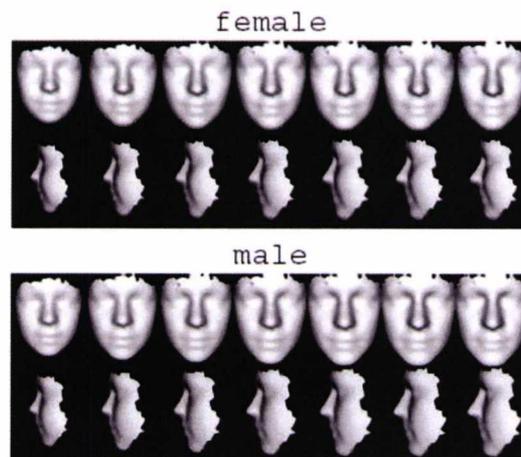
Figure 1.13: **Male and female faces synthesized at different ages along the aging trajectories in shape-space.** The aging trajectory for each gender was calculated (from dense surface point distribution models) as the path that joins the average face at different ages in the shape model space. Male and female faces may then be synthesized at ages 10, 15, 20, 25, 30, 35 and 40. Images courtesy of [48].

Hutton et al. also plotted the male and female aging trajectories in two dimensions. They found that, as age increases, the trajectories diverge, correlating well with accepted patterns of sexual dimorphism during growth (section 1.5). Additionally, they observed that the trajectories appeared approximately linear until about age 15, with changes in direction occurring after this point, demonstrating that over a lifetime, the change in the shape of the face is distinctly non-linear (see

---

[11]This technique has the effect of interpolating between examples and also averaging out unwanted variation between individuals.

also [12, 13, 14, 19, 17, 18]).

The main limitation with Hutton et al.'s approach is that only shape is considered in the aging procedure. As shown by Burt and Perrett [45], the modelling of textural changes is also desirable for the perception of age-progression.

1.7.5 Semi-automatic Facial Aging (Lanitis et al., 2002)

In perhaps the most rigorous approach to age-progression to date, Lanitis et al. [11] used a statistical model trained on appropriate images to simulate facial aging. The relationship between the model-based representation of an individual and its age (the "aging function") was calculated, allowing estimates of age to be made. The aim was to define a path within the model space, which explains most of the variation in age. Four different aging functions were tested in order to predict the age of a previously unseen individual. The so-called "weighted person-specific" aging function, which incorporated lifestyle factors as well as appearance (shape and texture) was observed to provide the most accurate prediction of age, suggesting that it may be advisable to include individual lifestyle factors when defining an aging trajectory through a model space.

To simulate the effect of aging on a face, the age of the subject was first estimated using the chosen aging function. The model-based description of the subject was then adjusted to the required target age. Figure 1.14 [11] shows example results of age-progression using this technique.

It was found that:

1. Observers correctly identified whether the age-progressed or regressed image was older or younger than the original subject.
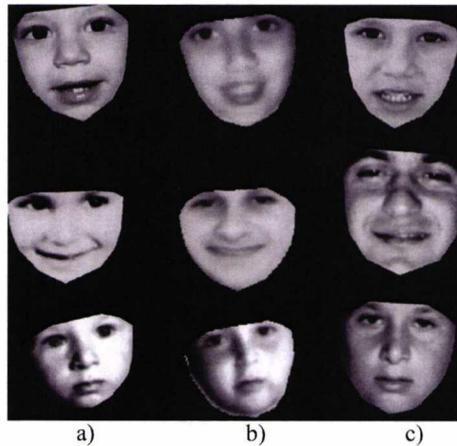
Figure 1.14: **Examples of age-progression (Lanitis et al., 2002)**. To simulate the effect of aging on a face, the age of the subject was first estimated using the chosen aging function. The model-based description of the subject was then adjusted to the required target age — a) Original image, b) Age-progressed image, c) Original image at the target age. Figure courtesy of [11].

2. A majority of observers accurately confirmed that the age-progressed or re-gressed image was of the same identity as the original subject.

3. When using the technique during face recognition tasks, the results showed improved classification rates for face recognition as compared to the case where age-progression is not performed.

The limitations of the approach taken by Lanitis et al. are as follows. Firstly, the model was constructed upon faces in unconstrained poses and expressions. That is, identity modes were not separated from those of accidental variation. Hence, dur-ing the age-progression, an uncontrolled variation may occur in the orientation or expression and comparison with the target face may prove problematic. Secondly, although lifestyle effects are incorporated, no attempt was made to model the sta-tistical relationship between the appearance of family members (such as parents

and siblings). More realistic age-progressed results could be obtained through the modelling of such influences.

1.8   Chapter Summary

In this chapter, a discussion of the physiological stages of growth and development was made, with particular emphasis placed on the facial changes which occur during childhood and adolescence. Artistic and anthropological approaches to age-progression were then outlined to examine the techniques in current use by missing persons bureaux [2, 4, 6, 28] and freelance artists [5, 29, 30, 31]. Although visually plausible results are obtained, the subjective, unverified nature of the techniques and the length of time required to complete each age-progression mean that there is a recognised need for semi-automatic aging software [36].

Scientific attempts at age-progression have also been presented. Despite the advances made by these approaches over subjective methods, several limitations do exist:

- The techniques are not robust to changes in head orientation or expression. Uncontrolled variations may occur if attempts are made to age a face at unconstrained pose or expression, making target comparisons difficult.

- Some approaches model shape only (cardioidal strain, Hutton et al.) and therefore do not provide a complete modelling of aging.

- None of the approaches attempt to model effects due to heredity or the previous appearance of an individual. These are important factors to consider in age-progression.

The primary aim of this thesis is an attempt to address these problems and develop a more scientifically rigorous approach. The theoretical basis of this method, presented in chapter 3, attempts to incorporate the changes in both shape and texture which are associated with aging. Learned statistical relationships originating from the physiology of growth, the genetic tendency of individual facial appearance to resemble that of parents and historical facial development trends are all incorporated within the framework of the model. First, however, the mathematical structure that underpins this approach needs to be discussed and this is presented in the next chapter.

Chapter 2

Mathematical Modelling of the Human Face

In this chapter, the mathematical basis for the age-progression techniques developed in this thesis are presented. In particular, the statistical technique of principal component analysis (PCA) is described in detail since it is used in the construction of the facial models in shape and texture and also permits a compact parametric representation of a face. The central role of Singular Value Decomposition (SVD) in the derivation of sample principal components is also discussed and the use of PCA in shape modelling demonstrated by applying the technique to an example set of images. The concept of texture is then introduced, such that facial aging may be modelled more completely, and a texture model is constructed using the example training set. The technique of Procrustes alignment and the methodology of image "warping" are also discussed.

2.1   Literature Review

Attempts to accurately model the human face and its constituent features using deformable models have long been extant in the field of Computer Vision. Specifically, PCA has been extensively discussed in the literature since it allows the formation of a model that permits only plausible, "legal" facial examples to be produced from a training set. Early work by Sirovich and Kirby [50, 51] described the intensity modelling of grey-scale facial images, showing that a face may be represented in terms of a low-dimensional coordinate system termed "eigenpictures" — these are the eigenfunctions of the covariance matrix of the ensemble of faces. Faces were successfully represented in terms of these basis vectors but no attempt was made to model facial shape. Since only rudimentary procedures were used to align the faces with respect to the eyes, the resulting images represented in terms of eigenpictures exhibited blurring since the eigenpictures necessarily averaged over the varying feature positions. Turk and Pentland [52, 53] performed a similar procedure on facial intensity vectors, compactly representing a training sample in terms of "eigenfaces". These eigenfaces were used in the classification of facial images as well as in the detection of faces within an image. Recognition experiments revealed that this approach was robust to changes in illumination but, by inference, was not robust to changes in pose or expression. To solve the registration problem, Craw and Cameron [54] suggested transforming the shape of each face in a training set to a reference shape prior to texture extraction and the application of PCA. They demonstrated that this "shape free" transformation improved recognition rates.

For the modelling of shape, Cootes et al. [55, 56] used PCA to describe the modes of shape variation in a training set of 2D heart images. In further work, Cootes et al. [57, 58, 59, 60, 61] also introduced Active Appearance Models (AAM), in order to combine shape and texture information. Using PCA, statistical models were built on

the shape and grey-level values of a set of training faces and the correlations between shape and texture learned to produce a combined model of appearance. Novel, photo-realistic faces were produced by adjusting the model parameters, such that the synthesized face appeared as close as possible to a given target. The AAM searches for faces in previously unseen images by using the difference between the current synthesized image and the target image to update its model parameters. The ability of AAMs to produce new, plausible examples of a training set has led to novel work on image segmentation [62] as well as to the development of emerging technologies for the generation of near photo-quality facial composites [63, 64, 65, 66, 67].

The work in this thesis is most closely related to that of AAMs. However, for reasons explained in chapter 3, shape and texture are modelled separately. Furthermore, the work herein involves facial *synthesis* as distinct from the most common applications of PCA for face recognition and image searching. The derivation and explicit use of PCA for facial modelling is described in the following sections.

2.2   Principal Component Analysis

There are multiple treatments of principal component analysis in the literature, [68, 69, 70, 71, 72, 73, 74] and the reader is referred to these for a detailed discussion of the technique. This section will outline the derivation of the method in the sample case and describe how it may be applied to digital images.

Linear principal component analysis is a statistical technique, which aims to produce an orthogonal basis from a set of correlated data variables. The new variables constituting the orthogonal basis, termed *principal components*, are related to the original data via a linear relationship and are derived in decreasing order of impor-

tance such that the first principal component accounts for the maximum possible variation in the training set [72]. From a geometrical perspective, PCA may be considered as a rotation to a new set of orthogonal axes whose orientation is such that the sum of squared errors (differences) between the actual data points and their perpendicular projections onto the first axis is minimised[1] (figure 2.1). The first axis then represents the first principal component.
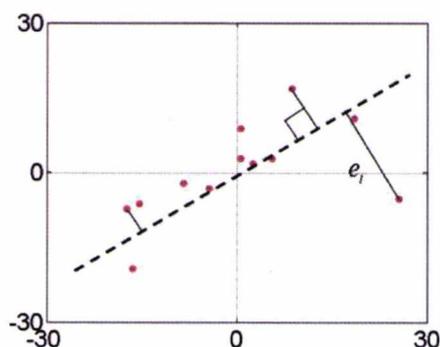
Figure 2.1: **An example of PCA using artificial 2D data to demonstrate the technique from a geometrical perspective**. The principal axis (dashed line) minimises the sum of the squared differences, $\sum_i e_i$, between the data and their orthogonal projections onto that axis. This maximises the variance of the data with respect to the chosen axis.

The technique was first proposed in 1901 by Karl Pearson [75], although it is often attributed to Harold Hotelling [76] who proposed it independently in 1933. The term *Karhunen-Loeve* transform has sometimes been used in the context of PCA since, in one of its most basic forms, it is identical [77]. However, some variations to the method, presented in the pattern recognition literature [78], differ from PCA in that they incorporate class information.

---

[1]An alternative but entirely analogous statement is that the principal components represent a set of orthogonal axes such that the sum of the squared projected lengths of the data variables onto the first axis is maximised.

2.2.1 Preliminary Definitions

Consider a *population* of random variables, $\mathbf{x}$, where $\mathbf{x} = [X_1, \ldots, X_q]^T$. The mean, μ, and covariance matrix, $\boldsymbol{\Sigma}$, of this population are defined as:

$$\mu = E(\mathbf{x}) \tag{2.1}$$

and,

$$\boldsymbol{\Sigma} = E(\{\mathbf{x} - \mu\}\{\mathbf{x} - \mu\}^T) \tag{2.2}$$

respectively, where E denotes the statistical expectation. The elements of $\boldsymbol{\Sigma}$, denoted by σ_{ij}, represent the covariances between the ith and j^{th} random variables. If two components of the data, X_i and X_j, are uncorrelated, their covariance is zero ($\sigma_{ij} = \sigma_{ji} = 0$). For a *sample* of n independent observations of $\mathbf{x}$, $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, the sample mean, $\overline{\mathbf{x}}$, is given by:

$$\overline{\mathbf{x}} = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \tag{2.3}$$

and the sample covariance matrix is given by:

$$\mathbf{S} = \frac{1}{n-1} \sum_{k=1}^{n} (\mathbf{x}_k - \overline{\mathbf{x}})(\mathbf{x}_k - \overline{\mathbf{x}})^T \tag{2.4}$$

where the factor $\left(\frac{1}{n-1}\right)$ is needed to ensure an unbiased estimate of the population covariance.

2.2.2 Derivation of Sample Principal Components

Consider the projection, z_k, of the kth observation, $\mathbf{x}_k$, onto the unit vector, $\mathbf{u}$, described by:

$$z_k = \mathbf{u}^T \mathbf{x}_k \tag{2.5}$$

The aim of PCA is to find the unit vector, such that the projection of all observations onto it will have the maximum possible variance. If $\mathbf{u}$ is determined, subject to the constraint $\mathbf{u}^T \mathbf{u} = 1$, then $\mathbf{u}$ is defined as the first *principal component* of the data. For the projection of all observations onto $\mathbf{u}$, the sample variance may be written as:

$$var(z) = \frac{1}{n-1}\sum_i (z_i - \bar{z})^2 = \frac{1}{n-1}\sum_i \left(\mathbf{u}^T\mathbf{x}_i - \frac{1}{n}\sum_j \mathbf{u}^T\mathbf{x}_j \right)^2 \tag{2.6}$$

where $\bar{z} = \frac{1}{n}\sum_{j=1}^n z_j$. Hence,

$$
\begin{aligned}
var(z) &= \frac{1}{n-1}\sum_i \left(\mathbf{u}^T\mathbf{x}_i - \mathbf{u}^T\frac{1}{n}\sum_j \mathbf{x}_j \right)^2 \\
&= \frac{1}{n-1}\sum_i \left(\mathbf{u}^T\mathbf{x}_i - \mathbf{u}^T\bar{\mathbf{x}} \right)^2
\end{aligned}
\tag{2.7}
$$

Therefore,

$$var(z) = \frac{1}{n-1}\sum_{i=1}^n \left(\mathbf{u}^T(\mathbf{x}_i - \bar{\mathbf{x}}) \right)^2 = \frac{1}{n-1}\sum_{i=1}^n \mathbf{u}^T(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T\mathbf{u} \tag{2.8}$$

Expanding this gives:

$$
\begin{aligned}
var(z) &= \frac{1}{n-1}\left(\mathbf{u}^T(\mathbf{x}_1 - \overline{\mathbf{x}})(\mathbf{x}_1 - \overline{\mathbf{x}})^T\mathbf{u} + \ldots + \mathbf{u}^T(\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^T\mathbf{u}\right) \\
&= \frac{1}{n-1}\mathbf{u}^T\left((\mathbf{x}_1 - \overline{\mathbf{x}})(\mathbf{x}_1 - \overline{\mathbf{x}})^T + \ldots + (\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^T\right)\mathbf{u} \\
&= \frac{1}{n-1}\mathbf{u}^T\mathbf{X}\mathbf{X}^T\mathbf{u} \tag{2.9}
\end{aligned}
$$

where matrix $\mathbf{X}$ contains as its columns the mean-subtracted observation vectors, $\{\mathbf{x}_i - \overline{\mathbf{x}}\}$. Since the sample covariance matrix is given by $\mathbf{S} = \frac{1}{n-1}\mathbf{X}\mathbf{X}^T$, then $var(z) = \mathbf{u}^T\mathbf{S}\mathbf{u}$. In order to find the first principal component, denoted $\mathbf{u}_1$, the optimisation problem of maximising the variance, $var(z)$, is solved via a standard use of Lagrange multipliers. The cost function may be defined by:

$$
Q = \mathbf{u}_1{}^T\mathbf{S}\mathbf{u}_1 - \lambda_1(\mathbf{u}_1{}^T\mathbf{u}_1 - 1) \tag{2.10}
$$

where λ_1 is a Lagrange multiplier (scalar) corresponding to the constraint equation. Differentiating with respect to $\mathbf{u}_1$ and setting to zero gives:

$$
\mathbf{S}\mathbf{u}_1 - \lambda_1\mathbf{u}_1 = \left(\mathbf{S} - \lambda_1\mathbf{I}_q\right)\mathbf{u}_1 = 0 \tag{2.11}
$$

where $\mathbf{I}_q$ is the $(q \times q)$ identity matrix. Multiplying equation (2.11) from the left by $\mathbf{u}_1^T$ and using the constraint equation, $\mathbf{u}_1{}^T\mathbf{u}_1 = 1$, gives:

$$
\mathbf{S}\mathbf{u}_1 = \lambda_1\mathbf{u}_1 \Rightarrow \mathbf{u}_1{}^T\mathbf{S}\mathbf{u}_1 = \lambda_1 \tag{2.12}
$$

This represents an eigenvalue/eigenvector decomposition of covariance matrix, $\mathbf{S}$. To ensure that the quantity $\mathbf{u}_1{}^T\mathbf{S}\mathbf{u}_1$ is a maximum (as required), λ_1 must necessarily be a maximum and represents the largest eigenvalue of $\mathbf{S}$. Correspondingly, $\mathbf{u}_1$ is the eigenvector corresponding to the largest eigenvalue. The second principal

component, $\mathbf{u}_2$, is derived similarly, with the additional constraint that $\mathbf{u}_2{}^T\mathbf{u}_1 = 0$. This is to ensure that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal and hence statistically uncorrelated. The new Lagrange cost function may be given by:

$$Q = \mathbf{u}_2{}^T\mathbf{S}\mathbf{u}_2 - \lambda_2(\mathbf{u}_2{}^T\mathbf{u}_2 - 1) - \phi\mathbf{u}_2{}^T\mathbf{u}_1 \tag{2.13}$$

where λ_2 and ϕ are the Lagrange multipliers corresponding to the two constraint equations respectively. Differentiating with respect to $\mathbf{u}_2$ and setting to zero yields:

$$2\mathbf{S}\mathbf{u}_2 - 2\lambda_2\mathbf{u}_2 - \phi\mathbf{u}_1 = 0 \tag{2.14}$$

Multiplying from the left by $\mathbf{u}_1^T$ yields:

$$\mathbf{u}_1{}^T\mathbf{S}\mathbf{u}_2 - \mathbf{u}_1{}^T\lambda_2\mathbf{u}_2 - \frac{\phi}{2}\mathbf{u}_1{}^T\mathbf{u}_1 = 0 \tag{2.15}$$

Since $\mathbf{u}_1{}^T\mathbf{u}_1 = 1$ and $\mathbf{S}$ is a symmetric matrix, this can be re-arranged as:

$$\phi = (\mathbf{S}\mathbf{u}_1)^T\mathbf{u}_2 - \lambda_2\mathbf{u}_1{}^T\mathbf{u}_2 \tag{2.16}$$

However, since $\mathbf{S}\mathbf{u}_1 = \lambda_1\mathbf{u}_1$ and $\mathbf{u}_1{}^T\mathbf{u}_2 = 0$, the following is obtained:

$$\phi = (\lambda_1 - \lambda_2)\mathbf{u}_1{}^T\mathbf{u}_2 = 0 \tag{2.17}$$

Equation (2.14) then reduces to:

$$\mathbf{S}\mathbf{u}_2 = \lambda_2\mathbf{u}_2 \tag{2.18}$$

Hence, $(\mathbf{S} - \lambda_2\mathbf{I}_q)\mathbf{u}_2 = 0$, where λ_2 is the second largest eigenvalue of $\mathbf{S}$ and $\mathbf{u}_2$ is the second largest principal component of $\mathbf{X}$. In general, up to q principal

components[2] may be derived by repeating this process, subject to the condition that $\mathbf{u}_i{}^T\mathbf{u}_j = \delta_{ij}$ where δ_{ij} is the Kronecker delta. Hence, a matrix of principal components may be ordered thus:

$$\mathbf{U} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_q \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

These principal components constitute a multivariate normal distribution and describe a hyper-ellipsoid in a multi-dimensional space.

Singular Value Decomposition (SVD)

Fundamental to the calculation of sample principal components is the technique of *Singular Value Decomposition (SVD)* [68]. Consider a matrix $\mathbf{X}$ in mean-deviation form (as described in section 2.2.2) of dimension $(q \times n)$, where $q < n$. SVD states that any matrix, $\mathbf{X}$, may be decomposed as:

$$\mathbf{X} = \mathbf{U}\mathbf{L}\mathbf{V}^T \tag{2.19}$$

where $\mathbf{U}$ and $\mathbf{V}$ are $(q \times r)$ and $(n \times r)$ matrices respectively, where r is the rank of $\mathbf{X}$. The columns $\{\mathbf{u}_i\}$ of the matrix $\mathbf{U}$ form an orthonormal basis for the columns of $\mathbf{X}$ whereas the orthonormal columns $\{\mathbf{v}_i\}$ of $\mathbf{V}$ span the row space of $\mathbf{X}$. The vectors $\{\mathbf{u}_i\}$ are called the left-singular vectors of $\mathbf{X}$ and the vectors $\{\mathbf{v}_i\}$ are the right-singular vectors. $\mathbf{L}$ is an $(r \times r)$ diagonal matrix and contains the corresponding "singular values" along its diagonal. In addition, $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

---

[2]Since $\mathbf{S}$ is of dimension $(q \times q)$, it is only possible to derive q principal components where $q < n$.

In PCA, the objective is to find the eigenvectors and eigenvalues of the covariance matrix, $\mathbf{S} = \frac{1}{n-1}\mathbf{X}\mathbf{X}^T$. Hence, multiplying equation (2.19) on the right by $\mathbf{X}^T$ yields:

$$\mathbf{X}\mathbf{X}^T = (\mathbf{U}\mathbf{L}\mathbf{V}^T)(\mathbf{U}\mathbf{L}\mathbf{V}^T)^T = (\mathbf{U}\mathbf{L}\mathbf{V}^T)(\mathbf{V}\mathbf{L}\mathbf{U}^T) = \mathbf{U}\mathbf{L}^2\mathbf{U}^T \qquad (2.20)$$

Hence, to within a scaling factor of $\frac{1}{n-1}$, calculating the eigenvectors of $\mathbf{S}$ yields the principal components as the columns of the orthonormal matrix $\mathbf{U}$ and diagonal matrix $\mathbf{L}$ additionally contains the *square roots* of the eigenvalues. That is,

$$\mathbf{L} = \begin{bmatrix} \lambda_1^{1/2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_q^{1/2} \end{bmatrix}$$

In certain cases (such as the observation of digital images), in which the number of pixels greatly exceeds the number of images, $q >> n$, where q is the number of elements of each observation vector (i.e. the number of pixels or landmarks) and n is the number of observations (images). From a computational perspective, calculating the eigenvectors and eigenvalues of the $(q \times q)$ matrix, $\mathbf{S}$, becomes problematic. However, it is possible to circumvent this problem by performing the SVD on the alternative covariance matrix, $\mathbf{S} = \frac{1}{n-1}\mathbf{X}^T\mathbf{X}$. Multiplying equation (2.19) from the left by $\mathbf{X}^T$ gives:

$$\mathbf{X}^T\mathbf{X} = (\mathbf{U}\mathbf{L}\mathbf{V}^T)^T(\mathbf{U}\mathbf{L}\mathbf{V}^T) = (\mathbf{V}\mathbf{L}\mathbf{U}^T)(\mathbf{U}\mathbf{L}\mathbf{V}^T) = \mathbf{V}\mathbf{L}^2\mathbf{V}^T \qquad (2.21)$$

Hence, the dimensionally smaller, $(n \times n)$ matrix, $\mathbf{V}$, is obtained by standard eigenvector decomposition and the orthonormal matrix of principal components, $\mathbf{U}$, can be achieved via a re-arrangement of equation (2.19):

$$\mathbf{U} = \mathbf{X}\mathbf{V}\mathbf{L}^{-1} \qquad (2.22)$$

Once the principal components have been found (through decomposition of the most appropriate form of the covariance matrix), it is possible to define a matrix $\mathbf{Z}$, where $\mathbf{Z} = \mathbf{U}^T \mathbf{X}$. The columns of $\mathbf{Z}$, $\{\mathbf{z}_i\}$, represent the projections of the mean-subtracted data onto the orthonormal principal components and hence give the novel representations of the data in a new, multi-dimensional space defined by the principal components. Therefore, any original data vector, $\mathbf{x}_j$, may be perfectly reconstructed from the principal components via:

$$\mathbf{x}_j = \overline{\mathbf{x}} + \sum_{k=1}^{m} \mathbf{u}_k \mathbf{z}_j(k) \qquad (2.23)$$

where m is the number of derived principal components and $\mathbf{z}_j(k)$ represents the kth component of the vector, $\mathbf{z}_j$ (the projection of the jth example onto all m principal components).

Truncation of Principal Component expansion

One of the main aims of PCA is to adequately represent a set of highly correlated data as a smaller number of uncorrelated variables. In this way, a reduction in dimensionality is achieved whilst the majority of the variance is maintained. Hence, the number of principal components to retain, t, must be chosen to achieve this. In other words, an optimal value for t must be determined such that the principal components may adequately represent an m-variable data set in $t < m$ dimensions [69]. This may be performed in several ways, with the simplest being to choose t such that a given, required proportion of the variance is explained in the original variables. For instance, since $> 98\%$ of samples in a normally distributed data set lie within ± 3 standard deviations of the mean (Chebyshev's Theorem [79]), 0.98 may be an appropriate value to choose. Since the SVD conveniently returns the principal components in order of decreasing significance, the chosen threshold, T

(for example, $T = 0.98$), may be set on the cumulative variance and t calculated such that the following is satisfied:

$$T = 100 \frac{\sum_{k=1}^{t} \lambda_k}{\sum_{j=1}^{m} \lambda_j} \tag{2.24}$$

If all m components are used, the jth observation in the original data set may be reconstructed as in equation (2.23). However, if $t < m$ components are retained, then only an approximate reconstruction may be effected:

$$\hat{\mathbf{x}}_j = \overline{\mathbf{x}} + \sum_{k=1}^{t} \mathbf{u}_k \mathbf{z}_j(k) \tag{2.25}$$

where the terms have the same meanings as in equation (2.23).

2.3 Modelling Shape

The first consideration when modelling shape variation for a set of objects must be a formal definition of the concept of shape. This was given by D.G. Kendall [62, 80]:

> *"Shape is defined as all the geometrical information that remains when location, scale and rotational effects are filtered out from an object."*

By this definition, shape is invariant to Euclidean transformations. In this thesis, the shape of an object is defined by a point model — a series of *landmark* points, each defining a specific position on the object. Landmarks may generally be classified into one of three categories [80], depending upon the rationale for their placement:

1. An *anatomical landmark* is point of biological relevance, high saliency and correspondence between organisms. They are assigned by an expert and tend to occur on the edges of an object, especially where there is a local maximum in

the curvature of the object's surface. An example of an anatomical landmark is the corner of the eye on a facial image.

2. A *mathematical landmark* is a point assigned on an object in accordance with some mathematical or geometrical property [81].

3. A *pseudo-landmark* is a point constructed on an object whose position is dictated by the locations of the anatomical or mathematical landmarks. Pseudo-landmarks are in general located around the outline of the object or in-between anatomical or mathematical landmarks.

Applied consistently across a sample of training images, a set of appropriate landmarks can successfully represent the variations over a pattern class[3] of shapes such as human faces. To achieve this, the landmarking procedure must ensure *correspondence* between the points in each image — that is, if the ith landmark is used to annotate a specific point in the jth image, the same landmark must be used to define the equivalent point on the kth image. This process is then repeated for all images in the training set. Once correspondence is established in this fashion, the x and y coordinates for each landmark point are placed into a data or *shape* vector, where $\mathbf{d} = [x_1, \cdots, x_q, y_1, \cdots, y_q]^T$. Given n training examples, n such vectors are generated. Before further analysis is performed, these vectors must be suitably aligned to remove the effects of location, scale and rotation in order that object shape is described in accordance with the stated definition.

Procrustes Alignment

Procrustes Alignment [62, 83] is a shape-preserving procedure designed to remove the effects of arbitrary rotation, scaling and offset in a data set — the alignment procedure will translate, scale and rotate a shape so as to minimise the sum of the

---

[3]A pattern class is a family of patterns that share some common properties [82].

squared distances between the boundary coordinates (landmarks) of the kth shape in a training set ($\mathbf{X}_k$) and some reference shape ($\mathbf{X}_r$). This is termed minimisation of the *Procrustes distance*, $P_d{}^2$, where:

$$P_d{}^2 = \sum_{j=1}^{q} \left[(x_k(j) - x_r(j))^2 + (y_k(j) - y_r(j))^2 \right] \tag{2.26}$$

where $\{x_k(j); y_k(j)\}$ are the x and y coordinates for the jth landmark on the kth shape to be aligned and $\{x_r(j); y_r(j)\}$ are the x and y coordinates for the corresponding landmark on the reference shape. The alignment procedure can be summarised as follows:

1. Compute the centroid (centre of mass) of each point set.

2. Subtract the respective centroid coordinates from each shape to translate its centre to the origin.

3. Re-scale each shape to have equal size, that is, their size-shape metric[4] must be identical. The Frobenius norm is the most commonly used shape-size metric and is given by:

$$S(\mathbf{d}_k) = \sqrt{\sum_{j=1}^{q} \left[(x_k(j) - \overline{x_k})^2 + (y_k(j) - \overline{y_k})^2 \right]} \tag{2.27}$$

where $\overline{x_k} = \frac{1}{q} \sum_{j=1}^{q} x_k(j)$ and $\overline{y_k} = \frac{1}{q} \sum_{j=1}^{q} y_k(j)$. The Frobenius norm is typically scaled to unity for all shapes.

4. Rotate each point set to align with the reference shape. The SVD can be used to determine the appropriate rotation matrix [84]:

---

[4]A shape-size metric, $S(\mathbf{d}_k)$, is defined as any positive, real-valued function of the shape vector, $\mathbf{d}_k$, that fulfils the property: $S(a\mathbf{d}_k) = aS(\mathbf{d}_k)$ [62]

$$\mathbf{U}\mathbf{L}\mathbf{V}^T = \mathbf{X}_k{}^T\mathbf{X}_r \qquad (2.28)$$

where,

$$\mathbf{X}_k = \begin{bmatrix} x_k(1) & y_k(1) \\ \vdots & \vdots \\ x_k(q) & y_k(q) \end{bmatrix}, \quad \mathbf{X}_r = \begin{bmatrix} x_r(1) & y_r(1) \\ \vdots & \vdots \\ x_r(q) & y_r(q) \end{bmatrix}$$

The matrix that rotates $\mathbf{X}_k$ to $\mathbf{X}_r$ can be shown to be:

$$\mathbf{R} = \mathbf{V}\mathbf{U}^T \qquad (2.29)$$

An alternative approach [58] is to translate and scale $\mathbf{X}_k$ and $\mathbf{X}_r$ according to steps 1-3 and then minimise E with respect to parameters a and b:

$$E = |\mathbf{X}_k\mathbf{R} - \mathbf{X}_r|^2 \qquad (2.30)$$

That is, to minimise the least squares error between the kth shape rotated by matrix $\mathbf{R}$ and the reference shape, where

$$\mathbf{R} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

In practice, an iterative procedure can be followed that allows the sample mean to be determined and aligns all training examples to the best estimate of the mean:

1. Choose the first training sample as the first estimate of the mean shape[5].

---

[5] The first training example is chosen by convention. However, this is arbitrary and any training example could be chosen equivalently.

2. Align all shapes to the mean estimate using Procrustes Alignment.

3. Re-calculate the estimate of the mean from the aligned shapes.

4. Repeat from step 2 if the mean estimate has changed.

This is repeated until there is convergence and no further changes to the mean estimate occur. Typically, only two iterations are sufficient for this to be achieved.

2.3.1 Formation of Shape Model

To illustrate the formation of a statistical shape model, we consider an example training set consisting of $n = 20$ RGB images of human lips, the shape of each being delineated by $q = 68$ corresponding landmarks. It is desirable to model this distribution such that the shape of any example within the training set may be represented compactly and that new, plausible examples may be generated, which are similar in appearance to the training examples. To this end, a parameterised model is sought of the form of equation (2.23). If the distribution of such parameters, $\mathbf{z}_j$, can be modelled, new examples similar to $\mathbf{d}_j$ can be produced by randomly sampling from the learned distribution function.

To perform the PCA, the columns of the data matrix, $\mathbf{D}$, were firstly aligned using the Procrustes procedure. The mean data vector, $\overline{\mathbf{d}}$, was then subtracted from every column to give new matrix, $\mathbf{D_s}$, so that variation was examined with respect to the mean shape. The result is depicted in figure 2.2.

The covariance matrix, $\mathbf{S} = \frac{1}{(n-1)}\mathbf{D_s}^T\mathbf{D_s}$ (since $q > n$, see section 2.2.2), was then computed. The eigenvectors and eigenvalues of $\mathbf{S}$ were calculated via an SVD using equation (2.21) and the matrix of orthonormal principal components, $\mathbf{P_s}$, computed
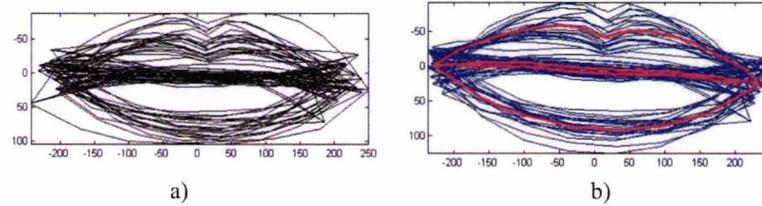
Figure 2.2: **Procrustes alignment of 20 lip shapes**. The shape vectors are initially aligned to the first example in the training set, which represents the initial estimate of the mean shape. The mean shape is then recalculated and the process repeated until there is convergence and no further changes to the mean estimate occur — a) Twenty unaligned lip shapes, translated to the origin, b) Procrustes alignment of twenty lip shapes. The red shape indicates the mean lip shape calculated subsequent to the alignment after two iterations.

using equation (2.22). Reconstruction of the data matrix in terms of the principal components is given by:

$$\mathbf{D_s} = \mathbf{P_s}\mathbf{B} \qquad (2.31)$$

where $\mathbf{B}$ is a matrix of dimension $(n \times n)$ whose columns consist of *shape model parameters*, each column containing a vector of coefficients, $\{\mathbf{b}_j\}$, which combine the principal components in the correct proportion to effect a perfect reconstruction of each data vector, $\{\mathbf{d}_j\}$. Equation (2.31) may be expressed in vector form as:

$$\mathbf{d}_j = \overline{\mathbf{d}} + \mathbf{P_s}\mathbf{b}_j \qquad (2.32)$$

Conversely, the projection of an example from the training set onto the principal components and hence the representation of the jth set of lips in the model space may be given by:

$$\mathbf{b}_j = \mathbf{P_s}^T(\mathbf{d}_j - \overline{\mathbf{d}}) \qquad (2.33)$$

From section 2.2.2 and equation (2.24), only $t = 3$ principal components need to be retained such that 90% of the shape variation in the training set is accounted for (figure 2.3).
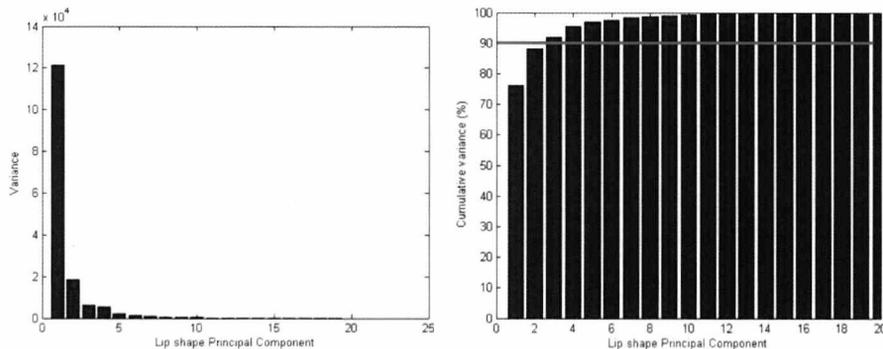
Figure 2.3: **The variance and cumulative variance explained by the shape principal components for a training set of 20 lips**. The variance associated with each shape principal component is shown on the left and the cumulative variance on the right — > 90% (red line) of the shape variation associated with the original data set can be expressed by the first three principal components.

A matrix $\mathbf{\Phi_s}$ may then be defined whose columns constitute these t principal components, hence $\mathbf{d}_j$ may be well approximated by:

$$\mathbf{d}_j \approx \overline{\mathbf{d}} + \mathbf{\Phi_s}\mathbf{b}_j \Rightarrow \mathbf{b}_j \approx \mathbf{\Phi_s}^T(\mathbf{d}_j - \overline{\mathbf{d}}) \tag{2.34}$$

By varying the elements of the jth parameter vector, $\mathbf{b}_j$, new lip shapes may be generated which are not members of the original training set but are statistically similar. If the standard deviation of $\mathbf{b}_j$ is given by σ_j, limits of $\pm 3\sigma_j$ may be applied to the random numbers to ensure that the generated novel shape is a typical example from the estimated distribution. In a similar fashion, it is possible to examine the *modes of variation* of the training set — that is, by adding a proportion of the jth

principal component to the mean shape in increments of 1 standard deviation, the variation explained by that principal component is revealed (figure 2.4). This may be described as follows:

$$\mathbf{d}_j{}' = \overline{\mathbf{d}} + \beta\sqrt{\lambda_j}\mathbf{P}_{\mathbf{s}j} \tag{2.35}$$

where $\mathbf{d}_j{}'$ is the shape associated with β standard deviations of the jth principal component, $\mathbf{P}_{\mathbf{s}j}$ and $\sqrt{\lambda_j}$ represents this standard deviation.

Subtraction of the mean shape vector from each aligned column of $\mathbf{D}$ is a necessary step prior to the application of PCA to ensure the removal of one degree of freedom from the analysis — if this is not performed, the first principal component will represent the mean of the data and this first mode may be disproportionately dominant.
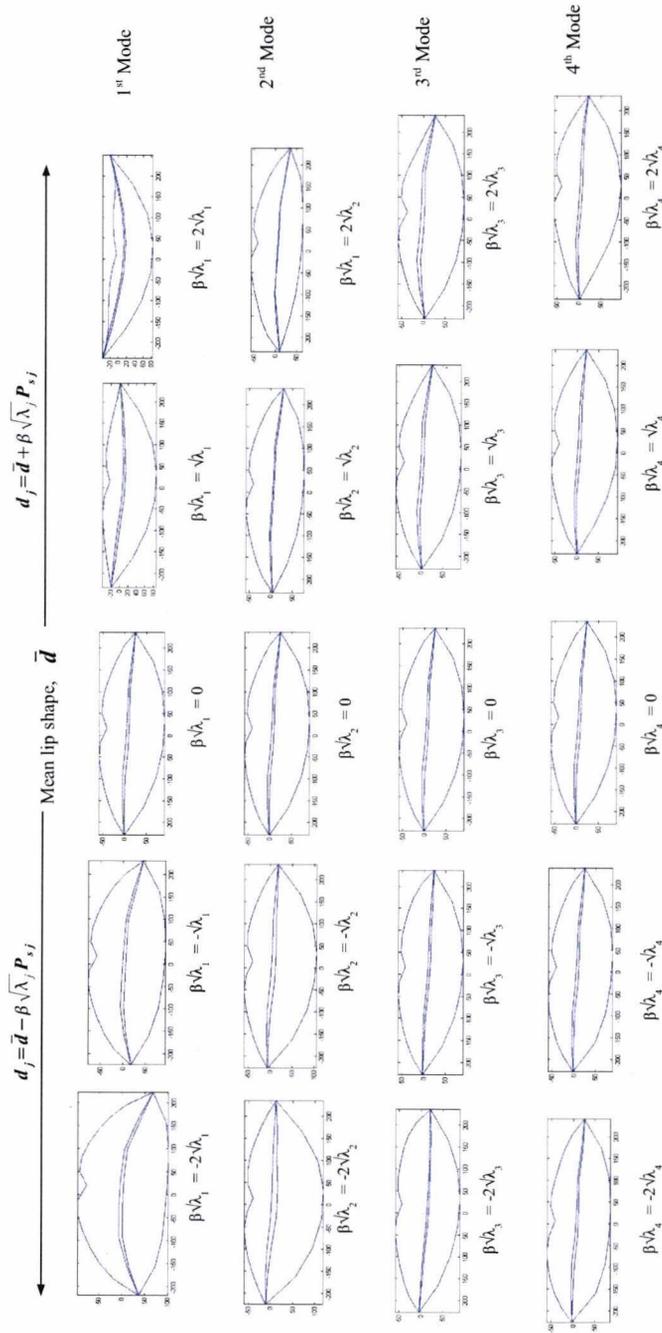
Figure 2.4: **The modes of shape variation for a training set of 20 lip images**. By varying the jth principal component between ± 2 standard deviations according to equation (2.35) for $j = 1 \ldots 4$, the variation embodied by the first four principal components may be observed. The four rows depict these modes of variation.

2.3.2 Modelling Texture

In order to model an object more completely, we must also consider intensity or *texture variation.* For an RGB image containing q pixels, the texture vector is given by the true colour pixel intensities [62]:

$$\mathbf{t}_j = [t_{1r}, \cdots, t_{qr}; t_{1g}, \cdots, t_{qg}; t_{1b}, \cdots, t_{qb}] \qquad (2.36)$$

where the subscripts r, g and b denote the true colour red, green and blue values respectively.

For texture to be described accurately, it is necessary that each texture map is extracted from a consistent shape configuration (arrangement of landmarks). If this is not performed, the application of PCA is not possible due to potential inconsistencies in pixel numbers across the training set. In addition, artificial texture variation may occur due to differences in shape thereby preventing the achievement of correspondence between texture maps. To address this, a geometric transformation is required to "warp" each example in the training set to a chosen reference shape (typically, the mean).

Geometric Transformations — Image "Warping"

Image warping is the general process by which one set of intensity or colour values (the input image) is mapped into another spatial configuration (the reference or base image). The process involves two distinct steps — a geometric transformation and a mapping of pixel values. The mapping function must be defined prior to the warping procedure and may be, for instance, linear, bilinear or polynomial in nature depending upon the desired application. Figure 2.5 depicts the warping process schematically, where the landmark points for the input image are given by $\{x_i^I; y_i^I\}$

and the corresponding points for the base image are given by $\{x_i{}^B; y_i{}^B\}$, where $i = 1 \ldots 4$.
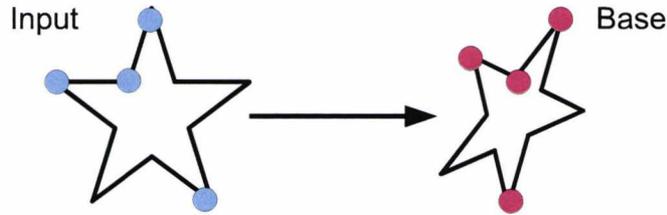
Figure 2.5: **A schematic representation of image warping** — the input coordinates must be transformed to the same spatial configuration as the base image via a geometric transformation.

To compute the warp, a bilinear mapping, for example, may be given by:

$$x_i{}^B = a_1 x_i{}^I + a_2 y_i{}^I + a_3 x_i{}^I y_i{}^I + a_4 \tag{2.37}$$

$$y_i{}^B = a_5 x_i{}^I + a_6 y_i{}^I + a_7 x_i{}^I y_i{}^I + a_8 \tag{2.38}$$

Solution of equations (2.37) and (2.38) for each set of corresponding points allows the free parameters $\{a_1 \ldots a_8\}$ to be calculated. In the example in figure 2.5, only four landmarks are employed to define the transformation. However, when the number of landmarks exceeds the number of degrees of freedom (typically the case for more complex shapes), the system becomes over-constrained and global transformations are of limited use.

As an alternative approach, Bookstein et al. [85] used thin-plate splines to compute a smooth warp. However, this method is relatively computationally expensive. Another approach employs a *piecewise* methodology, where the mapping function is assumed to be locally linear. This local linearity is embodied by joining the image

landmarks to divide the input and base images into a set of corresponding triangles through a process called *Delaunay triangulation*. For a set of points s in a plane, this is the triangulation, $DT(s)$, of s such that no point in s is inside the circumcircle of any triangle[6] [62]. That is, the Delaunay triangulation is a triangulation of the convex hull of points, in which every circumcircle of a triangle is an empty circle. Equivalently, the Delaunay triangulation of a discrete point set s is the geometric dual of the Voronoi tessellation for s [86].

To compute the piecewise affine warp[7], consider figure 2.6. The input and base landmarks are first joined to form corresponding triangles according to Delaunay triangulation. A pair of corresponding triangles in the input and base images is then selected, denoted t and t' respectively. The triangle vertices are denoted by $\mathbf{x}_j = [x_j, y_j]^T$ and $\mathbf{x}_j' = [x_j', y_j']^T$, $j = 1, 2, 3$. It is possible to describe any point, $\mathbf{x}$, within t as a linear combination of the triangle vertices:

$$\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

where $[\alpha, \beta, \gamma]^T$ are the coefficients of the x and y coordinates of the vertices. In matrix form, this may be given as:

$$\mathbf{T}\boldsymbol{\alpha} = \mathbf{x} \tag{2.39}$$

The corresponding point in t' may be given by the same linear combination of

---

[6]The circumcircle is the unique circle that contains all three vertices of the triangle.

[7]The piecewise transformation gives a relation between each pixel in the corresponding base and input triangles once triangulation is complete, such that the mapping is affine in nature (that is, it basically consists of scaling, translation, rotation and skewing) — the transformation is therefore rightly referred to as a *piecewise affine transformation*.

its triangle vertices. Hence:

$$\begin{bmatrix} x_1' & x_2' & x_3' \\ y_1' & y_2' & y_3' \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}$$

In matrix form:

$$\mathbf{S}\boldsymbol{\alpha} = \mathbf{x}' \tag{2.40}$$

Combining equations (2.39) and (2.40), it is possible to calculate the mapping between the coordinates of the input triangle and those of the base triangle as:

$$\mathbf{x}' = \mathbf{S}\boldsymbol{\alpha} = \mathbf{S}\mathbf{T}^{-1}\mathbf{x} = \mathbf{M}\mathbf{x} \tag{2.41}$$

Hence, to effect a warp, the mapping matrix $\mathbf{M}$ must be computed for every pair of corresponding triangles in the input and base images. Equation (2.41) is then applied to each and every pixel location within a given input triangle to find the corresponding locations in the base. The input pixel values are mapped to these calculated positions and the process is repeated for all pairs of triangles to complete the warp.

Conceptually, image warping is usually considered in the forward sense, that is, as a mapping *from* the input image *to* the base image. However, in general, there is not an exact one-to-one mapping between each pixel in the input and base images, hence some of the pixel values in the base image may remain undefined. In addition, there exist rounding errors in the computation of the base coordinates. These effects combine to produce "holes" or information gaps in the base image. Hence, it is often more effective to perform the *reverse* warp — that is, to substitute

$\mathbf{x}$ for $\mathbf{x}'$ and vice versa in equation (2.41) and apply the revised equation to every pixel location in a given base triangle. Once the corresponding input locations have been computed, the pixel values are transferred to the base and the process repeated for all triangle pairs. This reverse-warp ensures that the value of each pixel in the base image is defined. However, it does not ensure that the input coordinates are mapped to *integer* values in the base image and an interpolation method is required to calculate the base pixel values. The simplest method used is that of *nearest neighbour* interpolation or *point sampling*, whereby the mapped coordinate values are rounded to integer values. More accurate, yet computationally expensive, methods include *bilinear* and *bicubic* interpolation respectively.

For the training set of $n = 20$ lip images, the reference shape was chosen to be the mean lip shape and each shape was warped to this distribution using a Delaunay-based, piecewise affine warp. Pixel values or intensities were then extracted in a column-wise fashion from each shape-free image and placed into a texture vector for analysis.
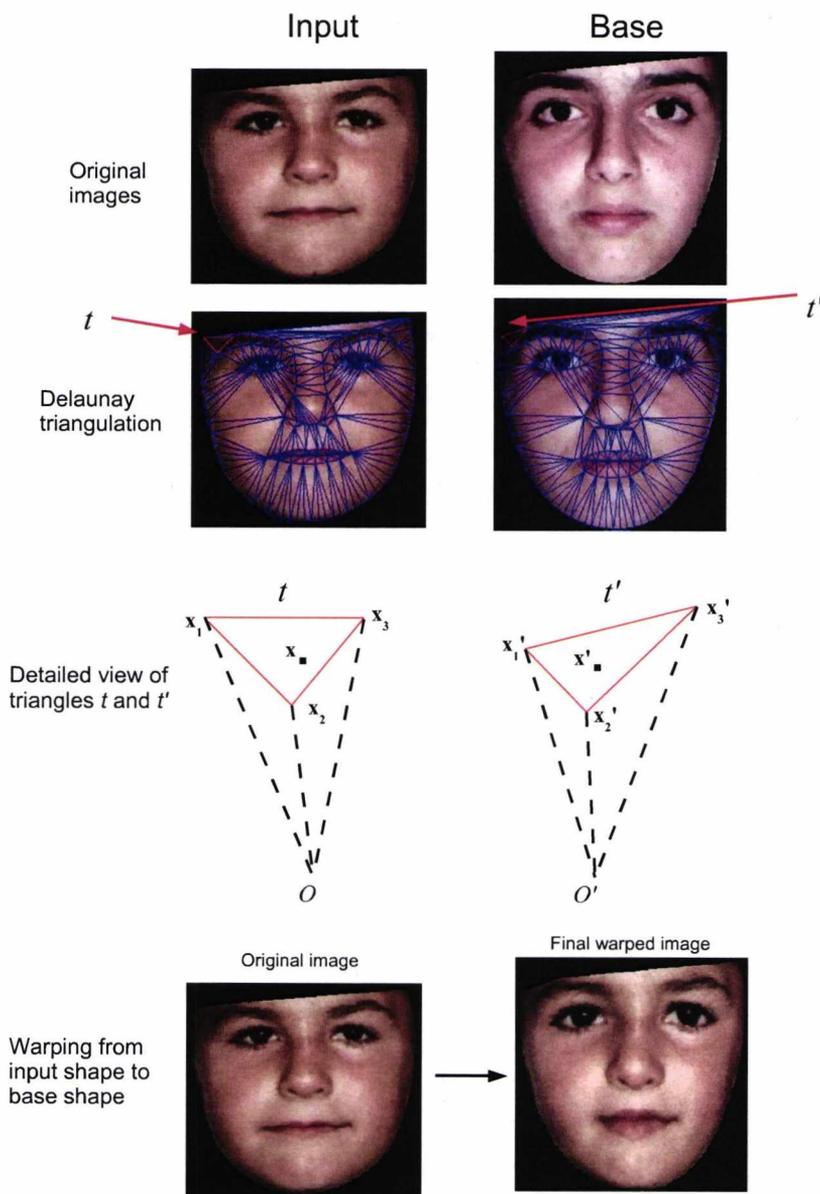
Figure 2.6: **An example of the piecewise, affine warp as applied to face images**. The input and base landmarks are joined to form corresponding triangles according to Delaunay triangulation. Corresponding triangles in the input and base images are denoted by t and t' respectively (origins O and O') with the triangle vertices depicted by vectors $\mathbf{x}_j$ and $\mathbf{x}_j'$, $j = 1, 2, 3$. Point $\mathbf{x} = [x, y]^T$ in the input triangle may be mapped uniquely onto point $\mathbf{x}' = [x', y']^T$ via the piecewise affine warp such that the pixel value of $\mathbf{x}$ may be mapped to the position of $\mathbf{x}'$. This is repeated for all such triangles in the input and base images such that the final, warped image is produced.

2.3.3 Formation of Texture Model

Upon construction of the texture matrix, $\mathbf{D_t}$ (whose columns comprise the mean-subtracted texture vectors), PCA was applied to this training set in an analogous fashion to the shape model. Since the number of pixels, q in an image is generally larger than number of images, n, in the training set, the covariance matrix, $\mathbf{S}$, is of dimension $(n \times n)$ and takes the form: $\mathbf{S} = \frac{1}{(n-1)}\mathbf{D_t}^T\mathbf{D_t}$. This ensures that a covariance matrix of a computationally viable size is used. The PCA produces a linear model:

$$\mathbf{t}_j = \overline{\mathbf{t}} + \mathbf{P_t}\mathbf{g}_j \tag{2.42}$$

where $\mathbf{t}_j$ is the jth texture vector in the training set[8], $\overline{\mathbf{t}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{t}_j$ represents the mean texture vector, $\mathbf{P_t}$ is a matrix of dimension $(q \times n)$ (whose columns are the orthonormal principal components) and $\mathbf{g}_j$ is a vector of *texture model parameters*. Alternatively, $\mathbf{g}_j$ may be viewed as the projection of the texture vector onto the principal components and hence provides the representation of the texture vector in the model space:

$$\mathbf{g}_j = \mathbf{P_t}^T(\mathbf{t}_j - \overline{\mathbf{t}}) \tag{2.43}$$

Once again, it is possible to truncate the number of principal components to explain the required amount of variation. The columns of matrix $\mathbf{\Phi_t}$ contain m PCs, where $m = 13$ to account for 90% of the variance (figure 2.7).

Hence:

$$\mathbf{t}_j \approx \overline{\mathbf{t}} + \mathbf{\Phi_t}\mathbf{g}_j \Rightarrow \mathbf{g}_j \approx \mathbf{\Phi_t}^T(\mathbf{t}_j - \overline{\mathbf{t}}) \tag{2.44}$$

---

[8]For normalisation purposes, the mean pixel value in a texture vector is sometimes subtracted along with the mean texture vector, $\overline{\mathbf{t}}$, prior to the application of PCA.
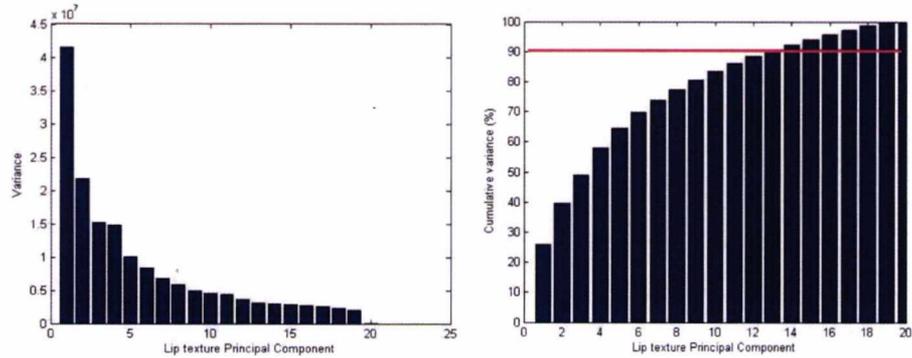
Figure 2.7: **The variance and cumulative variance explained by the texture principal components for a training set of 20 lips**. The variance associated with each texture principal component is shown on the left and the cumulative variance on the right — 90% (red line) of the texture variation associated with the original data set can be expressed by the first thirteen principal components.

Analogous to the shape model, it is possible to produce new, plausible (yet artificial) instances of texture by randomly sampling the elements of $\mathbf{g}_j$ from a normal distribution between $\pm 3\sigma_j$, where σ_j is the standard deviation $\mathbf{g}_j$. Some of the modes of textural variation are visually examined in figure 2.8.
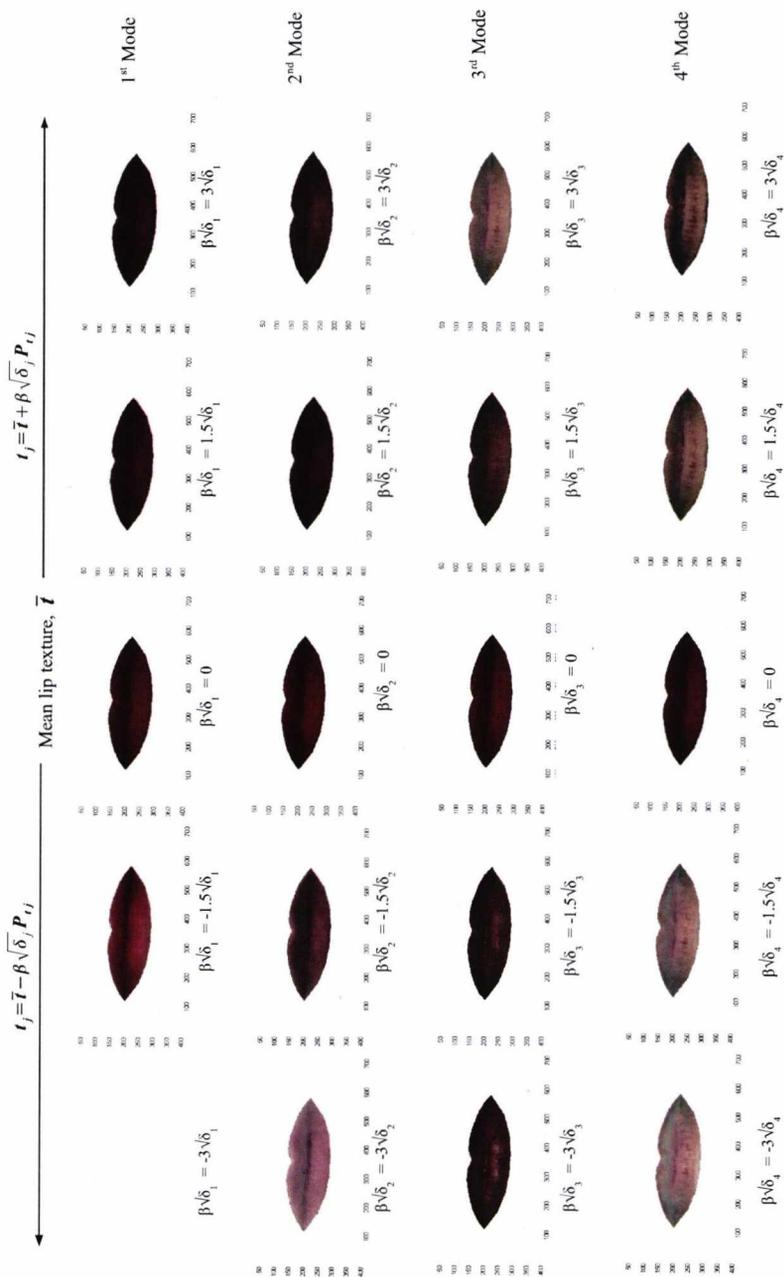
Figure 2.8: **The modes of texture variation for a training set of 20 lip images**. By varying the jth principal component between ± 3 standard deviations ($\sqrt{\delta_j}$ denotes the standard deviation for the jth principal component) according to an equivalent equation to (2.35) for $j = 1 \ldots 4$, the variation embodied by the first four principal components may be observed. The four rows depict these modes of variation.

2.4 Chapter Summary

In this chapter, the key mathematical methods underpinning the aging algorithms have been presented. To emphasise the central function of principal component analysis, the technique was formally introduced and its derivation shown in the sample case. To demonstrate its applicability to the modelling of shape and texture, an example training set was employed, consisting of 20 images of human lips. Initially, manually placed landmarks were used to delineate the lip shape and Procrustes Alignment was employed, prior to the application of PCA, to the mean-subtracted shape vectors. To model texture, each shape was first transformed to the mean shape using a Delaunay-based piecewise affine warping procedure. PCA was then applied to the extracted texture vectors.

In the next chapter, PCA is applied to the modelling of the human face and the theoretical basis for the aging algorithms is presented.

Chapter 3

Theory of Age-Progression Algorithms

In this chapter, several related techniques for age-progression are proposed. Initially, we discuss some necessary, preliminary image processing procedures. These include techniques for pose, expression and illumination compensation, which are employed prior to the application of the aging algorithms, to reduce artefacts in the computed results. A simple approach to age-progression is first introduced, which relies upon the calculation of a "consensus" aging trajectory through the model space, formed by weighting each model parameter vector by its age. A face is then age-progressed by translation in this direction. A "piecewise" linear approach is then presented, which takes better account of the non-linear stages of facial growth and development in children (0 to 20 years of age), as described in chapter 1. To further improve the aging technique, a theoretical framework is then described, which enables the incorporation of information provided by parental and sibling images. This methodology is extended to show how an aging axis may be constructed using the appearance of a subject at previous ages.

3.1 Training Sets and Demographics

Three databases were used in this work.

Database 1:

This database comprised 42 male Caucasian faces between the ages of 13 and 33 years and 25 female Caucasian faces between the ages of 14 and 52 years[1]. The database comprised RGB images [87], captured at a resolution of 3072×2048 pixels with the subject seated at a distance of 1.5 metres from a six million pixel digital camera. The subjects were required to face directly towards the camera to avoid potential complications in the aging algorithm due to pose variation and maintain a neutral expression to avoid similar problems with unconstrained expression. Images were captured in a laboratory without windows, preventing variations in lighting conditions due to daylight entering the room. Subjects were illuminated by a single fluorescent strip light and the camera flash.

Database 2:

A more extensive training set was acquired [88] to further model facial aging in children and teenagers. This was termed *Database 2* and comprised 1002 images of 47 male Caucasian and 35 female Caucasian subjects at a number of ages between 0 and 69 years. The images in this training database were of variable quality, resolution and illumination and the pose and expression of the faces were unconstrained. The data comprised a mixture of RGB, grey-scale and sepia images. 658 images were suitable for use in the shape model construction — these images consisted of 329 male faces and 329 female faces between the ages of 0 and 20 years, consistent with the accepted biological timespan of childhood and adolescence (section 1.4). Only

---

[1]These were the only available images at the time of performing this work and they were therefore used in the construction of shape and texture models, despite the paucity of examples.

236 male RGB images and 174 female RGB images were used for the construction of the texture models, since only these were perceived to be of sufficient quality and resolution. These were also the only images with a sufficiently small number of facial occlusions (hair, glasses, parts of clothing) for the texture models to be constructed viably. Despite the care taken with the selection of the images for inclusion in the shape and texture models, the unconstrained nature of the pose, expression and illumination of the faces required compensation techniques to be devised. These are discussed in detail in section 3.2.

Database 3:

To perform aging on a different racial group, a training set of 250 black children's facial images (250 different subjects) was acquired from the National Center for Missing and Exploited Children on-line database [89] (*Database 3*). Of these images, 90 male and 160 female were considered appropriate for the shape model with 87 male and 150 female utilised for the texture models. Analogous to the Caucasian training set, the images were unconstrained in terms of quality, resolution and illumination and in terms of pose and expression of the faces.

3.2 Pose and Expression Compensation

It is desirable to control the facial pose and expression when aging a face. If a model is constructed upon faces containing unconstrained pose or expression, the resultant modes of variation will be similarly unconstrained and uncontrolled artefacts may be introduced as a subject is age-progressed. To decouple the aging effect from such accidental variation, it is desirable to first compensate for these factors and construct a model based upon faces in an approximately frontal pose and neutral expression. A face presented to the system would need to be similarly compensated for aging to

proceed realistically. In applying such a pose or expression compensation algorithm, it is imperative that the identity of the subject is not altered.

A number of approaches to pose and expression compensation have been presented in the literature. Blanz and Vetter [90] used a 3D training set and learned expression transformations between pairs of images belonging to the same subject. Once correspondence had been established, changes in expression were mapped onto example faces. In similar work, Vetter [91] used a *linear object class* approach[2] to show that, for a training set of prototype faces at frontal and rotated views, any frontal view may be generated from a single rotated view.

Employing a different approach, Cootes et al. [58, 92] constructed three appearance (combined shape and texture) models of individuals at different head orientations (full profile, half-profile and frontal). A linear regression was used to learn the relationship between the model parameters and orientation angle and the appropriate model then used to synthesize new views at any allowed orientation. To generate significantly different views from those used to train the model, a coupled-view appearance model was produced using pairs of images taken at different orientations. The relationship was learned between the model parameters in different views and then used to generate alternative views of a subject to allow a frontal view to be determined from the corresponding profile. In related work, Gibson et al. [93] constructed a multiple-view appearance model based upon faces in frontal and 30° rotated views. This model was then used to estimate a missing view of a face when another was presented.

In the techniques discussed, the estimation of missing views is only possible

---

[2] A linear object class is defined as a 3D object class for which the 3D shape may be represented as a linear combination of a sufficiently small number of prototypical objects.

if a model is constructed upon corresponding frontal and rotated subject views, such that the relationship between the two orientations may be learned. Similarly, facial expression may only be "neutralised" if the model is trained upon subject faces, possessing both neutral and emotional expressions. In addition, the pose or expression is often controlled when selecting images for model training. This represents an idealised situation for examining pose and expression and formulating compensation procedures. Distinct from this, the work in this thesis requires novel methodologies to normalise images of subjects at completely *unconstrained* pose and expression and in which no corresponding images of different views or expressions exist for a given individual. The two approaches taken to this are described below.

3.2.1 Novel pose/expression compensation — Method 1

A pose and expression compensation technique was implemented, based upon the construction of a statistical shape model (using the technique outlined in section 2.3). Utilising all 658 faces in Database 2, a linear PCA was applied to the shape vectors and the modes of variation were visually examined to assess their influence on the facial shape. As implied from [11, 94, 95, 96], the identity, pose and expression were identifiable as approximately independent modes. This may be observed from the first, second and fifth modes of shape variation displayed in figure 3.1. The first mode depicts a dominant shaking of the head from right to left with a secondary perpendicular (nodding) motion present. The second mode shows an up-down nodding of the head with some slight rotation. Hence, shaking and nodding of the head may be considered as approximately independent modes. The fifth mode displays the face in a changing expression between a full smile and a serious expression with the mouth closed.

We posit the relationship:
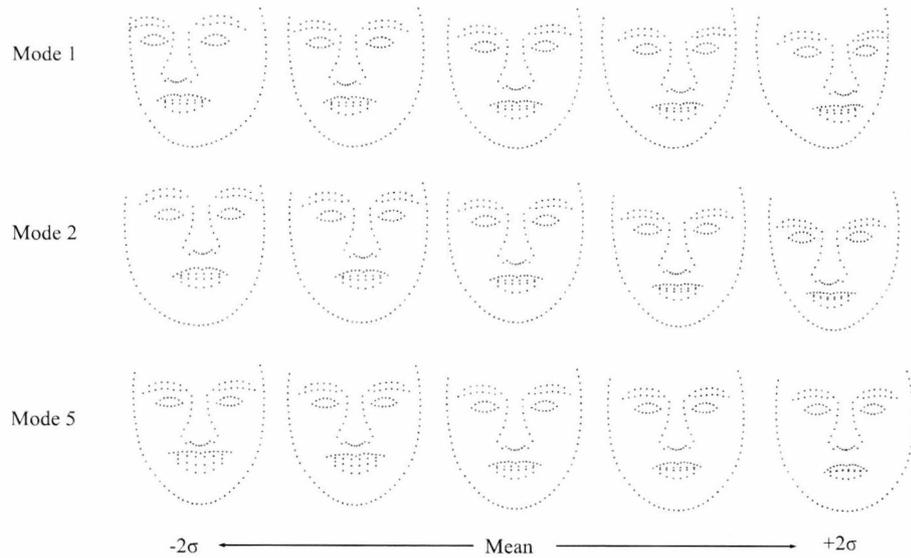
Figure 3.1: **The modes of facial shape variation.** The amounts of the first, second and fifth principal components of shape were varied between ± 2 standard deviations. Mode 1 shows head rotation coupled with minor nodding, mode 2 shows a dominant up-down nodding of the head with some slight rotation and mode 5 depicts changes in facial expression (a full smile to a more serious expression).

$$\mathbf{d}_j \approx \overline{\mathbf{d}} + \mathbf{P_{id}}b_{id} + \mathbf{P_{p1}}b_{p1} + \mathbf{P_{p2}}b_{p2} + \mathbf{P_e}b_e \tag{3.1}$$

where $\mathbf{d}_j$ is an approximation of the jth example in the training set, $\overline{\mathbf{d}}$ is the mean training shape and $\mathbf{P_{id}}$ represents the shape principal components related to identity. Analogously, $\mathbf{P_{p1}}$, $\mathbf{P_{p2}}$ and $\mathbf{P_e}$ represent the shape principal components related to the two different poses and the expression respectively (figure 3.1), where $\mathbf{P_{id}}$, $\mathbf{P_{p1}}$, $\mathbf{P_{p2}}$ and $\mathbf{P_e}$ have approximately mutual orthogonal columns. Similarly, b_{id}, b_{p1}, b_{p2} and b_e are the elements of the model parameter vector for the jth example, representative of the identity, pose and expression.

Since facial identity is observed to be approximately independent of pose and

expression, the jth example in the training set may be approximated using equation (3.1). Thus, the appropriate model parameters of a subject's shape vector (b_{p1}, b_{p2} and b_e, which correspond to the first, second and fifth principal components respectively) can be suitably modified. The face may then be reconstructed at the desired frontal pose and neutral expression. Figure 3.2 depicts example results of this method for pose compensation.

a)   b)   c)
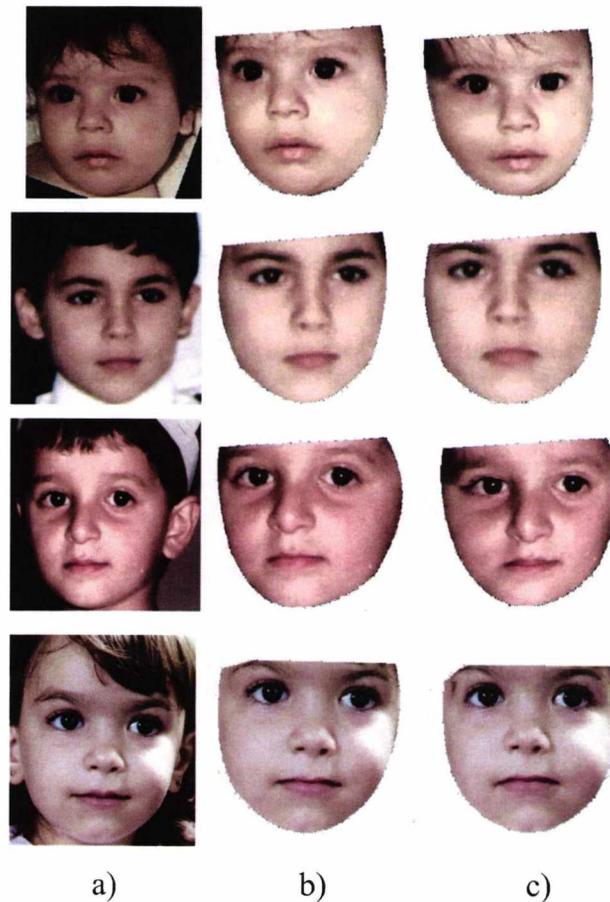
Figure 3.2: **Example results using Method 1 for pose compensation**. The first, second and fifth model parameters of the subject's shape vector (b_1, b_2 and b_5) were suitably modified to reconstruct the face at frontal pose and neutral expression — a) Original image, b) Shape model representation of the face including texture map, c) Pose compensated facial representation using Method 1.

For expression compensation to give plausible results, the facial texture must also be altered to account for the possibility of visible teeth [13]. This is achieved using a similar compensation method. In order that simply the mouth region was altered and not the entire face, a texture model was constructed based solely upon the extracted mouth region of each training face. The modes of variation were examined and the appropriate principal component coefficients adjusted to modify the texture such that the appearance of a closed mouth was adequately captured. In both gender cases, examination of the eighth mode of variation between ±2 standard deviations revealed a closing mouth with minimal changes occurring in lip colour. This mode was therefore used to perform mouth texture compensation on the necessary faces in the training set. However, it is important to state that the mouth is not the only facial region that alters when a person assumes an expression, such as a smile. Therefore, to obtain physiologically accurate results using this technique for expression compensation, the shape of the cheek region should also be altered in order to neutralise the expression. It was, however, not possible to do this, owing to the limitations of the point model used to landmark the face[3].

To incorporate the closed mouth into the entire face, a Delaunay-based, piece-wise, affine transformation was used to warp the closed mouth into the original image at the same location as the original mouth. Example results from this procedure are shown in figure 3.3.

Upon examination of figures 3.2 and 3.3, the results obtained using Method 1 for pose and expression compensation may be judged as reasonably poor and are not as perceptually accurate as the results obtained using equivalent artistic compensation

---

[3]It would be problematic to place landmarks around the cheek areas since these would likely not have good correspondence between images. However, it may be possible, in future work, to develop a more sophisticated point model such that these problems may be minimised.
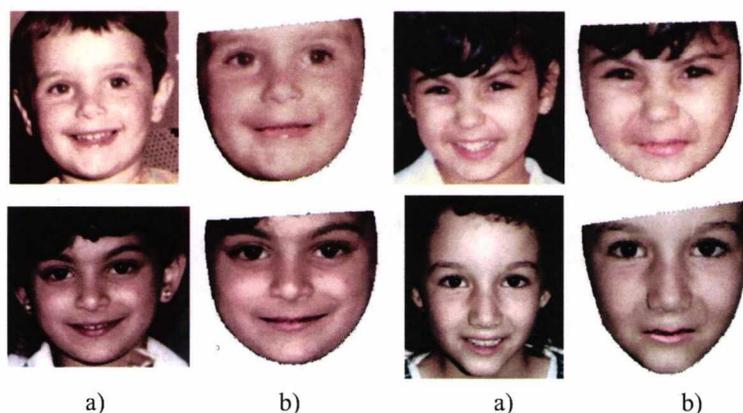
Figure 3.3: **Adjustment of the mouth texture to that of a closed mouth using Method 1**. The amount of the eighth principal component of mouth texture variation was adjusted to alter the mouth texture to that of a closed mouth — a) Original face, b) Mouth texture adjusted to that of an approximately closed mouth using Method 1 and the resulting texture map incorporated into the whole face.

procedures. In particular,there are problems around the eye areas and face width. These problems may be due to the pose compensation as a global technique that is applied over the whole face. Better results may be achieved by treating the eye area separately, for example, or by allowing the positions of particular features to be manually adjusted subsequent to the application of the algorithm. This may then allow more perceptually pleasing results to be obtained. There are additional problems regarding the expression compensation. As previously aforementioned, these issues may be resolved by adjusting other areas of the face (such as the cheek area and, furthermore, the eyes) when neutralising a smiling expression.

3.2.2 Novel pose/expression compensation — Method 2

In the second approach, the statistical model was constructed using the same unconstrained data as in Method 1. A shape parameter vector, compensated for pose and expression, was calculated according to:

$$\mathbf{b_{new}} = \mathbf{b} + \alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 + \alpha_3 \mathbf{e}_1 \tag{3.2}$$

where $\mathbf{b}$ is the original parameter vector, $\mathbf{p}_1$, $\mathbf{p}_2$ and $\mathbf{e}_1$ are the *pose and expression vectors* respectively and α_1, α_2 and α_3 are the appropriate scalars selected to multiply each vector. The pose and expression vectors were computed as follows:

1. Firstly, each training image was scored by an observer using a simple, 5-point, discrete scale (-2 to +2) according to the degree of head shaking (left or right), nodding (up or down) and smiling (none to broad) they exhibited (see Appendix C for a more detailed semantic description of the scale used).

2. The corresponding shape parameter vectors were then weighted by the appropriate score and pose vectors ($\mathbf{p}_1$ and $\mathbf{p}_2$) formed as a sum of these score-weighted shape parameters, according to:

$$\mathbf{p}_j = \sum_{i=1}^{n} \xi_i \mathbf{b}_i \tag{3.3}$$

for $j = \{1, 2\}$, where ξ_i is the score for the ith shape parameter vector, $\mathbf{b}_i$, for the jth pose (shaking or nodding) and there are n training examples. Similarly, the expression vector, $\mathbf{e}_1$, was calculated using:

$$\mathbf{e}_1 = \sum_{i=1}^{n} \psi_i \mathbf{b}_i \tag{3.4}$$

where ψ_i is the expression score for the ith shape parameter vector.

3. To alter the pose and expression of a face, scalar multiples of these vectors were added to the original shape parameter vector, as given in equation (3.2).

Figure 3.4 depicts example results of this method for pose compensation.

a)   b)   c)
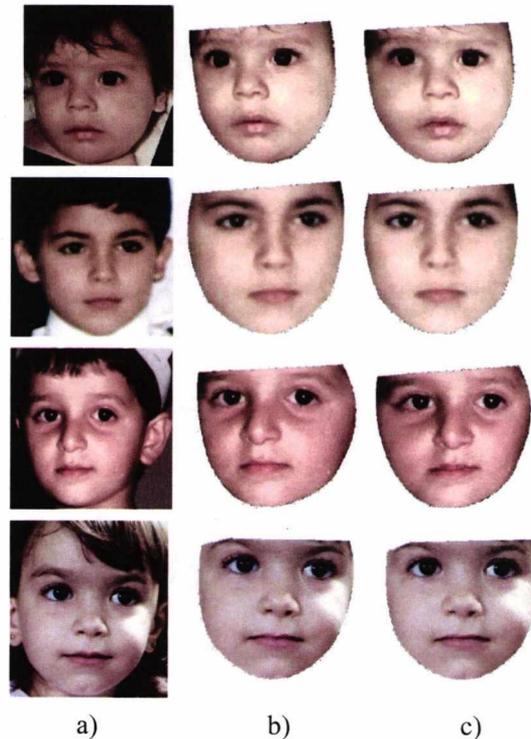
Figure 3.4: **Example results using Method 2 for pose compensation**. The pose and expression axes were calculated from score-weighted sums of the shape model parameter vectors. Scalar multiples of these axes were then used to adjust the face to frontal pose and neutral expression — a) Original image, b) Shape model representation of the face including texture map, c) Pose compensated facial representation using Method 2.

Once more, to ensure that expression compensation gives credible results, the mouth texture must be appropriately altered in cases where a smile is being neutralised[4]. To this end, a similar technique to Method 2 was employed. Equation (3.5) describes the procedure:

$$\mathbf{g_{new}} = \mathbf{g} + \gamma\mathbf{m} \tag{3.5}$$

where $\mathbf{g_{new}}$ is the vector of texture parameters of the closed mouth, $\mathbf{g}$ is the vector of original mouth texture parameters and γ is the scalar factor to multiply the *expression axis*, $\mathbf{m}$, in the correct proportion. The expression axis was calculated by a sum of score-weighted texture parameters, where the extracted mouth region of each training example was subjectively scored by ± 1 according to the visibility of the teeth. Once the texture map had been adjusted to resemble a closed mouth, the result was incorporated into the face using an analogous procedure to that of Method 1. Example results are shown in figure 3.5.

Upon examination of figures 3.4 and 3.5, the results obtained using Method 2 for pose and expression compensation are comparable to those obtained using Method 1. That is, they are perceptually poor, particularly around the eye area and facial width. Once more, akin to the discussion of results in section 3.2.1, these problems may be negated by treating the eye area independently to the rest of the face or by allowing suitable manual adjustment of the appropriate features.

---

[4]See section 3.2.1 for an explanation of why consideration of the mouth in isolation may not be sufficient to effect an accurate neutralisation of the smiling face.

a)   b)   a)   b)
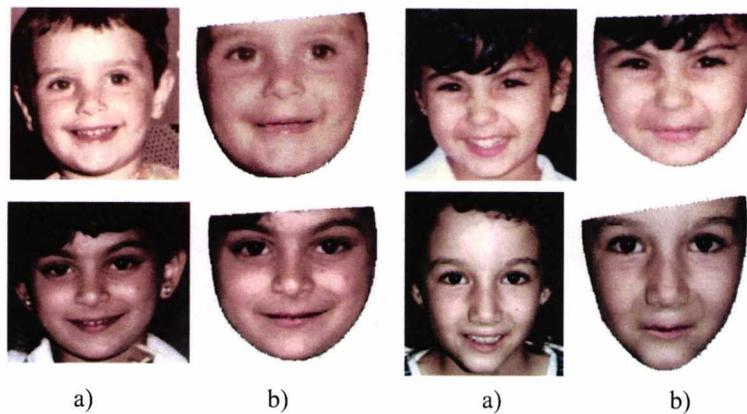
Figure 3.5: **Adjustment of the mouth texture to that of a closed mouth using Method 2**. The expression axis was calculated using a sum of score-weighted texture parameters and then scalar multiples of this axis used to adjust the texture to that of a closed mouth — a) Original face, b) Mouth texture adjusted to that of an approximately closed mouth using Method 2 and the resulting texture map incorporated into the whole face.

3.3 Texture Compensation

Plausible results are obtained from both methods of pose and expression compensation, hence either may be employed to produce faces of frontal pose and neutral expression. However, extraction of the texture map from an original facial image unconstrained in pose or expression may produce textural artefacts, which also must be compensated prior to texture model construction. In the case of modest head rotation of less than 25° to left or right (certainly the most common pose variation in Database 2), this was achieved by:

1. Reflecting the texture map about the vertical axis,

2. Calculating the average shape vector coordinates of the image pair, and

3. Mapping the average texture of corresponding triangles in the Delaunay tessellation of the images. This produces a texture map symmetrical about the vertical axis which may then be successfully used in the construction of the texture model[5]. It is important to note that this approach deteriorates as the head rotation angle increases beyond approximately 30° or in instances where the subject is looking to the extreme right or left.

Figure 3.6 demonstrates this procedure for example faces in Database 2. Figure 3.7 depicts results demonstrating the failures of the technique in the circumstances aforementioned. Texture maps such as these were not included in the texture models for aging.

---

[5]This compensation procedure may take place prior to or subsequently to any necessary alteration of the mouth texture during expression compensation.
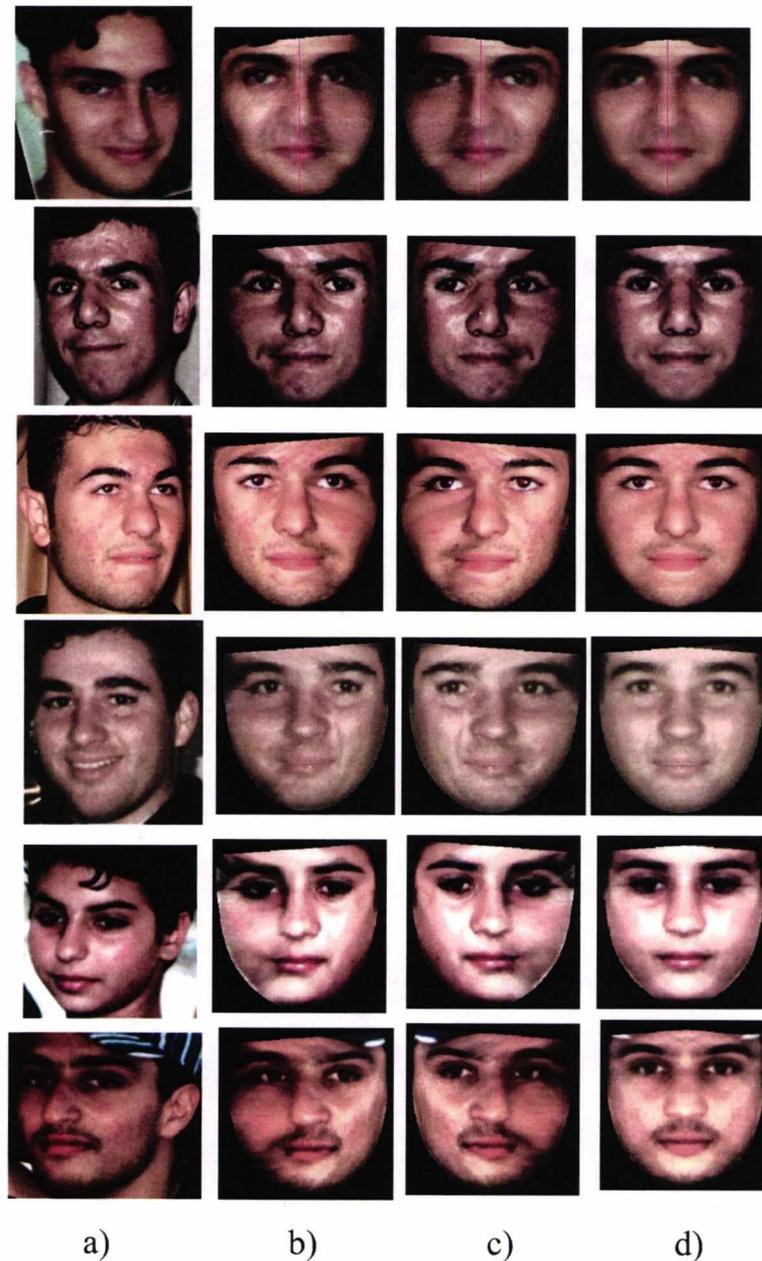
Figure 3.6: **Examples of texture compensation using Database 2**. The texture map is reflected about the vertical axis (exemplified by a red line for the first example) and the average texture of corresponding triangles in the Delaunay tessellation of the images is mapped — a) Original face, b) Shape-free texture map extracted from rotated face, c) Reflection of shape-free texture map about the vertical axis, d) Result of mapping the average texture of the two reflected images (b and c).

Figure 3.7: **Examples of unsatisfactory results using the texture map compensation.** In examples 1, 2 and 3, the result is unsatisfactory due to head rotation. In the 4th example, extreme eye movement is responsible for the undesirable result.

3.4 Illumination Compensation

A further issue to potentially affect the texture model is that of non-uniform illumination in the training images. This occurs throughout Database 2 and must be necessarily corrected before texture extraction and model construction. Common methodologies for illumination compensation include histogram matching of the non-uniformly illuminated image to that of a well-illuminated image [82, 97], homomorphic filtering [82] and region-based gamma intensity correction [98] as well as matching the energy (square of the intensity) embedded in the subject images to that of an idealised target [99].

The approach taken in this thesis considers image formation in terms of scene reflectivity and changes in illumination. For instance, if the response of the imaging device is linear then the recorded intensity of the image is given as:

$$s(x, y) = r(x, y)i(x, y) \tag{3.6}$$

where $r(x, y)$ represents the scene reflectivity and $i(x, y)$ represents the scene illumination. The illumination compensation procedure may then be summarised as follows:

1. A "gold standard" image was selected (figure 3.8), which depicts approximately uniform illumination.

2. This illumination was considered to vary smoothly (the simplest case) and thus was well approximated by a 2D polynomial function of the form $p_u(x, y)$. Points assumed to have identical reflectivity[6] were fitted to the polynomial function to form an estimate of the illumination.

3. Equivalent points of equal reflectivity were marked onto a non-uniformly illuminated subject image and the points fitted to a polynomial function, $p(x, y)$, in a similar fashion to step 2.

4. At corresponding points, i, on the images, $p_u(x, y)$ and $p(x, y)$ were used to estimate the intensity values, $\hat{s_u}(x_i, y_i)$ and $\hat{s}(x_i, y_i)$ respectively.

5. The ratio of these intensity values was calculated to give the so-called correction factor, $f_i = \hat{s}(x_i, y_i)/\hat{s_u}(x_i, y_i)$.

6. The correction factor was employed to compute the ith corrected pixel value in the subject image according to: $s'(x_i, y_i) = s(x_i, y_i)/f_i$.

For RGB images, this illumination compensation was performed on the sum of the red, green and blue pixel values (the intensity). The calculated correction factor was then applied to each of the three colour planes to compute the corrected image.

---

[6]Thirty points were marked onto the face, which were assumed to have equal reflectivity. Features such as the eyebrows, eyes and lips were avoided due to differences in reflectivity.
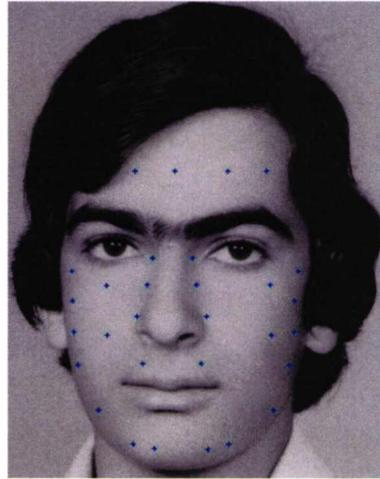
Figure 3.8: **The image chosen as the "gold standard" in the illumination compensation.** The blue markers represent points of assumed equal reflectivity across the face. A polynomial is then fitted between the intensity values and the (x, y) coordinates of these points to estimate the illumination field. Since the image is uniformly lit, the illumination field will be approximately flat.

Figure 3.9 show typical results using this illumination normalisation technique on both intensity and RGB images. This process was performed as necessary on the images of Database 2 prior to texture extraction and construction of the texture models for aging.

3.5 Wrinkle-maps

For the age-progression of adults, an aging model for texture constructed using PCA will yield results in which age-progressed subjects are perceived as younger than their true age [45]. Because PCA is a linear technique, the calculated components are produced by forming weighted averages of the images in the training sample (equation (2.22)). This process causes the fine details in the texture maps that exhibit low spatial correlation between the training faces (for example, wrinkles,

Figure 3.9: **Illumination compensation on intensity and RGB images**. Typical results of performing illumination compensation using a polynomial to model the illumination field — the images appear more uniformly illuminated than prior to the compensation.

freckles or other blemishes on the skin) to be averaged out. To obtain fully realistic aging effects in adults, these high-frequency details require preservation and even enhancement. Previous work by Tiddeman et al. [100] involved the decomposition of a prototype face (formed by averaging a sample of face images) using a wavelet analysis. The high order details were boosted to compensate for the inherent loss of high spatial frequency information in the averaging process and an enhanced texture prototype produced, which was more representative of the sample textures than the original prototype.

In this section, an ad-hoc procedure for enhancing high-frequency information is presented. The idea is to extract the fine facial details from a sample face A and apply those details to a subject face, B [101].

Let $\mathbf{I}_A$ be the image corresponding to subject A and let $\mathbf{I}_A^*$ represent the result

of convolving $\mathbf{I}_A$ with an averaging filter kernel, $\mathbf{G}$. If $\mathbf{I}_A{}^*$ is subtracted from $\mathbf{I}_A$, then only the high-frequency details of $\mathbf{I}_A$ (for example, wrinkles) will be preserved. This procedure is generally referred to as unsharp masking [82]. In this specific context, the difference image is termed a *"wrinkle-map"*, $\mathbf{W}$, where:

$$\mathbf{W} = \mathbf{I}_A - \mathbf{I}_A{}^* \tag{3.7}$$

Alternatively, $\mathbf{W}$ may be obtained by passing image $\mathbf{I}_A$ through a high-pass filter.

3.5.1 Applying a wrinkle-map to a subject face

Once it has been extracted from image $\mathbf{I}_A$, it is then desirable to apply the wrinkle-map to a subject, B. This is achieved by employing the following equation:

$$\mathbf{I}_B{}' = (\alpha - 1)\mathbf{I}_B{}^* + \mathbf{W}, \quad \alpha \geq 1 \tag{3.8}$$

where $\mathbf{I}_B{}^*$ is a smoothed version of image $\mathbf{I}_B$[7] (obtained by convolution with an averaging filter kernel) and α is a scalar[8]. This equation may be considered as a hybrid form of the standard *high-boost filtering* method commonly used in digital image processing (see [82]). In the normal implementation of this procedure, a detail image, $\mathbf{W}$, is constructed as in equation (3.7) but using the subject image, $\mathbf{I}_B$. This is then added to a fraction of the same subject image, $\mathbf{I}_B$, using a similar equation to (3.8). Equation (3.8) is a modified version of this, where the detail image is formed from a different image to that of the subject, namely, image $\mathbf{I}_A$.

---

[7]Prior to wrinkle-map addition, it is necessary to smooth the subject image to prevent the fine structures already present in the subject texture map from being visible through the wrinkle-map and producing artefacts.

[8]In practice, $\alpha \approx 2$ and the filter kernel used to form $\mathbf{I}_A{}^*$ in equation (3.7) should be small enough that only fine details are copied to subject B. If these precautions are not taken, the identity of subject B may be altered upon application of the wrinkle-map.

The image that results from the application of equation (3.8), $\mathbf{I}_B{}'$, contains the low and medium spatial frequencies of image $\mathbf{I}_B$ and the high frequencies belonging to image $\mathbf{I}_A$. Figure 3.10 shows the result of applying equation (3.8) to a subject, B, and the effect of varying the size of the averaging filter kernel, $\mathbf{G}$, used to produce $\mathbf{I}_A{}^*$ in equation (3.7). The larger the filter kernel used, the more low/medium frequencies are preserved and the wrinkles appear deeper and more numerous. Figure 3.11 shows the effect of varying the scalar quantity, α, in equation (3.8). As α increases, the proportion of $\mathbf{I}_B{}^*$ relative to $\mathbf{W}$ increases and more low-medium spatial frequencies are observed in $\mathbf{I}_B{}'$, which account for the greater number of wrinkles (fewer very fine wrinkles are observed, which are captured by the high spatial frequencies in $\mathbf{W}$). For ease of implementation, this approach to controlling the degree of wrinkling is often used in preference to adjusting the size of the filter, $\mathbf{G}$.

a) Filter size = 4 x 4   b) Filter size = 8 x 8   c) Filter size = 12 x 12
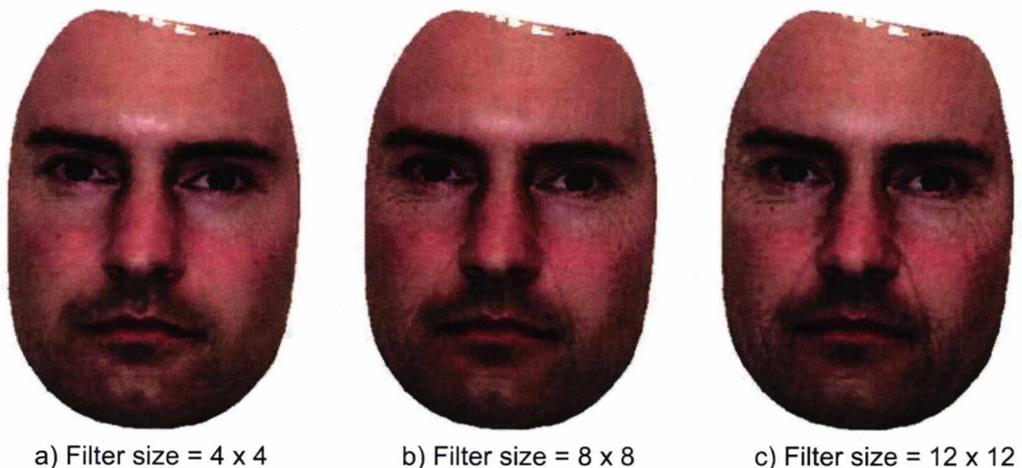
Figure 3.10: **An example of wrinkle-map application**. A wrinkle-map is generated from a sample face A and applied to a specific subject face B. As the size of the averaging filter used in the creation of the wrinkle-map is increased, more low frequency details are preserved and the wrinkles appear more pronounced. Images courtesy of [101].

a) α = 2, filter size = 8 x 8   b) α = 2.5, filter size = 8 x 8
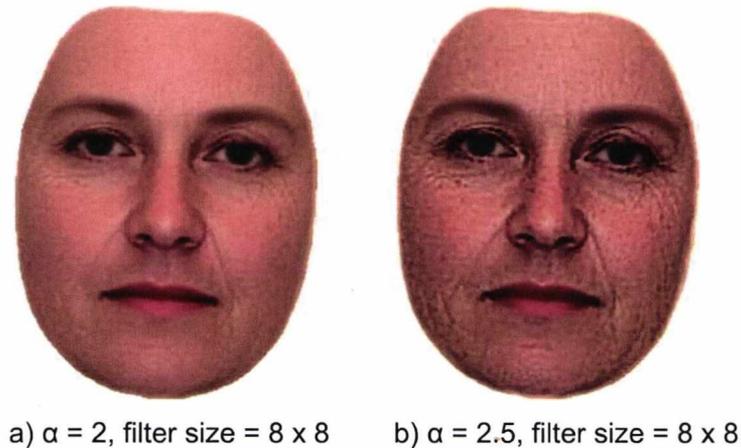
Figure 3.11: **Controlling the degree of wrinkling by varying the scalar, α.** The images show the effect of increasing the scalar quantity, α (which determines the relative proportions of the subject face and the wrinkle-map). In practice, the amount of wrinkling is controlled in this manner and not by varying the low-pass filter kernel size. Images courtesy of [101].

The wrinkle-map does not account for differences in facial type or for individual wrinkling patterns. For example, the presence or absence of nasolabial creases is not age-related. However, since such creases will be present in any wrinkle-map created from a sample face whose facial texture contains them, the creases will be universally applied as the wrinkle-map is added to the texture of a given subject. This may result in inaccurate wrinkling. For example, the subject in figure 3.10 originally had no nasolabial creases. Subsequent to wrinkle-map addition, nasolabial creases have been added, resulting in a wrinkled face that may not be indicative of the true wrinkling pattern for the subject.

It is possible to apply a chosen wrinkle-map to a localised region of the face to preferentially enhance the detail in that part of the face. This is achieved as follows:

$$\mathbf{W}' = \mathbf{W} \circ [\beta_1 \mathbf{M}_{forehead} + \beta_2 \mathbf{M}_{eyes} + \beta_3 \mathbf{M}_{mouth}] \tag{3.9}$$

where $\mathbf{W}'$ is the modified wrinkle-map, β_1, β_2 and β_3 are the (empirically chosen) weighting components[9] that determine the wrinkles appearing in the forehead, eyes and mouth regions respectively[10] and $\mathbf{M}_{forehead}$, $\mathbf{M}_{eyes}$ and $\mathbf{M}_{mouth}$ represent the respective *binary masks* defining these regions. The mask image for each region of interest is constructed (figure 3.12) such that each pixel in wrinkle-map $\mathbf{W}$ is multiplied by the corresponding (weighted) pixel value in the binary mask[11]. The three defined masks are then applied simultaneously to the face such that the wrinkles in the region of interest are emphasised to a greater degree than those in the remaining regions.

The output image is obtained by substituting $\mathbf{W}'$ from equation (3.9) for $\mathbf{W}$ in equation (3.8). Figure 3.13 shows example results of applying the wrinkle-map to different regions of the face.

---

[9]To ensure plausibility of wrinkling, the weighting component for the selected region is typically set to twice the magnitude of the other weightings. For example, if the eye region has been selected, then $\beta_1 = 2\beta_2 = \beta_3$.

[10]These are the regions where wrinkling tends to occur as an individual ages [5, 14, 21], section 1.4.

[11]The operator $\circ$ in equation (3.9) represents a Schur-Hadamard matrix product and indicates a pixel-wise multiplication of the wrinkle-map $\mathbf{W}$ with the sum of the weighted binary masks.

a) Forehead   b) Eye region   c) Mouth region
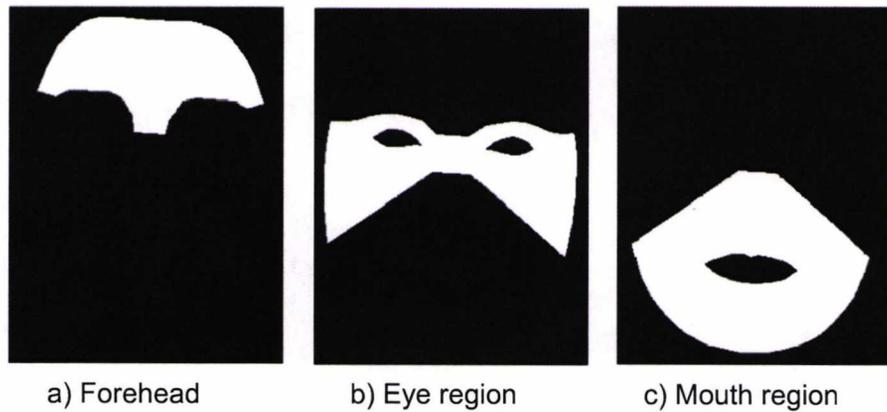
Figure 3.12: **Definition of masks for local wrinkle-map application**. A mask is defined for each region of interest on the face (forehead, eye region, mouth region) and each mask weighted by a scalar multiple, according to the desired relative proportions of wrinkling for the subject face. The wrinkle-map is then multiplied in a pixel-wise fashion by a sum of these weighted masks and a new wrinkle-map calculated.

Forehead   Eye region   Mouth region
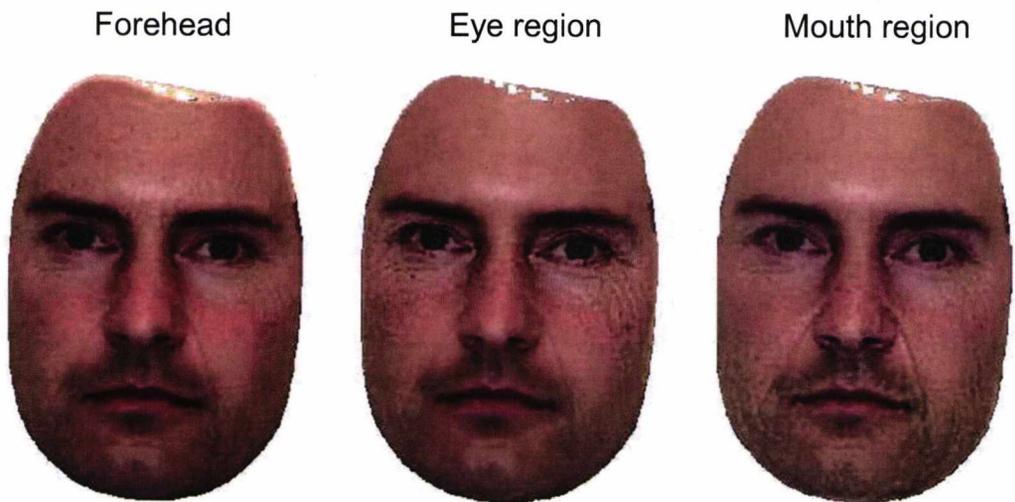
Figure 3.13: **Local wrinkle-map application**. The wrinkle-map may be applied to local regions of the face by varying the relative proportions of the forehead, eye region and mouth region binary masks. The result is more pronounced wrinkling in the area of interest and less pronounced wrinkling in the rest of the face.

3.6 Consensus Age-Progression

The construction of a statistical model of facial appearance within this thesis closely resembles that originally outlined by Cootes et al. [57]. However, in contrast to Cootes' model, the model herein retains two distinct parts — a PCA model of shape and a PCA model of texture. To assist in the construction of the shape model, a specially designed tool was produced [101] (see Appendix D) to increase the speed and accuracy of the landmarking. Figure 3.14 displays an example of the landmarks, whose coordinates constitute the shape vector of the face for examples in Database 1 (see Appendix E for the point model used for Databases 2 and 3).
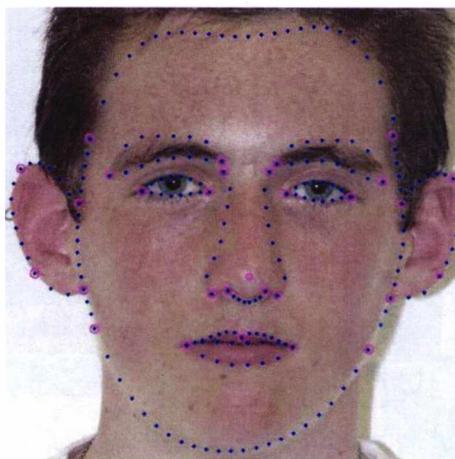
Figure 3.14: **Point model representing the shape of the face**. Magenta circular markers represent base landmarks and the blue points represent the interpolated landmarks, which follow the contours of the facial features (see Appendix D for details).

The N resultant shape vectors were pose and expression compensated (section 3.2) as necessary[12] and consecutively aligned to the mean face shape through an iterative Procrustes Alignment (section 2.3). A standard linear PCA was applied

---

[12]Owing to the nature of the training sets, pose and expression compensation were required on Databases 2 and 3 only, where necessary. The images in Database 1 were suitably constrained such that compensation was not required.

to the ensemble to extract the modes of variation and the orthonormal principal components were computed (via equation (2.20) or (2.21) and (2.22)). The jth example of a shape vector within the training set could then be described using an equation equivalent to (2.32):

$$\mathbf{d}_j = \overline{\mathbf{d}} + \mathbf{P_s}\mathbf{b}_j \tag{3.10}$$

where $\mathbf{P_s}$ is the matrix of orthonormal shape principal components.

The second part of the face model consists of a PCA model of the facial texture of the sample. This model was created by first performing illumination compensation, as necessary (section 3.4). Then all training examples were warped to the mean shape using an iterative piecewise, affine transform (section 2.3.2) and the RGB values were texture compensated as required (sections 3.3) and extracted to form an ensemble of shape-normalised, texture vectors. Linear PCA was again applied to produce a set of texture principal components. The jth texture vector may be expressed as a sum of orthonormal principal components using an equation equivalent to (2.42):

$$\mathbf{t}_j = \overline{\mathbf{t}} + \mathbf{P_t}\mathbf{g}_j \tag{3.11}$$

where $\mathbf{P_t}$ is the matrix of orthonormal texture principal components.

Crucially, the two sets of model parameters, $\mathbf{b}_j$ and $\mathbf{g}_j$, are relatively compact (approximately 30 components for shape and 100 components for texture) and enable a given facial appearance to be conveniently visualised as a point in a multidimensional vector space [66].

To calculate a *consensus aging axis* in shape, $\mathbf{V}_{\mathbf{age}}$, through the model space, the following equation was employed:

$$\mathbf{V}_{\mathbf{age}} = \frac{\sum_{j=1}^{N}(W_j - W_\mu)(\mathbf{b}_j - \mathbf{b}_\mu)}{\sum_{j=1}^{N}(W_j - W_\mu)} \tag{3.12}$$

where W_j and $\mathbf{b}_j$ are the age (in years) and the vector of shape model parameters respectively for the jth subject, W_μ is the mean age and $\mathbf{b}_\mu$ is the mean vector of shape model parameters. Since, in our specific case, $\mathbf{b}_\mu = 0$, we have[13]:

$$\mathbf{V}_{\mathbf{age}} = \sum_{j=1}^{N}(W_j - W_\mu)\mathbf{b}_j \tag{3.13}$$

The formation of the aging axis is displayed schematically in a hypothetical 2D space in figure 3.15.
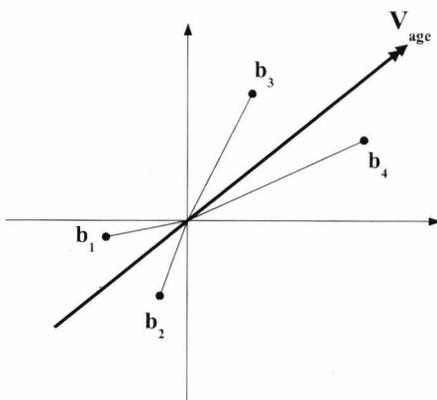
Figure 3.15: **The formation of the aging axis in shape model space.** The shape vectors, $\mathbf{b}_1$ to $\mathbf{b}_N$, are weighted by their corresponding mean-subtracted ages and summed to form the aging axis, $\mathbf{V}_{\mathbf{age}}$.

Equation (3.13) indicates that the axis is calculated by weighting the model

---

[13]The denominator in equation (3.12) may be discarded since, for practical use, the aging axis is normalised to unit length.

parameters of an individual face by their corresponding, *mean-subtracted* age. This mean-subtraction is considered necessary as follows. The formation of the space ensures that the shape model parameter vectors, $\{\mathbf{b}_j\}$, are zero-mean and thus, that the mean facial shape lies at the origin of the space:

$$\mathbf{b}_\mu = \frac{1}{N} \sum_{j=1}^{N} \mathbf{b}_j = 0 \tag{3.14}$$

However, this mean shape has a mean age $W_\mu \neq 0$:

$$W_\mu = \frac{1}{N} \sum_{j=1}^{N} W_j \neq 0 \tag{3.15}$$

To account for this discrepancy, the aging axis is formed by weighting the mean-subtracted shape model parameter vector by its corresponding mean-subtracted age.

The aging axis defines the average trajectory through the model space along which age increases. Thus, an example face translated in this direction should appear older. To proceed, a calibration procedure is first applied to the aging axis to produce a unit vector, $\hat{\mathbf{V}}_{\mathbf{age}}$. Appropriate scalar multiples of this vector may then be added to the model parameters to age the facial shape by the required number of years:

$$\mathbf{b}_j{}' = \mathbf{b}_j + \alpha \hat{\mathbf{V}}_{\mathbf{age}} \tag{3.16}$$

where $\mathbf{b}_j{}'$ is the set of aged shape parameters for the jth individual.

3.6.1 Calculation of age scaling parameter, α

The scalar multiple, α, in equation (3.16) effectively defines the number of years of aging and is calculated as follows:

Consider the orthogonal projection, $\mathbf{P}_i$, of the ith example in the training set (given by the parameter vector $\mathbf{b}_i$) onto the axis, $\hat{\mathbf{V}}_{\mathbf{age}}$, given by:

$$\mathbf{P}_i = (\mathbf{b}_i{}^T \hat{\mathbf{V}}_{\mathbf{age}}) \hat{\mathbf{V}}_{\mathbf{age}} \tag{3.17}$$

This gives the most accurate estimate of facial shape at age $A_i = W_i - W_\mu$ years. Similarly, the jth example of age $A_j = W_j - W_\mu$ years has a projection:

$$\mathbf{P}_j = (\mathbf{b}_j{}^T \hat{\mathbf{V}}_{\mathbf{age}}) \hat{\mathbf{V}}_{\mathbf{age}} \tag{3.18}$$

Thus, the distance in the direction of $\hat{\mathbf{V}}_{\mathbf{age}}$ corresponding to $(A_i - A_j)$ years is:

$$\mathbf{P}_i - \mathbf{P}_j = \left[(\mathbf{b}_i - \mathbf{b}_j)^T \hat{\mathbf{V}}_{\mathbf{age}} \right] \hat{\mathbf{V}}_{\mathbf{age}} = A_i - A_j \tag{3.19}$$

However, this calculation is for just two example faces. Defining:

$$a_{ij} = A_i - A_j \tag{3.20}$$

and

$$\Delta_{ij} = \left[(\mathbf{b}_i - \mathbf{b}_j)^T \hat{\mathbf{V}}_{\mathbf{age}} \right] \hat{\mathbf{V}}_{\mathbf{age}} \tag{3.21}$$

and using all permissible pairs of examples, α is defined as:

$$\alpha = M \frac{\sum_{i \neq j, i > j} \Delta_{ij}}{\sum_{i \neq j, i > j} a_{ij}} \tag{3.22}$$

where M is the number of years of aging required and the fraction $\frac{\sum_{i \neq j, i > j} \Delta_{ij}}{\sum_{i \neq j, i > j} a_{ij}}$ gives one year of aging in terms of displacement along the axis, relative to the origin.

Once the facial shape has been age-progressed according to equation (3.16), the vector of aged model parameters, $\mathbf{b}_j{'}$, may be combined with the shape principal components (equation (3.10)) to produce the aged facial shape, $\mathbf{d}_j{'}$.

3.6.2 Sexual Dimorphism

Sexual dimorphism in the aging of humans [14, 21] indicates a need to consider the two sexes separately. Hence, in the shape model, two aging axes are defined, one for each sex sub-space and an individual may be aged according to its respective Euclidean distance from these axes ($\mathbf{E_m}$ representing the Euclidean distance vector to the male axis and $\mathbf{E_f}$ representing the equivalent vector to the female axis), as depicted schematically in figure 3.16. This is necessary so that aging may proceed in accordance with both sex influences — for instance, a male face, by definition, will generally have closer proximity to the male axis than to the female ($\mathbf{E_m} < \mathbf{E_f}$). However, $\mathbf{E_f}$ may be sufficiently small that the female examples may have an effect on the face as it is age-progressed. This may be the case for a male of feminine appearance. Hence, the most realistic transformation is effected by employing a proportion of each sex axis, where $\mathbf{E_m}$ and $\mathbf{E_f}$ determine the appropriate weights to be used. Conversely, the same approach may be necessary to age a more masculine female (that is, a female more closely related in appearance to male faces than to other females [102, 103]) in the most appropriate fashion.

To age a face in this manner, equation (3.23) is used:

$$\mathbf{b}_j{'} = \mathbf{b}_j + \alpha_m \left(\frac{\mathbf{E_f}}{\mathbf{E_m} + \mathbf{E_f}} \right) \hat{\mathbf{V}}_m + \alpha_f \left(\frac{\mathbf{E_m}}{\mathbf{E_m} + \mathbf{E_f}} \right) \hat{\mathbf{V}}_f \qquad (3.23)$$
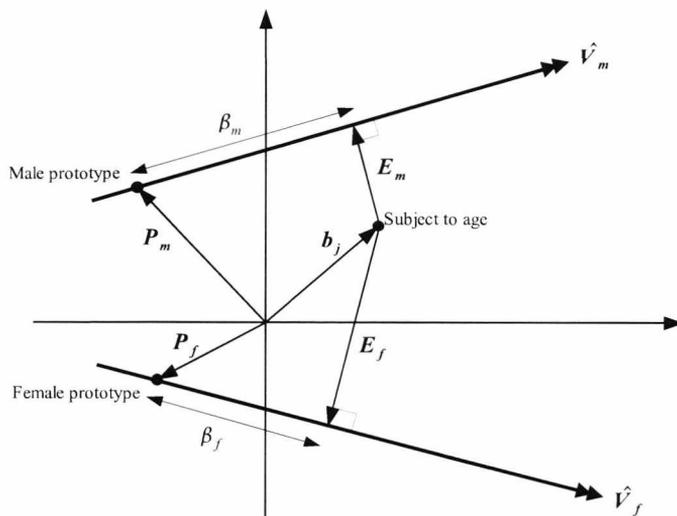
Figure 3.16: **A 2D representation demonstrating a subject's Euclidean distance vector in shape space from the male and female aging axes**. The length-normalised male and female aging axes are given by $\hat{\mathbf{V}}_{\mathbf{m}}$ and $\hat{\mathbf{V}}_{\mathbf{f}}$ respectively. The vectors $\mathbf{P}_{\mathbf{m}}$ and $\mathbf{P}_{\mathbf{f}}$ represent the male and female prototypes (averages) respectively, with $\mathbf{b}_j$ representing the jth subject to be aged. The subject is then aged according to its relative distance from the male and female axes.

where α_m and α_f are the scalar multiples to control the degree of aging (calculated from an equivalent expression to equation (3.22)) using the length-normalised male and female aging axes, $\hat{\mathbf{V}}_{\mathbf{m}}$ and $\hat{\mathbf{V}}_{\mathbf{f}}$, respectively. As required, this expression reduces to equation (3.16) in the case where $\mathbf{E}_{\mathbf{m}}$ or $\mathbf{E}_{\mathbf{f}}$ is zero.

To calculate $\mathbf{E}_{\mathbf{m}}$ and $\mathbf{E}_{\mathbf{f}}$, $\hat{\mathbf{V}}_{\mathbf{m}}$ and $\hat{\mathbf{V}}_{\mathbf{f}}$ are translated to pass through the male and female prototypes (or mean facial shapes), $\mathbf{P}_{\mathbf{m}}$ and $\mathbf{P}_{\mathbf{f}}$, respectively (figure 3.16). By enforcing the conditions $\mathbf{E}_{\mathbf{m}} \cdot \hat{\mathbf{V}}_{\mathbf{m}} = 0$ and $\mathbf{E}_{\mathbf{f}} \cdot \hat{\mathbf{V}}_{\mathbf{f}} = 0$, $\mathbf{E}_{\mathbf{m}}$ and $\mathbf{E}_{\mathbf{f}}$ may be calculated from geometrical considerations:

$$\mathbf{b}_j + \mathbf{E}_{\mathbf{m}} = \mathbf{P}_{\mathbf{m}} + \beta_m \hat{\mathbf{V}}_{\mathbf{m}} \tag{3.24}$$

$$\mathbf{b}_j + \mathbf{E_f} = \mathbf{P_f} + \beta_f \hat{\mathbf{V}}_\mathbf{f} \tag{3.25}$$

Using equation (3.24):

$$\beta_m = \frac{\mathbf{b}_j \cdot \hat{\mathbf{V}}_\mathbf{m} - \mathbf{P_m} \cdot \hat{\mathbf{V}}_\mathbf{m}}{\hat{\mathbf{V}}_\mathbf{m} \cdot \hat{\mathbf{V}}_\mathbf{m}} \tag{3.26}$$

Hence:

$$\mathbf{E_m} = \mathbf{P_m} + \left[\frac{\mathbf{b}_j \cdot \hat{\mathbf{V}}_\mathbf{m} - \mathbf{P_m} \cdot \hat{\mathbf{V}}_\mathbf{m}}{\hat{\mathbf{V}}_\mathbf{m} \cdot \hat{\mathbf{V}}_\mathbf{m}} \right] \hat{\mathbf{V}}_\mathbf{m} - \mathbf{b}_j \tag{3.27}$$

Similarly, using equation (3.25):

$$\beta_f = \frac{\mathbf{b}_j \cdot \hat{\mathbf{V}}_\mathbf{f} - \mathbf{P_f} \cdot \hat{\mathbf{V}}_\mathbf{f}}{\hat{\mathbf{V}}_\mathbf{f} \cdot \hat{\mathbf{V}}_\mathbf{f}} \tag{3.28}$$

Additionally:

$$\mathbf{E_f} = \mathbf{P_f} + \left[\frac{\mathbf{b}_j \cdot \hat{\mathbf{V}}_\mathbf{f} - \mathbf{P_f} \cdot \hat{\mathbf{V}}_\mathbf{f}}{\hat{\mathbf{V}}_\mathbf{f} \cdot \hat{\mathbf{V}}_\mathbf{f}} \right] \hat{\mathbf{V}}_\mathbf{f} - \mathbf{b}_j \tag{3.29}$$

Once these quantities have been calculated, it is possible to translate a face in the most appropriate direction in the shape space according to equation (3.23).

To visually demonstrate the existence of aging directions in the model space for male and female faces, the 2nd, 3rd and 4th components of each shape parameter vector for two groups of examples (ages 1 and 20 years respectively) are plotted in figure 3.17. A clear separation may be seen between the age groups due to clustering of individuals of similar age. In reality, the calculated aging axes exist in a high-dimensional space but this 3D approximation illustrates the principle by which they

are formed.
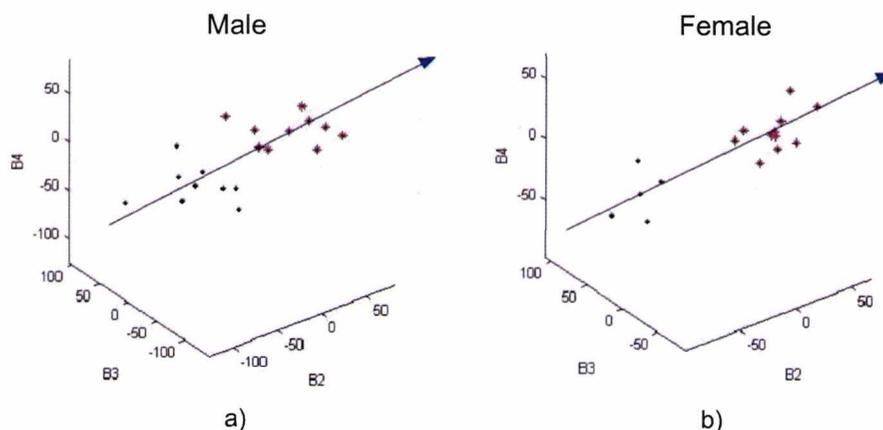
Figure 3.17: **A 3D illustration of the aging direction in the male and female sub-space.** The 2nd, 3rd and 4th components of the vector of shape parameters are plotted for *male* (a) and *female* (b) examples of age 1 (dots) and 20 years (stars). The line represents the aging direction, owing to the separation of the age groups within the space.

An aging axis for the texture variations was formed in an entirely analogous way to equation (3.13). A separate model for each sex was produced to avoid complications arising from a mixed sex model, such as the appearance of "ghost" facial hair on female subjects as aging proceeds. Subjects may then be aged according to the texture model as:

$$\mathbf{g}_j{}' = \mathbf{g}_j + \alpha_t \hat{\mathbf{V}}_{\mathbf{T}} \tag{3.30}$$

where $\mathbf{g}_j{}'$ is the vector of aged model parameters, $\mathbf{g}_j$ is the vector of original model parameters for the jth example, α_t controls the degree of aging and $\hat{\mathbf{V}}_{\mathbf{T}}$ is the length-normalised, texture aging axis. The vector of aged texture parameters, $\mathbf{g}_j{}'$, may then be recombined with the orthonormal principal components via equation (3.11). Once aging has proceeded according to both the shape and texture models,

the resultant aged texture map is warped to the aged shape using a Delaunay-based, piecewise, affine transformation (section 2.3.2) to produce an output face, age-progressed in both shape and texture.

3.6.3   Age-progression of Ethnic Groups

To age a subject according to their ethnic heritage, an analogous procedure to sections 3.6 and 3.6.2 must be followed. That is, if the shape model is constructed from a training set of different ethnicities, aging axes are formed for each of the represented ethnic groups. Hence, a White European face (see Appendix F), for example, should be aged according to the White European aging direction and a Black African according to the Black African aging direction so that an individual is age-progressed in the most appropriate manner for their racial group. In addition, sex axes are constructed and employed as previously. If a person of mixed ethnic heritage is presented, the Euclidean distance vector is calculated from each of the appropriate aging axes and the person aged according to these distances using an equivalent expression to equation (3.23). To minimise complications in this instance, it is necessary to separate the sexes. In an equivalent manner to the formation of separate sex texture models, separate ethnic texture models must also be constructed for each sex to ensure that aging proceeds realistically[14].

---

[14]A supplementary texture model, constructed upon both racial groups for each sex, may be employed if a subject is of mixed heritage so that age-progression may proceed in an analogous fashion to the shape, using the Euclidean distance from each axis.

3.7 Piecewise Aging

The approach to aging described in the previous section is based upon the assumption that the aging process is inherently linear. However, facial change with time is known to occur in a non-uniform fashion [12, 13, 14, 19, 17, 48, 18]. In particular, adolescence causes significant changes over a relatively short period of time with pronounced differences occurring between the sexes [14, 18, 22] (see sections 1.4 and 1.5). Bogin [12] identifies four main, developmental periods that occur between the ages of 0 and 20 years and describes in detail the changes and rates of growth and development occurring during these stages. These four stages are termed **Infancy** (0-3 years), **Childhood** (3-7 years), **Juvenile** (7-10 years) and **Adolescence** (10-20 years).

To produce a more accurate model for the facial aging of children and teenagers, a more comprehensive approach was devised, based upon these known developmental stages. In essence, a "piecewise" model, consisting of several linear aging directions, was calculated in shape space using faces restricted to specified age ranges. Using Database 2, the same facial images used in the model construction described in section 3.6 were pose and expression compensated using Method 2[15]. The resulting shape vectors were subsequently used to construct a mixed sex, shape model and equation (3.13) used to calculate four aging axes corresponding to the physiological developmental stages. To demonstrate the formation of these axes and their approximate directions in the model space, the prototype (mean) shape parameter vector for each age group was calculated and the 2nd, 3rd and 4th components plotted in three dimensions. This was performed for the male and female sub-spaces of Database 2 separately and figure 3.18 displays the respective results. The consensus

---

[15]Method 2 is arbitrarily used for pose and expression compensation since both Method 1 and 2 yield comparableresults.

axis is also shown to illustrate the differences in direction between the piecewise and consensus vectors and hence the possibility of a different outcome using this approach.
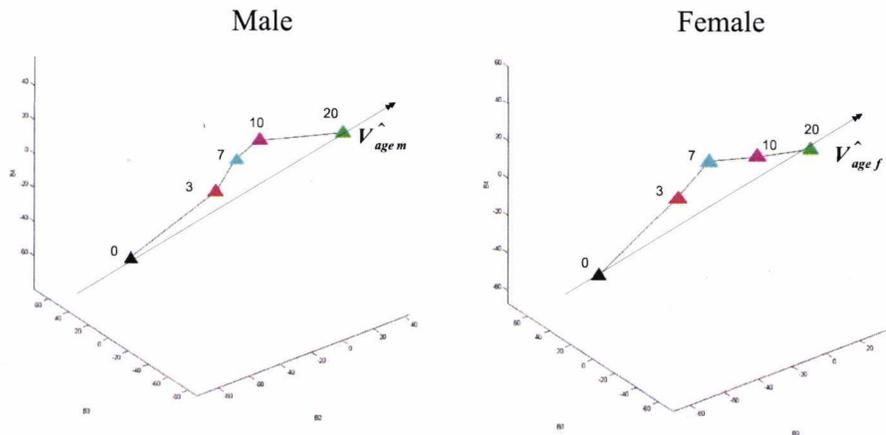
Figure 3.18: **A 3D illustration of the piecewise aging directions in the male and female sub-spaces respectively**. The 2nd, 3rd and 4th components of the vector of prototype shape parameters are plotted for each of the age groups in the *male* (left) and *female* (right) sub-spaces respectively. The lines depict the aging directions for each of the age groups with the consensus axis also given for comparison. This illustrates the differences in direction between the piecewise and consensus vectors and hence the possibility of a different outcome using the piecewise approach.

To understand how aging is achieved using the piecewise model, consider a 2 year old subject required to be aged by 8 years:

1. The subject, $\mathbf{b}_i$, is aged by 1 year according to the Infancy axis:

$$\mathbf{b}_i{}' = \mathbf{b}_i + \alpha_I \hat{\mathbf{V}}_{\mathbf{I}}$$

where α_I represents 1 year of aging in the direction of the Infancy axis, $\hat{\mathbf{V}}_{\mathbf{I}}$.

2. The resultant aged shape from the previous step is then aged by 4 years according to the Childhood axis:

$$\mathbf{b}_i{}'' = \mathbf{b}_i{}' + \alpha_C \hat{\mathbf{V}}_\mathbf{C}$$

where α_C represents 4 years of aging in the direction of the Childhood axis, $\hat{\mathbf{V}}_\mathbf{C}$.

3. Lastly, $\mathbf{b}_i{}''$ is aged by a further 3 years using the Juvenile axis to produce the final shape, $\mathbf{b}_i{}'''$, aged by 8 years in total using the piecewise approach:

$$\mathbf{b}_i{}''' = \mathbf{b}_i{}'' + \alpha_J \hat{\mathbf{V}}_\mathbf{J}$$

where α_J represents 3 years of aging in the direction of the Juvenile axis, $\hat{\mathbf{V}}_\mathbf{J}$.

A more sophisticated approach would employ equation (3.23) at each step to weight the male and female influences. However, this was not performed owing to a paucity of training examples in the formation of the piecewise axes. Therefore, examples were age-progressed according to their appropriate sex axes only.

For texture, separate piecewise models were formed for the male and female subjects. Piecewise axes were calculated in an equivalent manner to the shape model and subjects aged according to the relevant sex model. The resultant aged texture maps were then warped to the corresponding aged shapes using a Delaunay-based, piecewise, affine transformation as previously.

3.8 Person-Specific Aging Model

To age a face in the most accurate manner, *person-specific* factors must also be considered in addition to the average tendencies over a peer group sample. Person-specific factors include, for example, previous facial development trends for the given individual subject (which may reasonably be expected to have some correlation with future growth and development patterns) and the facial appearance and development of close relations (in particular, parents and siblings), which again may be useful predictors for aging [5, 12, 13]. Person-specific factors such as these are ultimately biological in origin, but no attempt to build a predictive model based on directly measurable biological factors, such as DNA, is currently possible.

In this section, a statistical approach is presented to incorporate such individual factors. This technique effectively combines statistically derived knowledge on these factors, together with the average trends expressed by the model described in sections 3.6 and 3.7. The basic philosophy of the approach is to treat the different influences as competing directions in the model space. That is, vectors are calculated for the *historical* (previous appearance), *consensus* and *parental* factors and the age-progressed vector computed as a weighted linear combination of these axes. Owing to the modelling of all available influences on aging, this more sophisticated method may reasonably be expected to produce more accurate results than either the consensus or piecewise approaches. The theoretical framework is described as follows.

Consider a training sample consisting of facial images of a large number of individuals at various ages. The aim is to progress a given image from its current age y_c to a target age, y_T. Figure 3.19 (a schematic representation of the model space) depicts the groups of faces required for this person-specific transformation, which

may be summarised as follows:

- Group O contains all faces in the sample at the same current age, y_c.

- Group T contains all faces in the sample at the same target age, y_T. In an ideal situation, Groups O and T will contain faces of the same sex and ethnic origin.

- Group I consists of the predicted appearances (from the previous appearances) at the target age, y_T. The corresponding true appearances at this age comprise group T.

- Group P consists of parental images of the individuals in group T[16]. Group P also contains the preferably same-sex parent of the subject to be aged.

In addition to possessing images of the training subjects at ages y_c and y_T, it is assumed that more than one image is available of each subject at ages $\leq y_c$. These are termed the *historical* images. Denoting the shape model vectors of the kth such image by $\mathbf{b}_k$, an average vector (displayed as a unit vector) over a previous time period leading up to y_c years may easily be calculated as:

$$\hat{\mathbf{V}}_\mathbf{i} = \sum_{k=1}^{c}(y_k - W_\mu)(\mathbf{b}_k - \boldsymbol{b_\mu}) \qquad (3.31)$$

where y_k is the age of the kth example, W_μ is the mean age of all training examples, $\boldsymbol{b_\mu}$ is the set of mean shape parameters and c is the number of historical images.

The vector $\hat{\mathbf{V}}_\mathbf{i}$ thus functions as the historical trend (termed the *historical vector*), indicating the average direction through the model space in previous years

---

[16]Ideally, to allow for the difference between sexes during aging, the parental images will be of the same sex as members of group T. However, it may be permissible to use the opposite-sex parent if an image of the same-sex one is unavailable.
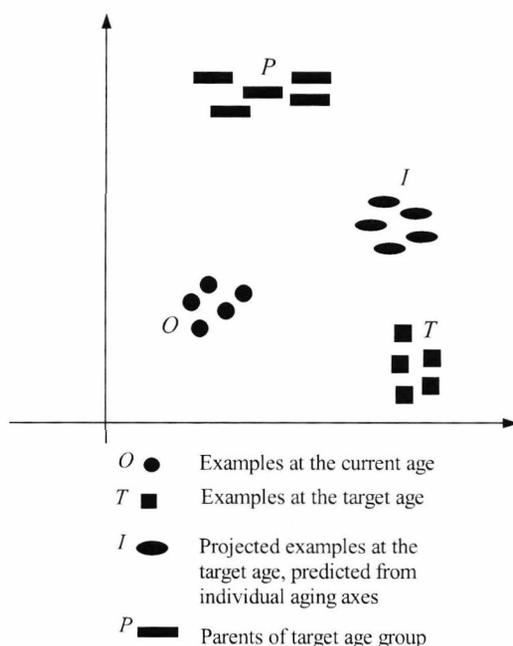
Figure 3.19: **Definition of groups in person-specific aging model**. Group O consists of images at the current age of y_c years, group T consists of images at the target age of y_T years, group I contains the predicted appearance of images in group O at the target age and group P contains parental images of those individuals in group T.

for each subject in group O. Calculation of the vector $\hat{\mathbf{V}}_i$ and subsequent use of equation (3.16) enables a prediction to be made of the appearance of each member in the training sample at the target age y_T, were the historical trend to be faithfully followed. Hence, group I is constructed as each member of group O is advanced to the target age using the predictions of the individual historical vectors.

An average vector, $\hat{\mathbf{V}}_c$, for the change between y_c and y_T years, which represents the mean change in facial appearance between the two groups O and T is calculated as:

$$\hat{\mathbf{V}}_{\mathbf{c}} = \sum_{i=c}^{T} (y_i - W_\mu)(\mathbf{b}_i - \boldsymbol{b_\mu}) \qquad (3.32)$$

Thus, every example between y_c and y_T years is weighted by its mean-subtracted age and a sum formed. The direction $\hat{\mathbf{V}}_{\mathbf{c}}$ is referred to as the *consensus vector*.

Finally, the *parental vector* denoted by $\hat{\mathbf{V}}_{\mathbf{p}}$, is given by the vector difference between the subject to be aged, $\mathbf{z}$ (from group O) and its parent at the target age, $\mathbf{p}$ (from group P):

$$\hat{\mathbf{V}}_{\mathbf{p}} = \mathbf{p} - \mathbf{z} \qquad (3.33)$$

To depict these axes schematically in 2D, consider figure 3.20. In this figure, $\mathbf{z}$ represents the subject face at the current age (from group O), $\mathbf{z_h}$ represents $\mathbf{z}$ at some previous age[17], $\mathbf{c}$ is the actual (unknown) appearance at the target age, $\mathbf{i}$ is the projected appearance at the target age calculated using the historical vector and $\mathbf{p}$ represents the parent of $\mathbf{z}$.

---

[17]It is desirable to include as many historical examples as possible in order to calculate a well-defined historical vector.
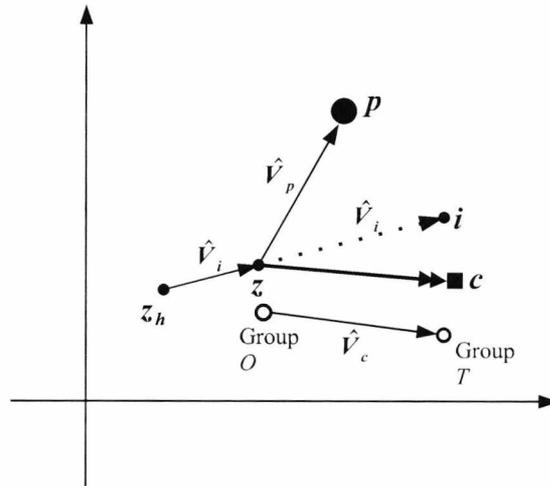
Figure 3.20: **Formation and use of historical, consensus and parental vectors employed for aging a subject** — z_h represents the subject, z, at some previous age, c is the actual (unknown) appearance at the target age, i is the projected appearance at the target age calculated using the historical vector and p represents the parent of z.

3.8.1 Linear Model

The age-progressed vector, c, is modelled as a weighted linear combination of $\hat{V}_c$, $\hat{V}_i$ and $\hat{V}_p$:

$$c = z + w_1 \hat{V}_c + w_2 \hat{V}_i + w_3 \hat{V}_p \tag{3.34}$$

That is, the aged face may be estimated by using a combination of the consensus information between the initial and target ages (embodied by $\hat{V}_c$), the individual information from the historical images (embodied by $\hat{V}_i$) and the parental information specific to the subject and its parent at the target age (embodied by $\hat{V}_p$).

To optimise the estimate of w_1, w_2 and w_3, it is considered that, as far as possible, the age-progressed vector, c, should simultaneously satisfy three criteria:

- It should be a "typical" member of the distribution of faces at the target age
  — this encompasses the tendency of an individual to follow a prototypical
  trend based on the aging patterns of a relevant peer group of faces of the same
  sex and ethnic origin.

- It should be a "typical" member of the predicted distribution resulting when
  the sample is projected to the target age according to the historical trend —
  this utilises the propensity of an individual to continue along an aging direction
  determined by its appearance at previous ages.

- It should be a "typical" member of the distribution produced by considering
  the difference vectors between individuals at the target age (group T) and their
  respective parents (group P), termed the "parent-to-child" distribution — this
  describes the predisposition of an individual to resemble close genetic relatives
  such as a parent. The parental information specific to the subject is embodied
  by the parental vector, $\hat{\mathbf{V}}_{\mathbf{p}}$. However, to calculate the weighting, w_3, for this
  axis and therefore a measure of the parental influence on the subject, example
  differences between members of group T and their corresponding parents in
  group P are used. This is because it is desirable for the age-progressed vector,
  $\mathbf{c}$, to resemble its parent in the typical manner in which members of the target
  group resemble their parents[18].

By considering members of the groups depicted in figure 3.19, it is possible
to form probability density estimates, which allow precise, maximum-likelihood
interpretation to be applied to these criteria. The very purpose of PCA is to
produce model parameters which are *independently* distributed within the vector

---

[18]Further incorporation of person-specific parental influences could reasonably include consideration of the vector difference between the child at the current age and its parent at the same age. This could then be used to provide additional person-specific aging.

space. The multivariate normal distribution for independent variables, denoted by $\mathbf{y} = [y_1, y_2 \cdots y_N]$, is given by:

$$p(\mathbf{y}; \boldsymbol{\mu_y}, \mathbf{C_y}) = A \exp \left\{ -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu_y})^T \mathbf{C_y}^{-1}(\mathbf{y} - \boldsymbol{\mu_y}) \right\} \tag{3.35}$$

where $A = \frac{1}{2\pi^{N/2}|\mathbf{C_y}|^{1/2}}$, $\mathbf{C_y}$ is a *diagonal* covariance matrix and $\boldsymbol{\mu_y}$ is the set of mean model parameters.

Similarly:

- the *target* distribution of the model vectors within group T is denoted $p_T(\mathbf{c}; \boldsymbol{\mu_T}, \mathbf{C_T})$, where $\boldsymbol{\mu_T}$ and $\mathbf{C_T}$ are the set of mean model parameters and covariance matrix of group T respectively. This distribution essentially provides the probability of the aged vector $\mathbf{c}$ belonging to the target age group.

- the *target difference* distribution, given by the difference vectors between the vector of a subject at the target age (group T) and its corresponding predicted value at the target age (group I), is denoted $p_D(\boldsymbol{\delta}; \boldsymbol{\mu_\delta}, \mathbf{C_\delta})$, where $\boldsymbol{\delta} = \mathbf{c} - \mathbf{i}$ and $\boldsymbol{\mu_\delta}$ and $\mathbf{C_\delta}$ are the set of mean model parameters and covariance matrix of the group respectively. This distribution essentially provides the probability of a given difference vector, $\boldsymbol{\delta}$, being a suitable difference between a member of the target age group T and the equivalent, predicted member in group I, determined from the historical vector.

- the *parent-to-child* distribution, given by the difference vectors between a parent model parameter vector in group P and its corresponding offspring in the target group, T, is denoted $p_D(\boldsymbol{\Delta}; \boldsymbol{\mu_\Delta}, \mathbf{C_\Delta})$, where $\boldsymbol{\Delta} = \mathbf{c} - \mathbf{p}$ and $\boldsymbol{\mu_\Delta}$ and $\mathbf{C_\Delta}$ are the set of mean model parameters and covariance matrix of the group respectively. This distribution essentially provides the probability of a

given difference vector, $\boldsymbol{\Delta}$, being a suitable difference between a member of the target age group and its parent.

The task is to estimate the free parameters w_1, w_2 and w_3 in equation (3.34) so that $\mathbf{c}$ may belong equally to all distributions in a maximum likelihood sense. Accordingly, the joint probability (also known as the "likelihood") is required, which corresponds to the product of the probability density functions, $L = p_T(\mathbf{c})p_D(\boldsymbol{\delta})p_D(\boldsymbol{\Delta})$. The appropriate vector $\mathbf{c}$ is then found such that this product is *maximised*. That is, we seek to maximise:

$$L = k\exp\left\{-\frac{1}{2}\phi - \frac{1}{2}\gamma - \frac{1}{2}\kappa\right\} \tag{3.36}$$

where $\phi = (\mathbf{c} - \boldsymbol{\mu_T})^T \mathbf{C_T}^{-1}(\mathbf{c} - \boldsymbol{\mu_T})$, $\gamma = (\boldsymbol{\delta} - \boldsymbol{\mu_\delta})^T C_\delta^{-1}(\boldsymbol{\delta} - \boldsymbol{\mu_\delta})$, $\kappa = (\boldsymbol{\Delta} - \boldsymbol{\mu_\Delta})^T C_\Delta^{-1}(\boldsymbol{\Delta} - \boldsymbol{\mu_\Delta})$ and k is a multiplicative constant. Taking logarithms:

$$-2\ln L = k\left\{\phi + \gamma + \kappa\right\} \tag{3.37}$$

Maximising the product of the distributions is equivalent to minimising the sum of the exponential terms and hence minimising the following cost function:

$$Q = -2\ln L = -2\ln\left[p_T(\mathbf{c})p_D(\boldsymbol{\delta})p_D(\boldsymbol{\Delta})\right] \tag{3.38}$$

Substituting our known expressions for $p_T(\mathbf{c})$, $p_D(\boldsymbol{\delta})$ and $p_D(\boldsymbol{\Delta})$ into equation (3.38) yields:

$$Q(w_1, w_2, w_3) = k[\phi + \gamma + \kappa] \tag{3.39}$$

To find the optimal solution for $\mathbf{c}$, standard calculus methods are used, taking partial derivatives of Q with respect to w_1, w_2 and w_3 respectively and setting these to zero. We obtain:

$$\frac{\partial Q}{\partial w_1} = 2\hat{\mathbf{V}}_{\mathbf{c}}^T \mathbf{C_T}^{-1}(\mathbf{c} - \boldsymbol{\mu}_T) + 2\hat{\mathbf{V}}_{\mathbf{c}}^T \boldsymbol{C}_{\boldsymbol{\delta}}^{-1}(\boldsymbol{\delta} - \boldsymbol{\mu}_{\boldsymbol{\delta}}) + 2\hat{\mathbf{V}}_{\mathbf{c}}^T \boldsymbol{C}_{\boldsymbol{\Delta}}^{-1}(\boldsymbol{\Delta} - \boldsymbol{\mu}_{\boldsymbol{\Delta}}) = 0 \quad (3.40)$$

Similarly,

$$\frac{\partial Q}{\partial w_2} = 2\hat{\mathbf{V}}_{\mathbf{i}}^T \mathbf{C_T}^{-1}(\mathbf{c} - \boldsymbol{\mu}_T) + 2\hat{\mathbf{V}}_{\mathbf{i}}^T \boldsymbol{C}_{\boldsymbol{\delta}}^{-1}(\boldsymbol{\delta} - \boldsymbol{\mu}_{\boldsymbol{\delta}}) + 2\hat{\mathbf{V}}_{\mathbf{i}}^T \boldsymbol{C}_{\boldsymbol{\Delta}}^{-1}(\boldsymbol{\Delta} - \boldsymbol{\mu}_{\boldsymbol{\Delta}}) = 0 \quad (3.41)$$

and:

$$\frac{\partial Q}{\partial w_3} = 2\hat{\mathbf{V}}_{\mathbf{p}}^T \mathbf{C_T}^{-1}(\mathbf{c} - \boldsymbol{\mu}_T) + 2\hat{\mathbf{V}}_{\mathbf{p}}^T \boldsymbol{C}_{\boldsymbol{\delta}}^{-1}(\boldsymbol{\delta} - \boldsymbol{\mu}_{\boldsymbol{\delta}}) + 2\hat{\mathbf{V}}_{\mathbf{p}}^T \boldsymbol{C}_{\boldsymbol{\Delta}}^{-1}(\boldsymbol{\Delta} - \boldsymbol{\mu}_{\boldsymbol{\Delta}}) = 0 \quad (3.42)$$

Using equation (3.34) in addition to $\boldsymbol{\delta} = \mathbf{c} - \mathbf{i}$ and $\boldsymbol{\Delta} = \mathbf{c} - \mathbf{p}$, solving (3.40), (3.41) and (3.42) yields equations of the form:

$$aw_1 + bw_2 + cw_3 = \gamma_1 \quad (3.43)$$

$$bw_1 + dw_2 + ew_3 = \gamma_2 \quad (3.44)$$

$$cw_1 + ew_2 + fw_3 = \gamma_3 \quad (3.45)$$

where $a = \hat{\mathbf{V}}_{\mathbf{c}}^T \Sigma \hat{\mathbf{V}}_{\mathbf{c}}$, $b = \hat{\mathbf{V}}_{\mathbf{c}}^T \Sigma \hat{\mathbf{V}}_{\mathbf{i}}$, $c = \hat{\mathbf{V}}_{\mathbf{c}}^T \Sigma \hat{\mathbf{V}}_{\mathbf{p}}$, $d = \hat{\mathbf{V}}_{\mathbf{i}}^T \Sigma \hat{\mathbf{V}}_{\mathbf{i}}$, $e = \hat{\mathbf{V}}_{\mathbf{i}}^T \Sigma \hat{\mathbf{V}}_{\mathbf{p}}$, $f = \hat{\mathbf{V}}_{\mathbf{p}}^T \Sigma \hat{\mathbf{V}}_{\mathbf{p}}$ and $\Sigma = \mathbf{C_T}^{-1} + \boldsymbol{C}_{\boldsymbol{\delta}}^{-1} + \boldsymbol{C}_{\boldsymbol{\Delta}}^{-1}$.

In addition:

$$\gamma_1 = \tag{3.46}$$

$$\hat{\mathbf{V}}_{\mathbf{c}}^T \mathbf{C_T}^{-1}(\boldsymbol{\mu}_T - \mathbf{z}) + \hat{\mathbf{V}}_{\mathbf{c}}^T C_\delta^{-1}(\boldsymbol{\mu}_\delta + \mathbf{i} - \mathbf{z}) +$$

$$\hat{\mathbf{V}}_{\mathbf{c}}^T C_\Delta^{-1}(\boldsymbol{\mu}_\Delta + \mathbf{p} - \mathbf{z})$$

$$\gamma_2 = \tag{3.47}$$

$$\hat{\mathbf{V}}_{\mathbf{i}}^T \mathbf{C_T}^{-1}(\boldsymbol{\mu}_T - \mathbf{z}) + \hat{\mathbf{V}}_{\mathbf{i}}^T C_\delta^{-1}(\boldsymbol{\mu}_\delta + \mathbf{i} - \mathbf{z}) +$$

$$\hat{\mathbf{V}}_{\mathbf{i}}^T C_\Delta^{-1}(\boldsymbol{\mu}_\Delta + \mathbf{p} - \mathbf{z})$$

$$\gamma_3 = \tag{3.48}$$

$$\hat{\mathbf{V}}_{\mathbf{p}}^T \mathbf{C_T}^{-1}(\boldsymbol{\mu}_T - \mathbf{z}) + \hat{\mathbf{V}}_{\mathbf{p}}^T C_\delta^{-1}(\boldsymbol{\mu}_\delta + \mathbf{i} - \mathbf{z}) +$$

$$\hat{\mathbf{V}}_{\mathbf{p}}^T C_\Delta^{-1}(\boldsymbol{\mu}_\Delta + \mathbf{p} - \mathbf{z})$$

Equations (3.43) to (3.48) are the main results of this analysis. The linear equations (3.43), (3.44) and (3.45) may be solved simultaneously to calculate weighting factors w_1, w_2 and w_3 in terms of the known parameters expressed by equations (3.46) to (3.48). Equation (3.34) may then be used to produce the age-progressed appearance. Shape and texture are treated independently with the final aged texture map warped to the aged shape to complete the facial synthesis.

3.8.2 Historical/consensus approach

In the absence of appropriate parental data, the use of another close relative, such as a full-blood sibling, may be permissible to capture any familial influence on aging. If no parental or sibling images are available, the age-progression can only be modelled with the historical and consensus vectors and equation (3.34) reduces to a two variable model:

$$\mathbf{c} = \mathbf{z} + w_1 \hat{\mathbf{V}}_\mathbf{c} + w_2 \hat{\mathbf{V}}_\mathbf{i} \tag{3.49}$$

Upon minimisation of an equivalent cost function to (3.39), $Q(w_1, w_2)$, w_1 and w_2 may be found from the simultaneous solution of two equations:

$$a w_1 + b w_2 = \gamma_1 \tag{3.50}$$

$$b w_1 + c w_2 = \gamma_2 \tag{3.51}$$

where, in this instance, $a = \hat{\mathbf{V}}_\mathbf{c}^T \Sigma \hat{\mathbf{V}}_\mathbf{c}$, $b = \hat{\mathbf{V}}_\mathbf{c}^T \Sigma \hat{\mathbf{V}}_\mathbf{i}$, $c = \hat{\mathbf{V}}_\mathbf{i}^T \Sigma \hat{\mathbf{V}}_\mathbf{i}$ and $\Sigma = \mathbf{C_T}^{-1} + \mathbf{C}_\delta^{-1}$. In addition:

$$\gamma_1 = \hat{\mathbf{V}}_\mathbf{c}^T \mathbf{C_T}^{-1} (\boldsymbol{\mu}_T - \mathbf{z}) + \hat{\mathbf{V}}_\mathbf{c}^T \mathbf{C}_\delta^{-1} (\boldsymbol{\mu}_\delta + \mathbf{i} - \mathbf{z}) \tag{3.52}$$

$$\gamma_2 = \hat{\mathbf{V}}_\mathbf{i}^T \mathbf{C_T}^{-1} (\boldsymbol{\mu}_T - \mathbf{z}) + \hat{\mathbf{V}}_\mathbf{i}^T \mathbf{C}_\delta^{-1} (\boldsymbol{\mu}_\delta + \mathbf{i} - \mathbf{z}) \tag{3.53}$$

It is straightforward to show that w_1 and w_2 are given by:

$$w_1 = \frac{c\gamma_1 - b\gamma_2}{(ac - b^2)} \tag{3.54}$$

$$w_2 = \frac{a\gamma_2 - b\gamma_1}{(ac - b^2)} \tag{3.55}$$

For purposes of comparison, a model based only on the historical axis can be considered. This is easily achieved by assigning a value of zero to the consensus axis, $\hat{\mathbf{V}}_{\mathbf{c}}$, in equation (3.49), yielding:

$$\mathbf{c} = \mathbf{z} + w_2\hat{\mathbf{V}}_{\mathbf{i}} \tag{3.56}$$

where $w_2 = \gamma_2/c$, $c = \hat{\mathbf{V}}_{\mathbf{i}}^T\mathbf{C}_\delta^{-1}\hat{\mathbf{V}}_{\mathbf{i}}$ and $\gamma_2 = \hat{\mathbf{V}}_{\mathbf{i}}^T\mathbf{C}_\delta^{-1}(\boldsymbol{\mu}_\delta + \mathbf{i} - \mathbf{z})$.

3.8.3 Age-Regression

The aging techniques described for children may equally well be applied in reverse to make a face appear younger than its current age. To achieve such an age-regression, the aging coefficients in the models described by equations (3.16), (3.23) and (3.30) are set to negative values.

3.9 Chapter Summary

In this chapter, three models of varying sophistication for age-progression were presented. Each model requires preliminary image processing to be carried out on the images prior to model production. Accordingly, two novel methods for pose and expression compensation were firstly discussed. Furthermore, to reduce the textural artefacts from these procedures, an approach to texture compensation was outlined. As a final stage in the pre-processing, a compensation technique for illumination was described to allow adjustment of the texture maps to appear more uniformly lit.

In the first approach to age-progression, the *consensus* methodology was outlined and the aging axis computed via a weighted sum of model parameter vectors. To model the influence of both genders, the Euclidean distance of a subject from each

gender axis was employed. A more sophisticated approach (accounting for the non-linear nature of the aging process) was then outlined in the form of the *piecewise* technique, in which four linear axes were constructed based upon the periods of developmental growth in children and teenagers. Since the piecewise axes are derived from physiological influences, the results obtained may reasonably be expected to produce more accurate results than the consensus approach.

Finally, the most comprehensive approach to age-progression was presented to incorporate person-specific factors. This accounted for previous facial development trends and the tendency of individuals to resemble close (genetic) relatives. A theoretical framework was outlined to combine these factors, along with consensus information derived from a population. By considering these influences as competing directions in the model space, expressions were derived for the optimal amount of each component to be added to a subject vector to effect the age-progression.

In the next chapter, the computational implementation of the three age-progression algorithms is presented and their accuracy quantitatively assessed.

Chapter 4

Implementation of Age-Progression Algorithms

In this chapter, the approaches outlined in sections 3.6, 3.7 and 3.8 are implemented to demonstrate the production of age-progressed images. The results of these procedures are visually and quantitatively compared with the true appearance at the target age (the *target image*) to assess the degree to which the subject retains its unique identity as it is translated through the model space.

To demonstrate the flexibility of the PCA model for age-progression, related work on age *prediction* is presented. The consensus axis is used to predict the age of a subject from its respective shape and texture model representation. The technique of Fisher Linear Discriminant analysis is then introduced as an alternative methodology for age prediction. The results of the two approaches are compared and conclusions drawn regarding their relative suitability for the application.

4.1 Consensus Approach — Database 1

Figures 4.1 and 4.2 depict typical results of aging example faces from Database 1 (section 3.1) using the *consensus* approach. Equation (3.16) was applied in increments of two years, up to and including twenty years[1]. This range was chosen to examine the effect of both modest and larger amounts of aging. To confirm the reversibility of the method, figure 4.3 depicts the results of a typical age-regression (section 3.8.3).
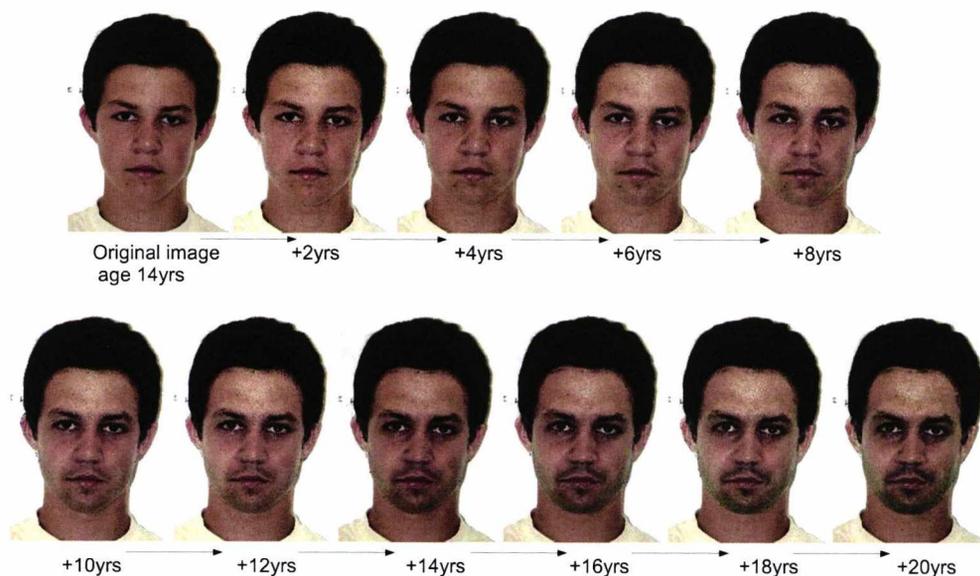
Figure 4.1: **Example result of aging an in-sample male**. Face 1.1 (14 years of age) is age-progressed by 2 year increments, up to age 34.

---

[1]The resultant aged faces have been incorporated into the original images to provide context in terms of hair, clothing and background. This was achieved by a Delaunay-based warping procedure of the aged facial shape and texture patch into the original image at the face location. A standard median filter was used to reduce the appearance of the join lines during this process.
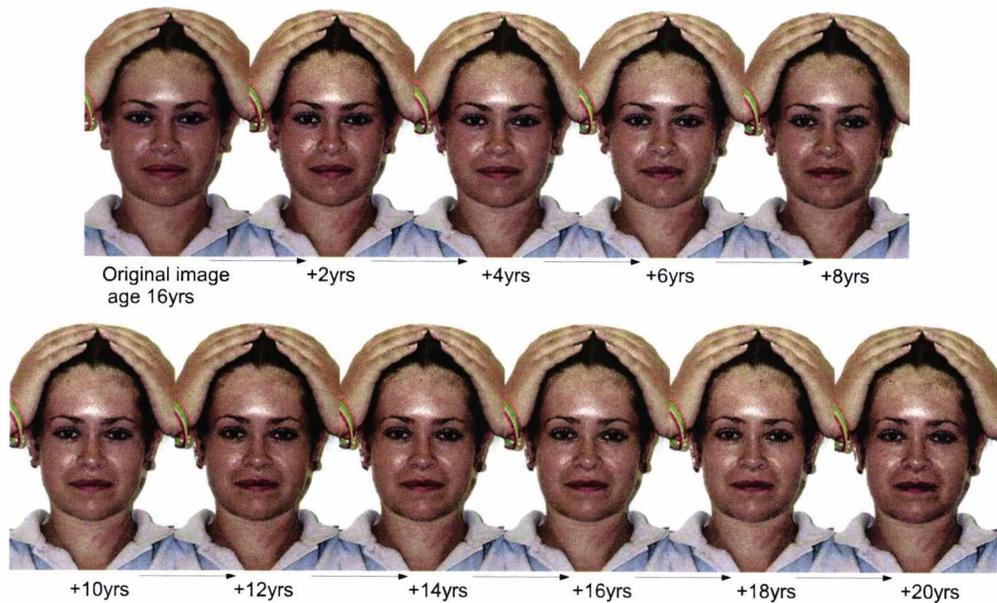
Figure 4.2: **Example result of aging an in-sample female**. Face 1.2 (16 years of age) is age-progressed by 2 year increments, up to age 36. The subject was asked to hold their hair back from the forehead to prevent partial occlusion of the face.
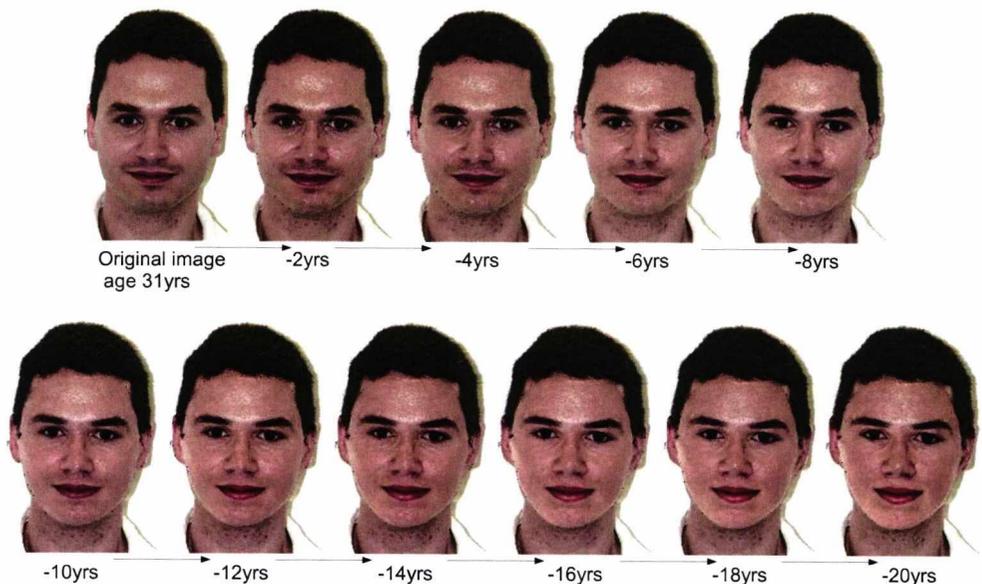
Figure 4.3: **Example result of age-regression on an in-sample male**. Face 1.3 (31 years of age) is age-regressed by 2 year increments, up to age 11.

Each face in figures 4.1 and 4.2 subjectively appears to have increased in age over the twenty year period. This is in the opinion of this author only[2] and is indicated by several well-documented visual cues. For instance, the eyes appear smaller and more deeply set with the skin under the eyes appearing thinner and darker [5], the lips have become thinner [23, 24] and the lower face has become generally broader. Conversely, the age-regression (figure 4.3) results in a face that appears progressively younger.

It is important to note that significant visual changes may only be explicitly observed for relatively substantial amounts of age-progression or regression (for instance, ten or more years). This is because the aging axis is defined using a small number of training faces with a marked spread in age — hence, since the axis is relatively well-defined near the end-points, large increments of aging yield notable changes to facial appearance. However, since the axis defines an *average* direction through the model space between young adolescents and mature adults, age-progression by successive two year increments produces approximately uniform, subtle changes. This may be seen in figure 4.1, where the face age-progressed by twenty years appears more age-accurate than the face age-progressed by ten years (which appears younger, in the opinion of this author[3], than its target age of twenty four years due to the uniform changes modelled at each two year increment of aging). Since, in reality, facial changes with age are *non-uniform* (particularly in adolescence [13, 14, 19, 17]), Database 1 is not considered to be the optimal training set for the age-progression of subjects under the age of twenty years. In addition, the consensus model itself requires revision in order to more accurately model the aging of children

---

[2]There were no additional checks performed on these results. This is because the study performed using Database 1 was an initial feasibility study only to check the basic viability of the consensus age-progression approach.

[3]Again, this was not checked with other observers, owing to the purpose of using Database 1 as a basic check that the consensus approach yielded sensible results.

and teenagers.

To address these issues, a database was required containing more training examples at intermediate ages. Owing to the formation of the axis as a weighted sum of model parameter vectors, inclusion of such examples would change the axis direction in the space (see figure 4.4 for a 2D graphical depiction) and this would allow more age-accurate results to be achieved[4]. In addition, such a training set would permit age-progression according to the piecewise approach.
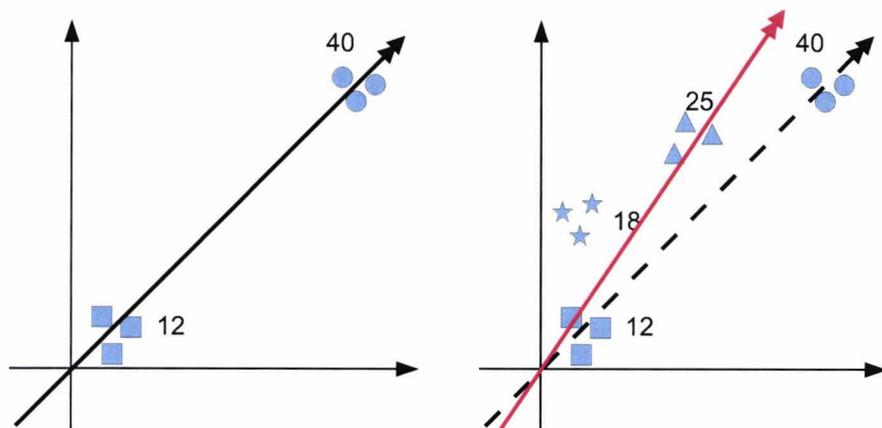
Figure 4.4: **Inclusion of examples at intermediate ages changes the direction of the aging axis**. Database 1 contains examples of young adolescents and mature adults (left), such that the calculated aging axis allows the average facial changes to be computed. However, owing to the formation of the axis as a weighted sum of model parameter vectors, inclusion of examples of intermediate age will change its direction (right) and its use will produce more accurate results.

---

[4]In reality, a database was required containing subjects between 0 and 20 years of age, such that the facial changes of childhood and adolescence could be modelled.

4.2 Consensus, Piecewise & Historical/Consensus Approaches — Database 2

In the course of this research, an alternative training set, referred to as Database 2, was obtained (see section 3.1), containing 658 faces, aged 0 to 20 years. Despite the lack of uniformity in these images[5], Database 2 was considered more appropriate for modelling the age-progression of children than Database 1, owing to the more suitable age-ranges therein. Hence, the construction of more accurate aging axes was possible (see section 4.1).

To implement the consensus approach on this training set, equation (3.16) was applied to example faces. Similarly, analogous equations to (3.16) were applied to the same subjects to implement the piecewise model (using the Infancy, Childhood, Juvenile and Adolescence axes — see section 3.7). To examine the application of the historical/consensus approach, equation (3.49) was employed (see section 3.8.2) and, finally, equation (3.56) was used to compare the results obtained using the historical approach alone with the aforementioned approaches. Figures 4.5 to 4.7 display the results from these procedures. Where possible and appropriate, aging sequences are displayed, along with the corresponding targets.

---

[5]Pose, expression, illumination, resolution and image quality varied considerably as compared to Database 1.

Figure 4.5: **Example results of age-progression using the consensus, piecewise, historical/consensus and historical approaches respectively**. Column A): Original image, B) Face aged using the consensus model, C) Face aged using the piecewise model, D) Face aged using the historical/consensus approach, E) Face aged using the historical direction only, F) Subject at the target age.
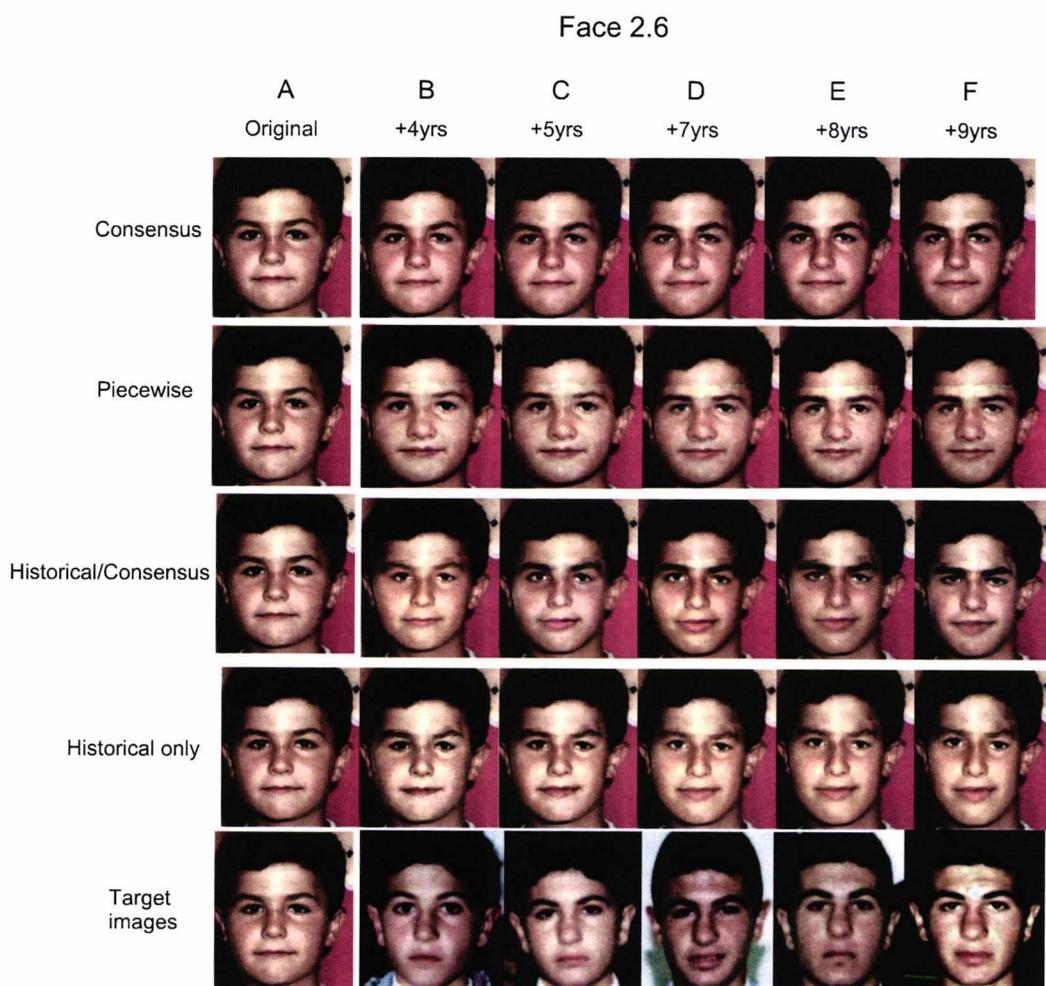
Figure 4.6: **Sequential age-progression of an in-sample male using the consensus, piecewise, historical/consensus and historical approaches respectively**. Column A) Original image, Face 2.6, 7 years old, B) Result of age-progression by 4 years, C) 5 years of aging, D) 7 years of aging, E) 8 years of aging, F) 9 years of aging. The rows represent the different age-progression methodologies, with the bottom row displaying the target images for comparison with each age-progressed image.
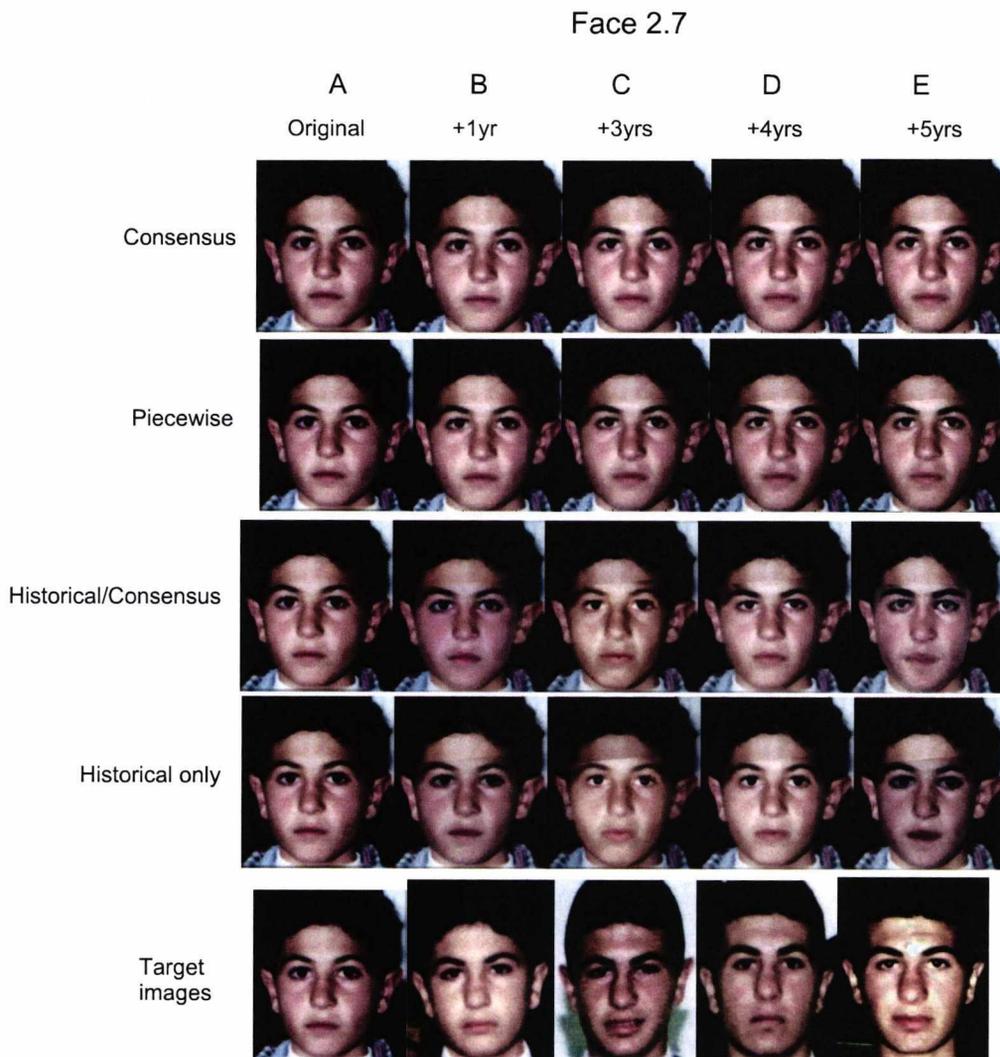
Figure 4.7: **Sequential age-progression of an in-sample male using the consensus, piecewise, historical/consensus and historical approaches respectively.** Column A) Original image, Face 2.7, 11 years old, B) Result of age-progression by 1 year, C) 3 years of aging, D) 4 years of aging, E) 5 years of aging. The rows represent the different age-progression methodologies, with the bottom row displaying the target images for comparison with each age-progressed image.

4.2.1   Discussion

In this author's opinion, some of the results produced using the approaches presented in sections 3.6 to 3.8.2 are plausible and realistic. However, some of the results are much poorer, for reasons that will be explained below.

A priori, the historical/consensus approach is expected to be the most accurate method since it incorporates person-specific information in addition to trends from a peer group. However, in practice, this is often not the case. Due to a paucity of historical examples, the historical axis may not be well-defined and therefore may not allow accurate prediction of a subject's future appearance. This effect can be seen particularly in figure 4.5, face 2.5, image d. In addition, if the historical images are of varying quality and captured under different illumination conditions, the problem is compounded and the historical path through the model space will again be poorly defined. This may lead to inaccuracies in the age-progression of the texture, as can be seen in figure 4.5, face 2.4 and 2.5, image e. In conclusion, reliable results may only be achieved using this method if a sufficient number of good quality historical images are available.

Similarly, the piecewise approach is expected to produce more accurate results than the consensus approach since the axes are constructed upon known physiological growth stages. However, the piecewise axes may also be poorly defined compared to the consensus axis due to the relative sparsity of training examples. Hence, the results obtained using this methodology may not be as reliable or as visually pleasing as counterpart results achieved using the consensus approach (in figure 4.5, face 2.5, compare image c with image b).

Owing to the consistently plausible nature of the consensus results, *superficial*

examination[6] would seem to imply that this approach is the most appropriate to use for age-progression. However, if a larger number of suitable training examples[7] was available, the piecewise approach should be employed in preference, owing to its more rigorous modelling of the non-linear nature of aging. More importantly, where suitable individual information (such as numerous historical, parental and/or sibling images) is available, this should always be additionally incorporated to ensure that aging proceeds in accordance with person-specific influences.

To visually assess the accuracy of the approaches, the results were compared to the target images using the author's perception and were found to be broadly similar, particularly for the consensus approach. However, owing to the diverse nature of the training images in terms of image type, resolution and non-uniform illumination, these comparisons were not straightforward. Additionally, despite careful efforts to neutralise pose and expression (see sections 3.2.1 and 3.2.2), perfect compensation was not possible, further exacerbating the problem of comparison with target faces. The results of *quantitative* comparisons over a sample of images will be given in section 4.3.

4.3 Quantitative Assessment of the Results

4.3.1 Shape comparison

To obtain an objective measure of the accuracy of the age-progressed images, a quantitative comparison was made between them and their corresponding target images, beginning with a shape comparison, as follows:

---

[6]The term "superficial examination" is used to indicate the conclusions drawn upon initial examination of the results and comparison of the results from the different techniques. In these circumstances, the consensus approach appears to produce the most plausible results.

[7]Ideally images of uniform illumination and high resolution containing faces at frontal pose and neutral expression.

1. The root mean square error (RMSE) was calculated between the shape vector of an age-progressed subject ($\mathbf{d_{aged}}$) and the aligned shape vector of the corresponding target face,[8] ($\mathbf{d_{tar}}$), according to:

$$e_t = \sqrt{\frac{1}{q}\sum_{i=1}^{q}(|\mathbf{d_{aged}} - \mathbf{d_{tar}}|^2)} \tag{4.1}$$

   where each shape vector comprises q elements.

2. An equivalent equation to (4.1) was used to compute the RMSE, e_o, between $\mathbf{d_{aged}}$ and the shape vector of a variety of other faces at the target age, known to be of different identity ($\mathbf{d_{other}}$). These faces were chosen to be all examples in the training set (Database 2) at the target age, which were of the same sex as the age-progressed subject but of different identity, such that a variety of different subject faces were included[9].

3. The calculated value, e_t, was compared with each obtained value of e_o.

For the preservation of subject identity throughout the age-progression, e_t should be consistently *smaller* than e_o. To determine whether this was the case, the rank of e_t in relation to each e_o was plotted. This is given in figure 4.8.

---

[8]For this calculation to be effected, an initial pose and expression compensation (see sections 3.2.1 and 3.2.2) of the target face was necessary to orient its shape vector to frontal pose and neutral expression. The age-progressed face automatically appears in this orientation due to initial pose and expression compensation prior to application of the desired aging algorithm.

[9]This may have included some examples that were of similar appearance to the true target face. However, it is hoped that the effect of this was minimal, owing to the inclusion of many other, non-similar examples.
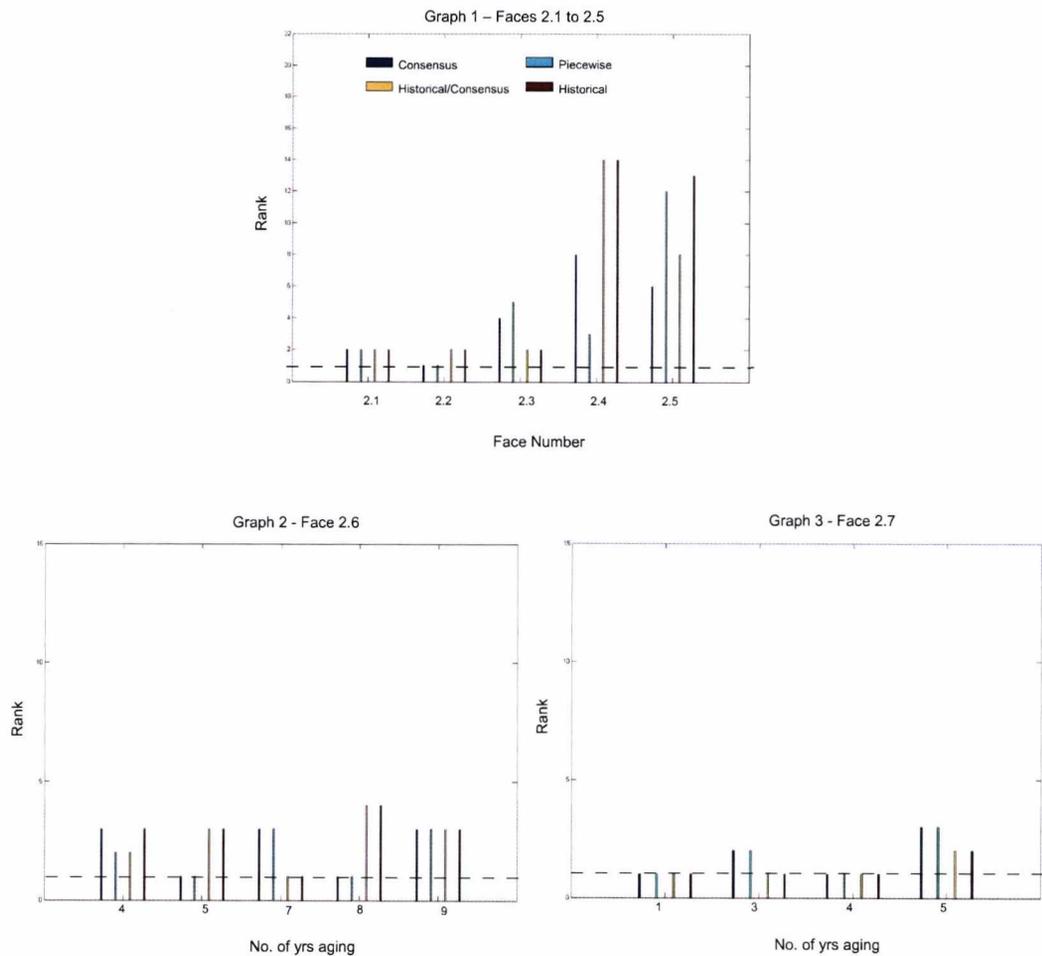
Figure 4.8: **Graphical results of the rank of e_t in relation to each e_o (shape comparisons) for faces 2.1 to 2.7**. Graph 1 relates to the age-progressed images depicted in figure 4.5 (faces 2.1 to 2.5). Similarly, Graphs 2 and 3 relate to the sequences of age-progression displayed in figures 4.6 (face 2.6) and 4.7 (face 2.7) respectively. The dashed line represents a rank equal to unity (the ideal rank of e_t compared to each e_o).

From examination of figure 4.8, it may be observed that there is generally a smaller RMSE between $\mathbf{d_{aged}}$ and $\mathbf{d_{tar}}$ than between $\mathbf{d_{aged}}$ and $\mathbf{d_{other}}$ — that is, in general, $e_t < e_o$ and the calculated rank is low. Hence, the faces aged using the models are closer in shape to the corresponding target faces than to other faces of different identities at the target age within the limits of this study and a number of uncontrollable factors. *We broadly conclude that the faces successfully retain their identity throughout the age-progression.*

Certain anomalous results in figure 4.8 may be explained as follows. The RMSE is a simple, global measure of the difference between shape vectors in which all landmarks in the point model (used to construct the shape model) are treated with equal weight. Some landmarks have weak correspondence across the sample and the RMSE metric does not provide an exact discernment between facial shapes. For example, the corners of the eyes have very high inter-subject correspondence whereas certain landmarks around the jaw-line do not. These particular landmarks may cause unduly high RMSE values to be calculated. Secondly, the faces are pose and expression compensated where necessary (see sections 3.2.1 and 3.2.2), prior to the calculation of the RMSE values. This process introduces error into the shape vectors since perfect compensation to frontal pose and neutral expression is not possible. This will affect accurate calculation of the RMSE. Thirdly, as explained in previous sections, the aging axes (in particular, the piecewise and the historical axes) may not be sufficiently well-defined to give accurate results for age-progression. Therefore, the value of e_t in such cases might be artificially high.

4.3.2 Texture comparison

Comparison of the age-progressed or regressed texture map with a target face is a non-trivial task, owing to the large variations within the training set, with respect to illumination, noise, resolution and image type. Figure 4.9 illustrates this point by displaying a series of images for one subject, taken from the training set.

Figure 4.9: **A sequence of training images at different ages for one subject.** These images demonstrate the non-uniformity of the images in Database 2, in which the illumination, noise, resolution and image type vary significantly. These factors mean that textural comparisons of an age-progressed face with its target are problematic.

In an attempt to address this problem and to facilitate comparisons between the age-progressed face and its corresponding target, the following procedure was carried out:

1. Each training image was cast to grey-scale and the texture maps extracted (section 2.3.2).

2. A *normalisation* procedure was applied to produce a mean pixel value of zero and a standard deviation of unity. That is, for each image, **I**, containing q pixels, $\{I_i \dots I_q\}$:

$$\mu = \frac{1}{q} \sum_{i=1}^{q} I_i \qquad (4.2)$$

$$\sigma = \sqrt{\frac{1}{(q-1)} \sum_{i=1}^{q} (I_i - \mu)^2} \tag{4.3}$$

The image, **I**, was then normalised by applying the following equation to each pixel:

$$I_i^n = \frac{(I_i - \mu)}{\sigma} \tag{4.4}$$

for $i = \{1 \ldots q\}$ pixels. The normalised image was therefore given in vector form by:

$$\mathbf{I}^n = [I_i^n \ldots I_q^n]^T \tag{4.5}$$

This procedure was implemented to ensure that each texture map had the same global statistics in terms of intensity (mean) and contrast (standard deviation).

3. The statistical texture models were reconstructed and the aging axes defined using an equivalent equation to (3.13).

4. Analogous to the shape comparisons, the RMSE, e_t, was computed between each aged texture vector ($\mathbf{t_{aged}}$) and its appropriate target texture vector ($\mathbf{t_{tar}}$) using an equivalent equation to (4.1).

5. The RMSE, e_o, was computed between $\mathbf{t_{aged}}$ and other real faces at the target age ($\mathbf{t_{other}}$) to compare with the error computed in Step 4.

6. Finally, the correlation coefficient (ρ_t) was calculated between $\mathbf{t_{aged}}$ and $\mathbf{t_{tar}}$, according to equation (4.6):

$$\rho_t = \frac{\mathbf{t_{aged}}^T \mathbf{t_{tar}}}{\sqrt{(\mathbf{t_{aged}}^T \mathbf{t_{aged}})(\mathbf{t_{tar}}^T \mathbf{t_{tar}})}} \tag{4.6}$$

An equivalent equation was then used to calculate the correlation coefficient, ρ_o, between $\mathbf{t_{aged}}$ and $\mathbf{t_{other}}$. This was to determine whether the texture map of the aged face was more highly correlated with the target face than with test faces.

To examine the results of these procedures, the rank of e_t in relation to each e_o was plotted. In a similar fashion, the rank of ρ_t was plotted in relation to each ρ_o. These results are depicted in figures 4.10 and 4.11.

Figure 4.10: **Graphical results of the rank of e_t in relation to each e_o (texture comparisons) for faces 2.1 to 2.7**. Graph 1 relates to the age-progressed images depicted in figure 4.5 (faces 2.1 to 2.5). Similarly, Graphs 2 and 3 relate to the sequences of age-progression displayed in figures 4.6 (face 2.6) and 4.7 (face 2.7) respectively. The dashed line represents a rank equal to unity (the ideal rank of e_t compared to each e_o).
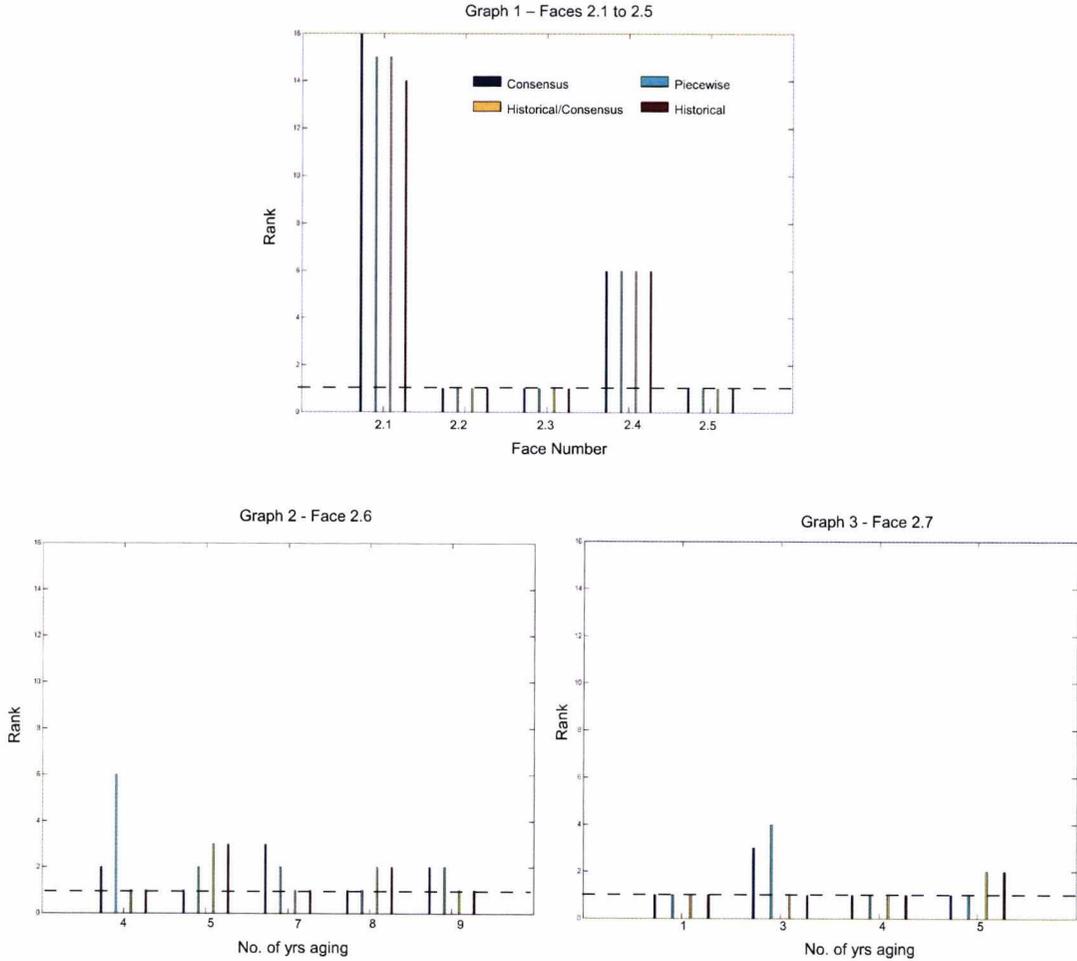
Figure 4.11: **Graphical results of the rank of ρ_t (correlation coefficient between t_{aged} and t_{tar}) in relation to each ρ_o for faces 2.1 to 2.7.** Graph 1 relates to the age-progressed images depicted in figure 4.5 (faces 2.1 to 2.5). Similarly, Graphs 2 and 3 relate to the sequences of age-progression displayed in figures 4.6 (face 2.6) and 4.7 (face 2.7) respectively. The dashed line represents a rank equal to unity (the ideal rank of ρ_t compared to each ρ_o). In this calculation, a lower rank indicates a *higher* value of ρ_t compared to each ρ_o, which is desirable.

In general, there is a smaller RMSE between $\mathbf{t_{aged}}$ and $\mathbf{t_{tar}}$ than between $\mathbf{t_{aged}}$ and $\mathbf{t_{other}}$ — that is, $e_t < e_o$ and the computed rank is low. Within the limits of this study, *this broadly implies that a face retains its identity in texture as aging proceeds.* However, this conclusion must be taken in context with the fact that the low rank does not occur for all examples and is not consistent across all approaches. This is the result of implicit textural variations within the training set (such as non-uniform illumination and disparity of image resolution), which cannot be compensated for. Therefore, in a quantitative sense, it is unclear to what extent a face will successfully retain its identity in texture as it is age-progressed. Furthermore, although a larger correlation coefficient is observed between $\mathbf{t_{aged}}$ and $\mathbf{t_{tar}}$ than between $\mathbf{t_{aged}}$ and $\mathbf{t_{other}}$ for some individuals ($\rho_t > \rho_o$), this is not a consistent trend. This finding further implies that additional tests are necessary (for example, perceptual tests, which will be discussed in section 5.1) to establish the consistency of subject appearance once an aging algorithm has been applied.

4.4 Age prediction — Fisher Linear Discriminant

A related aspect to the work on age-progression is that of age *prediction* — that is, prediction of the age of a subject from the shape and/or texture space representation. A primary motivation is to use automatic face recognition (AFR) algorithms to determine the age of children and young adults. Although it is unlikely that absolute determination of age could be made using AFR, evidence to support the likely age range of a child might be useful as part of systems to limit the supply of age-related goods or services to entitled persons, or to refute a claimed age in the case of persons attempting to enter a country. Work related to this is that of age classification. Kanno et al. [104] used Neural Networks to perform classification of young male faces into one of four age groups, based upon their facial shape. Additionally, Ueki

et al. [105] presented age-group classification of facial images under different lighting conditions using 2DPCA[10] and Linear Discriminant Analysis (LDA).

In this section, the technique of the *Fisher Linear Discriminant* (FLD) for age prediction is examined and its use for this purpose compared with that of the previously derived consensus aging axis. Discriminant analysis differs from PCA in that the latter seeks directions that are efficient for representation whereas the former seeks directions that are efficient for discrimination [107]. That is, PCA finds the directions of maximum variance to represent a data set efficiently — however, these directions are not necessarily effective for discrimination and classification. The goal of classical discriminant analysis is to find a line onto which the data set may be projected, whose orientation is such that the projected samples are well separated into distinct classes. Consider a set of n d-dimensional samples $\{\mathbf{x}_1 \ldots \mathbf{x}_n\}$, where it is desirable to separate the samples into two distinct classes, ω_1 and ω_2 (the *class labels*), containing n_1 and n_2 samples respectively. If a matrix, $\mathbf{X}$, containing the samples $\{\mathbf{x}_1 \ldots \mathbf{x}_n\}$ on its columns, is defined as:

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1d} & \cdots & x_{nd} \end{bmatrix}$$

then the dot product may be calculated between the matrix $\mathbf{X}$ and a new vector $\mathbf{w}$:

$$\mathbf{y} = \mathbf{w}^T \mathbf{X} \tag{4.7}$$

where the projected set of n samples $\{y_1 \ldots y_n\}$ is divided into the subsets Y_1 and Y_2, which correspond to the classes ω_1 and ω_2. Geometrically, if $||\mathbf{w}|| = 1$,

---

[10]Two-dimensional PCA is based on 2D image matrices rather than 1D vectors [106].

each y_i is the projection of the corresponding $\mathbf{X}_i$ onto a line of unit length in the direction of $\mathbf{w}$. If the samples labelled ω_1 approximately comprise one cluster while those labelled ω_2 comprise another, the aim of the analysis is to find the line $\mathbf{w}$ in such an orientation that the projections $\{y_i\}$ are maximally separated. Figure 4.12 depicts the results of projecting example 2D data samples onto the line $\mathbf{w}$, drawn at two different orientations.
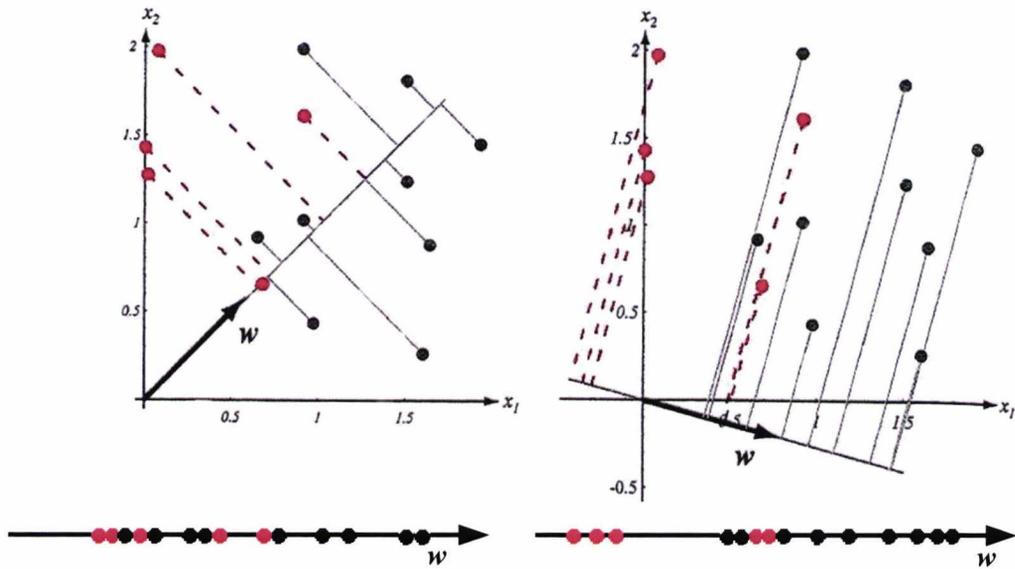
Figure 4.12: **Classification of example 2D data**. The same set of data points is projected onto two different lines in the directions marked $\mathbf{w}$. The figure on the right shows greater separation between the red and black projected points (recreated from [107]).

To find the direction of $\mathbf{w}$, a measure of the separation between the projected points is calculated. This is given by the difference between the sample means. If $\mathbf{m}_i$ is the d-dimensional sample mean given by $\mathbf{m}_i = \frac{1}{n_i}\sum_{\mathbf{X}\in D_i}\mathbf{X}$, then the sample mean for the projected points is given by:

$$\tilde{m}_i = \frac{1}{n_i}\sum_{\mathbf{y}\in Y_i}\mathbf{y} = \frac{1}{n_i}\sum_{\mathbf{X}\in D_i}\mathbf{w}^T\mathbf{X} = \mathbf{w}^T\mathbf{m}_i \tag{4.8}$$

This is the projection of the sample mean, $\mathbf{m}_i$, onto $\mathbf{w}$. It follows that the distance between the projected means is:

$$|\tilde{m}_1 - \tilde{m}_2| = |\mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)| \tag{4.9}$$

To obtain good separation of the projected data, the difference between the means needs to be large relative to the standard deviation for each class. Instead of forming sample variances, the *scatter* for the projected samples is used and is given by:

$$\tilde{s}_i^2 = \sum_{y \in Y_i} (y - \tilde{m}_i)^2 \tag{4.10}$$

The sum $\tilde{s}_1^2 + \tilde{s}_2^2$ is called the total *within-class scatter* of the projected samples. The *Fisher Linear Discriminant* thus employs that linear function $\mathbf{w}^T\mathbf{X}$ for which the criterion function,

$$J(\mathbf{w}) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

is a maximum. It can be shown (see [107]) that this criterion function may be equivalently written as:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{S}_B\mathbf{w}}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}} \tag{4.11}$$

where $\mathbf{S}_B$ is termed the *between-class scatter matrix* and is given by $\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$. In addition, $\mathbf{S}_W$ represents the *within-class scatter matrix* and is given by $\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2$, where the scatter matrix for each class is given by $\mathbf{S}_i = \sum_{\mathbf{X} \in D_i} (\mathbf{X} - \mathbf{m}_i)(\mathbf{X} - \mathbf{m}_i)^T$.

The vector $\mathbf{w}$ that maximises $J(\mathbf{w})$ in equation (4.11) can be shown to be [107]:

$$\mathbf{w} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \tag{4.12}$$

Hence, $\mathbf{w}$ is obtained for Fisher's linear discriminant — the linear function yielding the maximum ratio of between-class to within-class scatter. This is the line of maximum separation between the two classes of data.

To demonstrate the use of FLD as a classifier, the PCA shape parameters calculated from a male sample in Database 2 (section 3.1) were selected. Examples at the ages of one and twenty years were extracted (termed the "young" and "old" groups respectively) and the FLD line, $\mathbf{w}$, calculated according to equation (4.12). The examples were then projected onto the line via equation (4.7). This process was duplicated for equivalent young and old female groups. Finally, the process was repeated using the PCA texture parameters[11] of the same male and female subjects. Figure 4.13 depicts the projections of each set of parameters onto the calculated FLD line. For shape, the two groups are easily separated (a and b). Hence, examples may easily be classified as belonging to one of the two groups. Conversely, for texture, there is weaker differentiation between the two groups (c and d) and therefore classification on the basis of texture is more problematic. This is not unexpected since facial changes between the ages of one and twenty years are more predominantly structural (shape-related) than textural, therefore discrimination between individuals at these ages is expected *a priori* to be more successful in shape than in texture. This difficulty in classification for texture is compounded by the unavoidable differences in the texture maps in terms of illumination, image quality and resolution despite efforts to standardise them by using the normalised grey-scale images.

---

[11]where the texture parameters were calculated from a PCA of normalised grey-scale texture maps (see section 4.3.2).
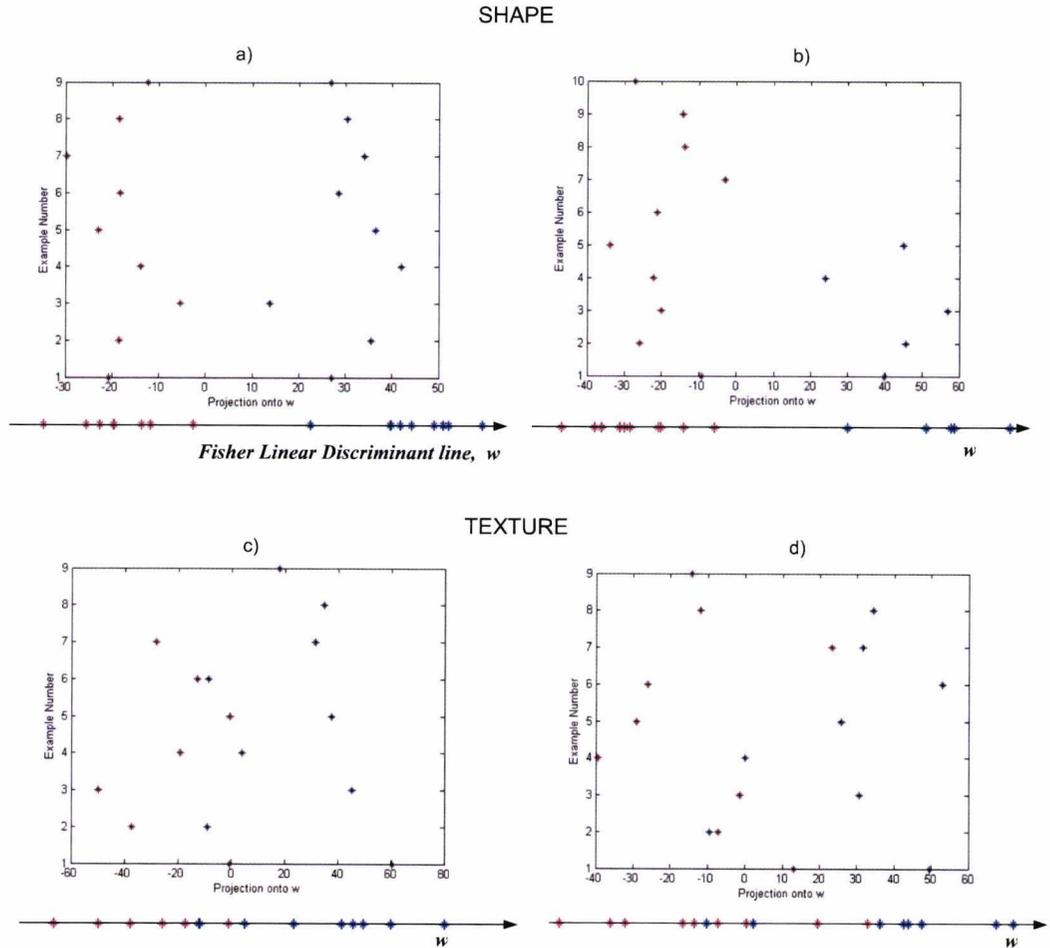
Figure 4.13: **Projection of shape and texture model parameters onto the FLD line**. This yields good separation for the shape case (a) male, b) female), where the young group is shown in blue and the old group in red. For the texture (c) male, d) female), the separation is poorer and the two groups cannot be easily classified.

4.4.1 Age prediction using consensus aging axis and FLD

To examine the differences between the consensus aging axis and the FLD line, a revised consensus aging axis, $\hat{\mathbf{V}}_{\mathbf{age}}$, was first calculated via equation (3.13) using only the one and twenty year old examples. This axis and the FLD line may both be used to predict the age of a subject from its shape or texture parameters. This was achieved as follows. Firstly, the mean projection of the young age group in shape onto the revised consensus aging axis was calculated according to equation (4.13). This mean projection lies directly on the axis.

$$\boldsymbol{\mu_y} = \frac{1}{n} \sum_{i=1}^{n} \left[\mathbf{B_y}_i{}^{T} \hat{\mathbf{V}}_{\mathbf{age}} \right] \hat{\mathbf{V}}_{\mathbf{age}} \qquad (4.13)$$

where n is the number of young examples and $\mathbf{B_y}_i$ is the ith shape parameter vector in the young age group, that is, the ith column of the matrix:

$$\mathbf{B_y} = \begin{bmatrix} b_{11} & \cdots & b_{n1} \\ \vdots & \ddots & \vdots \\ b_{1q} & \cdots & b_{nq} \end{bmatrix}$$

Similarly, the mean projection of the old age group onto the axis was calculated using:

$$\boldsymbol{\mu_o} = \frac{1}{m} \sum_{i=1}^{m} \left[\mathbf{B_o}_i{}^{T} \hat{\mathbf{V}}_{\mathbf{age}} \right] \hat{\mathbf{V}}_{\mathbf{age}} \qquad (4.14)$$

where m is the number of old examples and $\mathbf{B_o}$ is the matrix of shape parameter vectors for the old age group. To find the mean projection of the two groups onto the FLD line, $\mathbf{w}$, equivalent equations to (4.13) and (4.14) were employed by substituting $\mathbf{w}$ for $\hat{\mathbf{V}}_{\mathbf{age}}$.

From these values, the mean projection onto each axis of every year of age

between one and twenty years was also computed[12]. The age of an out-of-sample example was then predicted by projecting onto the axis and determining the mean to which it lay closest geometrically. This was performed for all shape vectors using the aging axis and FLD line as independent age predictors. The mean predicted age was then calculated for each year of aging between one and twenty years (figure 4.14) and the equivalent analysis performed for the texture models.

---

[12]Although the aging axis and FLD have both been calculated using only examples of one and twenty years, hypothetical means for each age in between may be calculated by using the projected means of the young and old age groups.
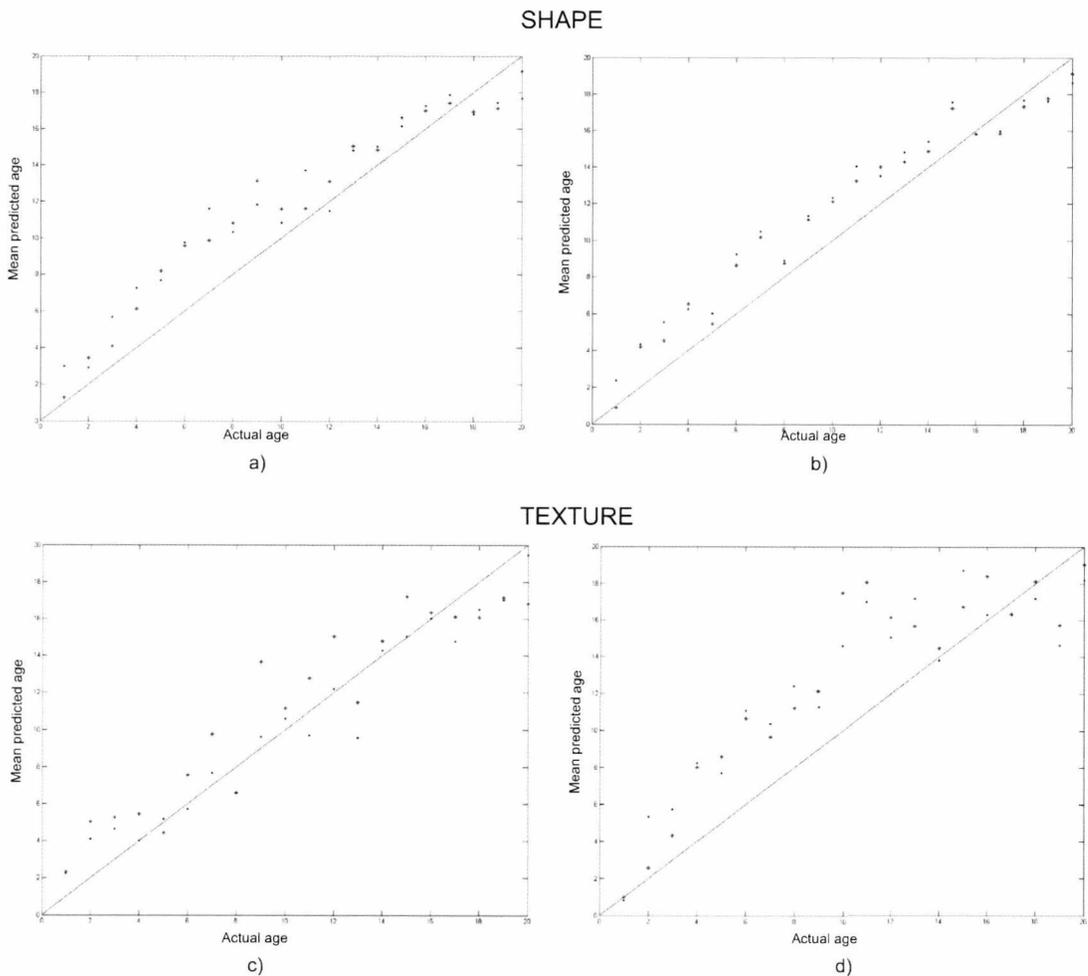
Figure 4.14: **Predicting age from the model parameters by employing the aging axis and the FLD line**. The mean predicted age calculated using the aging axis (blue points) and the FLD line (red stars) is plotted against the actual age for the male and female shape parameters (a and b) and for the male and female texture parameters (c and d) respectively.

The graphs in figure 4.14 show the following. Firstly, for the male and female shape model parameters (graphs a and b), the mean predicted age obtained using both the consensus aging axis and the FLD line is consistently higher than the actual age for the majority of examples. This may be due to the paucity of one and twenty year old examples used to construct the aging axis and the FLD line. If more examples were available for inclusion, the aging axis and FLD line would be more well-defined and this may therefore allow a higher degree of accuracy for age prediction. In addition, the fact that each example face has been pose and expression compensated prior to any attempt at age prediction may have introduced error into the shape parameter vector such that its age could not be accurately predicted.

For the male texture model parameters (graph c), there is no persistent trend to the age prediction using either the aging axis or the FLD line. However, for the female texture model parameters (graph d), the mean predicted ages are consistently higher than the actual ages for both approaches for the majority of examples. These observations imply that neither approach is particularly successful for the task of age prediction for male and female textures. Analogous to the results in shape, this may be due to a paucity of examples used to construct the aging axis and the FLD line. Additionally, the unconstrained nature of the texture maps in terms of illumination (despite compensation efforts), image quality and resolution may have compounded the inability of either approach to accurately perform age prediction.

To assess which approach was superior[13], the frequency of the smallest errors in age prediction was recorded for each approach, for each of the mean predicted ages (1-20 years). The results are given in Table 4.1.

---

[13]It must be noted that, as discussed above, neither approach performed outstandingly.

	Frequency of smallest error	
	Fisher Linear Discriminant	*Consensus axis*
Shape — male	11	9
Shape — female	14	6
Texture — male	3	17
Texture — female	12	8

Table 4.1: The frequency of the smallest error in age prediction for each approach (FLD and consensus aging axis), out of a total of 20 age predictions (the rows sum to 20).

	Calculated χ^2 value	*Standard χ^2 value*
Shape — male	0.25	
Shape — female	3.25	5%: $\chi^2 = 3.84$
Texture — male	9.85	1%: $\chi^2 = 6.64$
Texture — female	0.85	0.1%: $\chi^2 = 10.83$

Table 4.2: Values of χ^2 calculated under the assumption that the null hypothesis is true. These values are compared against standard χ^2 values to determine whether it is possible to reject the null hypothesis.

To determine the statistical significance of these findings, a null hypothesis was postulated that the probability of accurate age prediction is equal for the two approaches. A standard chi-squared statistic was calculated[14], where the expected variable, $E_i = N/2$ if the null hypothesis is true ($N = 20$ since there are twenty mean predicted ages). The results of this analysis and standard χ^2 results [79] at varying significance levels are given in Table 4.2.

From these values, the null hypothesis cannot be resolutely rejected at any of the given significance levels in the case of the male and female shape model parameters, since the calculated χ^2 values are less than the standard values. This indicates that age prediction is comparable using the FLD line and the aging axis. However, the

---

[14]See chapter 5, section 5.1 for a description of the chi-squared test and the associated equation, (5.1).

largest error in prediction is *smaller* using the FLD line than using the aging axis (4.13 years compared to 4.6 years), indicating that the FLD line could be used in preference for age prediction.

The results from the shape calculations may be explained as follows. If the FLD line and the aging axis lie in a similar direction in the model space, the age prediction will give similar results for both approaches. This is expected for the shape since the two age groups (one and twenty years) are well separated in the model space, owing to the craniofacial changes that occur between one and twenty years of age. Hence, *a priori*, the FLD line and the aging axis were expected to produce similar results for age prediction.

However, for texture, since fewer changes occur between one and twenty years than in shape, the groups are not expected to be well separated and, therefore, the FLD line to compute optimal classification may be in a different direction to the aging axis. This means that different age prediction results would be expected using the different approaches. Age prediction using texture may be further compounded by inconsistencies in illumination and image quality. From Table 4.2, it is possible to reject the null hypothesis[15] at the 0.01 significance level for the case of the male parameters. Therefore, it is possible to conclude that the probability of the null hypothesis being correct is 1% or less and that there is a significant difference between the observed and expected values in the chi-squared test. This indicates that age prediction is more accurate for the aging axis approach than for the FLD approach. However, for the female parameters, it is not possible to reject the null hypothesis at any significance level and, therefore, the approaches are comparable for age prediction.

---

[15]The null hypothesis can be rejected in favour of the alternative hypothesis that one approach is more accurate than the other for age prediction.

4.5   Chapter Summary

In this chapter, the results of implementing the novel aging algorithms were presented and discussed. A restricted sample of images was first used to test the consensus approach and the basic validity of the method was confirmed. However, due to the insufficiency of this data for modelling aging in children and teenagers, a more comprehensive database of images was instead employed and the consensus, piecewise and historical/consensus approaches were applied. In general, satisfactory results were produced. The reasons for unsuccessful age-progressions and artefacts were also discussed.

To quantitatively assess the performance of the algorithms, a measure of the error between the age-progressed face and the target face was computed and an equivalent calculation subsequently performed using other faces at the target age (known to be of different identity). It was found that subjects were more similar in both shape and texture to the target face than to the alternative faces, indicating that identity is successfully retained upon application of the aging algorithms.

Finally, the Fisher Linear Discriminant was employed for the task of age prediction and its performance compared against the use of the consensus aging axis. It was found that the FLD line and the consensus axis were statistically comparable as age predictors. However, the largest error in prediction was smaller using the FLD line, indicating that this approach could be preferentially used.

In the next chapter, human perceptual evaluations of the age-progression algorithms are presented. To assess the suitability of the algorithms to replace or augment artistic age-progression methods, typical results are also compared to those produced by professional forensic artists.

Chapter 5

Evaluation of Age-Progression Algorithms

In addition to the quantitative analysis of the age-progressed results presented in chapter 4, it is important to consider human perceptual measures as a means of assessing their reliability and accuracy. Indeed, human perception must be considered the most important measure since the generated images would, in reality, be used in forensic scenarios by human beings. To this end, a series of perceptual tests is described, in which participants were required to estimate the age of a face and to then match the aged face to its target.

To demonstrate the value of pursuing novel approaches to age-progression, the results are contrasted, both visually and quantitatively, with images that have been age-progressed using artistic methods. Comparable results would indicate that the algorithms could be employed as a replacement for (or as a supplement to) forensic artistry. However, owing to the increase in speed of image production, a greater number of age-progressions could be performed for missing children (a primary motive of the work presented in this thesis).

145

Finally, to demonstrate the applicability of the algorithms to real-world scenarios, a number of commissions were undertaken from the media and from forensic artists. The production of age-progressed results from supplied images is a good test of the approaches since the ultimate aim is to apply them to genuine forensic cases.

5.1 Perceptual Tests

A group of observers unfamiliar with the age-progressed subjects were asked to examine the age-progressed results and perform tasks, based on their perception of the images. This is essential since, in real-life applications, human beings will make the judgements on the age-progressed images. To this end, a series of perceptual tests was designed and uploaded to the internet. The tests remained on-line for a period of approximately four months and observers were repeatedly required to perform two basic tasks:

1. To estimate the age of a presented subject face to the nearest year.

2. To match the subject face to one of six presented target faces, based on the extent of perceived similarity.

Figures 5.1 and 5.2 depict examples of both elements of the test. A subject face, age-progressed using either the consensus, piecewise or historical/consensus approach, was presented and the observer required to estimate its age to the nearest year and type this into the box provided. The observer was then required to match the face to one of six presented in the face pool and again type their answer into the given box (a whole number between 1 and 6, corresponding to the face, which they considered to be the closest perceptual match to the age-progressed face). One of these face pool examples was the true appearance of the subject at the target

age (the target face) and the other five were foil faces. These foils were chosen at random from an appropriate sub-sample of the training set (Database 2). That is, from a sub-sample of the same sex and target age as the age-progressed subject. To ensure that the target face was not more distinctive than the foil faces, all faces were cast to grey-scale and normalised according to the procedure outlined in section 4.3.2. Presentation of the images in a consistent format aimed to prevent observer bias towards a particular face pool member based on textural similarities between the subject and face pool images. In addition, all faces were pose and expression compensated to frontal pose and neutral expression (using method 2 outlined in section 3.2.2) to prevent any possible cues to identity being perceived through particular characteristic expressions.

In total, fourteen aged subjects were presented to each observer (a total of 376 observers) — however, analysis of all subjects was not required to ensure optimal concentration. The observers were required, however, to examine the same subject face aged according to the different approaches (consensus, piecewise and historical/consensus respectively).

A.

Estimate the age of this face: ☐

Which of the faces below most closely matches this face in appearance? ☐

1.

2.

3.

4.

5.

6.

Figure 5.1: **Perceptual tests of the aging algorithms**. Example of a page presented to an observer as part of the on-line perceptual tests of the aging approaches.

Figure 5.2: **Perceptual tests of the aging algorithms**. A further example of a page presented to an observer as part of the on-line perceptual tests of the aging approaches.

5.1.1   Results — Task 1

In the first instance, the mean predicted age was plotted against the actual age. This was performed for each of the three aging approaches described in sections 3.6, 3.7 and 3.8.2. The plots are given in figure 5.3.
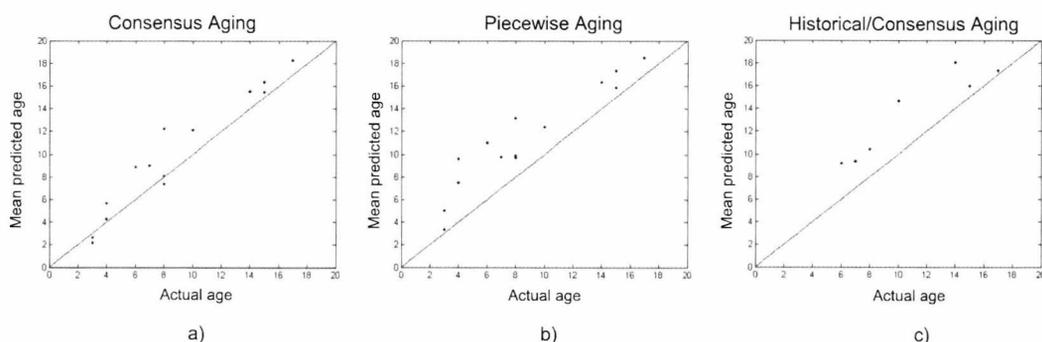
Figure 5.3: **Graph of mean observer predicted age against actual age**. The observers predicted the age of each subject face and the mean predicted age was calculated. This was performed for subjects aged according to the consensus (a), piecewise (b) and historical/consensus (c) approaches respectively.

The mean predicted ages are observed to be systematically *higher* than the actual ages for all three techniques. However, there is a maximum error of no more than ±5 years. In addition, the mean error in age prediction (Table 5.1) has a maximum value of 2.70 years (the piecewise approach).

To put these results into a meaningful context, it is important to consider the innate ability of the observers in estimating age from ordinary photographs. There have been many psychological studies performed to this effect [108, 109, 110, 111] as well as more recent studies in automatic verification of identity across age-progression [112]. The findings of the psychological studies must be considered when assessing and explaining the results of Task 1 (described previously) to pre-

vent under- or over-estimation of their importance. For example, it has been shown that the accuracy of age estimation by humans is very low, approximately 30% [111]. Therefore, the results of Task 1 are unsurprising and may be due to innate human difficulties in judging age.

Conversely, other studies have shown age estimation to be very robust to image distortions [109]. Therefore, it is unlikely that the unconstrained quality of the age-progressed images had a negative effect on the ability of the observers to perform age estimation. This suggests other factors may have been responsible for the errors, in addition to the previously discussed natural human difficulties in assessing facial age. One particular factor could be that the ability to judge age has been shown to be dependent on the racial group of the observer and that of the subject face [108]. That is, observers are able to more accurately judge the ages of subjects in their own racial group than the ages of subjects in other racial groups. This finding may impact the results of Task 1 since the age-progressed subjects presented to the observers were all of White European origin. Since the test was presented in a web-based format, it is conceivable that not all of the observers were themselves members of this racial group and may therefore have been innately unable to perform accurate age estimation. This could have affected the mean predicted ages calculated from all observer data.

Finally, it has been demonstrated that training can improve age estimation accuracy [110]. Since the observers described in this thesis were given no training, this may have affected their ability to accurately judge the ages of the age-progressed subjects.

To compare the age estimation of the observers across the three age-progression

Approach	μ (years)	e_p
Consensus	1.42	1.81
Piecewise	2.70	3.11
Historical/Consensus	2.57	2.95

Table 5.1: Calculation of the mean error in age prediction (μ) and the RMSE between the mean predicted age and the true age (e_p) for each approach.

approaches, the RMSE (e_p) was calculated between the mean predicted ages and the true ages for each approach (Table 5.1). The lowest value is observed for the consensus approach, indicating that this methodology gives more age-accurate results than do the other approaches, in agreement with the quantitative tests (section 4.3). This is confirmed by observation of the smallest mean error in age prediction for the consensus approach. Furthermore, perception of age is more accurate for faces aged using the historical/consensus approach than for those aged in a piecewise fashion (concluded from the smaller μ and e_p values for the former compared to the latter). This is encouraging since it implies that the use of historical images provides more perceptually accurate results than employing a piecewise model constructed on a peer group sample alone.

5.1.2 Results — Task 2

The results of the face pool identification are given in Table 5.2. The expected number of correct matches due to chance, E, is given for each subject as well as the observed number of correct matches, O. For the majority of subjects, $E > O$. However, for some subjects, notably Faces 2 and 13, E is comparable to O and this may indicate that the observers were only matching the subject face to the target with a probability comparable to chance. The significance of these results is assessed in a chi-squared test, described below. In addition, the results will be examined with respect to psychological studies on the recognition of unfamiliar faces.

Subject Number	N	E	O		
			Consensus	Piecewise	Hist./Cons.
1	211	35.17	86	75	-
2	196	32.67	59	44	40
3	180	30	84	107	42
4	186	31	170	167	160
5	196	32.67	133	97	-
6	212	35.33	111	121	-
7	210	35	89	71	37
8	213	35.5	160	155	-
9	216	36	140	132	147
10	223	37.17	182	186	-
11	228	38	126	123	134
12	232	38.67	126	75	-
13	223	37.17	57	51	41
14	224	37.33	185	155	-

Table 5.2: Observed number of correct matches (O) for each subject face in the face pool identification task for each of the three age-progression approaches. The number of observers for each subject is given by the column titled N. The expected number of correct matches due to chance ($N/6$) is given by the column titled E.

To analyse the significance of the results from the image-matching task, a standard chi-squared statistic was calculated:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i - 0.5)^2}{E_i} \tag{5.1}$$

where k is the number of cases. For this task, $k = 2$ since the outcome for each observer is either a correct ($i = 1$) or an incorrect ($i = 2$) match of the aged face to its target[1]. The null hypothesis may then be stated as follows:

H_0: The observed number of correct matches (O_1) is no greater than the number expected to occur due to chance (the expected variable, E_1).

Hence, in equation (5.1), O_i is the *observed* variable — that is, the number of correct or incorrect matches for a subject, out of a total of N observations. Finally, E_i is the *expected* variable — that is, the expected number of correct or incorrect matches. Since there are six potential targets for each aged face, $E_1 = N/6$ (correct matches) and $E_2 = 5N/6$ (incorrect matches).

Therefore, if O_1 is *significantly* greater than E_1, the null hypothesis can be rejected. This significance is determined by comparing the calculated χ^2 value with a standard value at the required significance level [79]. Table 5.3 displays the calculated χ^2 values for each subject face in addition to the standard χ^2 values for one degree of freedom at significance levels of 0.05, 0.01 and 0.001.

---

[1] Since $k = 2$, the system has $(k - 1) = 1$ degree of freedom. Therefore, equation (5.1) contains the Yates correction for continuity in the subtraction of 0.5 from the $(O_i - E_i)$ value. This correction factor prevents overestimation of statistical significance in the case where there is only one degree of freedom.

Subject Number	Calculated χ^2 values			Standard χ^2 values
	Consensus	*Piecewise*	*Hist./Cons.*	
1	87.03	53.25	-	
2	24.84	4.45	2.00	
3	115.21	237.46	5.64	
4	744.33	701.44	651.18	
5	367.35	152.03	-	
6	192.74	256.02	-	5%: $\chi^2 = 3.84$
7	98.75	43.62	0.1	1%: $\chi^2 = 6.64$
8	521.15	480.03	-	0.1%: $\chi^2 = 10.83$
9	358.23	305.07	411.38	
10	674.17	707.24	-	
11	242.70	232.11	298.29	
12	234.90	40.77	-	
13	12.28	5.89	0.44	
14	697.73	442.52	-	

Table 5.3: Calculation of the χ^2 value for each age-progressed subject face based on the numbers of expected and observed correct matches to the target face. In addition, the standard χ^2 values are shown at significance levels of 0.05, 0.01 and 0.001 respectively for one degree of freedom.

Upon examination of the results in Table 5.3, the null hypothesis may be legitimately rejected at the chosen significance level of 0.001 for the majority of subjects age-progressed according to the consensus and piecewise approaches. This implies that, for these methodologies, an individual will successfully retain its identity as it is age-progressed. In addition, the χ^2 values are greater for the consensus approach than the piecewise approach, indicating once more the perceptual validity of the consensus technique. For the historical/consensus approach, the null hypothesis cannot be resolutely rejected in general and hence a subject does not retain its identity as convincingly as in the other approaches.

These results, however, must be considered in light of findings in psychology relating to the innate human ability to recognise unfamiliar faces, since the observers in the study were all unfamiliar with the age-progressed subjects and members of the face pool. If the psychological data is not taken into account and a baseline established for the recognition of unfamiliar faces then it is possible that the significance of the results of Task 2, as previously described, might be misinterpreted. It has been shown that observers unfamiliar with target individuals perform very poorly in recognition experiments using poor quality video footage [113]. In this study, Burton et al. found that a group of unfamiliar observers found it more difficult to distinguish between target faces they had encountered before in an experiment and those they had not, compared to the case for a group of familiar observers. This implies that human beings have an innate inability to recognise unfamiliar faces and this may help to explain the imperfect results of Task 2.

Other studies have shown that, while recognition rates are still poor, person identification can tolerate large discrepancies in image quality between the images to be matched [114]. This implies that discrepancies between the image quality of

the age-progressed subjects and those of the face pool in Task 2 should not have impeded the image matching procedure. Other studies of image quality in face recognition include those by Kemp et al. [115] and Pike et al. [116], who found that recognition rates of unfamiliar observers was still very low, even when using high quality images, such as those printed onto a credit card. Furthermore, results obtained by Bruce et al. [117] also emphasise the difficulties of unfamiliar face matching. In their study, observers were shown high quality video of unfamiliar target subjects and then asked to identify the same targets from an array of high quality photographs. The error rate was high, of the order of 25%. Hence, even if high quality imagery had been available for Task 2, the recognition rates of the correct target from the face pool may not have been improved.

Other psychological studies have shown similar results on face recognition. Roark et al. [118] found that familiarising subjects with high resolution images or videos of faces was sufficient to improve recognition from low resolution, whole-body images. This implies that familiarity with a target face is important for recognition.

Recent work by Megreya and Burton [119] showed that there are large individual differences on unfamiliar face matching, concluding that unfamiliar faces are processed for identity in a qualitatively different way than are familiar faces.

These studies all describe the base level for the recognition of unfamiliar faces as very low. Hence, it is unsurprising that errors in recognition occurred in the results of Task 2. Despite these errors, age-progressed images produced using the consensus and piecewise approaches were correctly matched to the target at a rate significantly above chance in the majority of cases, implying that the subject face has retained its identity throughout the age-progression. Errors in the target matching

of faces age-progressed according to the historical/consensus approach could be due to innate human difficulties in face recognition. However, the errors could also be caused by inaccuracies in the age-progression due to a paucity of historical examples (see discussion of results in section 4.2.1).

5.2 Comparison with Forensic Artist results

The work presented in this thesis is intended to be complementary to (or a potential replacement for) the artistic age-progression techniques currently used for missing persons investigations. As such, it is necessary to compare the results from the novel approaches with such artistically aged images. Images of missing children were obtained from the National Center for Missing and Exploited Children (NCMEC) on-line database [89] and age-progressed using the consensus and piecewise approaches. The results were then assessed, both visually and quantitatively, by comparison with the artistically produced progressions, which are also available from the database. Figure 5.4 depicts the results of the age-progression, with the artistic progressions shown alongside.
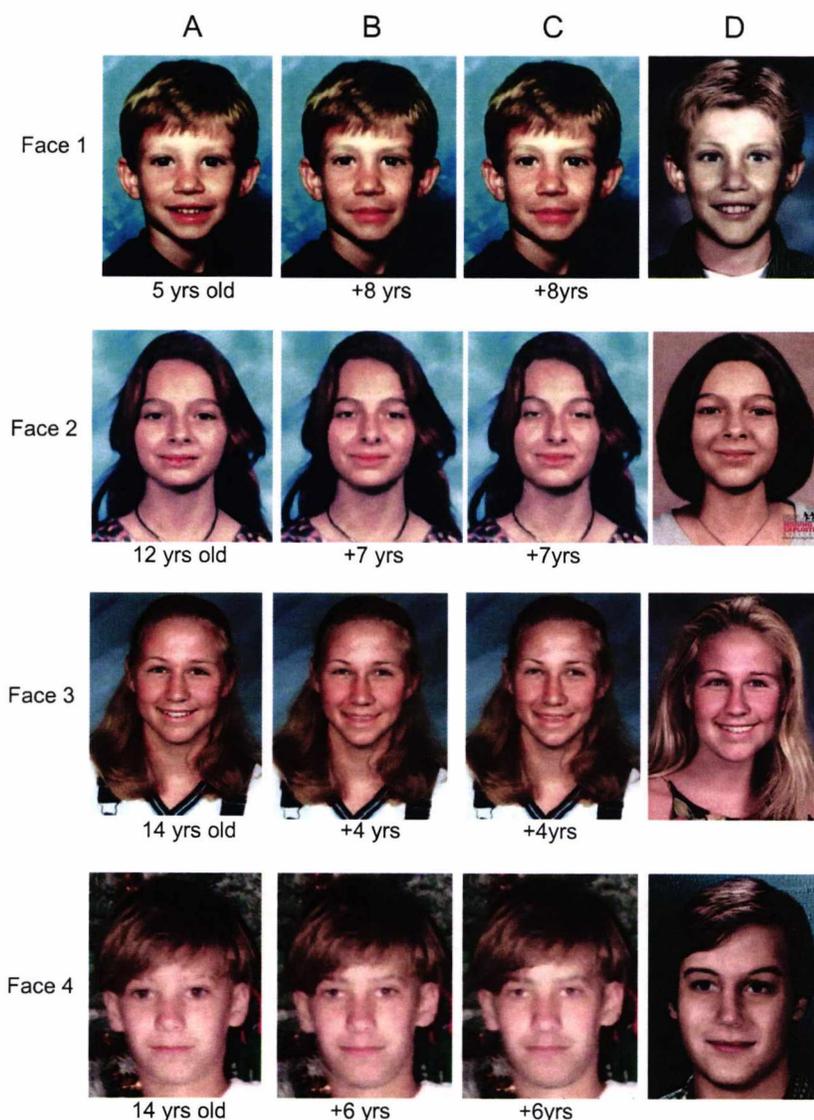
Figure 5.4: **Comparison with artistic age-progressions**. The results from the consensus and piecewise approaches are compared to the resultant images from the forensic artistry technique. Column A) Original image, B) Subject aged according to the *consensus* approach, C) Subject aged according to the *piecewise* approach, D) Age-progressed image produced by a forensic artist.

From examination of figure 5.4, it is clear that the consensus and piecewise techniques produce age-progressed images which are perceptually similar (in the opinion of this author) to those of the artistic approach. However, to determine the most appropriate method to use, the results from all approaches would ideally require comparison with the true face at the target age. Since the test images used were those from real missing persons enquiries, the target face is not available and therefore such comparisons may not be drawn.

Quantitatively, the RMSE (e_a) was calculated (using an equivalent equation to (4.1)) between the shape vector of the age-progressed face ($\mathbf{d_{aged}}$) and that of the artistically aged face ($\mathbf{d_{art}}$). This was then compared to the RMSE (e_o) calculated between $\mathbf{d_{aged}}$ and alternative target faces, known to be of different identity ($\mathbf{d_{other}}$). In general, $e_a < e_o$ for both rigorous techniques. This may be observed from figure 5.5, in which the rank of e_a in relation to each e_o is plotted. The result implies that statistically rigorous age-progression produces results more similar in shape (and, arguably, in identity) to the artistically aged face than to other target faces of different identity. This in turn indicates that the aging procedures herein produce broadly similar results to the work of forensic artists although, encouragingly, the procedures are much less time-consuming and require no artistic skill.

The rationale behind the use of unfamiliar faces in the comparison of the rigorous and artistic approaches to age-progression is as follows. It was desirable to perform as realistic and as fair a study between the rigorous and artistic approaches as possible. Hence, genuine missing children were selected and the artistic age-progressions that had already been produced and published on-line were found. These were then used in a comparison test with the age-progressed results produced using the rigorous approaches to assess the degree of perceptual similarity between them and,

hence, the potential to supplement the work of forensic artists with that of the work presented in this thesis. An improvement to this comparison approach could be to have a sample of observers assess the images to determine perceptual similarity, such that the sole reliance on the opinion of one individual (this author) could be avoided. A further improvement could be made using a familiar observer group. That is, performing an age-progression on a subject from some previous age up to the current age using the approaches outlined in this thesis and commissioning a forensic artist to produce an equivalent age-progression subjectively. A group of observers known to the subject (for example, family members) could then be asked to view the images and comment on their perceptual similarity to the subject's current, known appearance. This may allow a more meaningful comparison to be made between the rigorous and artistic methodologies.

Rank of e_a compared to each e_o (shape comparison) for subjects age-progressed according to the consensus and piecewise approaches
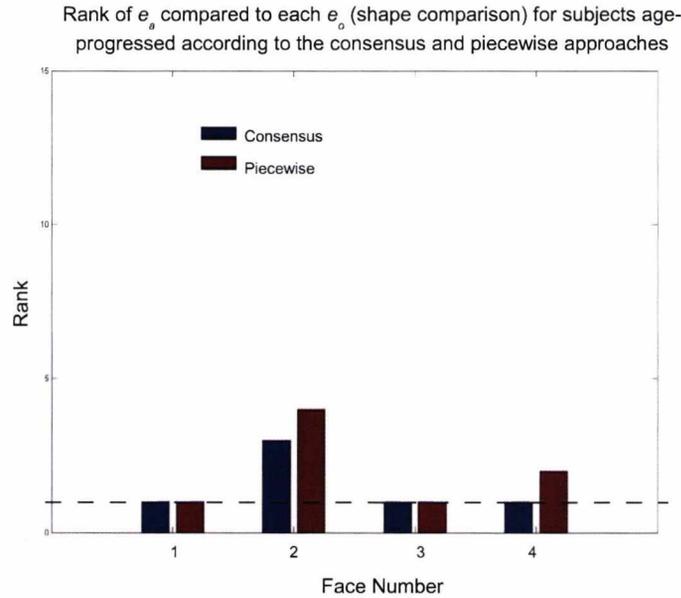
Figure 5.5: **Graphical results of the rank of e_a in relation to each e_o.** The RMSE was calculated between d_{aged} and d_{art} (e_a) and between d_{aged} and d_{other} (e_o). The rank of e_a in relation to each e_o is plotted for Faces 1 to 4, aged according to the consensus and piecewise approaches. The dashed line represents a rank equal to unity (the ideal rank of e_a compared to each e_o).

5.2.1 Results from Database 3 using Consensus Approach to Aging

To extend the rigorous aging approaches and to further test their flexibility, a sample of 250 black children's faces was obtained from the NCMEC on-line database [89] (this sample was termed Database 3). Statistical shape and texture models were constructed using this sample as described in sections 2.3.1 and 2.3.3 and example subjects age-progressed by defining a consensus axis. Comparisons were then performed between the results and available artistic progressions, in a similar way to those presented in section 5.2.

Figure 5.6: **Comparison with artistic age-progressions for Database 3**. The results obtained from consensus age-progression are compared to the resultant images from the forensic artistry technique. Column A) Original image, B) Subject aged according to the *consensus* approach, C) Age-progressed image produced by a forensic artist.

The age-progression of both the male and female examples gives plausible results (figure 5.6). In addition, the results are perceptually comparable (in the opinion of this author) to those achieved using the forensic artistry techniques, once more implying that the use of the consensus approach may be considered as a real alternative to the artistic approach. The results from the RMSE analysis also add credence to this since, in general, the RMSE value between $\mathbf{d_{aged}}$ and $\mathbf{d_{art}}$ (e_a) is lower than between $\mathbf{d_{aged}}$ and $\mathbf{d_{other}}$ (e_o). This may be concluded from figure 5.7.
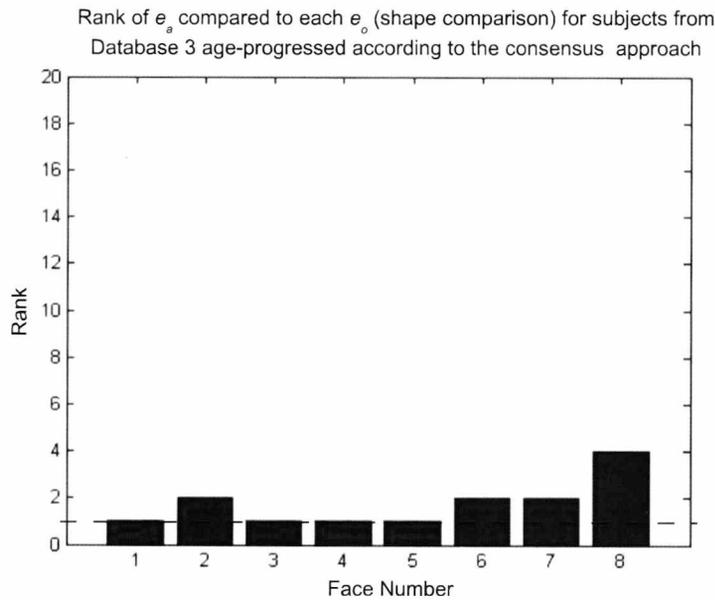
Figure 5.7: **Graphical results of the rank of $\mathbf{e_a}$ in relation to each $\mathbf{e_o}$ for Database 3**. The RMSE was calculated between $\mathbf{d_{aged}}$ and $\mathbf{d_{art}}$ (e_a) and between $\mathbf{d_{aged}}$ and $\mathbf{d_{other}}$ (e_o). The rank of e_a in relation to each e_o is plotted for Faces 1 to 8, aged according to the consensus approach. The dashed line represents a rank equal to unity (the ideal rank of e_a compared to each e_o).

5.3   Chapter Summary

In this chapter, the results obtained from the age-progression algorithms were evaluated by implementation of perceptual tests, in which observers were required to match an age-progressed face to one of six targets. An hypothesis test revealed that the probability of matching the age-progressed subject to the correct target with the recorded success rate on the basis of chance was negligible. These findings imply a perceptual plausibility for the age-progressed results using these methodologies.

Since the work in this thesis is intended to supplement or replace forensic artistry, the results from the consensus and piecewise approaches were compared against artistic age-progressions. Subsequent qualitative and quantitative assessment showed similarities between the results, implying that the algorithms may reasonably be employed for forensic age-progressions, with equivalent results to artistry achieved more quickly.

In the next chapter, appropriate conclusions about this work are drawn and future work is discussed.

Chapter 6

Summary & Conclusions

In this thesis, new methods have been introduced for the rigorous age-progression or regression of digital facial images. Principal component analysis provides the model space and the methods rely on identifying paths or trajectories within the space using rigorous statistical methods. Several models have been presented. The consensus and piecewise methods rely on average developmental trends to predict growth patterns for the age-progression of missing children. In addition, the theoretical basis for an extension of these models to include person-specific factors, such as historical and familial correlations, has been outlined and implemented computationally. This statistical framework is quite general and can, in principle, be extended to include other influences such as lifestyle effects. Since plausible, near photo-quality results are obtained over short time-scales, the presented methods are proposed as a replacement for (or a supplement to) the currently used forensic artistry techniques, which are time-consuming to implement and subjective in nature. A summary and conclusions on the work presented in this thesis is given. Suggestions for future work are then suggested in the final section.

6.1 Summary

This thesis began by highlighting the need for a new approach to facial age-progression for the purpose of aging missing persons. The limitations of the currently used artistic techniques were discussed in terms of the level of artistic skill required and the length of time typically needed for completion. The concept of a statistical learning model as a more rigorous methodology was briefly introduced to address these issues such that age-progressed images could be produced quickly and accurately by a non-specialist. Since the proposed model is constructed upon the analysis of real facial images, it was expected *a priori* that the results would be more objective and reliable than the artistically produced equivalent.

In the first chapter, the artistic [2, 4, 6, 29, 30] and anthropologically-based [5, 13, 32] approaches to age-progression were presented to give an overview of the techniques currently used in practice. Previous scientific approaches [9, 10, 11, 45, 46] were then outlined and their limitations discussed. For example, no attempt has been made in existing work to rigorously model the tendency of an individual to resemble its close genetic relatives, which is considered an important influence for age-progression. In addition, the models are not robust to variations in pose and expression. Hence, a need was identified for a more comprehensive system.

Chapter 2 provides the necessary mathematical background to the new techniques developed in this thesis. The techniques of PCA, geometric transformations (warping) and the construction of shape and texture models were discussed in depth and details of their computational implementation provided through an illustrative application.

In the first half of chapter 3, the necessary pre-processing steps for the aging

algorithms were discussed, namely pose and expression normalisation techniques as well as compensation procedures for texture and illumination. In the second half of the chapter, the initial approach to age-progression, the *consensus* method, was presented and the aging axis derived. To theoretically improve the accuracy of such an approach, a *piecewise* technique was described, where the axes were constructed upon smaller, physiologically relevant age-ranges. Finally, a more sophisticated theoretical model was presented, incorporating parental influences and previous developmental trends in addition to consensus information.

Chapter 4 presents the results of the application of all three of the aging models expounded in chapter 3. Results were compared to the target images, both visually and through a quantitative assessment of their similarity in terms of the root mean square error. Finally, in chapter 5, psychological tests were employed to assess the perceptual validity of the results. The results were also compared against equivalent images produced by professional forensic artists.

6.2 Conclusions

The conclusions that may be drawn from the work outlined in section 6.1 are as follows:

Firstly, the results from the aging algorithms were found to give broadly comparable results to those from forensic artistry after a visual and quantitative assessment. Since these results may be achieved in a more time-efficient fashion and with a higher throughput, it is reasonable to consider the novel algorithms as techniques that may supplement the forensic artistry approach. Secondly, the techniques described in this thesis are based upon the rigorous statistical analysis of real facial images as opposed to artistic interpretation. The results are therefore more objective in nature. However, the results are not always accurate or plausible, owing to the paucity of training examples used to construct the models and the unconstrained nature of the training images in terms of illumination, image quality and resolution. Thirdly, the algorithms provide more flexibility than the artistic approach, owing to the pose and expression estimation techniques incorporated into the models. It must be noted, however, that the compensation techniques outlined in this thesis do not always work well compared to artistic interpretation. Finally, artistic skill is only required to add cosmetic detail to the results produced by the algorithms, therefore a non-specialist may easily produce reasonably plausible, near photographic-quality age-progressed or regressed images.

A further conclusion that may be drawn from the results is that the age-progressed subject appears to successfully retain its identity as it is translated through the model space. This may be concluded from quantitative comparison of the aged face with its target and from the results of the perceptual tests since, in the majority of cases, the aged face was successfully matched to its correct target.

Finally, from both the quantitative and qualitative assessments, the consensus approach was found to be the most reliable and accurate method for age-progression using the limited data available for this work. However, the piecewise and historical/consensus methodologies cannot be discounted since both may produce more reasonable results upon collection of a larger and more appropriate training set.

6.3 Future work

A number of directions for future work are suggested:

- Collection of a more extensive training set. High resolution images of individuals between 0 and 20 years of age would allow more accurate piecewise axes to be constructed. Such a training set should incorporate individuals of many races, together with high quality parental and sibling images, allowing the person-specific model to be implemented more fully.

- Collection of lifestyle and health information for individuals in the training set. This would allow these influences on aging to be modelled in the person-specific approach.

- Dedicated software development with a Graphical User Interface. This would allow systematic testing of the accuracy and efficacy of the novel algorithms for age-progression since many images could be produced rapidly for evaluation.

- Age-related changes in facial proportions are well-documented in anthropology [13]. By taking detailed measurements on the original facial images and on the corresponding results obtained from the aging algorithms, the extracted facial proportions could be compared against known effects. This could then be used to determine whether the modelled changes are consistent with physical growth and development.

- More extensive perceptual testing. To attach greater statistical significance to the results from the perceptual tests, a greater number of observers would be required to participate. In addition, results from the age-progression of a different racial group should be presented to the observers in a set of analogous tests.

- Extension to three-dimensional models. This would allow the growth of features to be more easily identifiable and may produce more accurately aged results from a perceptual perspective. In addition, the age-progressed face could be rotated, which might aid recognition.

- Adult aging. Employing a more detailed point model for the landmarking procedure may allow the shape of adult facial features to be more accurately modelled during age-progression. For example, incorporating features such as aging lines around the mouth and eyes into the point model [120] is expected to permit improved reliability in the aging of adult faces.

- Increased realism. To more accurately model the textural details captured by the high spatial frequencies (for example, wrinkles, blemishes), a measure of "complexion" might reasonably be added to the PCA model of texture. This might include a local measure of entropy (statistical measure of randomness) to indicate the inherent smoothness of the skin. This would allow features of the skin, which exhibit low spatial correlation between subjects, to be included in the model rather than added as an overlay (wrinkle-map), subsequent to the age-progression.

A time-efficient and rigorous age-progression system to potentially supersede the currently used artistic methodologies was the central motivation behind the work presented in this thesis. To address the issues highlighted in the literature, three approaches of increasing complexity and flexibility were presented. In particular, this

thesis has contributed a person-specific approach, which statistically models previous appearance trends in addition to those of a peer group sample. The theoretical extension to include the tendency of an individual to resemble its parents or siblings is straightforward, as is the incorporation of lifestyle factors. Additionally, the algorithms provide greater flexibility than other approaches to age-progression, in terms of the pose and expression compensation techniques. Finally, the increase in speed afforded by the algorithms means that a greater number of age-progressed images could be produced than is currently practical, using artistic techniques. It is hoped that this may eventually greatly contribute to the search for missing persons.

Appendix A

In order to determine how a system for age-progression may help in the search for missing children, it is vital to consider exactly what constitutes a missing child and the dynamics that exist in such a situation. A marketing study [3] was performed in order to assess the demand and requirements for an age-progression system with respect to missing children. The findings of the study are given herein.

For the purposes of considering missing children, a child is defined as an individual who is less than 18 years of age. It is generally accepted that the number of children reported missing each year is much fewer than the number who go missing in actuality. Instead of considering numbers in absolute terms, it is necessary to make an assessment of how many children are considered to be *long-term missing*. This term refers to children who have been missing for more than one year. A very long-term missing child is classed as one who has been missing for more than five years. These are the categories who would require or benefit from age-progression techniques, since it is reasonable to assume that their faces may have altered due to processes of growth and development during these time periods. From obtained NISMART-2 data[1], it is estimated that the rate of recovery for missing children is greater than 99%, with long-term missing children representing less than 1% of

---

[1]National Incidence Studies of Missing, Abducted, Runaway and Thrownaway Children October 2002 reports, which are based on missing episodes in North America during 1999.

cases. This represents approximately 111 children per one million children in the population.

Of these 111 children, it is estimated that approximately 80% are runaways or throwaways. It is assumed that in some of these cases, there will be no desire for reconciliation and in such circumstances there would be no need of a forensic age-progression tool. However, it is estimated that in approximately 50% of cases, the parent is likely to seek reconciliation and hence age-progression may indeed aid them in their search. From these assumptions, it can be stated that *about 55 children per million are considered long-term missing and sought for,* with such cases likely to benefit from an age-progression tool. For validation, the NISMART-2 figures can be cross-referenced with two other sources. Figures from PACT (Parents and Abducted Children Together) in 2000-2001 estimate that there were 546 cases of trans-national child abduction. Since these cases are termed trans-national, it is reasonable to assume that they are long-term missing children cases. For a child population of approximately 22 million in the United Kingdom, these figures may be estimated to represent about 25 cases per million. Additionally, the Canadian Case Summary of 2004 Missing Children Reports document reported 358 parental abductions. Canada has an estimated child population of 10 million, giving an estimate of long-term missing children at 36 cases per million. The results from these three independent sources are of the same order of magnitude. Given that less than 1% of missing children are classified as long-term missing, this lends credence to the accuracy of the estimates.

The key weakness of the analysis of long-term missing children cases is that there are no available data on the recovery rates of these children beyond one year. It is estimated that of the approximately 55 children per million who are long-term

missing and sought for, about 50% of these will become very long-term missing. *This gives roughly 30 very long-term and sought-for missing children per million who would benefit from an age-progression tool.*

Appendix B

Age-progression has been used in commercial applications in the media and has recently been employed in advertising, most notably by insurance company Norwich Union as depicted in Figure B.1.

Figure B.1: **Commercial use of age-progression in advertising**. The first image (far left) is of the subject to be aged and the fourth image (far right) is that of her mother. The middle images are an age-progressed version of the subject and an age-regressed version of the mother (Norwich Union advert, courtesy of http://news.bbc.co.uk/1/hi/magazine/4360482.stm).

The first image (far left) is of the subject to be aged and the fourth image (far right) is that of her mother. The middle images are an age-progressed version of the subject and an age-regressed version of the mother[1].

---

[1] Although Norwich Union failed to disclose the exact nature of the aging technique used, the author speculates that appropriate adjustments were made to the shape and texture of the facial images to produce intermediate incarnations of both subjects.

The BBC have also used age-progression in their programmes, principally in their broadcast entitled "Honey, we're killing the kids" from 2005/6 [121]. Figure B.2 shows examples of such facial age-progressions. To achieve the results, knowledge of craniofacial development and growth statistics for an appropriate population are used. Lifestyle choices are incorporated by adjusting the components of the face to reflect known, long-term physiological effects of habits such as smoking and unhealthy eating.

a)   b)
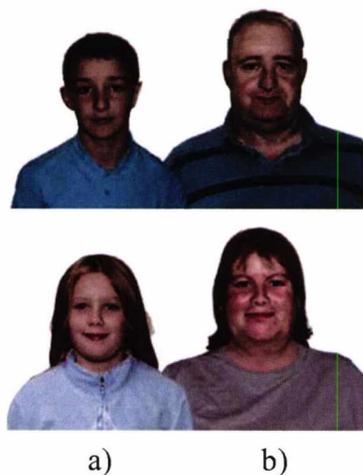
Figure B.2: **The use of age-progression for education in the media**. Age-progression may be used to demonstrate the effects of an unhealthy lifestyle for future appearance — a) Original subject, b) Subject age-progressed in accordance with current unhealthy lifestyle habits. Images courtesy of http://www.bbc.co.uk/bbcthree/tv/killing_the_kids/pilot.shtml.

Appendix C

In order to perform pose and expression compensation using Method 2, the following subjective scale relating to semantic descriptions of pose and expression was used by an observer to score the faces in a training set.

Attribute 1 — Head shaking:

- $+2$ — Head rotated fully to the left, partial occlusions of some features possible (approximately $+30°$ rotation)

- $+1$ — Head rotated slightly to the left (between approximately $+10$ and $+20°$ rotation)

- 0 — Head in approximately frontal view (approximately $0°$ rotation)

- -1 — Head rotated slightly to the right (between approximately -10 and -20° rotation)

- -2 — Head rotated fully to the right, partial occlusions of some features possible (approximately -30° rotation)

Attribute 2 — Head nodding:

- +2 — Head pointed fully downwards (approximately +30° pitch)

- +1 — Head pointed slightly downwards (between approximately +10° and +20° pitch)

- 0 — Head in approximately frontal view (approximately 0° pitch)

- -1 — Head pointed slightly upwards (between approximately -10° and -20° pitch)

- -2 — Head pointed fully upwards (approximately -30° pitch)

Attribute 3 — Face smiling:

- +2 — Very broad smile with an open mouth, both rows of teeth visible

- +1 — Broad smile, upper row of teeth visible

- 0 — Smile with closed mouth, no teeth visible

- -1 — Neutral expression

- -2 — Mouth down-turned

Appendix D

The landmarking tool is based upon a least squares polynomial fitting procedure, where one or more polynomial line segment(s) was used to describe the shape of each facial feature. The shape of each polynomial line segment was determined via a set of control points, $\{x, y\}$, with the purpose of finding the n degree polynomial in x that best fits y in a least squares sense. This provides a smooth curve, $\hat{y}(x)$, that can be manipulated manually by the user to follow the contour of a given feature. The least squares problem may be written in terms of the Vandermonde matrix, $\mathbf{V}$ - an $[(n+1) \times (n+1)]$ matrix where the jth column vector is the vector $[x_1^{j-1}, x_2^{j-1}, \cdots, x_n^{j-1}]^T$ for $j = 1 \cdots n + 1$:

$$\mathbf{V}\mathbf{p} \cong \mathbf{y} = \hat{\mathbf{y}} \tag{D.1}$$

or, in matrix form:

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n+1} \end{bmatrix} \cong \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n+1} \end{bmatrix} \tag{D.2}$$

The elements of $\mathbf{V}$ are powers of x_i^{j-1} and the coefficients that constitute $\mathbf{p}$ may

be determined by least squares methods. Vector **y** contains the y-coordinates of the control points. The order of each polynomial curve was set according to the nature of each feature — for example, some features exhibit more curvature than others and are therefore more accurately represented by a higher-order polynomial curve. The two end points of each line segment are classified as anatomical landmarks because they were positioned at salient points on the feature boundary (e.g. at each corner of the eye). Pseudo landmarks for each feature were obtained by sampling the coordinates of equidistant points along the interpolated curve. The number of these interpolated landmarks (between the control points and along the polynomial curves) was determined by the length of the curve section and its shape. Control points may be positioned by using the left mouse button to 'click and drag' such that the curve follows the contour of a particular feature. In addition, the zoom function may be used to examine a chosen feature in greater detail, such that the landmarks may be placed in the most accurate positions possible. The following describes the landmarking procedure in greater depth for completeness and depicts the graphical user interface designed to guide a user through the semi-automatic landmarking procedure:

1. Upon activating the load file push button, the user selects an image file from a listbox. This then appears in the work area located on the right hand side of the interface.

2. Firstly, the user is required to locate three landmarks; at the outer corner of the left eye[1], at the outer corner of the right eye and at the base of the nose. After the third point has been located, the remaining landmarks are automatically placed in their approximate positions. This is achieved by computing the transformation required to map the three landmarks, contained in the pre-

---

[1]The convention used here is that left refers to the left hand side of the displayed face from the perspective of the user, i.e. not the subject's left eye.

viously determined mean face shape, to the initial three landmarks located by the user. The computed transformation is then applied to mean face shape as a whole, providing an affine transform that defines the preliminary positions of all the points in the face shape.

3. The landmark points control a set of spline curves (also plotted) that delineate the perimeter of the internal facial features and the head itself.

4. Landmarks can subsequently be moved from their approximate positions to their correct locations using a 'click and drag' technique, whereby the user selects a landmark point via the left mouse button and drags it to its correct position, holding the left button down during the procedure. When the left mouse button is released, the spline curve(s) associated with the translated point is redrawn, updating the face shape in response to the user's action (see figure D.1). Selected or active landmarks and their associated spline curves are plotted as red graphics objects, whereas the inactive landmarks and spline curves are plotted in blue. When a spline curve becomes active, its description, e.g. chin, appears in a frame on the left hand side of the interface. This is particularly helpful when adjusting landmarks around the mouth, where there are many spline curves that could otherwise become confused. Landmarks that define the end of one spline curve and the beginning of another connected curve are referred to as base landmarks and are plotted as magenta circles, distinguishing them from the ordinary landmarks plotted in either red or blue.

5. For landmarking purposes the images are displayed at full resolution (2048 × 3072 pixels, 300dpi). Regions of the face can be enlarged using the zoom mode push button located under the image. When the zoom mode is set to on, placing the mouse cursor over a region of interest in the face image and clicking the left mouse button will enlarge that area, making it easier to place

landmarks/spline curves accurately. Selecting the zoom mode button again allows the user to exit zoom mode and return to the standard landmarking mode.

6. Once the user is satisfied that all the spline curves are correctly located, the face shape can be saved using the save file pushbutton on the left hand side of the interface. Depressing the save file button starts an interpolation process whereby pseudo-landmarks are generated that lie at equidistant positions along a spline curve. Spacing of the pseudo-landmarks for each curve section is predetermined on an empirical basis according to the likely curvature of the section. For instance, the perimeter of the mouth exhibits a higher degree of curvature than the boundary of the head, hence the densities of pseudo-landmarks in curve sections delineating the mouth are relatively high. Pseudo-landmarks are saved to a MATLAB '.mat' file, as are the original landmarks, which are required if the saved shape is to be reloaded for modification in the future.

Figure D.1 depicts the graphical user interface designed to guide a user through this semi-automatic landmarking procedure.

Step 1

Template image showing correct position of landmarks

Instruction box to guide user through landmarking process

Loads file to be landmarked

Saves landmarks to file

Least-squares polynomial fitting procedure estimates curves to represent facial features

Zoom to enlarge specific features

Step 2

Blue objects represent **inactive** control points and curve sections

Red objects display **active** control points and curve sections

Magenta markers indicate base landmarks associated with one or more curve sections

`Click and drag' procedure to move points corresponding to positions of features

Figure D.1: **The use of the Graphical User Interface for facial shape delineation**. The GUI is designed to guide a user through the steps of the semi-automatic landmarking procedure for a face. Step 1 indicates the initial estimate for the positions of the curve sections using the least-squares fitting procedure. Step 2 displays the results once the 'click and drag' procedure has been initiated to move the points to produce correspondence with the facial features. The example face is deliberately shown at low contrast to enhance the appearance of the landmarks.

Appendix E

Database 2 comprises a training set of faces between the ages of 0 and 20 years, consistent with the accepted biological timespan of childhood and adolescence (section 1.4). Many images comprise faces which are partially occluded, specifically by the subject's hair. In particular, the forehead and ears are most commonly occluded. Since any landmarks used to delineate these regions must be considered unreliable, a new point model was employed (figure E.1), in which the forehead and ears were excluded. This ensured that only reliably placed landmarks were included in the shape model. This point model was also used for Database 3.
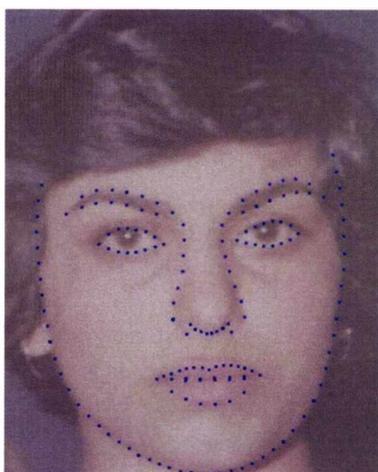
Figure E.1: **New point model for use with Databases 2 and 3**. The occluded features of the forehead and ears are excluded such that only those features capable of yielding reliable landmarks are included in the shape model.

Appendix F

The term "race" is usually defined as in [5] — "one of the major zoological subdivisions of mankind, regarded as having a common origin and exhibiting a relatively constant set of physical traits, such as pigmentation, hair form, and facial and bodily proportions." When considering race from this biological perspective as opposed to a cultural one, there are three traditionally accepted major racial groups — *Caucasoid* (European-derived), *Negroid* (African-derived) and *Mongoloid* (Asian-derived). These three groups exhibit unique facial characteristics and, as such, growth and development with time may proceed differently for the individuals contained therein. Figure F.1 [5] depicts the lateral view of a skull for each of the three racial groups to demonstrate the facial differences between them.
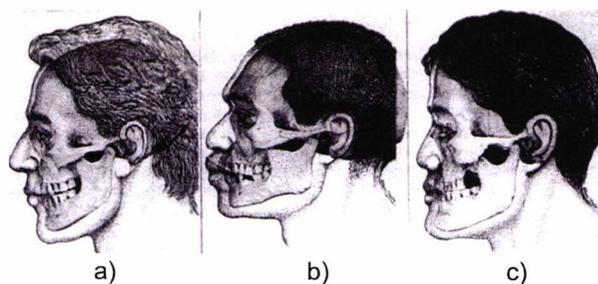
Figure F.1: **The unique structure of the skulls of different racial groups.** The skulls of the Caucasoid, Negroid and Mongoloid races are shown — a) Caucasoid (European-derived) skull, b) Negroid (African-derived) skull, c) Mongoloid (Asian-derived) skull. Images courtesy of [5].

From figure F.1, individuals of the Caucasoid race tend to exhibit a flat face with longer and narrower nasal openings than those of the Negroids and Mongoloids. In addition, the skin is usually light in colour. The Negroid skull typically exhibits a pronounced projection of the lower face with wider and shorter nasal openings than the Caucasoids and Mongoloids. The bridge of the nose is also broader and flatter and the mouth broader with fuller lips. The eyes of Negroid individuals tend to be wider set than those of the other groups. The skin colour varies on a continuum from light to dark brown. In the Mongoloid skull, the face appears flat with a short cranial vault or distance from front to back. The width of the mouth and nasal openings are typically between those of the equivalent features in the Caucasoids and Negroids.

Appendix G

The longer-term objective of the work presented in this thesis is to produce a semi-automatic, integrated system for facial aging, presented in a Graphical User Interface (GUI) format. This would allow a user (from the Police or missing persons' bureaux) to achieve a rigorously age-progressed image for use in investigations. As such, no artistic or specialised computer skills would be required to quickly and effortlessly produce an age-progressed face[1]. Figure G.1 shows a flow diagram indicating the functionality of such a GUI for age-progression and regression.

---

[1]See Appendix B for a discussion of other potential uses for the proposed GUI.
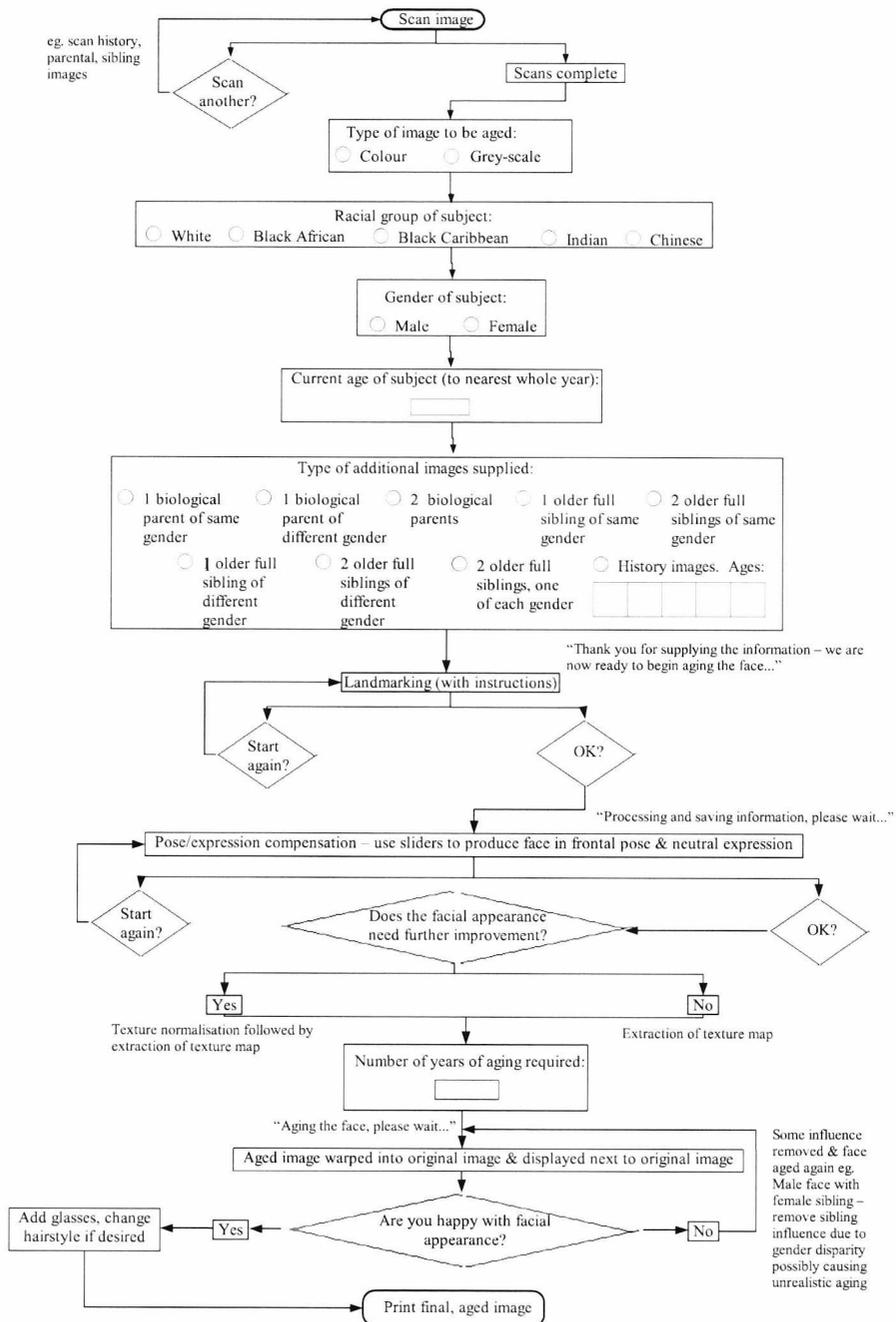
Figure G.1: **Flow diagram depicting the functionality of the proposed GUI.** Such a system would allow a user to age a face in a rigorous manner quickly and effortlessly. The ethnic groups and images of particular relatives are for illustrative purposes only and may be changed or supplemented as required.

The functionality of the GUI may be outlined as follows:

1. The subject image is scanned into the system. Once the scan is complete, two buttons appear: "Scan another?" and "Scans complete". If there are available history images, parental and older sibling images to scan, the user clicks the former button until all scans are completed. Once "Scans complete" is depressed, the user is asked to select, via a set of radio buttons, whether the subject image to be aged is grey-scale or RGB[2].

2. Next, the user is asked to select the racial group of the subject to be aged using radio buttons (as many racial groups as possible in different training sets — for example, White Caucasian, Black African, Black Caribbean, Indian, Chinese). There should also be the option to depress more than one radio button for individuals of mixed race. Once racial selection has been made, the user is asked to select the gender of the subject to be aged, again from radio buttons. Once this is completed, the user is asked to type the current age of the subject (to the nearest whole year) into a box.

3. The user is then asked to indicate the type of additional images supplied – for instance, parental, sibling and historical images. If historical images are selected, the user is then asked to type the ages of the images, to the nearest whole year, into a series of boxes. All of the information supplied by the user (from steps 1 to 3 inclusively) is saved for future use in the program.

4. The subject face then appears on the screen with instructions on how to apply the landmarks to delineate facial shape (a condensed version of the steps outlined in Appendix D). The user then landmarks the face, following these instructions. There is a button for a "Start again" option if the user is unsatisfied

---

[2]If the subject image is grey-scale, the parental/sibling/history images must also be grey-scale. An equivalent statement may be made for an RGB subject image, such that the type of the image is consistent across all images.

with the results. Once the user has completed the landmarking successfully, the "OK" button may be depressed.

5. Once "OK" is clicked on the previous step, a wait-bar comes up displaying: "Processing and saving information, please wait...". The landmark coordinates are placed into a data vector and the file is saved. The data vector is also aligned to the mean of the training data (the appropriate training data used depending upon which racial group was selected in step 2. Both genders are included in the training set for the shape model[3]). The face is then warped to the mean shape and the texture vector extracted. All information is saved without the necessity for the user to be aware of the procedures taking place.

6. The next step involves pose and expression compensation of the shape vector. The subject face appears again and the user is instructed to use sliders to alter the face until it appears at frontal pose and neutral expression with the mouth closed and no teeth visible. There are sliders to control head rotation, nodding, tilting and smiling. Each time the slider is moved and a new shape is produced, the texture of the face is warped to the new shape. The sliders are greyed-out so it is only possible to use one at a time and view the effects on the face gradually. At each stage (each time the user moves to a different slider), the shape vector is saved.

There is also a slider to make the teeth invisible if they are visible once the smile has been normalised[4] Once the user is satisfied with the results, the

---

[3]The exception to this may be in the case whereby a person of mixed racial heritage is presented to the system — to minimise complications, it may prove useful to separate the genders for the shape model as well as for the texture.

[4]Each time the slider is moved, the mouth is incorporated back into the face and the whole texture map including the new mouth is warped to the latest saved shape produced by the other sliders. Hence, the whole face and the effect of the slider is visible to the user.

"OK" button may be depressed. Alternatively, the "Start again" button may be clicked to start the pose and expression compensation procedure once more. Once "OK" is clicked, the user is asked whether the face requires any further improvement. If "N" is clicked, the texture is extracted and saved. If "Y" is depressed, the texture map is normalised (via the reflection technique of section 3.3) and then extracted and saved.

7. The new pose/expression compensated face appears on the screen and the user is asked to type the required amount of aging into a box. The button named "Age the face" may then be depressed and a wait-bar appears, displaying the phrase "Aging the face, please wait...".

The program then uses all of the saved information (type of image, racial group, gender, current age of subject, presence of parental, sibling and/or history images) in order to age the face in shape and texture. The aged texture is then warped to the aged shape. For example, if the user had indicated that the subject was a five year old white male to be aged by five years and that also present was an image of his father, older sister and history images at ages 1, 3 and 4, the program would do the following. Firstly, the white training set would be selected (containing both male and female faces) for the Procrustes alignment phase and the pose/expression compensation phase. For the aging, the male sub-space of the white training set would be used and the history images used to form the historical axis, with this then used to predict the appearance of the subject at the target age of ten years. The consensus axis is formed by considering all male subjects in the current age group and target age group. The parental axis is formed between subject child and parent with a child-to-parent distribution formed using the training set. A similar

sibling axis is formed between the child and his sister. Again, a child-to-sibling distribution may be formed by considering the males in the current age group who have sisters available for use[5]. The maximum likelihood analysis is then performed and the subject aged according to all influences appropriately.

If a mixed race person was presented for aging, it would be necessary to have two consensus axes, one for each racial group. For example, if a person of mixed white/black race was presented, a consensus axis would be formed from the current and target age groups for the white faces and a similar consensus axis for the black faces. Along with the other influences, the face would then be aged according to these racial consensus axes. In this situation, it would be beneficial to have images of both parents.

8. The aged result is then displayed on screen (warped into the original image, as described in section 4.1) adjacent to the original image. If the user is satisfied with the aged appearance, the "Y" button is depressed and there would then be the option of adding a different hairstyle (difficult with the current training set, Database 2) and/or glasses if necessary (cosmetic details). If "N" is depressed, aging could proceed again, this time in a slightly different manner. For instance, if the subject is male and the available parental image is male but the sibling is female, the subject could be aged again excluding the sibling influence (if there is a gender discrepancy such as this, this should perhaps be the first influence to be removed for the aging to proceed once more). Once the user is satisfied with the outcome, cosmetic details such as hairstyle and glasses could be added to complete the age-progressed image.

---

[5]This description of using the algorithm for aging assumes very extensive training sets with parents and siblings of the subjects in the training set available for usage.

Appendix H

For any semi-automatic age-progression system to be considered viable, it must demonstrate the ability to age a face reliably for real-world applications. A commission was made by the BBC to produce extreme age-progressed images of particular celebrities for use in a television programme. The images provided were added to Database 1 and the models recalculated and employed to perform age-progression on the subjects using the consensus approach. Figure H.1 depicts the results of aging the faces of Richard Hammond (Face 1) and Charlotte Church (Face 2) by large increments.
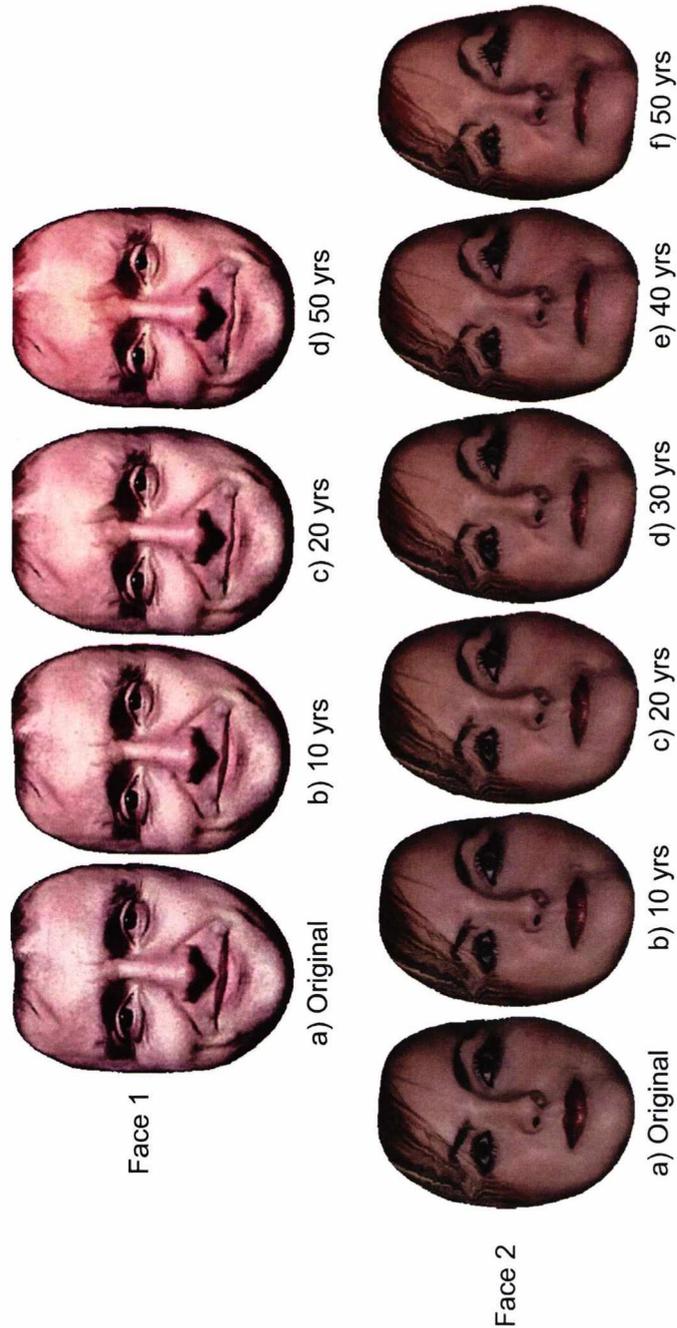
Figure H.1: **Age-progression for use in the media**. The consensus approach was used to perform sequential age-progression on example celebrity faces, provided by the BBC, for use in a television programme. The original age of Richard Hammond (Face 1) is approximately 35 years and the original age of Charlotte Church (Face 2) is approximately 18 years.

Catherine Mary Scandrett   196

From observation of figure H.1, it is clear that the faces have aged yet still retain the subject's unique identity. However, owing to effects outlined in section 3.5 (such as the smoothing of wrinkles and other high-frequency detail in the texture), the appearance of the age-progressed faces does not appear consistent with the degree of aging applied. Hence, in these cases, the addition of a wrinkle-map is necessary subsequent to the age-progression. Figure H.2 depicts the effect of this procedure for subject faces aged by fifty years. From this, it is clear that wrinkle-map addition causes the face to appear older and more consistent in appearance with the specified amount of aging[1]. However, the presence of artefacts in the age-progressed image of Face 2 is notable and may be explained as follows. Firstly, in the original image, the hair is partially occluding the forehead, right eye and eyebrow. Since these textural aberrations are not well-modelled by the rest of the training set, age-progression of the texture map produces artefacts in these facial regions. Additionally, the target ages for the progressions were 85 years for Face 1 and 68 years for Face 2. Since these ages extend beyond the range of the training set, extrapolation beyond the data may also have produced artefacts.

---

[1]The effect is less pronounced for Face 1 than for Face 2. This is because the original Face 1 image already contained high-frequency detail, which may have reduced the visual effect of the wrinkle-map.

a)   b)

Figure H.2: **Consensus age-progression with wrinkle-map addition**. Age-progression using the consensus model alone produces results which appear younger than the target age due to the smoothing of wrinkles and other high-frequency detail in the texture. The addition of a wrinkle-map produces more age-accurate results — a) Original image, b) Result of fifty years of aging using the consensus model with wrinkle-map addition.

As a further example of a real-world application, Beyond International Ltd. production company [122] requested an age-progressed image of the presenter for their television programme, "Beyond Tomorrow". The consensus aging algorithm (trained on Database 1) was applied and the resultant image combined with a wrinkle-map. Figure H.3 [101] depicts the results of this procedure. Some artistic manipulation of the hair (the addition of grey streaks) has been applied.
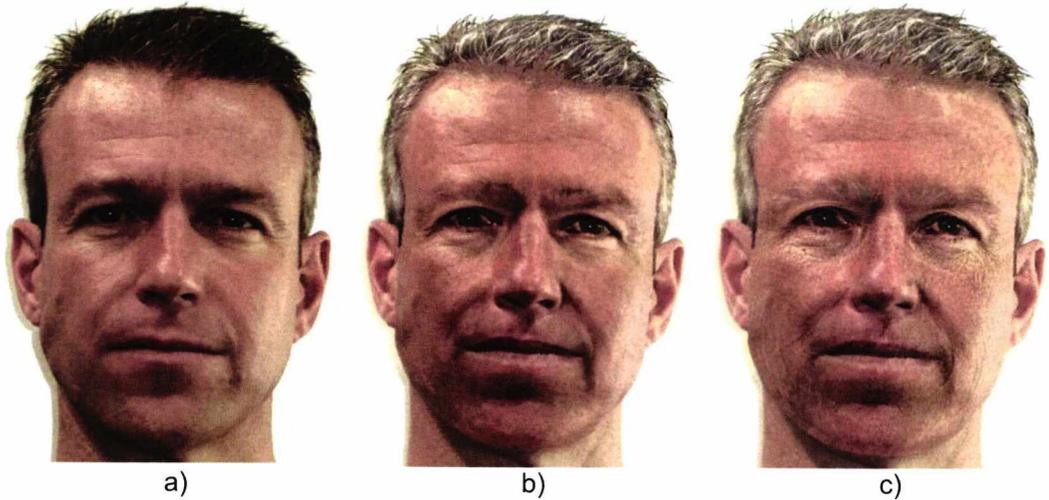
Figure H.3: **Age-progression with and without wrinkle-map addition for a real-world application in the media** — a) Original image of Hayden Turner, presenter for "Beyond Tomorrow" television programme for Beyond International Limited, b) Result of age-progression using the consensus model, c) Result of age-progression using the consensus model with wrinkle-map addition.

H.0.1   Forensic artist commission

The ideas and some of the work presented in this thesis have been discussed with forensic artist Cathy Charsley of the Metropolitan Police. Subsequently, a request was made for an age-progression using the novel techniques for comparison with her own artistically produced image. Two attempts were made. In the first, the consensus axis ($\hat{\mathbf{V}}_{\mathbf{age}}$) only was initially employed to age the face. Since an image of the subject's brother was provided, a second attempt was made, in which the following equation was employed:

$$\mathbf{b}' = \mathbf{b} + \beta_c \alpha_c \hat{\mathbf{V}}_{\mathbf{age}} + \beta_s \alpha_s \hat{\mathbf{V}}_{\mathbf{s}} \qquad (\text{H.1})$$

where α_s is the appropriate scalar multiple for the required degree of aging in the direction of the *sibling* axis, $\hat{\mathbf{V}}_{\mathbf{s}}$. This may be calculated using:

$$\hat{\mathbf{V}}_{\mathbf{s}} = \mathbf{b}_{\mathbf{s}} - \mathbf{b} \tag{H.2}$$

where $\mathbf{b}_{\mathbf{s}}$ is the vector of shape model parameters for the sibling and $\mathbf{b}$ is the equivalent vector for the subject. The coefficients β_c and β_s were used to weight the influence of the consensus and sibling axes appropriately and, in the absence of sufficient pairs of blood relatives in the training sample, their values were determined through subjective experimentation. Appropriate values were found to be: $\beta_c = 1$ and $\beta_s = 0.2$.

Figure H.4 depicts the results of the age-progression using both approaches[2]. From inspection, *the use of the empirical consensus/sibling approach produces more perceptually accurate results than does the consensus approach alone.* That is, the consensus/sibling result appears closer in appearance to a "typical" member of the target age group than does the consensus result. Indeed, the consensus result appears perceptually younger than the target age. There may be several reasons for this, as follows:

1. The consensus axis is formed by effectively weighting the model parameter vectors of the training set by their respective ages, according to equation (3.13). However, since some training examples may appear older or younger than their true age, these discrepancies may cause an age-progressed subject to be perceived similarly. Therefore, it may be more appropriate to weight the parameter vectors by their *perceived* ages.

2. The prototype for each age may not be sufficiently well-defined for accurate age-progression to proceed due to training images which are unconstrained in

---

[2]The artistic age-progression could not be displayed due to denied permission by artist Cathy Charsley on the basis of privacy for the missing child's family.

quality and resolution. If the training set contained a larger number of high resolution images with limited artefacts, more accurate prototypes could be calculated and a more well-defined aging axis computed. This would lead to the production of more accurate progressions.

3. The incorporation of the aged face into the original image may cause it to be perceived as younger than the target age. This is because the image contains the original ears, hair and clothing, which are important external cues to age. Incorporation into an alternative image (for example, the sibling image) could therefore permit more accurate age-perception of the face. This is the case in figure H.4, where image g appears older than image e.

These observations about perceived age reinforce the notion that familial influences must be modelled, where possible, for a more realistic age-progressed result to be achieved.
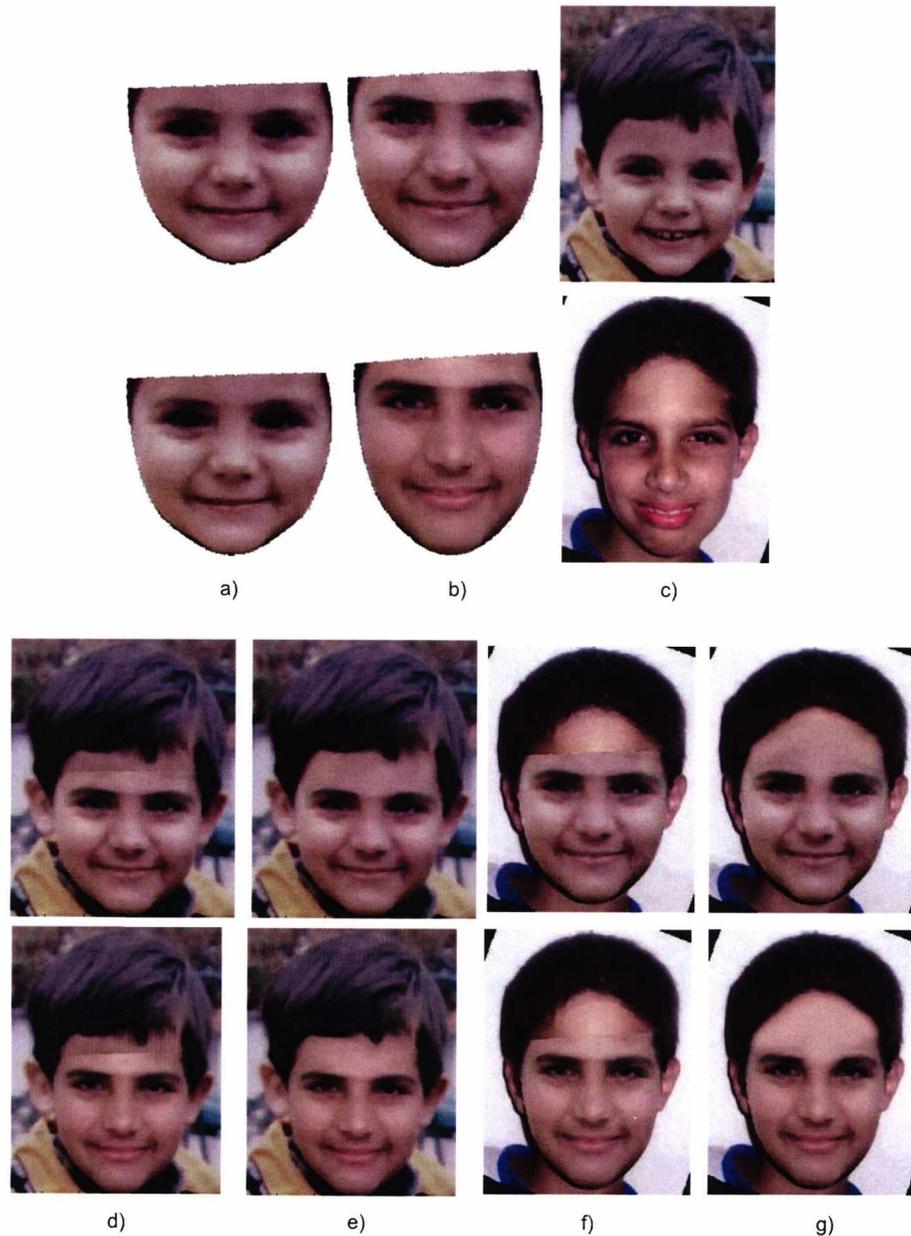
Figure H.4: **Age-progression of a missing subject using the consensus approach and an empirical consensus/sibling approach** — a) Original 2 year old subject, b) Row 1 — age-progression by 10 years using consensus approach, Row 2 — age-progression using consensus/sibling approach, c) Row 1 — original subject, Row 2 — 9 year old sibling, d) Equivalent to b with original hair, ears and clothing, e) Equivalent to d with join line blended artistically, f) Equivalent to d with sibling hair, ears and clothing, g) Equivalent to f with join line blended artistically.

Bibliography

[1] G.J. Y'Edynak and M.Y. Iscan. *Forensic Analysis of the Skull.* Wiley Liss, Inc., 1993.

[2] T. Blythe. National Missing Persons Helpline, Roebuck House, 284 Upper Richmond Road West, London, SW14 7JE. http://www.missingpersons. org, 2003.

[3] R. Perkins. Market research project to support the forensic aging project. Technical report, Marketing on a Shoestring, Kent, UK, 2004.

[4] M.H. Manhein. FACES laboratory, 227 Howe-Russell Geoscience Complex, Department of Geography & Anthropology, Louisiana State University, Baton Rouge, Louisiana, 70803, USA. http://www.lsu.edu/faceslab/law/ ageprogression.htm, 1995.

[5] Karen T. Taylor. *Forensic Art & Illustration.* CRC Press, 2000.

[6] Glenn Miller. National Center for Missing and Exploited Children, Charles B. Wang International Children's Building, 699 Prince Street, Alexandria, VA, 22314-3175, USA. http://www.lsu.edu/faceslab/law/ageprogression. htm, 1995.

[7] National Missing Persons Helpline, PO Box 28908, London, SW14 7ZU. http: //www.missingpersons.org/unidentified_age.asp, 2003.

[8] G. Davies, H. Ellis, and J. Shepherd. Face Recognition Accuracy as a Function of Mode of Representation. *Journal of Applied Psychology*, 63:180–187, 1978.

[9] L.S. Mark, R.E. Shaw, and J.B. Pittenger. *Social and Applied Aspects of Perceiving Faces*, chapter 2: Natural Constraints, Scales of Analysis, and Information for the Perception of Growing Faces, pages 11–49. Lawrence Erlbaum Associates, Inc., 1988.

[10] J.B. Pittenger and R.E. Shaw. Aging Faces as Viscal-Elastic Events: Implications for a Theory of Nonrigid Shape Perception. *Experimental Psychology: Human Perception and Performance*, 1(4):374–382, 1975.

[11] A. Lanitis, C.J. Taylor, and T.F. Cootes. Toward Automatic Simulation of Aging Effects on Face Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):442–455, 2002.

[12] B. Bogin. *Patterns of Human Growth*. Cambridge University Press, 2nd edition, 1999.

[13] L.G. Farkas. *Anthropometry of the Head and Face*. Raven Press, Ltd., 2nd edition, 1994.

[14] D.H. Enlow and M.G. Hans. *Essentials of Facial Growth*. W.B. Saunders Company, 1996.

[15] J. Penry. *Looking at Faces and Remembering Them - A Guide to Facial Identification*. Elek Books, 1971.

[16] C. Wilkinson. *Forensic Facial Reconstruction*. Cambridge University Press, 2004.

[17] M.Y. Iscan and R.P. Helmer, editors. *Forensic Analysis of the Skull*. Wiley-Liss, Inc., 1993.

[18] F. Falkner and J.M Tanner. *Human Growth - A Comprehensive Treatise*, volume 2. Plenum Press, 2 edition, 1986.

[19] J.M. Tanner. *Growth at Adolescence*. Blackwell Scientific Publications, Oxford, 2nd edition, 1962.

[20] S.E. Bishara, L.C. Peterson, and E.C. Bishara. Changes in Facial Dimensions and Relationships between the Ages of 5 and 25 years. *American Journal of Orthodontics*, 85(3):238–252, 1984.

[21] M.A. Taister, S.D. Holliday, and H.I.M. Borrman. Comments on Facial Aging in Law Enforcement Investigation. *Forensic Science Communications*, 2(2), 2000.

[22] J. Lewis. The Physiological and Psychological Development of the Adolescent. Yale-New Haven Teacher's Institute, 1991.

[23] Bernstein Cosmetic Surgery Clinic, 1801 NW Market St., Suite 107, Seattle, Washington, 98107, USA. `http://www.bernsteincosmetic.com`, 2003.

[24] The Franklynn Centre, Franklynn Suite, The Priory, Haywards Heath, Sussex, RH16 3LB, UK. `http://www.franklynncentre.co.uk/nonsurgical.htm`, 2003.

[25] R. Atkins. Nancy Burson: Making Faces. Contemporanea, January 1991.

[26] N. Burson and T. Schneider. Method and Apparatus for Producing an Image of a Person's Face at a Different Age. In *Biological Sciences United States Patent 4276570, 1355*, pages 137–143, June 1981.

[27] K. Ramsland. Forensic Art and how it has been used to apprehend criminals. Online article, 2004.

[28] National Centre for Policing Excellence, Bramshill, Hook, Hampshire, RG27 0JW. *Guidance on the Management, Recording and Investigation of Missing Persons*, 2005.

[29] W. Neville. Florence County Sheriff's Office, Florence, SC, USA. `http://www.forensicartist.com/`, 2001.

[30] S.A. Fusco. Orange County Sheriff's Office, 2500 W. Colonial Drive, Orlando, FL, 32803, USA. `http://www.ocsoartist.com/age_progression_drawings.htm`, 1997.

[31] D'L. Waldron. Ocean Drive, Santa Monica, CA, 90402, USA. `http://www.dlynnwaldron.com/ageprogression.html`, 2005.

[32] B. Clark. *Anthropometry of the Head and Face*, chapter 14, pages 184–186. Raven Press, Ltd., 2nd edition, 1994.

[33] B. Hoddinott. Forensic Artist. `http://www.hoddinott.com`, 2004.

[34] H. McLaughlin. Aprilage Development, Inc. `http://www.aprilage.com`, 1998.

[35] N. Moore. Presenting the Absent - Visualization, Post-Photography and the Real. In *Invencao Symposium, Brazil*, 1999.

[36] S.A. Feik and J.E. Glover. *Craniofacial Identification in Forensic Medicine*, chapter 3: Growth of Children's Faces, pages 203–225. Oxford University Press, 1998.

[37] D.W. Thompson. *On Growth and Form*. Cambridge University Press, 1961. Earlier editions published in 1917 and 1942.

[38] 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. *Experimental Psychology: Human Perception and Performance*, 5(3):478–493, 1979.

[39] L.S. Mark and J.T. Todd. Describing Perceptual Information about Human Growth in terms of Geometric Invariants. *Perception and Psychophysics*, 37(3):249–256, 1985.

[40] J.T. Todd, L.S. Mark, R.E. Shaw, and J.B. Pittenger. The Perception of Human Growth. *Scientific American*, 242(2):132–144, 1980.

[41] L.S. Mark, J.T. Todd, and R.E. Shaw. Perception of Growth: A Geometric Analysis of how Different Styles of Change are Distinguished. *Journal of Experimental Psychology: Human Perception and Performance*, 7:855–868, 1981.

[42] L.S. Mark and J.T. Todd. The Perception of Growth in Three Dimensions. *Perception and Psychophysics*, 33:193–196, 1983.

[43] V. Bruce, M. Burton, T. Doyle, and N. Dench. Further Experiments on the Perception of Growth in Three Dimensions. *Perception and Psychophysics*, 46(6):528–536, 1989.

[44] N. Ramanathan and R. Chellappa. Modelling Age Progression in Young Faces. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 387–394. IEEE, IEEE Computer Society, 2006.

[45] D.M. Burt and D.I. Perrett. Perception of Age in adult Caucasian Male Faces: Computer Graphic manipulation of Shape and Colour information. *Royal Society of London Series B*, 259:137–143, 1995.

[46] A.J. O'Toole, T. Vetter, H. Volz, and E.M. Salter. 3D Caricatures of Human Heads: Distinctiveness of the Perception of Facial Age. *Perception*, 26:719–732, 1997.

[47] A.J. O'Toole, T. Price, T. Vetter, J.C. Bartlett, and V. Blanz. 3D shape and 2D surface textures of human faces: the role of "averages" in attractiveness and age. *Image and Vision Computing*, 18:9–19, 1999.

[48] T.J. Hutton, B.F. Buxton, P. Hammond, and H.W.W. Potts. Estimating Average Growth Trajectories in Shape-Space using Kernel Smoothing. *IEEE Transactions on Medical Imaging*, 22(6):747–753, 2003.

[49] T.J. Hutton, B.F. Buxton, and P. Hammond. Dense Surface Point Distribution Models of the Human Face. In L. Staib, editor, *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (part of Computer Vision and Pattern Recognition)*, pages 153–160, Kauai, Hawaii, USA, December 2001. IEEE.

[50] L. Sirovich and M. Kirby. Low-dimensional Procedure for the Characterisation of Human Faces. *Journal of the Optical Society of America*, 4(3):519–524, March 1987.

[51] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve Procedure for the Characterisation of Human Faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, 1990.

[52] M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[53] M. Turk and A. Pentland. Face Recognition using Eigenfaces. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–590, Maui, Hawaii, USA, December 1991. IEEE, IEEE Computer Society Press.

[54] I. Craw and P. Cameron. Parameterising Images for Recognition and Reconstruction. In Peter Mowforth, editor, *British Machine Vision Conference*, pages 367–370. Springer Verlag, 1991.

[55] T.F. Cootes, D.H. Cooper, C.J. Taylor, and J. Graham. A Trainable Method of Parametric Shape Description. In *2nd British Machine Vision Conference*, pages 54–61. British Machine Vision Association, Springer Verlag, September 1991.

[56] T.F. Cootes, D.H. Cooper, C.J. Taylor, and J. Graham. A Trainable Method of Parametric Shape Description. *Image and Vision Computing*, 10(5):289–294, June 1992.

[57] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active Appearance Models. *IEEE Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[58] T.F. Cootes and C.J. Taylor. *Statistical Models of Appearance for Computer Vision*. University of Manchester, Imaging Science and Biomedical Engineering Stopford Building Oxford Road Manchester M13 9PT, March 2004. This is a continuing work.

[59] T.F. Cootes and C.J. Taylor. Constrained Active Appearance Models. In *8th IEEE International Conference on Computer Vision*, volume 1, pages 748–754, Vancouver, Canada, July 2001. IEEE, IEEE Computer Society Press.

[60] T.F. Cootes and C.J. Taylor. Modelling Object Appearance using the Grey-Level Surface. In E. Hancock, editor, *5th British Machine Vision Conference*, pages 479–488, York, England, September 1994. British Machine Vision Association, BMVA Press.

[61] A. Lanitis, C.J. Taylor, and T.F. Cootes. Automatic Tracking, Coding and Reconstruction of Human Faces using Flexible Appearance Models. *IEE Electronic Letters*, 30:1578–1579, 1994.

[62] M.B. Stegmann. Active Appearance Models — theory, extensions & cases, 2nd Edition. Master's thesis, Technical University of Denmark, 2000.

[63] C.J. Solomon, S.J. Gibson, and A. Pallares-Bejarano. EigenFit - The Generation of Photographic Quality Facial Composites. *Journal of Forensic Science*, In press, 2005.

[64] C.J. Solomon, S.J. Gibson, and A. Pallares-Bejarano. Photographic Quality Facial Composites for Criminal Investigations. In M. Colonna, M. Belviso, and A. Addante, editors, *The 10th Biennial Scientific Meeting of the International Association for Craniofacial Identification (IACI)*, pages 52–60. IACI, 2002.

[65] S.J Gibson, C.J Solomon, and A. Pallares-Bejarano. Synthesis of Photographic Quality Facial Composites using Evolutionary Algorithms. In R. Harvey and J. A. Bangham, editors, *British Machine Vision Conference*, pages 221–230, University of East Anglia, Norwich, UK, 9th-11th September 2003. Britsh Machine Vision Association, BMVA Press.

[66] S.J. Gibson, C.J. Solomon, M.I.S. Maylin, and A. Pallares-Bejarano. Facial Attribute Manipulation for Composite Systems and Computer Graphics Applications. In M.H. Hamza, editor, *3rd IASTED International Conference on Visualisation, Imaging and Imaging Processing*, volume 1, pages 146–151, Benalmadena, Spain, 8th-10th September 2003. International Association of Science and Technology for Development, Acta Press.

[67] C.D. Frowd, P.J.B. Hancock, and D. Carson. EvoFIT: A holisitc, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Perception*, 1(1):19–39, 2004.

[68] I.T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 2nd edition, 2002.

[69] J.E. Jackson. *A User's Guide to Principal Components*. John Wiley & Sons, Inc., 1991.

[70] B.R. Frieden. *Probability, Statistical Optics and Data Testing*. Springer-Verlag, 1991.

[71] S.F. Arnold. *The Theory of Linear Models and Multivariate Analysis*. Wiley & Sons, 1981.

[72] C. Chatfield and A.J. Collins. *Introduction to Multivariate Analysis*. Chapman and Hall, 1980.

[73] M.S. Srivastava and C.G. Khatri. *Introduction to Multivariate Statisitcs*. Elsevier, 1979.

[74] D.C. Lay. *Linear Algebra and its Applications*. Addison-Wesley, 1993.

[75] K. Pearson. On Lines and Planes of Closest Fit to Systems of Points in Sace. *Philosophical Magazine*, 6(2):559–572, 1901.

[76] H. Hotelling. Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology*, 24:417–441, 498–520, 1933.

[77] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 2nd edition, 1990.

[78] A. Webb. *Statistical Pattern Recognition*. Wiley & Sons, 2002.

[79] D.R. Anderson, D.J. Sweeney, and T.A. Williams. *Introduction to Statistics - Concepts and Applications*. West Publishing Company, 3rd edition, 1994.

[80] I.L. Dryden and K.V. Mardia. *Statistical Shape Analysis*. John Wiley, 1998.

[81] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. *International Journal of Computer Vision*, 1(4):321–331, 1987.

[82] R.C. Gonzalez and R.E. Woods. *Digital Image Processing*. Prentice-Hall, Inc., 2nd edition, 2002.

[83] C. Goodall. Procrustes Methods in the Statistical Analysis of Shape. *Journal of the Royal Statistical Society, Series B*, 53(2):285–339, 1991.

[84] F.L. Bookstein. Landmark Methods for Forms without Landmarks: Localising group differences in outline shape. *Medical Image Analysis*, 1(3):225–244, 1997.

[85] F.L. Bookstein. Principal Warps: Thin Plate Splines and the Decomposition of Deformations. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 11:567–585, 1989.

[86] A. Okabe, B. Boots, K. Sugihara, and S.N. Chiu. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. John Wiley & Sons, 2nd edition, 2000.

[87] Images used in initial aging study captured and provided by Dr. Matthew Maylin, 2002.

[88] Image database courtesy of the FG-NET consortium, with kind permission of Dr. Andreas Lanitis. `http://sting.cycollege.ac.cy/~alanitis/fgnetaging/index.htm`, 2004.

[89] National Center for Missing and Exploited Children, Charles B. Wang International Children's Building, 699 Prince Street, Alexandria, VA, 22314-3175, USA. http://www.missingkids.com/, 2004.

[90] V. Blanz and T. Vetter. A Morphable Model for the Synthesis of 3D Faces. In *International Conference on Computer Graphics and Interactive Techniques*, pages 187–194, Los Angeles, 8th-13th August 1999. ACM Press/Addison-Wesley Publishing Co.

[91] T. Vetter. Synthesis of Novel Views from a Single Face Image. *International Journal of Computer Vision*, 28(2):103–116, 1998.

[92] T.F. Cootes, G.V. Wheeler, K.N. Walker, and C.J. Taylor. Coupled-view Active Appearance Models. In M. Mirmehdi and B. Thomas, editors, *The 11th British Machine Vision Conference*, volume 1, pages 52–61, University of Bristol, 11th-14th September 2000. British Machine Vision Association, ILES Central Press.

[93] S.J. Gibson, C.J. Solomon, and A. Pallares-Bejarano. Estimation of Missing Facial Views. One day BMVA Symposium - "Advancing Biometric Technologies", 6th March 2002.

[94] G.J. Edwards, A. Lanitis, C.J. Taylor, and T.F. Cootes. Statistical Models of Face Images - Improving Specificity. *Image and Vision Computing*, 16(3):203–211, 1998.

[95] G.J. Edwards, C.J. Taylor, and T.F. Cootes. Learning to Identify and Track Faces in Image Sequences. In *International Conference on Face and Gesture Recognition*, pages 260–265, 1998.

[96] A. Lanitis, C.J. Taylor, and T.F. Cootes. A Unified Approach to Coding and Interpreting Face Images. In *5th International Conference on Computer*

Vision, volume 5, pages 368–373, Massachusetts, USA, 20th-23rd June 1995. IEEE Computer Society.

[97] T.S. Jebara. 3D Pose Estimation and Normalisation for Face Recognition. Undergraduate thesis submitted for bachelor of engineering, McGill University, May 1996.

[98] S. Shan, W. Gao, B. Cao, and D. Zhao. Illumination Normalization for Robust Face Recognition Against Varying Lighting Conditions. In *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pages 157–164, 2003.

[99] C. Leung. Real Time Face Recognition. Master's thesis, University of Queensland, 2001.

[100] B.P. Tiddeman, D.M. Burt, and D.I. Perrett. Prototyping and Transforming Facial Textures for Perception Research. *IEEE Computer Graphics and Applications*, 21(5):42–50, 2001.

[101] S.J. Gibson. *EigenFIT: A Statistical Learning Approach to Facial Composites*. PhD thesis, School of Physical Sciences, University of Kent, 2006.

[102] K.A. Deffenbacher, C. Hendrickson, A.J. O'Toole, D.P. Huff, and H. Abdi. Manipulating Face Gender: Effects on Categorisation and Recognition Judgments. *Journal of Biological Systems*, 6(3):219–240, September 1998.

[103] H. Abdi, D. Valentin, B. Edelman, and A.J. O'Toole. More about the difference between men and women: Evidence from linear neural networks and the Principal Component approach. *Perception*, 24:539–562, 1995.

[104] T. Kanno, M. Akiba, Y. Termachi, H. Nagahashi, and T. Agui. Classification of Age Group Based on Facial Images of Young Males by Using Neural

Networks. *IEICE Transactions on Information and Systems*, E84-D(8):1094–1101, 2001.

[105] K. Ueki, T. Hayashida, and T. Kobayashi. Subspace-based Age-group Classification Using Facial Images under Various Lighting Conditions. In *7th International Conference on Automatic Face and Gesture Recognition*, pages 43–48, Southampton, UK, April 2006. IEEE, IEEE Computer Society Press.

[106] J. Yang, D. Zhang, A.F. Frangi, and J y. Yang. Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):131–137, January 2004.

[107] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., 2nd edition, 2001.

[108] H. Dehon and S. Bredart. An 'Other-Race' Effect in Age Estimation from Faces. *Perception*, 30(9):1107–1113, 2001.

[109] P.A. George and G.J. Hole. The Role of Spatial and Surface Cues in the Age-Processing of Unfamiliar Faces. *Visual Cognition*, 7(4):485–509, April 2000.

[110] P. Sorqvist and M. Eriksson. Effects of Training on Age Estimation. *Applied Cognitive Psychology*, 21(1):131–135, 2006.

[111] Y. Mamoru and K. Hiroyasu. Gender and Age Estimation from Face. In *2nd Annual HOIP Workshop*, pages 14–17, 2001.

[112] N. Ramanathan and R. Chellappa. Face Verification across Age Progression. *IEEE Transactions on Image Processing*, 15(11):3349–3361, November 2006.

[113] A.M. Burton, S. Wilson, M. Cowan, and V. Bruce. Face Recognition in Poor-Quality Video: Evidence from Security Surveillance. *Psychological Science*, 10(3):243–248, May 1999.

[114] C.H. Liu, H. Seetzen, A.M. Burton, and A. Chaudhuri. Face Recognition is Robust with Incongruent Image Resolution: Relationship to Security Video Images. *Journal of Experimental Psychology: Applied*, 9:33–41, 2003.

[115] R. Kemp, N. Towell, and G. Pike. When Seeing Should Not be Believing: Photographs, Credit Cards and Fraud. *Applied Cognitive Psychology*, 11:211–222, 1997.

[116] G. Pike, R. Kemp, and N. Brace. The Psychology of Human Face Recognition. In *IEE Symposium on Visual Biometrics*, pages 111–116, 2000.

[117] V. Bruce, Z. Henderson, K. Greenwood, P.J.B. Hancock, A.M. Burton, and P. Miller. Verification of Face Identities from Images Captured on Video. *Journal of Experimental Psychology: Applied*, 5(4):339–360, December 1999.

[118] D.A. Roark, A.J. O'Toole, and H. Abdi. Human Recognition of Familiar and Unfamiliar People in Naturalistic Video. In *IEEE International Workshop on Analysis and Modelling of Faces*, pages 36–43, Nice, France, 2003.

[119] A.M. Megreya and A.M. Burton. Unfamiliar Faces are not Faces: Evidence from a Matching Task. *Memory & Cognition*, 34(4):865–876, June 2006.

[120] B. George and C.J. Solomon. An Intelligent Landmarking System for the Generation of High Quality Facial Composites. Forensic Imaging Group, University of Kent, September 2006.

[121] British Broadcasting Corporation, "Honey, we're killing the kids". `http://www.bbc.co.uk/health/tv_and_radio/honey/`, `http://www.honeywerekillingthekids.com`, 2005/6.

[122] Beyond International Limited, 109 Reserve Road, Artarmon NSW 2064, Australia. `http://www.beyond.com.au/index.html`, 2005.