

A Method for Automatic Synthesis of Aged Human Facial Images

Maulin R. Gandhi



Department of Electrical & Computer Engineering
McGill University, Montreal, Canada

August 2004

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of
the requirements of the degree of Master of Engineering

© 2004 Maulin R. Gandhi

“In the attitude of silence the soul finds the path in a clearer light, and what is elusive and deceptive resolves itself into crystal clearness. Our life is a long and arduous quest after Truth.” – Mahatma Gandhi.

*To my parents,
Dr. & Mrs. Rajesh N. Gandhi*

Abstract

Age changes cause major variations in the appearance of human faces. Due to many lifestyle factors, it is difficult to precisely predict how individuals may look with advancing years or how they looked with "retreating" years. In this thesis, a new automatic aging scheme is proposed. The method requires only one input image of a subject and produces an age progressed or age regressed image of the person at a target age. This thesis also presents a complete facial image normalization scheme, compensating an image for variability due to lighting effects, face pose and minor expressions. We also propose a new texture enhancement system to transfer geometric details from multiple surfaces onto a new surface.

Over 800 high quality images are collected from the Internet, ranging from 15 years to 99 years in age. These data are used to train a Support Vector Machine to derive an accurate relationship between a facial image and the age of the subject. This relationship is known as the age prediction function. First, an image warping technique is proposed to localize the major facial features of all the images to the same points. Then, a Support Vector Regression Machine is trained with the warped images to learn the age prediction function.

Signature images for various age groups from our database are created. Using these signature images known as age prototypes, and the Support Vector Regression function for age prediction, we present a new age simulation technique. The proposed method preserves the identity of the subject while enforcing realistic simulation of aging effects on adult face images between 20 and 90 years old.

Sommaire

Les changements d'âge provoquent des variations importantes dans l'apparition de visages humains. En raison de beaucoup de facteurs de style de vie, il est difficile de prédire avec précision comment les individus peuvent apparaître avec les années avançantes ou comment ils apparaissaient avec les années "se retirant". Dans cette thèse, un nouveau projet de vieillissement automatique est proposé. La méthode exige seulement une image d'un sujet et produit un âge progressé ou un âge régressé de l'image de la personne à un âge prévu. Cette thèse présente aussi un projet de normalisation complète de l'image du visage, en compensant une image pour la variabilité en raison de l'éclairage, la pose du visage et les expressions mineures. Nous proposons aussi un nouveau système d'amélioration de texture pour transférer des détails géométriques des surfaces multiples sur une nouvelle surface.

Plus de 800 images de haute qualité sont collectionnées de l'Internet, avec des limites d'âge de 15 ans à 99 ans. Ces données sont utilisées pour entraîner une Machine de Vecteur de Soutien afin de tirer un rapport exact entre une image du visage et l'âge du sujet. Ce rapport est connu comme la fonction de prédiction d'âge. D'abord, une image déformant la technique est proposée pour localiser les caractéristiques importantes du visage de toutes les images aux mêmes points. Ensuite, une Machine de Rétrogradation de Vecteur de Soutien est entraînée avec les images déformées pour apprendre la fonction de prédiction de l'âge. Les images de signature pour les tranches d'âges différentes de notre base de données sont créées.

Utilisant ces images de signature connues comme les prototypes d'âge et la Rétrogradation de Vecteur de Soutien pour la prédiction d'âge, nous présentons une nouvelle technique de simulation d'âge. La méthode proposée préserve l'identité du sujet en appliquant la simulation réaliste des effets du vieillissement sur les images de visages adultes entre 20 et 90 ans.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Martin D. Levine, for his support, advice and guidance in matters of practical, theoretical and financial nature throughout this Masters. Dr. Levine gave me the opportunity to benefit from his wide knowledge and experience. His ideas and fashion of thinking contributed considerably to this thesis.

Special thanks to Karthik Sundaresan for helping me collect images from the Internet and to D'Lynn Waldron for providing us with high quality age progression portraits. I would also like to thank Dr. Levine's research group, Jisnu Bhattacharyya, Bhavin Shastri, Gurman Gill, Harkirat Sahambi and Ajit Rajwade, for their support and discussions during our regular meetings.

Without the support and encouragement from my family and friends, this thesis would have been impossible. I especially want to thank my parents for encouraging me to strive for excellence and advising me to pursue what matters most to me. I am grateful to my sister, Ami Gandhi, for all her love and support, her husband, Nikhil Angra, for his words of wisdom and advices, and Akhil Angra, for keeping us entertained at all times.

Last but not least, I would like to express my lifelong gratitude to my girlfriend, Anju Sharma, for her love, patience, help, encouragement and all the other things which accompany me from the beginning to the end of thesis. Thank you for being there when I needed you most.

Table of Contents

Chapter 1 Introduction	1
1.1 Motivation	3
1.1.1 Capture of wanted fugitives	3
1.1.2 Missing children	3
1.1.3 Updating employee databases	3
1.1.4 Age invariant face recognition	4
1.2 Related issues	4
1.2.1 Diversity of aging variation	4
1.2.2 Collecting training data	5
1.3 Overview	6
1.3.1 Age prediction	7
1.3.2 Age Simulation	8
1.4 Thesis contributions	9
1.5 Structure of the thesis	9
Chapter 2 Previous Work	10
2.1 Facial age prediction	11
2.2 Facial Age Simulation	13
2.2.1 Machine learning for age simulation	13
2.2.2 Medical face measurements for aging	20
2.3 Summary	24
Chapter 3 Image Normalization	25
3.1 Illumination Correction	27
3.1.1 The Single Scale Retinex	28
3.1.2 Histogram fitting	29
3.1.3 Experiments on illumination invariance	33
3.1.4 Summary	37
3.2 Image Warping	38
3.2.1 Feature Specification	39
3.2.2 Warp Generation	41
3.2.3 Pixel mapping and interpolation	44

3.2.4 Results of image warping.....	49
3.3 Summary.....	52
Chapter 4 Age Prediction.....	53
4.1 Source and description of training database	54
4.1.1 Database normalization.....	55
4.2 Support Vector Machines	58
4.2.1 Optimal separating hyperplanes.....	58
4.2.2 Support Vector Regression and the ' ϵ -insensitive zone'	59
4.2.3 Kernel functions for non-linearity.....	60
4.2.4 Convergence tolerance (e)	61
4.2.5 Benefits of SVMs.....	61
4.3 Experiments on the CBIR Aging Database.....	61
4.3.1 Optimum feature space	63
4.3.2 Best kernel and parameters (K , P).....	68
4.3.3 Best regression parameters (ϵ , e)	69
4.4 Summary.....	72
Chapter 5 Age Simulation.....	73
5.1 Image Based Surface Detail Transfer (IBSDT)	74
5.2 Age Prototypes.....	78
5.2.1 Extension to IBSDT.....	78
5.2.2 Texture enhancement results.....	80
5.2.3 Age prototyping	81
5.3 Image synthesis.....	86
5.3.1 Approach.....	86
5.3.2 Experimental Results	88
5.4 Summary.....	96
Chapter 6 Conclusion and Future Directions.....	97

List of Figures

Figure 1.1: Effects of makeup and plastic surgery on perceived age.....	6
Figure 1.2: Block diagram of the normalization process	7
Figure 1.3: Block diagram of the training process.....	8
Figure 1.4: Block diagram of the age progression/regression.....	8
Figure 2.1: Different facial regions used for age estimation [Lanitis, 2002].....	12
Figure 2.2: Comparison of transformed and actual images [Lanitis et al., 2002].....	16
Figure 2.3: Transferring texture to simulate aging effects [Shan et al., 2001]	17
Figure 2.4: Texture processing by wavelet prototyping [Tiddeman et al., 2001].....	19
Figure 2.5: Age transformation examples [Tiddeman et al., 2001].....	19
Figure 2.6: Results depicting facial deformation [Hussein, 2002].....	22
Figure 2.7: Results of combining wrinkle simulation and facial deformation techniques [Hussein, 2002].....	23
Figure 3.1: Images of the same person under different lighting conditions.....	25
Figure 3.2: Images of the same person with different head poses	26
Figure 3.3: Effect of scale c , on Single Scale Retinex processing of a face image	29
Figure 3.4: Unnatural illumination caused by histogram equalization of the R-Image	30
Figure 3.5: Histogram fitting process on a sample image.....	32
Figure 3.6: Training images used for the first experiment.....	34
Figure 3.7: Scale, c , versus recognition accuracy	34
Figure 3.8: Scale, c , versus average recognition accuracy.....	35
Figure 3.9: Scale, c , versus average recognition accuracy for more training images	36
Figure 3.10: Illumination correction flowchart.....	37
Figure 3.11 Deforming an image by warping to a new shape	38
Figure 3.12: Feature specification on a face image	40
Figure 3.13: Mapping the source shape to the target shape	42
Figure 3.14: Sampling the contours to produce points across the face region.....	42
Figure 3.15: Forward versus reverse mapping.....	45
Figure 3.16: Nearest neighbour versus bilinear interpolation.....	46
Figure 3.17: Derivation of bilinear interpolation.....	47
Figure 3.18: Warping program flowchart	48

Figure 3.19: The original three test images with landmark points (red crosses) connected by smooth cubic splines (green curves).....	49
Figure 3.20: Original images warped to average shape.....	50
Figure 3.21: Average image produced by warped face images.....	50
Figure 3.22: Original images scaled and rotated to align eye positions.....	51
Figure 3.23: Average image obtained by aligning eye positions.....	51
Figure 4.1: Histogram of male and female images in the database.....	55
Figure 4.2: Block diagram of the normalization process.....	56
Figure 4.3: Sample cropped images and their age from the CBIR Aging Database.....	57
Figure 4.4: Notion of the optimal separating hyperplane.....	58
Figure 4.5: The ' ϵ -insensitive zone' in linear Support Vector Regression.....	60
Figure 4.6: Results of experiments on various image sizes.....	67
Figure 4.7: Graph of Absolute Error Vs e and ϵ	70
Figure 4.8: Number of Support Vectors versus e and ϵ	71
Figure 4.9: Squared Cross Correlation Coefficient versus e and ϵ	72
Figure 5.1: Two synthetic images of the same person at different ages.....	76
Figure 5.2: Age progression with various values of σ	77
Figure 5.3: Age regression with various values of σ	77
Figure 5.4: Sample images from the 40 - 44 year age group and their average.....	80
Figure 5.5: Texture enhanced male age prototype with different values of σ	80
Figure 5.6: Flowchart of age prototype creation.....	82
Figure 5.7: The male age prototypes.....	83
Figure 5.8: The female age prototypes.....	84
Figure 5.9: Typical graphs from the age prototypes creation process.....	85
Figure 5.10: Flowchart of age simulation.....	88
Figure 5.11: Age simulation results on subject 1.....	89
Figure 5.12: Age simulation results on subject 2.....	90
Figure 5.13: Age simulation results on subject 3.....	91
Figure 5.14: Age simulation results on subject 4.....	92
Figure 5.15: Comparison of age simulations to actual images from Subject 5.....	94
Figure 5.16: Comparison of age simulations to actual images from Subject 6.....	95
Figure 6.1: Attaching hair information to the age progression.....	99

List of Tables

Table 2.1: Summary of age prediction results	24
Table 3.1: Description of the landmark points and curves	40
Table 4.1: Examples of Kernel functions	60
Table 4.2: Spatial Vs Frequency Domain as a feature space	64
Table 4.3: Masking Vs Not Masking on the feature spaces.....	65
Table 4.4: Relative and actual image sizes	66
Table 4.5: Results of applying different kernels for age prediction.....	68

Chapter 1

INTRODUCTION

“I often get e-mails inquiring about software into which one can put a photograph and have it automatically turned into an age-progressed portrait. There is no such software.” [Waldron, 2003]¹. Our objective in this thesis is to develop software that automatically simulates age changes on a facial image. Given a *single* image of the subject and a *target age*, we construct a new image of the subject to reflect the new age.

Aging is an inevitable process. Age changes cause major variations in the appearance of human faces. Some aspects of aging are fairly uncontrollable and are largely based on hereditary factors; others are somewhat controllable, resulting from many social factors including smoking, stress and lifestyle, among others. For instance, according to data provided by the Roswell Park Cancer Institute [Aprilage, 2002], a person who has been smoking a pack of cigarettes each day takes on the wrinkles of a non-smoker 1.4 times older. Thus a 30 year old smoker would look 42.

¹ D'Lynn Waldron is an author, artist and photographer. She is a classically trained artist who has also studied physical anthropology which allows her to do a forensic analysis of the structure of a face and understand how age changes it. As a scientist, she helped develop the computer technology used in publishing, the graphic arts, and the movies. Digital age progression is one of the many services she offers. We personally communicated with her.

One of the fundamental objectives of this thesis was to explore the use of Support Vector Machines (SVMs) [Vapnik, 1995] for predicting the age of a given facial image. SVMs have had significant success in the past few years on a wide range of classification problems in Computer Vision [Osuna, 1997], but haven't been applied to the age prediction problem yet. Our approach is based on a new result demonstrating that SVMs can learn the aging function to a reasonable margin of error. A global aging function for the entire population² is learnt, irrespective of lifestyle, smoking habits or health of the individuals. There may be several confounding issues related to this approach, but it is the first stage towards tackling the problem of human aging. Our approach is based on the presence of a common trend in the aging process that occurs across the whole population. We determine this global trend. Given a set of images³ that have been normalized for illumination effects, facial expression, pose and shape variability, SVMs are trained to learn the global aging function.

In this thesis, we also address the synthesis of new facial images of subjects for a given target age. We present a novel scheme that uses the global aging function to perform this synthesis. Our scheme enforces perceptually realistic images by preserving the identity of the subject.

Our method has several advantages. Primarily, it is an automatic system that does not involve the expertise of a sketch artist to create aged portraits. It saves the time and money of users of this system since it produces the output within minutes. Secondly, it is based on a simple and clean theoretical foundation with the clearly stated limitations and algorithm breaking points. Thirdly, only a single image of the subject is required to perform the synthesis. Lastly, the synthesized faces are visually realistic.

A good quality image⁴ is essential for the method to work well. If the face is clearly visible in the image, the age prediction and synthesis system will be more accurate.

² The population is represented by a database of images

³ The set of images are labelled by age of the subject

⁴ Good quality indicates high resolution (at least 100 x 150), a well lit face with little or no expression or rotation.

1.1 Motivation

Various real-life applications could benefit from an automated aging system like ours. The following examples demonstrate some uses of the aging software.

1.1.1 Capture of wanted fugitives

The most common usage for such a system would be to assist in the capture of wanted fugitives, [CrimeLib, 2002]. Often, suspects who are wanted for crimes and have not been located, and the only available photographs are outdated. This system would be useful to predict what their current or future portraits would look like.

Some fugitives change their identities with facial hair, weight gain, weight loss, a disguise or hair loss. These visual effects will not be captured by our system. However, many criminals are predictable, rigid and conservative and do not make much effort to disguise themselves [CrimeLib, 2002]. Hence, our system would be useful in such cases.

1.1.2 Missing children

A special case of this application would be to predict the current facial appearance of persons missing for several years.

According to the NISMART-2 research [Flores, 2002], an estimated 797,500 children were reported missing from home in the year 1999. The National Center for Missing and Exploited Children [NCMEC, 2001] offers age progression and facial reconstruction services which take up to 7 days to produce. With the enormous number of missing children, automated software to produce the age progressed portraits would be practical and useful.

1.1.3 Updating employee databases

Numerous companies have employee records with photographs of the employees stored in their database. After a certain number of years, the employees have to update their photographs to match their current appearance. Automatic software such as ours might eliminate the need for this, by automatically aging the entire database of employees [Lanitis et al., 2002].

1.1.4 Age invariant face recognition

Image variability due to changes in age can easily impair current face recognition systems. Faces of subjects would look more 'similar' to faces in their own age group causing misclassifications during the recognition phase. This concern could easily be handled with a system such as ours, by normalizing all the faces to a single age. Normalization would eliminate any variability due to age changes.

1.2 Related issues

There were several issues encountered when attempting to develop the method. Compared to other sources of variation such as illumination, pose and facial expressions, variability due to aging has its own unique characteristics that make it difficult to model. These issues are outlined in the following sections.

1.2.1 Diversity of aging variation

Aging is essentially similar for all individuals. It produces the appearance of fine lines and wrinkles on the face, lips thin out, hair colour changes, hairlines recede and causes balding in some men [CrimeLib, 2002]. However, the individual diversity is broad. Factors such as race, gender, genetics and lifestyle come into play during the aging process.

It has been established that the individual's style of living considerably affects the aging process. As mentioned earlier, smoking is known to be a leading cause for faster aging. A healthy lifestyle and exercise can slow down the aging process while stress and extreme weather conditions such sun exposure can accelerate aging. Skin loses elasticity, starts to sag and wrinkle, and becomes leathery due to repeated exposure to the sun [Smith, 1997].

We focused on the global aging trend in this research, simply finding a relationship between the appearance of the subjects to their age and gender. In order to produce a more accurate model, it would be essential to take additional factors such as race, genetics and lifestyle into account. Primarily due to a lack of information on our subjects in our database, we did not incorporate these factors.

A further study would be required to incorporate race, genetics and lifestyle into the aging model. These factors would have to be studied in depth to quantify their effect in terms of

years. Then, when creating an aged portrait, the factors will be incorporated to determine the facial appearance at a target age.

1.2.2 Collecting training data

While collecting training data, it is important to distinguish between the actual age and perceived age of an individual. Effects such as heavy make-up, plastic surgery or more recently the widespread use of Botulinum Toxin Type A injections⁵ can make a person look a lot younger than they actually are. This effect comes into play when collecting face images whose actual age does not match the perceived age.

Figure 1.1 illustrates how facial plastic surgery and makeup can change the perceived age of an individual [Holcomb 2003]. In Figure 1.1(a), subject 1 has concealed numerous cues that reveal her actual age after the application of makeup. It is even more evident in Figure 1.1(c) and (d) where subject 2 has undergone plastic surgery to make her look youthful.

This phenomenon was a major set-back while collecting images from the Internet for our aging database. We simply discarded images of females whose actual age did not reflect the perceived age, hence our database consisted of mainly male subjects. Further details on our database can be found in Section 4.1.

⁵ More commonly known as botox injections [Botox, 2004]



Figure 1.1: Effects of makeup and plastic surgery on perceived age

1.3 Overview

Our approach is based on learning a global aging function. We use a number of training images to learn the relationship between the coded representation of the face and the actual age of the subjects. Once this relationship is established, it is possible to estimate the age of a previously unseen image. The relationship is also used to estimate the ages of synthesized images. Some terms need to be defined first.

Age Progression: Age progression is the computer-generated manipulation of a photograph to make the subject appear several years older, [LSU Faces Laboratory, 2003]. In other words, it is the process of synthesizing new images of an individual at a future age given their current appearance.

Age Regression: Age regression is the process of synthesizing new images of an individual at an earlier age given their current appearance.

Age Prototype: An age prototype is a visual facial model of a particular age or age group. It is a texture enhanced ‘average’ image of a set of images, capturing the main characteristics of the set.

Our method is broken down into two main categories, age prediction and age simulation outlined next.

1.3.1 Age prediction

Age prediction involves the use of Support Vector Machines for predicting the age of a given face image. A database of facial images labelled by age is used to train the SVM for this estimation (See Section 4.1 for a detailed description of the database).

To ensure that the variability of the images in the training system was only due to age of the subject, we carried out a pre-processing step of normalizing all the images in the database for illumination, pose and facial expression. All the images are also warped to the same shape in order to localize the facial features (and wrinkles) at roughly the same points.

The block diagram in Figure 1.2 summarizes image normalization with an example of the transformation.

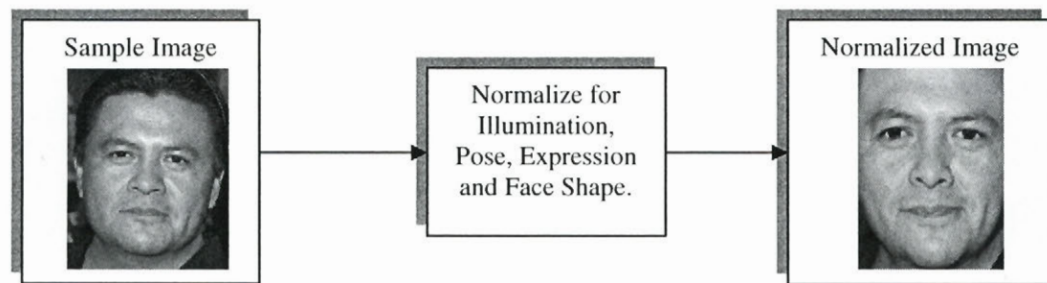


Figure 1.2: Block diagram of the normalization process

Then, a SVM is trained with the normalized facial images to learn the age prediction function. Equation 1.1 contains the function, F , to be learnt.

$$A_p = F(I)$$

1.1

where I is the coded representation of the face image and A_p is the predicted age.

A block diagram of the training process is shown in Figure 1.3.

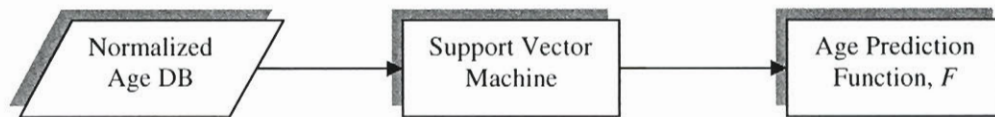


Figure 1.3: Block diagram of the training process

1.3.2 Age Simulation

To simulate age changes in new face images, we combined the aging function and a method to transfer wrinkle information between subjects. Given the original image and a target age, we use an optimization approach to perform age progression or regression. The optimization is carried out in the image synthesis process, illustrated by the flowchart in Figure 1.4.

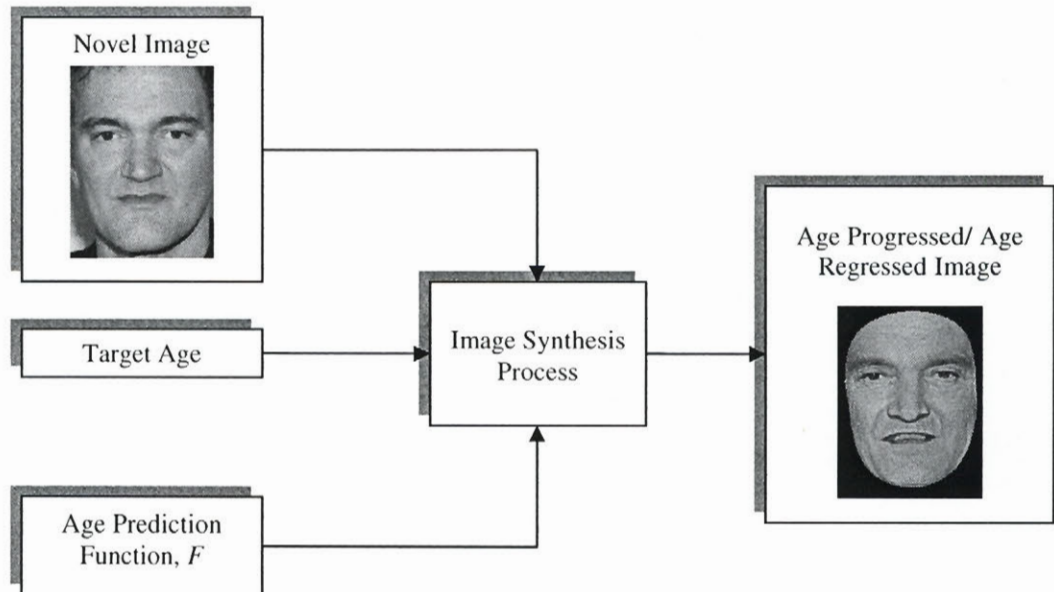


Figure 1.4: Block diagram of the age progression/regression

1.4 Thesis contributions

Significant contributions of this thesis include:

- A novel framework for automatic age progression in human face images using Support Vector Machines (SVMs) for age prediction. This thesis describes the first successful application of SVMs to facial age prediction.
- A novel extension to the Image Based Surface Detail Method to handle multiple images with useful applications in computer graphics for texture enhancement
- A unique approach that involves the combination of the Retinex and Histogram Equalization techniques to carry out illumination normalizations that outperforms all other methods in the literature. We have come up with a very close ideal solution to the illumination compensation problem.

1.5 Structure of the thesis

The next chapter provides background information on the state-of-the-art in age prediction and age simulation.

Chapter 3 describes the image normalization process. An in-depth view is presented on the techniques used to perform illumination invariance and methods to eliminate variability due to face shape, expression and pose. The goal is to normalize the facial image so that the only variation is the age of the subjects.

Chapter 4 explains the core technology investigated in this thesis, Support Vector Machines. This chapter gives a brief detailed mathematical formulation of SVMs and their application to the age prediction problem.

Chapter 5 deals with the scheme used for the creation of ‘age prototypes’ and the method behind synthesis of new face images. This chapter is the heart of our system and combines results from Chapters 3 and 4 to achieve our goal of automatic age progression. Experiments and results on various images for age progression and regression are also presented.

Finally Chapter 6 summarizes our work and suggests some possible directions for future research.

Chapter 2

PREVIOUS WORK

A limited amount of research has been carried out, since only recently have researchers begun studying this phenomenon. Previous work on aging can be broken down into two major categories; (i) Facial age prediction and (ii) facial age simulation. As the name suggests, age prediction involves the use of training and test sets to generate a model that can estimate the age of a given face in an image. Simulation of facial aging effects on the other hand determines on how a subject's appearance changes with age, using this information to create age progressed (or regressed) images of a previously unseen individual.

Facial age simulation is further broken down into two categories. The first uses computer vision techniques, performing age progression via learning schemes and image-based methods. Typically, face shape and texture changes are modelled by a machine learning system. The models are used to synthesize new images to simulate age changes. The second category models facial aging by means of a face anthropometry theory⁶. Aging individuals then involves using measurements provided by the theory to modify their facial appearance.

⁶ Face anthropometry theory is the science dedicated to the measurement of the human face. Medical face anthropometry theory has founded the answer and determined all the features that vary with advancing years.[Farkas, 1987; Farkas, 1994]

2.1 Facial age prediction

Age prediction is concerned with the use of a training set to train a model that can estimate the age of unseen facial images. The estimation is heavily dependent on the training images (i.e. the database in use). Among the first to research age prediction were Kwon and Lobo [Kwon and Lobo, 1993; Kwon and Lobo, 1999] who proposed a method to classify input face images into one of the following three age groups: babies, young adults and senior adults. Their study was based on geometric ratios and skin wrinkle analysis. Geometric ratios were first computed from facial features to distinguish babies from adults. The ratio of the distance between the eyes to the distance between the eyes and nose worked best for separating babies from adults. After that, seniors were distinguished from young adults and babies by detecting and measuring skin wrinkles using snakelets. Finally, a fusion rule was used to combine the ratios and wrinkle information to judge age category. Their method was tested on a database of only 47 high resolution face images containing babies, young and middle aged adults. Kwon and Lobo reported 100% classification accuracy on these data. Several issues are pertinent to their research. First, the method was tested on only 47 images, making it difficult to evaluate. Second, just three age groups were tested. Finally, the method required a relatively high resolution image (at least 256 x 256) to perform wrinkle analysis.

Realizing the shortcomings of Kwon and Lobo, Horng et al. [Horng et al., 2001] proposed a faster and more robust system for age group classification. Their method was composed of the same fundamentals but different approaches to solving geometric ratios and wrinkle analysis. Neural Networks were employed to perform classification, achieving 81.6% accuracy for 230 experimental images. However, still only three age groups were used.

Hayashi et al. [Hayashi et al., 2002] focused their study on facial wrinkles for the estimation of age and gender. They used a database of 300 individuals ranging from 15 years to 64 years of age, taken under controlled conditions. Skin regions were first extracted from the face images, followed by histogram equalization to enhance wrinkles. Then, a special Hough transform, DTHT (Digital Template Hough Transform) was used to extract both the shorter and longer wrinkles on the face. Finally, a look-up table was used for age and gender estimation (classified by a 10 year age group). Their experiments were not very successful on the age classification task though, achieving only 27% accuracy on age estimation, and

83% on gender classification. It is important to note that they did not mention the size or source of their test set to generate their accuracy values. Moreover, it worked on Japanese faces only since the training database consisted of Japanese faces. Hayashi et al. also note the difficulty of extracting wrinkles from females' ages between 20 and 30 due to presence of makeup.

Lanitis [Lanitis, 2002] empirically studied the significance of different facial parts for automatic age estimation. The algorithm is based on his previous work on statistical face models (described in Section 2.2). His investigation involved the assessment of the following face regions: the whole face (including the hairline), the whole internal face, the upper part of the face and the lower part of the face. Figure 2.1 (taken from [Lanitis, 2002]), illustrates the different facial regions used.

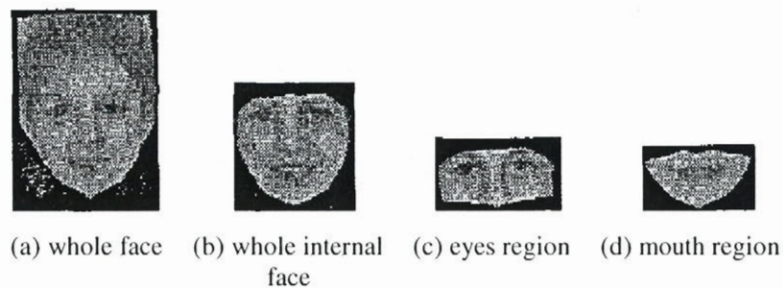


Figure 2.1: Different facial regions used for age estimation [Lanitis, 2002]

Experimental results revealed that the area around the eyes Figure 2.1(c), proved to be the most significant for age estimation. The model of the upper facial part minimized estimation error and standard deviation resulting in a mean error of 3.83 years and a standard deviation of 3.70 years. Lanitis claims that introduction of the hairline (when using the whole face) has a negative effect on the results. He claims that the increased variability of the hair region distorts the age estimation task. His study was limited to subjects ranging from 0 to 35 years old, and contained 330 images, of which only 80 were used for testing purposes. Evidently, faces with more wrinkles weren't used, leaving in doubt his ability to estimate the age of subjects older than 35 years.

Most of the research in this area is very limited by the size and quality of the database used. Some researchers have focused on particular age groups only, while others use an extremely wide classification range. Primarily due to the lack of a good database, a global

age estimation function, covering an extensive range of ages has yet to be developed. From the research thus far, we can conclude that the region around the eyes would be most useful for age estimation and that the faces should be cropped at the forehead to avoid the hairline that could impede age estimation.

2.2 Facial Age Simulation

Facial age simulation involves the reconstruction of a given facial image by simulating aging effects on the image to reflect how an individual might look like in the future, or in the past. There are two main approaches to solve this problem; (i) using machine learning methods to learn the changes in face parameters or (ii) using scientific measurements to capture the aging process in face images. These approaches are discussed in the following sections.

2.2.1 Machine learning for age simulation

Lanitis et al. [Lanitis et al., 2002] describe how the effects of aging on facial appearance can be explained using learned age transformations. Their study presents experimental results to show that reasonably accurate estimates of age can be made for unseen images. Furthermore, Lanitis et al. demonstrate how their proposed framework can be used for simulating aging effects on new face images. The basis of their approach is a statistical face model [Edwards et al., 1998], [Lanitis et al., 1997] that allows compact and reversible coding of face images. Training images are used to learn the relationship between the coded representation of face images and the actual ages of subjects. Once this relationship has been established, it is possible to estimate the age of a previously unseen individual, based on the coded representation and to reconstruct the appearance of an individual at any age. The face model consisted of the shape (3D) and the texture model (2D intensity model).

The *shape model* is presented with training examples represented by the 2D coordinates of 50 landmarks visually placed at predefined positions. Using these data, a pre-processing step of alignment is carried out to minimize shape variation due to rotation, scaling and translation. The shape model explains shape variation in individual object features and the spatial relationship between them. The *intensity model* on the other hand is trained with examples that have been warped to the mean shape (to eliminate shape variation in the training set) using 3600 pixels from each image to train the model. The resulting shape-

intensity representation required 50 model parameters (eigenvectors) to explain 95% of the variance.

A statistical face model is generated from the training examples represented by a set of variables. The mean face is calculated and deviation of each face from the mean is established (by PCA). As a result, training examples can be reconstructed using Equation 2.1.

$$X = X_m + Pb, \quad 2.1$$

where X is a training face, X_m is the mean face, P is the matrix of eigenvectors and b the eigenvalues (weights or model parameters).

It is therefore possible to calculate the set of model parameters (or eigenvalues) corresponding to a given example by Equation 2.2:

$$b = P^{-1}(X - X_m) \quad 2.2$$

Active shape model search [Cootes et al., 1995] is used for automatically locating the facial landmarks on a new image so that it is possible to obtain a model-based parametric description of a new face image. However, Lanitis et al. [Lanitis et al., 2002] assumed that the landmarks were already available for their set of experiments, which may have been located manually.

Next, a genetic algorithm is used to calculate the optimum parameters of the aging function responsible for aging variation in the images. The equation of the required function is:

$$age = F(b), \quad 2.3$$

where age is the actual age of an individual, b is the vector containing the 50 raw model parameters (eigenvalues) and F is the aging function

A function that maps the model parameters to age is now available; a quadratic function was empirically determined to minimize the mean estimated age. Given certain model parameters of a new image, the age can be computed via the function in Equation 2.3.

In order to synthesize a new image given the age, the inverse function has to be computed. Lanitis et al. [Lanitis et al., 2002], adopted a simulation approach, where a large number of plausible combinations of the elements of b are generated. For each combination, the age is estimated using the aging function. As a result, a number of b vectors corresponding to each age among the range of ages were generated, which were then used to create a look up table showing the most typical set of b parameters for each age within the age range of interest. This permitted real-time conversion from age to the model parameters b .

Lanitis et al. [Lanitis et al., 2002] also discussed four ways of deriving an aging function for a new face image for an individual from a given set of face images:

- *Global aging function*: The simplest approach where all individuals are assumed to undergo similar changes in appearance due to aging.
- *Appearance specific aging function*: Exploiting the observation that people who look similar tend to age in a similar way.
- *Weighted aging function*: Since the training data is limited, the previous method will fail in the absence of a face in the database that is similar to the new face. Therefore aging is performed by combining aging functions of all training individuals by weights that describe how similar these training individuals look to the new face.
- *Weighted person-specific aging function*: Integrating life-style profile of each individual with facial appearance into a weighted aging function. This definition extended their earlier assumption to individuals who looked similar and lived similar lifestyles who were likely to age in a similar manner.

Hence, given a new face image, its aging function is defined using one of the four methods above. Calculating a new set of feature parameters from the aging function could then simulate the appearance of this subject.

The work by Lanitis et al. [Lanitis et al., 2002; Lanitis, 2002; Lanitis and Taylor, 2000] is encouraging and provides important experience for further study in this topic. The main issue related to their work is the use of a relatively small database containing only 500 face images of 60 subjects which display other types of variation besides aging, such as significant changes in 3D pose, lighting, expression, image quality, and resolution. Their database consisted of scanned photographs from the subjects' albums accounting for the

poor quality. A database of face images taken under fairly controlled conditions would help improve performance. Furthermore, their attention is mainly focused on age variation between childhood and adulthood (less than 30 yrs old), due to the availability of images. It would be interesting to see the result of their method on older faces with more wrinkles and stubble. Additional texture processing would be necessary when dealing with wrinkles, since their method would most likely average these out due to their fine structure. The results weren't visually appealing either, possibly due to poor database quality. Some age transformation results are shown in Figure 2.2. Figure 2.2(b) displays the reconstruction of the unseen image in Figure 2.2(a) at the target age and Figure 2.2(c) is what the reconstruction is actually supposed to look like. It is clear from these results that the reconstructed and target image are quite different, and this method may not be a very good solution to the problem mainly due to smoothing of wrinkles.

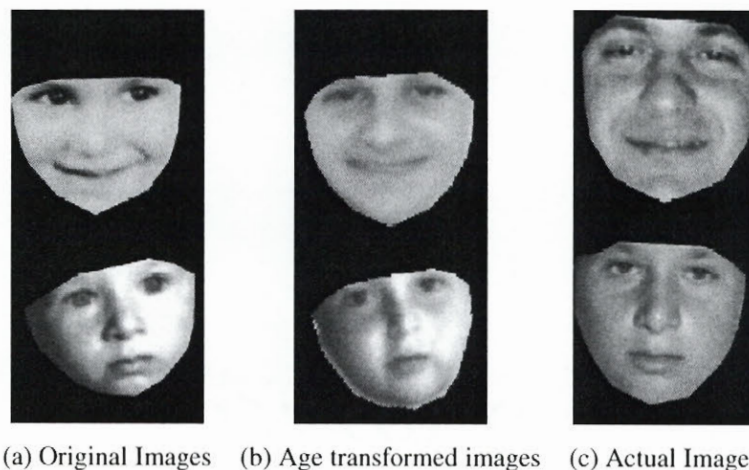


Figure 2.2: Comparison of transformed and actual images [Lanitis et al., 2002]

Shan et al. [Shan et al., 2001] proposed a novel method known as Image-Based Surface Detail Transfer (IBSDT), by which the geometric details of an object are captured from a single image in such a way that it is independent of its reflectance property. These captured geometric details can then be transferred to another surface to produce the appearance of the new surface with added geometrical details while its reflectance properties are preserved. A Gaussian filter controls the scale of the surface details to be transferred. The process is very similar to transferring high frequency components from a source image to a destination

image. They proposed an application of this algorithm to age simulation. Geometrically, the difference between the skin surface of an old person and a young person's is that the old person's skin surface has more bumps (wrinkles) on the face. The basis of Shan et al.'s approach is that if the bumps of an old person are transferred to a young person, then the young person's face will look old. Results from their work shown in Figure 2.3.

Alignment is performed by first marking face boundaries and face features such as eyes, nose and mouth. A triangulation-based image warping is then used to warp the source to the destination. IBSDT is then applied to pixels inside the face boundary while the pixels in the regions of the two brows, the two eyeballs, nose top and mouth are not modified by IBSDT. Different values of σ (width of the Gaussian filter) in Figure 2.3(b), (c), (e) and (f). The original images of a young and senior adult are shown in Figure 2.3 (a) and (d) respectively.

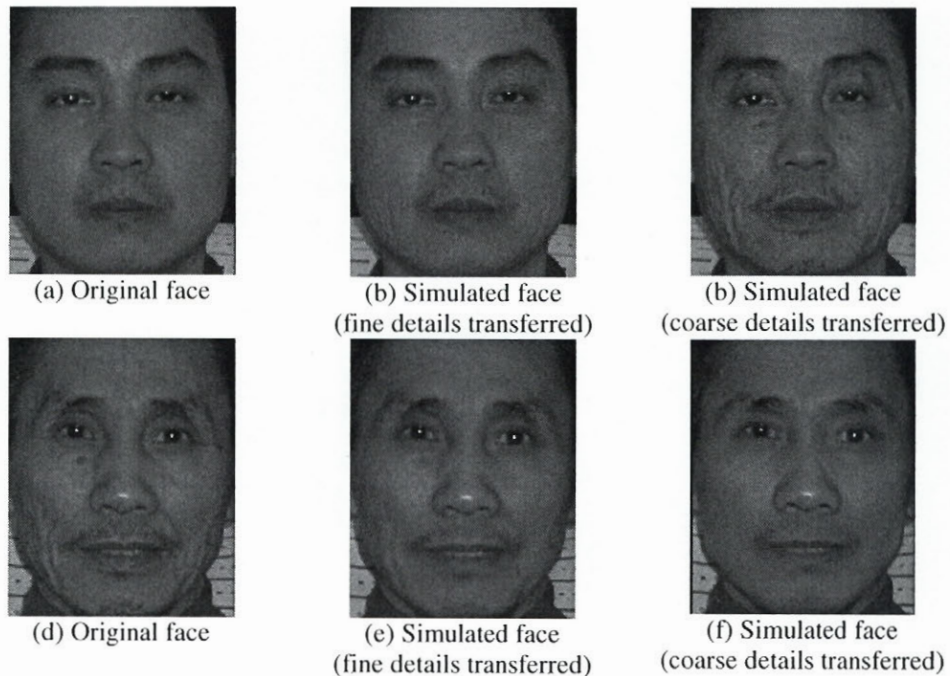


Figure 2.3: Transferring texture to simulate aging effects [Shan et al., 2001]

The main advantage of the IBSDT method is that it requires only a single image for each object to produce aesthetically pleasing results. The major drawback is the lack of an automatic age progression scheme to determine how much surface detail to transfer.

Tiddeman et al. [Tiddeman et al., 2001] produced very convincing results using a new, wavelet-based method for prototyping facial textures and for artificially transforming the age of facial images. Facial prototypes are used to define the differences between two sets of images which can be applied to a new subject's face image. They create a facial prototype for every age group using these differences to define an axis of transformation along a perceived dimension such as age, gender or race. The main problem encountered earlier was the loss of facial texture (such as stubble and wrinkles) in the prototypes due to the blending process. Wrinkles weren't captured well in the previous methods and the facial textures appear smoother than they do in real faces⁷. Localizing facial wrinkles is very difficult due to achieve due to the wide variability in their fine structure. A smoothness constraint is imposed even in automated warping methods such as Blanz and Vetter [Blanz and Vetter, 1999].

Consequently, a wavelet-based method is used to improve the textures of the facial prototypes. They adjust the amplitude of the edges (changes in pixel intensity) in the prototype image so that it matches the average edge amplitudes in the sample at different locations, orientations, and spatial scales. The edges in the image are captured using biologically inspired Gabor wavelets. This additional processing yields much sharper images that don't exhibit smoothing effects, producing excellent results. Experimental validation of age prototypes was carried out and confirmed that the additional wavelet prototyping accurately captures aging information lost in the blending process. Also, age transformations that used the additional wavelet processing methods proved to be five times more effective than without the processing. Figure 2.4 and Figure 2.5 show results from Tiddeman et al. [Tiddeman et al., 2001] that demonstrate texture enhancement and age transformation.

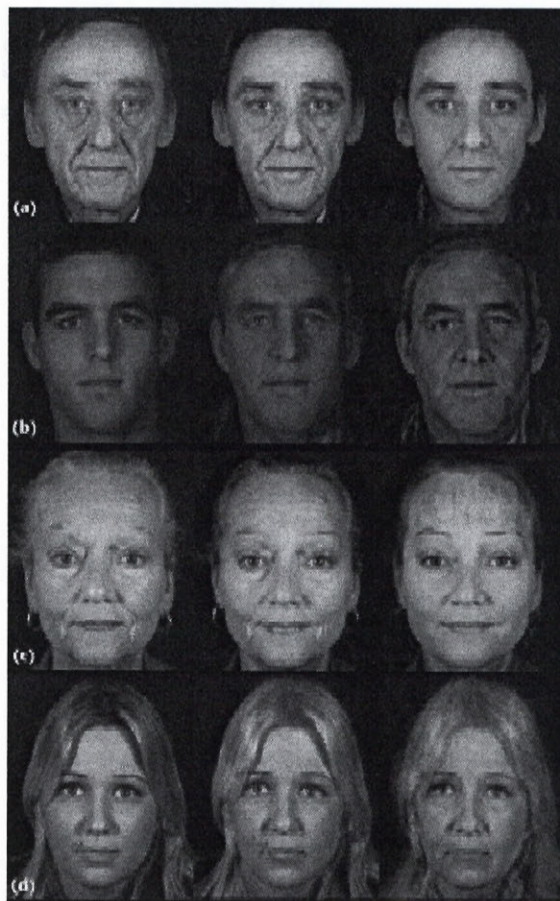
⁷ This is because the warping stage does not align the wrinkles precisely



(a) Original images



(b) Corresponding texture enhanced images

Figure 2.4: Texture processing by wavelet prototyping [Tiddeman et al., 2001]**Figure 2.5: Age transformation examples [Tiddeman et al., 2001]**

The results seen in Figure 2.4 and Figure 2.5 look realistic, in that the identity of the subject is maintained while age transformations are performed. Thus, this approach could definitely prove to be useful for enhancing the facial textures after reconstruction. The work by Tiddeman et al. [Tiddeman et al., 2001] was primarily focused on texture enhancement and did not use any aging model or learning scheme to perform the transformations.

Comparing the methods by Shan et al. [Shan et al., 2001] and Tiddeman et al. [Tiddeman et al., 2001], the former is superior since it achieves similar results by using just one target image while the latter needs two target images⁸. However, the work by Tiddeman et al. introduced the notion of prototypes that was useful for our study. The general drawback in this area is once again the absence of a good aging database. In order to devise a learning mechanism that would quantify how much of the wrinkles to transfer to age any subject to a specific age or age group, a good database is essential.

2.2.2 Medical face measurements for aging

The modeling of facial aging has been attempted based on medical statistics and the medical aspects of the facial deformation. The face structure, size of bones, skin texture change, appearance of wrinkles and muscular change are all greatly influenced by the growth and aging process.

By measuring facial parameters, such as height of forehead, height and width of nose, height of central mid-face and height of lips of female patients aging from 25 to 65, Pitanguy et al. [Pitanguy et al., 1996; Pitanguy et al., 1998] reported a strong correlation between age and the variation of these parameters. As a result, a numerical model for facial aging was built based on geometric ratios and a warping technique was used to simulate aging changes in female facial images [Pitanguy et al., 1998; Leta et al., 2000]. However, for the most part, shape changes were captured and the results are not visually convincing. With the ultimate purpose of simulating facial surgery, Koch et al. [Koch et al., 1996] combined and extended various methods from geometric modelling, finite element analysis, and image processing to render realistic 3D images of a post surgical situation. It is claimed that, due to its versatility,

⁸ The difference of the two images is superimposed on a new face to simulate aging effects.

this method could be employed in different applications including the simulation of facial aging. However, no further details were given.

Pessa et al. [Pessa et al., 1999] showed that skeletal remodelling was such that the ratio of the maxillary height to orbital height was greatest during youth, but smaller during infancy and old age. This finding suggested a non-linear or multi-modal manner in the process of facial aging. These medical statistics give us a better understanding of the facial aging process from a global perspective, although their applicability to automatic facial aging would be difficult due to the precise measurements required. In addition, mainly shape changes are recorded while changes in texture (wrinkle, stubble, etc) are ignored. Therefore, reconstruction does not appear to be realistic.

Hussein [Hussein, 2002] recently proposed a novel facial animation algorithm. It employs two techniques which integrate facial deformation based on the face anthropometry theory and wrinkle simulation, which is called the BRDF⁹ (Bidirectional Reflectance Distribution function) quotient image¹⁰ technique. Given some neutral face, F , the idea is to capture two characteristics of F with advancing years. The first is geometric deformation details like skin texture given in F after passing years. The second is anthropometric data change that developed in the face anthropometry measurements theory. Then, the two techniques are warped together to any particular person's face in order to carry out facial aging. Hussein states that wrinkles are important for understanding and interpreting facial expression besides playing a major role in determining age. His work focuses on simulation of realistic facial aging by concentrating on the medical aspects of the facial deformation and appearance of the wrinkle.

The three main elements considered for facial aging were wrinkles, skin texture variation and facial shape change. Hussein shows how to simulate and predict aging accounting for these three factors [Hussein, 2002] and proposes two techniques – one for facial deformation and one for simulation of wrinkles.

⁹ The bidirectional reflectance distribution function (BRDF) of a surface describes how light is scattered at its surface. [Hussein, 2002]

¹⁰ The quotient image of two coinciding image objects is defined by the ratio of their albedo functions. [Shashua and Riklin-Raviv, 2001]

The facial deformation simulation technique is based on integrating aspects such as increase in wrinkle creases, bags and wattles, progressive changes in skin texture and hair colour, as well as balding with advancing years. The idea is similar to computing the difference between an old and young face, warping the difference onto a new face to make it look older. It is formulated mathematically via the aid of the quotient image [Shashua and Riklin-Raviv, 2001]. The results from the facial deformation technique only are shown in Figure 2.6.

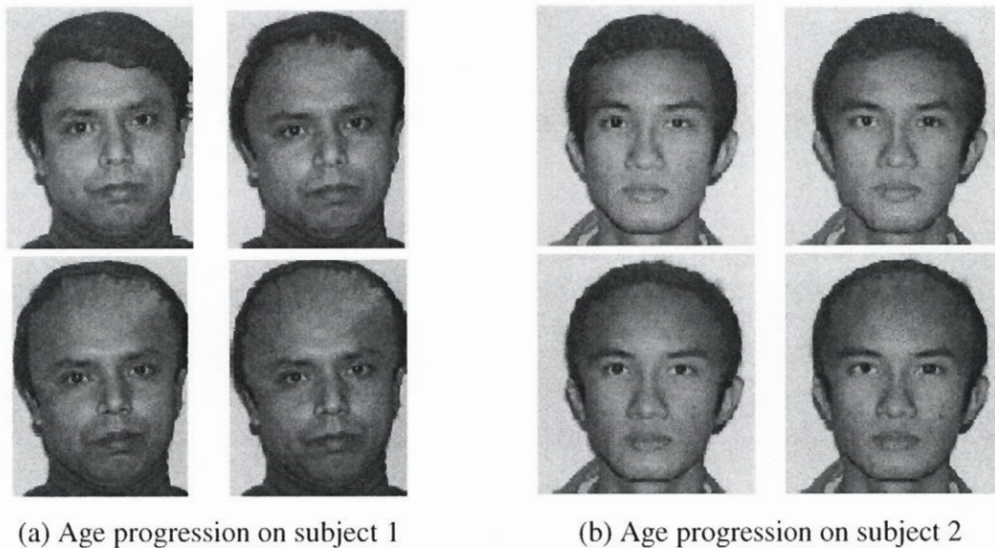


Figure 2.6: Results depicting facial deformation [Hussein, 2002]

Next, simulation of facial texture such as stubble and wrinkles is considered. Hussein shows how one can capture the image space created by the aging process, from a single image. He uses the BRDF (Bidirectional Reflectance Distribution function) quotient image technique to accomplish this task. Given two images of coinciding shapes, the first step is identification of landmark points over the features points, followed by a pixel alignment by linear warping over the feature points obtained by Delaunay triangulation. The final step is to compute the BRDF quotient image. This process yields wrinkles simulation without the need to model each wrinkle as a single entity. Figure 2.7 shows the result of combining both the techniques to create an aging simulation.

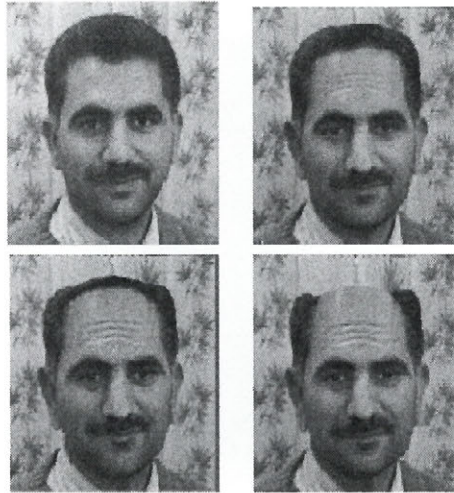


Figure 2.7: Results of combining wrinkle simulation and facial deformation techniques [Hussein, 2002]

The aging simulations don't look very realistic and the facial deformations hardly capture any major visible changes in shape and texture, besides the balding process. The combination of wrinkle simulation to the facial deformation simply added more wrinkles on the forehead, while the rest of the face remained almost unchanged. The use of the face anthropometry measurements is not evident from visual inspection of the results either. A drawback of the method is the absence of any learning mechanism to determine aging, but this research provided important experience in the simulation of aging on a particular face image.

Overall, using anthropometric data to simulate aging on a face is visually unsatisfactory from the results thus far. Changes in face shape, and other shape parameters are not as evident as changes in the face texture (wrinkles and stubble). The state-of-the-art using medical anthropometry theory for age progression would not be very practical when applied to face images, as seen in the result obtained by Hussein [Hussein, 2002] in Figure 2.6 and Figure 2.7.

2.3 Summary

For the age prediction task, the size and quality of an aging database is a major set-back as mentioned in previous sections. Table 4.1 summarizes the various age prediction results obtained by researchers, including the details on each database in use.

Researcher	Database	Training Data	Testing Data	Age Prediction Accuracy
Kwon and Lobo [Kwon and Lobo, 1999]	47 images, classified by babies, young adults and senior adults	-	15 images	100 %
Horng et al. [Horng et al., 2001]	230 images, classified by babies, young adults and senior adults	-	230 images	81.6%
Hayashi et al. [Hayashi et al., 2002]	300 images ranging from 15 – 64 years	-	300 images	27%
Lanitis [Lanitis, 2002]	330 images ranging from 0 to 35 years old	250 images	80 images	Mean Error: 3.83 years

Table 2.1: Summary of age prediction results

In comparison, we collected over 800 high quality images ranging from 15 – 99 years in age. Our database contains the widest age margin with the most number of images in the research thus far which solves the problem of size and quality of training database.

On the other hand, the primary deficiency of age simulation is the lack of realistic simulation of aging effects. Face anthropometric data used to capture aging effects don't produce realistic wrinkles as seen in Section 2.2. Researchers, Shan et al. [Shan et al., 2001] and Tiddeman et al. [Tiddeman et al., 2001] produced very good age simulations, however they focused primarily on age simulation and not age prediction.

We extended the concepts by Shan et al. [Shan et al., 2001] and Tiddeman et al. [Tiddeman et al., 2001] for age simulation by combining an age prediction function and an optimization approach. This approach will ensure accuracy of age progression to a specific age or age group.

Chapter 3

IMAGE NORMALIZATION

Face recognition accuracy heavily depends on how well the input images have been compensated for pose, illumination and facial expression. This chapter presents a complete system for the normalization of human face images. The output image of this system could be used directly as an input to any face recognition system, or for the age prediction system, as in our case.

Variations among images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity [Moses et al., 1991]. For instance, illumination changes caused by light sources at arbitrary positions and intensities contribute to a significant amount of variability as seen in Figure 3.1 (Images are taken from the Harvard Face Database).



Figure 3.1: Images of the same person under different lighting conditions.

Simultaneously, the three rotations¹¹ of a rigid face affect facial pose, leading to another major source of variance in the object's 2D image. Figure 3.2 depicts how the 2D image of a person varies with rotation. *In-plane rotation* occurs along the image plane and is easily eliminated by aligning eye positions. Conversely, *out-of-plane rotation* is more complicated, since it involves the loss of information during projection from 3D to 2D due to occluded facial parts. In addition, non-rigid deformations of the face caused by facial expressions produce subtle differences among images of the same face with different facial expression. Therefore, to study aging on a face image, we have to normalize the image database for illumination, pose and facial expression in order to isolate just the aging effects.

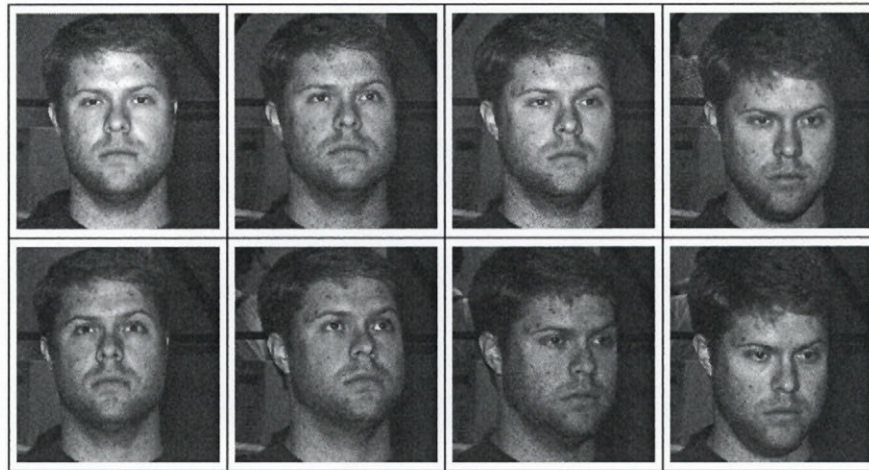


Figure 3.2: Images of the same person with different head poses¹²

For common face recognition systems, it is often assumed to be sufficient to compensate for only face pose and illumination effects, as an individual's identity is preserved with minor changes in facial expression. Changes in expression cannot be tolerated for age prediction because certain expressions can cause wrinkles to appear on the face making subjects look older than they actually are.

To study the facial texture due to wrinkles, we have to ensure that all wrinkles are located at approximately the same position. This restriction is achieved by localizing the major facial

¹¹ The three rotations occur in the Cartesian planes, X, Y and Z

¹² The images are taken from the Yale B face database [Georghiades et al., 2001]

features to the same areas. As a result, the facial shape, pose and minor expressions will also be normalized.

This chapter presents the methods used to perform image normalization. Section 3.1 introduces the histogram fitted Retinex image method for illumination correction. Section 3.2 contains details on the image warping method to eliminate in-plane rotation, minor out-of-plane rotation and expressions, thereby constructing uniform face shapes.

3.1 Illumination Correction

Researchers have proposed various methods to handle the illumination correction or normalization problem. As part of the Content Based Image Retrieval¹³ group, we have studied numerous methods from the literature to resolve this problem. None of these techniques have worked to our satisfaction.

For an in-depth literature survey and background on illumination normalization, the reader is referred to [Bhattacharyya, 2004]. Bhattacharyya investigated the Retinex [Land, 1977] method to remove shadows and specularities¹⁴ from images. Noting the proficiency of this method to eliminate shadows and specularities, we combined the Retinex with histogram fitting to bring all images to the same dynamic range. Face recognition results obtained by applying our normalization scheme on standard illumination databases were better than any other normalization technique reported in the literature. In some cases, using only a single training image of each individual, we realized 100% accuracy under variable lighting conditions. The methodology and experiments are outlined in subsequent sections.

¹³ The Content-Based Image Retrieval Group (CBIR) is concerned with searching large image databases for specific visual objects or visual object classes. CBIR is part of the Centre for Intelligent Machines, McGill University.

¹⁴ Specularities are glare components of surface properties, making the side directly facing the light source brighter.

3.1.1 The Single Scale Retinex

When the dynamic range of a scene exceeds the dynamic range of the recording medium, the visibility of colour and detail will usually be quite poor in the recorded image. *Dynamic range compression* attempts to correct this situation by mapping a large input dynamic range to a relatively small output dynamic range. Simultaneously, the colours recorded from a scene vary as the scene illumination changes. *Colour constancy* aims for the colour to look the same under widely different viewing conditions and illuminants. The Retinex is an image enhancement algorithm that provides a high level of dynamic range compression and colour constancy [Jobson et al., 1997].

Many variants of the Retinex have been published over the years. The last version that Land [Land, 1977] proposed is now referred to as the Single Scale Retinex (SSR) [Jobson et al., 1997]. The Single Scale Retinex for a point (x, y) in an image is defined as being:

$$R_i(x, y) = \log I_i(x, y) - \log[F(x, y) \otimes I_i(x, y)], \quad 3.1$$

where $R_i(x, y)$ is the Retinex output and $I_i(x, y)$ is the image distribution in the i^{th} spectral band. There are three spectral bands – one each for red, green and blue channels in a colour image.

In Equation 3.1, the symbol \otimes represents the convolution operator and $F(x, y)$ is the Gaussian surround function shown in Equation 3.2. The final image produced by Retinex processing is known as the *R-Image* and denoted I_R .

$$F(x, y) = Ke^{-r^2/c}, \quad 3.2$$

where $r^2 = x^2 + y^2$, and c is the Gaussian surround constant - analogous to the σ generally used to represent standard deviation.

The Gaussian surround constant c is referred to as the scale of the Retinex. A small value of c provides very good dynamic range compression, but at the cost of poorer colour rendition, causing greying of the image in uniform zones of colour. Conversely, a large scale provides better colour rendition but at the cost of dynamic range compression [Jobson et al., 1997].

For our study, we are not concerned about loss of colour, since face recognition is ultimately performed on grey-scale images. Moreover, the dynamic range compression gained by small scales is the essence of our illumination normalization process. All the shadowed regions are greyed out to a uniform colour, eliminating soft shadows and specularities¹⁴. Consequently, an illumination invariant signature of the original image is created. Figure 3.3 illustrates the effect of Retinex processing on a face image, I , with different values of c . As c increases, the R-Image I_R , contains reduced greying and lesser loss of colour, as seen in Figure 3.3(c) and Figure 3.3(d). However, at larger values of c , the shadow is still visible. On the other hand, with $c=6$ in Figure 3.3(b), the resulting image has greyed out the shadow region to blend in with the rest of the face.

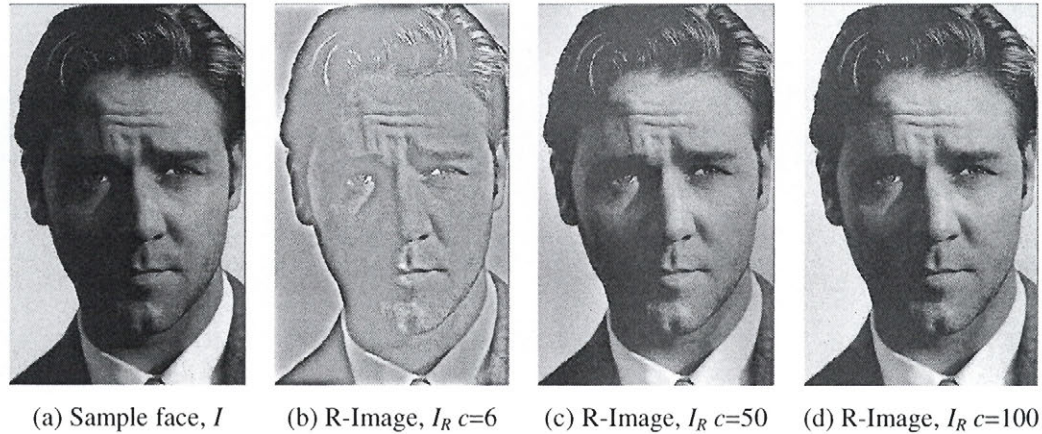


Figure 3.3: Effect of scale c , on Single Scale Retinex processing of a face image

3.1.2 Histogram fitting

Histogram fitting is necessary to bring all the images that have been Retinex processed to the same intensity dynamic range. The histogram of the sample R-Image, I_R , is modified to match a histogram of a target R-Image, \hat{I}_R . It is possible to merely apply *histogram equalization*¹⁵ to the R-Image. However, a well lit R-Image does not have a uniform histogram distribution and this process gives a surreal, unnatural illumination to the face, as shown in Figure 3.4.

¹⁵ Histogram equalization maps the pixels of the input image to a uniform intensity distribution

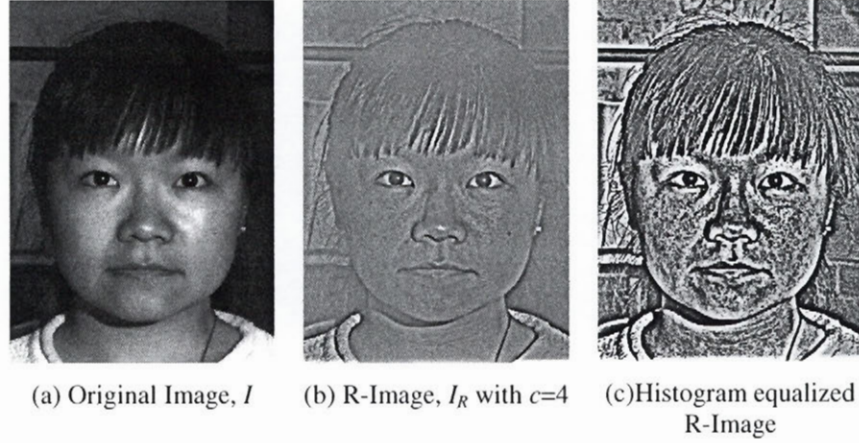


Figure 3.4: Unnatural illumination caused by histogram equalization of the R-Image

Texts such as [Gonzalez and Woods, 1992] encourage the normalization of a poorly illuminated image via *histogram fitting* to a similar, well illuminated image.

Let $H(i)$ be the histogram function of an image, and $G(i)$ the desired histogram we wish to map to via a transfer function $f_{H \rightarrow G}(i)$. We first compute a transfer function for both $H(i)$ and $G(i)$ that will map the histogram to a uniform distribution, $U(i)$. These functions are $f_{H \rightarrow U}(i)$ and $f_{G \rightarrow U}(i)$ respectively. Equation 3.3 and Equation 3.4 depict the mapping to a uniform distribution which is also known as histogram equalization [Gonzalez and Woods, 1992].

$$f_{H \rightarrow U}(i) = \frac{\sum_{j=0}^i H(j)}{\sum_{j=0}^{n-1} H(j)} \quad 3.3$$

$$f_{G \rightarrow U}(i) = \frac{\sum_{j=0}^i G(j)}{\sum_{j=0}^{n-1} G(j)}, \quad 3.4$$

where n is the number of discrete intensity levels. For 8-bit images, $n=256$.

To find the mapping function, $f_{H \rightarrow G}(i)$, we invert the function $f_{G \rightarrow U}(i)$ to obtain $f_{U \rightarrow G}(i)$. Since the domain and the range of the functions of this form are identical, the inverse mapping is trivial and is found by cycling through all values of the function. However, due

to the discrete nature of these functions, inverting can yield a function which is undefined for certain values. Thus, we use linear interpolation and assume smoothness to fill undefined points of the inverse function according to the values of well-defined points in the function. As a result, we generate a fully defined mapping $f_{U \rightarrow G}(i)$ which transforms a uniform histogram distribution to the distribution found in histogram $G(i)$. The mapping $f_{H \rightarrow G}(i)$ can then be defined as in Equation 3.5.

$$f_{H \rightarrow G}(i) = f_{U \rightarrow G}(f_{H \rightarrow U}(i)) \quad 3.5$$

Figure 3.5 demonstrates the histogram fitting process to a sample image. The original image (source) is shown in Figure 3.5(a) and the corresponding R -Image is shown in Figure 3.5(b). The target image (well illuminated face) and its corresponding R -Image, \hat{I}_R , are shown in Figure 3.5(d) and Figure 3.5(e) respectively. The histograms of the source and target R -Image are shown in Figure 3.5(c) and Figure 3.5(f). On application of histogram fitting to the target histogram, the resulting source image and its histogram are shown in Figure 3.5(g) and Figure 3.5(h).

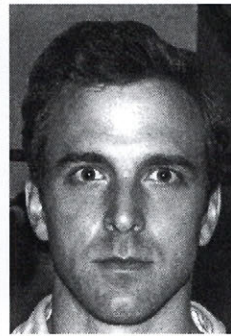
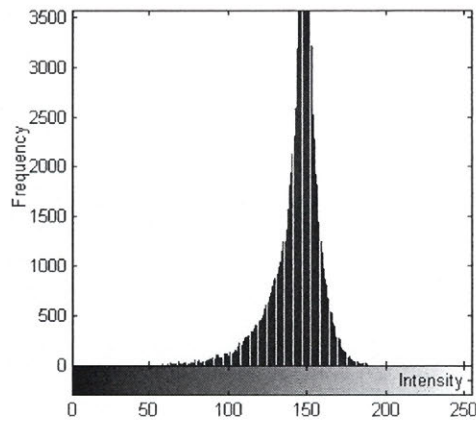
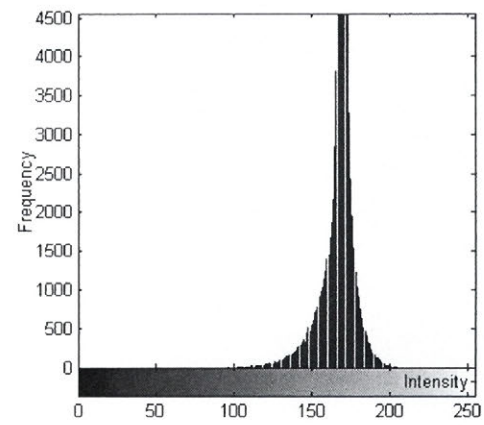
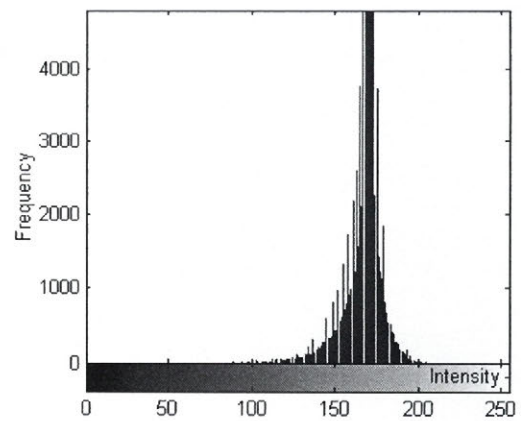
(a) Original Face, I (b) R-Image, I_R with $c=4$ (d) Well-lit Face, \hat{I} (e) R-Image, \hat{I}_R with $c=4$ (c) Source R-Image Histogram, $H(i)$ (f) Target R-Image Histogram, $G(i)$ (g) Histogram fitted R-Image, $(I_R)_{FIT}$ (h) Histogram of $(I_R)_{FIT}$

Figure 3.5: Histogram fitting process on a sample image

3.1.3 Experiments on illumination invariance

We carried out several experiments to examine the performance of our method for illumination invariance. The Yale B face database¹⁶ [Georghiades et al., 2001] was used for all face recognition experiments. Each subject in the database has 65 images under different lighting conditions, resulting in a total of 650 images. Images of subjects under ambient lighting were discarded.

For face recognition, Support Vector Machines (SVMs) are used as the learning scheme. Details on SVMs can be found in Chapter 4 and Vapnik [Vapnik, 1995]. Since there are 10 subjects in total, we execute a 10-class classification using SVMs. A SVM is trained for each set of experiments with a linear kernel and default parameters¹⁷ for classification. Our proposed illumination correction method is used to normalize the database before the experiments are carried out.

In the first experiment, we illustrate the effect of the Gaussian surround constant c on face recognition accuracy. The objective is to find a good value, or range of values for c that would achieve the best illumination invariance. A SVM was trained with only 10 images (one image per subject) and tested with the remaining 640 images. Images with frontal lighting were selected as the training images. Figure 3.6 contains the histogram fitted R-Images at scale, $c=2$, that were used for training.

¹⁶ The Yale B face database contains grey-level images of 10 subjects from different ages and races, with different hairstyles and facial hair, taken under a wide range of carefully measured illumination directions.

¹⁷ Default parameters are provided by Chang and Lin [Chang and Lin, 2001] in their implementation of Support Vector Machines

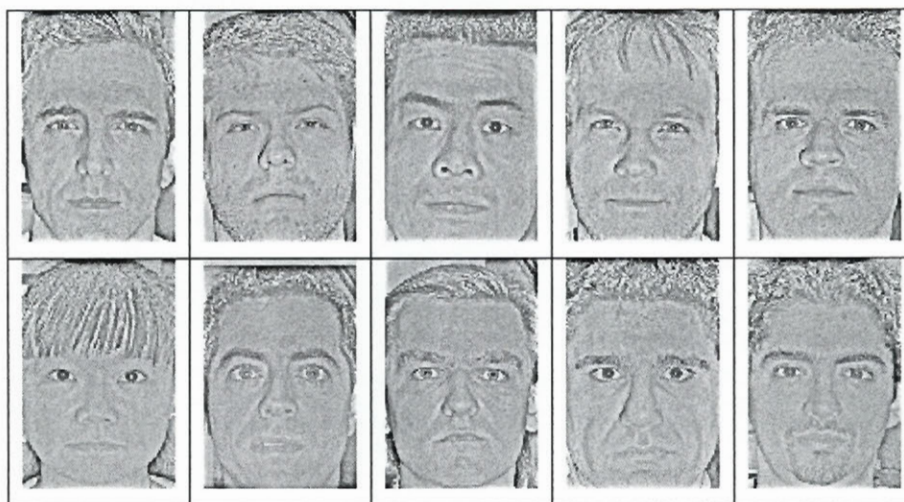


Figure 3.6: Training images used for the first experiment¹⁸

The recognition accuracy for unseen data (the test set) with different values of c is shown in Figure 3.7.

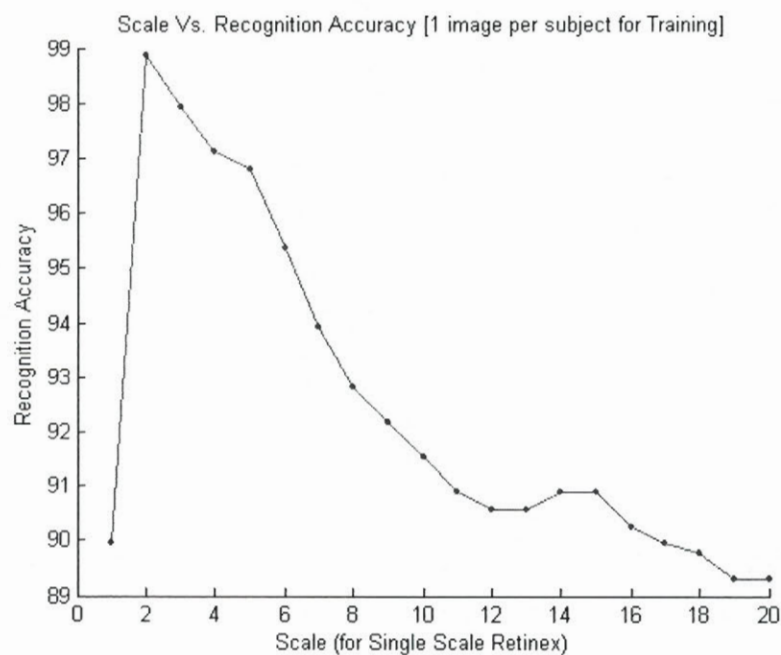


Figure 3.7: Scale, c , versus recognition accuracy

¹⁸ The contrast in the images has been stretched for viewing.

Clearly, the histogram fitted R-image is indeed a powerful means for illumination correction. With $c=2$, only 7 images were misclassified achieving almost 99% accuracy. By comparison, histogram fitting only yielded 80.2% accuracy (124 misclassified images). It is evident that Retinex processing significantly improved recognition rates. Lower values of c are better for illumination correction and as c increases, recognition rates decrease. The only exception occurs when the value of $c=1$, where the recognition rates are much lower. This exception is explained by the fact that the images are overly greyed out with very small c , hindering the SVM from classifying images correctly. We can safely conclude that illumination correction is best at Retinex scales between $c=2$ and $c=6$.

For the first experiment, we selected the training images manually. In the second experiment, each set of training images was selected randomly from the database, and the remaining images used for the test set. Once again, only one image per subject was chosen, and each set of experiments (for every scale) repeated 20 times. The graph illustrating how the average face recognition accuracy changes with c is depicted in Figure 3.8.

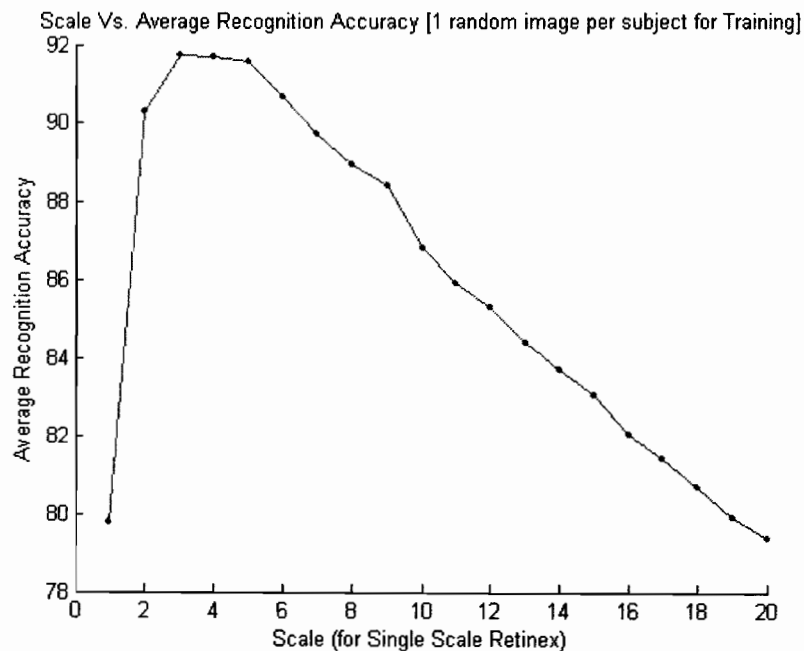


Figure 3.8: Scale, c , versus average recognition accuracy

The curve in Figure 3.8 summarizes the effect of c on face recognition performance. As determined previously, values of c between 2 and 6 still prove to give the best performance. We note an almost linear fall in the recognition performance as c increases after a value of 5. Even when the training images are selected at random, the histogram fitted R-Images still outperform standard histogram fitting, gaining a high recognition rate of almost 92% (at $c=3$).

In the third experiment, we examine the performance of our method, when more than one image per subject is selected at random for training. We initially used two images per subject and compared it with the graph in Figure 3.8. Again, the experiments for each scale were carried out 20 times for a random pair of training images. The results are summarized in Figure 3.9.

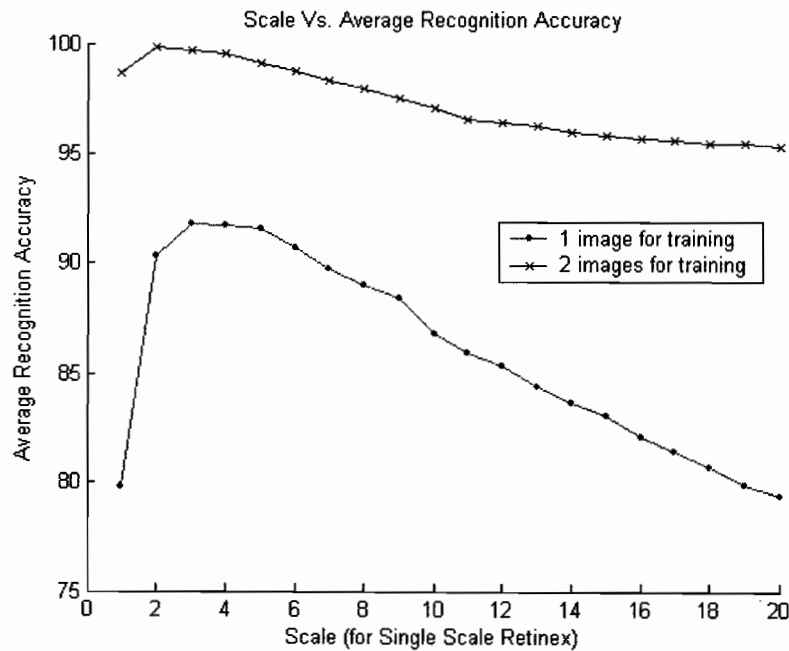


Figure 3.9: Scale, c , versus average recognition accuracy for more training images

As expected, an increase in the number of training images resulted in better performance as observed by the curves in Figure 3.9. For every scale, the recognition accuracy was

consistently better when two images per subject rather than only one image were used. Furthermore, the recognition accuracy fell at a much lower rate. It is important to note that for a scale of $c=2$, the recognition accuracy was almost always 100% over the unseen data when two images were selected at random for every subject. The average of over 20 experiments was 99.84% which is exceptional, considering the size of training set.

Finally, using more than 2 images per subject for training was also evaluated and always exhibited 100% accuracy over the test set.

3.1.4 Summary

The histogram fitted R-Image is a new illumination invariant signature, whose exceptional performance is related to the high level of dynamic range compression produced by the Single Scale Retinex.

From the experiments, we concluded that a good value for the Retinex scale c would be between 2 and 6. In addition, the process of creating an R-Image is extremely fast, taking only a few milliseconds per image. The program flowchart is depicted in Figure 3.10.

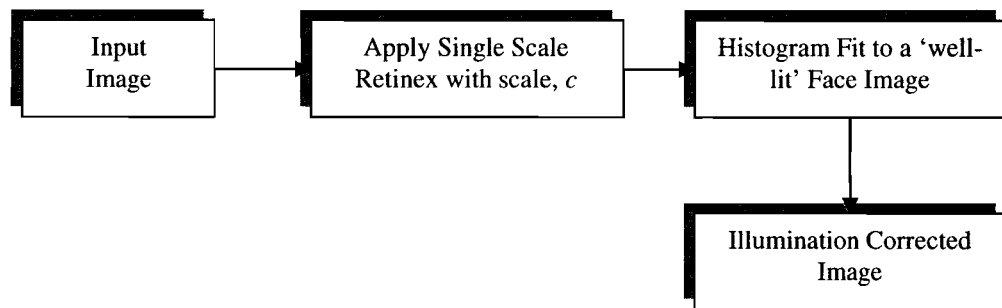
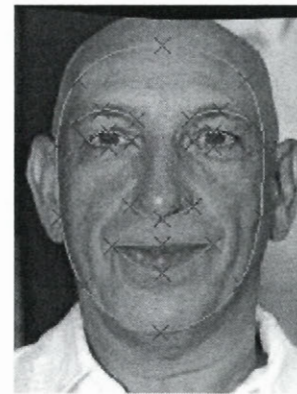
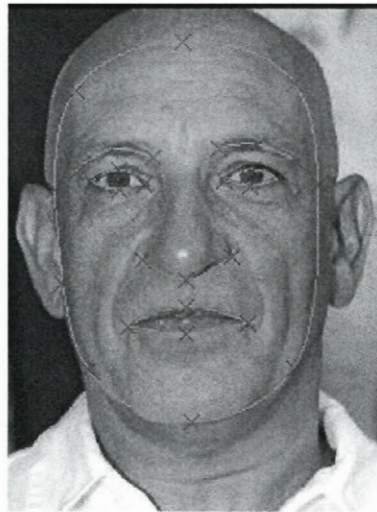


Figure 3.10: Illumination correction flowchart

3.2 Image Warping

Image warping is the geometrical mapping of an image from one spatial domain (the source shape) onto another (the target shape) [Chan and Lau, 1998]. The source and target shapes are defined by important points, known as *feature points*, on the image. Given the feature correspondence constraints¹⁹ between the two shapes, a function over the whole image must be derived that maps the source image to the target image. Consider the source and target shapes on the face images depicted in Figure 3.11(a) and (b). The desired output of the warping process is the original face image fitted to the target shape as seen in Figure 3.11.



(a) Original image defined by source shape (b) Resulting image deformed to target shape

Figure 3.11 Deforming an image by warping to a new shape

Various image warping techniques have been proposed by researchers and the reader is referred to [Wolberg 1990] for a detailed introduction to image warping and [Ruprect and Muller, 1995] for an excellent review. Triangulation based methods are popular and involve breaking the region down into many small triangles, interpolating each triangle

¹⁹ The correspondence constraints determine *how* the face image is warped based on new positions for these points.

independently. The only problem faced by triangulation is that, the triangles must be small enough to create a smooth interpolation.

All warping processes consist of two main steps:

- (i) Computing the desired displacements of all pixels in the source image, and
- (ii) Re-sampling the image to create the output image.

The first step is done by applying a global analytic function, W , to the image pixel positions. W is a function that maps a source shape to the target shape. The process of deriving W is essentially an interpolation problem [Lee et al., 1996]. We generate W by means of affine transformations, described in Section 3.2.2.

For the second step, a number of algorithms have been developed in recent years. We used an approach that takes coordinates of four neighbouring pixels and computes their deformed coordinates with the mapping function, W . Ultimately, a *bilinear interpolation* or *fill* [Gonzalez and Wintz, 1987 p. 249] is carried out for all the pixels within the obtained rectangle. Section 3.2.3 discusses the detailed methodology.

From the above discussion, it is obvious that problem of image warping is effectively the problem of interpolating two functions simultaneously [Tiddeman et al., 2001]. We have added an extra step to the above mentioned two. The three stages are given as follows:

1. Feature specification
2. Warp generation
3. Pixel mapping and interpolation

3.2.1 Feature Specification

Feature specification is the most tedious aspect of warping. Our feature space includes 33 landmarks on the face, connected by smooth curves. Figure 3.12(a) shows the location of these landmarks (marked by red crosses) on a face image from our database. The landmarks are used as control points to generate a total of 13 smooth curves across the face region (seen by the green curves connecting the landmarks in Figure 3.12(b)). As seen in the figures,

these landmarks and curves are adequate to describe the shape of any face²⁰. Table 3.1 contains a list of all the curves and landmark points used.

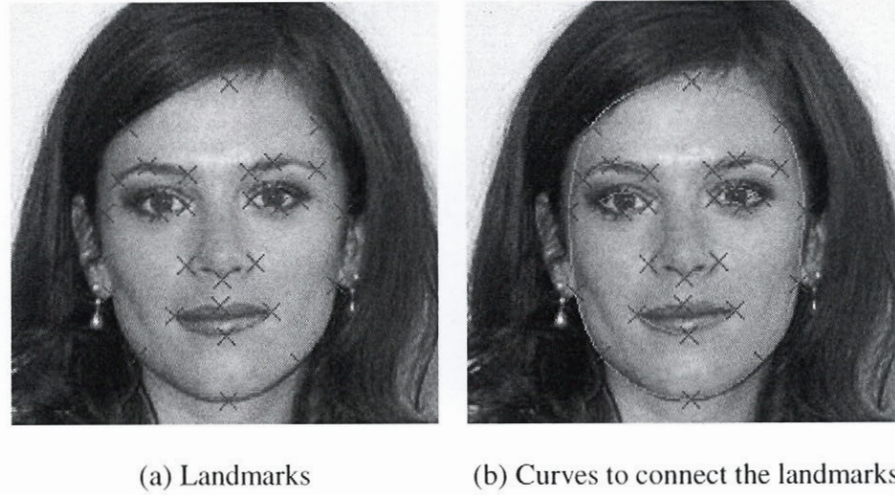


Figure 3.12: Feature specification on a face image

	Name of curve	Number of control points	Landmarks
1	Forehead	5	3, 4, 5, 6, 7
2	Left Brow	3	13, 14, 15
3	Right Brow	3	16, 17, 18
4	Left eye (top)	3	19, 20, 21
5	Left eye (bottom)	3	19, 22, 21
6	Right eye (top)	3	23, 24, 25
7	Right eye (bottom)	3	23, 26, 25
8	Nose	3	27, 28, 29
9	Left ear	2	3, 8
10	Right ear	2	7, 12
11	Upper lip	3	30, 31, 32
12	Lower lip	3	30, 33, 32
13	Chin	5	8, 9, 10, 11, 12
14	Left pupil	1	1
15	Right pupil	1	2

Table 3.1: Description of the landmark points and curves

²⁰ Although this may seem to be a tedious task, it has to be done once for the training data. This activity is done off-line. Given a test image to be aged, only the landmarks for this face must be done. The provided GUI makes this a very fast operation.

The landmarks are *manually* selected with the help of an interactive user interface. It is important to note that they could be found automatically by a program; but this is not an insignificant task, given the current state-of-the-art. In general, warping approaches require precise placement of landmarks that current automated systems would not be able to provide.

The curves that connect the landmarks are generated by interpolating the control points using *cubic spline polynomials* which encompass several desirable features [Eckhoff, 2000]:

- They present smooth-looking resultant curves.
- They have good stability properties, passing through each control point without any “wiggles”.
- They possess the property of minimum curvature, that is, a perfectly elastic bar placed on the data points would assume the shape of a cubic spline.

Cubic splines are piecewise third-order polynomials which pass through a set of m control points. The second derivative of each polynomial is set to 0 at the end points which provides a boundary condition that completes the system of $m-2$ equations. This system of equations gives rise to a simple tridiagonal matrix in an equation that is easily solved to give the coefficients of the polynomials [Bartels et al., 1998]. See Appendix A for mathematical details.

3.2.2 Warp Generation

As mentioned earlier, the warp function W maps the source shape to the target shape. The correspondence between these two shapes is established by two sets of points, P and Q , in the source and target shapes respectively. For each point p in P , there exists a corresponding point q in Q . The points are explicitly obtained by sampling the cubic splines that connect the landmark positions on the face shape. Figure 3.13 illustrates how the function W maps a set of points P to Q . Figure 3.14 demonstrates the sampling on face contours to define sets P and Q . The yellow crosses in the figure indicate samples.

A total of 1000 samples were taken across the face region. It is possible that some samples are taken at the same point, but this *over-sampling* ensures that the face shape is covered entirely.

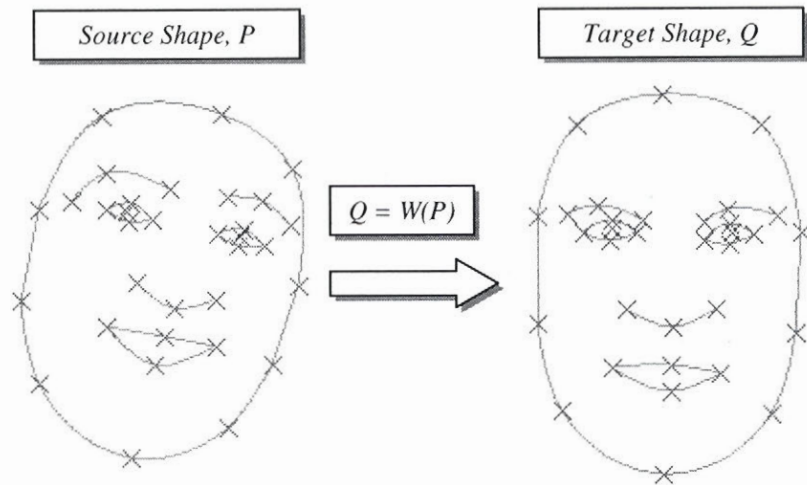


Figure 3.13: Mapping the source shape to the target shape



Figure 3.14: Sampling the contours to produce points across the face region

The function W must have several desirable features. First, it must be smooth to avoid distorting images by creating discontinuities or abrupt deformations. Second, it must be a one-to-one function that guarantees that the image does not fold back upon itself [Lee et al., 1996]. The one-to-one property indicates each point in the transformed image should correspond to only one point in the original image and vice versa. Hence, the warp generation may be formulated as follows:

Given a set of point pairs $(P, Q) = \{(p_i, q_i)\}$, find a smooth one-to-one function $w : R^2 \rightarrow R^2$ such that $w(p_i) = q_i$ for each point p_i in P and the corresponding point q_i in Q .

The *affine transform* offers a simple yet elegant solution to the warp generation problem. It satisfies the one-to-one property, produces a smooth function and is fast since it is a linear problem [Jähne, 1997]. Affine transformations are linear coordinate transforms that include the elementary transformations of translation, rotation, scaling, stretching, and shearing. The transformation can be written as a single matrix multiplication from the source point (x, y) to the destination point (x', y') as in Equation 3.6. It has six degrees of freedom; two for translation in the x and y directions (t_x, t_y), and one each for rotation, scaling, stretching, and shearing ($a_{11}, a_{12}, a_{13}, a_{14}$).

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad 3.6$$

where (x, y) are the coordinates of the source point, (x', y') are the coordinates of the destination point, and $(a_{11}, a_{12}, a_{13}, a_{14}, t_x, t_y)$ are the coefficients of the affine transform to be determined.

The affine transform is also referred to as *three-point mapping*, because it maps a triangle to a triangle. Since we have a set of corresponding points between the source image and the target shape, the coefficients must be determined from these points. For an affine transform three non-collinear points (to map a triangle into a triangle) are necessary resulting in Equation 3.7.

$$\begin{bmatrix} x'_1 & x'_2 & x'_3 \\ y'_1 & y'_2 & y'_3 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{bmatrix} \quad 3.7$$

or

$$Q = WP \quad 3.8$$

From Equation 3.8, W can easily be computed as shown in Equation 3.9.

$$W = QP^{-1} \quad 3.9$$

The inverse of the matrix P exists when the three points (x_1, y_1) , (x_2, y_2) and (x_3, y_3) are linearly independent (that is, geometrically cannot lie on one line). With more than three points (as is the case for the face shape), the parameters of the affine transform can be solved in a least square sense²¹. It is evident from Equation 3.10 that the main computation involved is the inversion of a 3 x 3 matrix while the rest of the formulation is not computationally intensive.

$$W = QP^T (PP^T)^{-1} \quad 3.10$$

$$QP^T = \begin{bmatrix} \sum x'_n x_n & \sum x'_n y_n & \sum x'_n \\ \sum y'_n x_n & \sum y'_n y_n & \sum y'_n \\ \sum x_n & \sum y_n & N \end{bmatrix} \quad 3.11$$

$$PP^T = \begin{bmatrix} \sum x_n^2 & \sum x_n y_n & \sum x_n \\ \sum x_n y_n & \sum y_n^2 & \sum y_n \\ \sum x_n & \sum y_n & N \end{bmatrix} \quad 3.12$$

Since the affine transform is one-to-one, its inverse is also affine and is denoted by W^I . Now that the warp function W has been derived, the next step involves generating the target image by pixel mapping and interpolation.

3.2.3 Pixel mapping and interpolation

Since the transformed grid points of the input image no longer coincide with the grid points of the output image, pixel mapping and interpolation are required. The source image I_s is mapped and interpolated to the destination image I_D by applying the warping function W .

In computer graphics, there are two kinds of pixel mapping; *forward mapping* and *reverse mapping*. Forward pixel mapping, scans I_s pixel by pixel, and copies them onto I_D at positions determined by W . Since the input pixels are mapped from the set of integers to real

²¹ See *Solution of Overdetermined Linear Inverse Problems*, Section 3.4.3 in [Jähne, 1997]

numbers, there is a possibility of “holes” in I_D , that is, some pixels in the destination images may not be mapped. This is a major drawback of forward mapping.

Reverse or *inverse* mapping on the other hand goes through I_D pixel by pixel, and samples the correct pixel positions from I_S . By scanning the destination image, we guarantee that all the pixels in I_D are filled. Pixel positions are determined by the inverse transform, W^{-1} . Figure 3.15 illustrates the two mappings. The forward mapping scheme may map two pixels from I_S to the same pixel in I_D causing a hole in I_D , as seen in the figure. In reverse mapping, it doesn't matter if two destination pixels are mapped to the same source pixel, also shown in Figure 3.15(b).

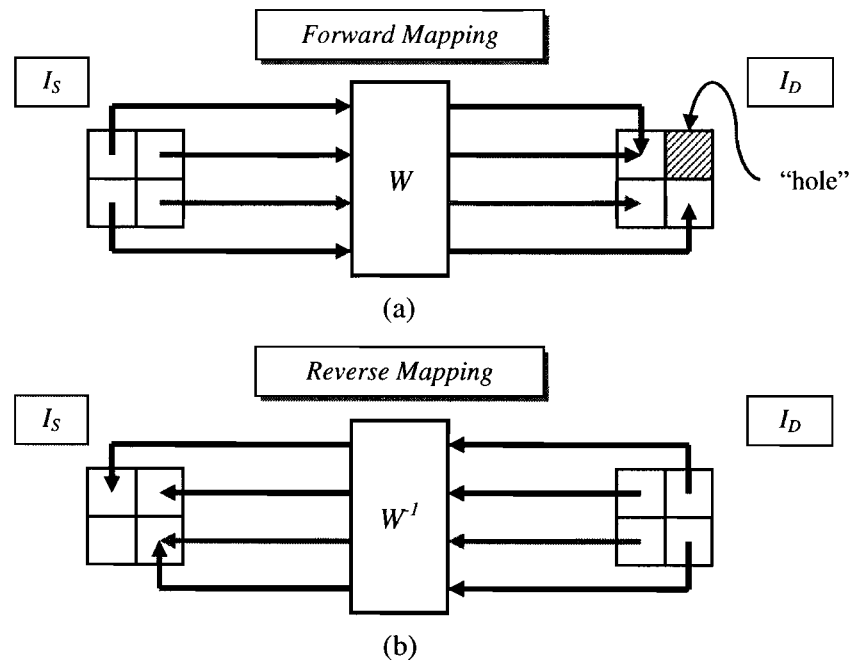


Figure 3.15: Forward versus reverse mapping

We used reverse mapping to guarantee complete coverage of the destination image I_D . The inverse warping function W^{-1} is easily determined by finding the inverse of the 3 x 3 matrix, W . Reverse mapping finds the corresponding source pixels involved in forming the intensity of a destination pixel. However, the calculated source pixel coordinates may be fractional coordinates since the transformation matrix contains real numbers. To handle these

calculated fractional pixel “addresses”, it is up to the interpolation to produce a reasonable intensity value for a position not falling exactly on integer coordinates.

The colour assigned to the new pixel is computed as an interpolated combination of the pixel values closest to the pixel in the source image. A naïve approach, known as *nearest neighbour interpolation*, simply truncates the calculated fractional addresses to the nearest integer pixel address. Although fast, the results would not be visually satisfactory due to discontinuities or distortions in the destination image. A better method would be to sample all four nearest pixels that fall around the fractional coordinates resulting in smoother output images. This method is known as *bilinear interpolation*, demonstrated in Figure 3.16.

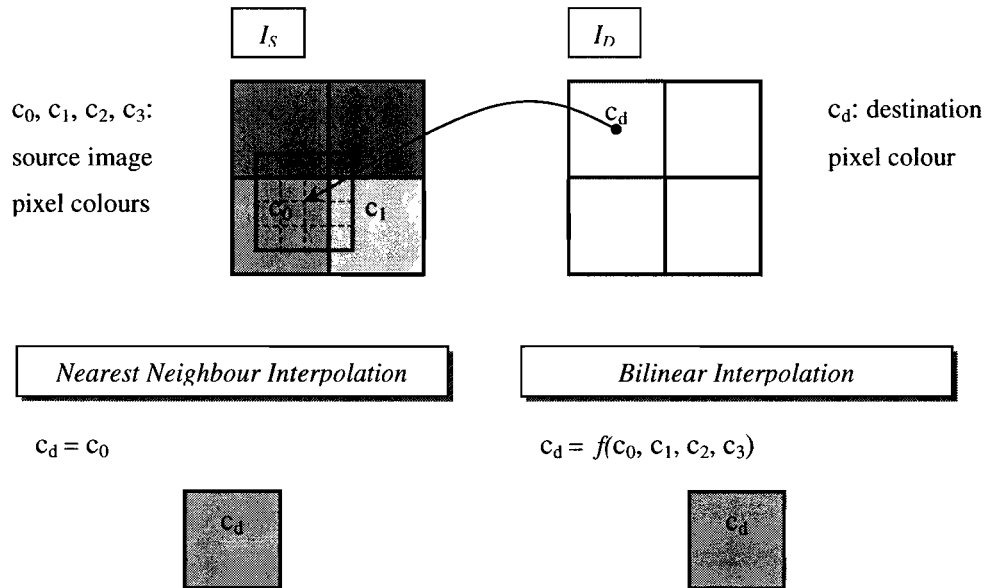


Figure 3.16: Nearest neighbour versus bilinear interpolation

Bilinear interpolation uses a distance weighted average of the four nearest pixels in the image domain, obtained by the transformation function W , to determine the intensity of each pixel in the target image. The closest neighbour has the largest contribution to the destination pixel, while neighbouring pixels farther away have a lesser effect. In other words, the contribution of a neighbouring pixel is inversely proportional to the distance between the

neighbour and the mapped point. The weighting is carried out by the formula in Equation 3.13 and illustrated in Figure 3.17.

$$c(x', y') = c(x_0, y_0) + [c(x_1, y_0) - c(x_0, y_0)]\Delta x + [c(x_0, y_1) - c(x_0, y_0)]\Delta y + [c(x_1, y_1) + c(x_0, y_0) - c(x_0, y_1) - c(x_1, y_0)]\Delta x\Delta y \quad 3.13$$

Where $\Delta x = (x - x_0)$, $\Delta y = (y - y_0)$ and $c(x, y)$ is the colour of the pixel at location (x, y) .

Notice from Equation 3.13, as Δx increases (that is, the mapped point moves further away from $c(x_0, y_0)$ in the x direction), $c(x_0, y_0)$ contributes less to the final colour of the mapped point. The same notion applies to the y direction.

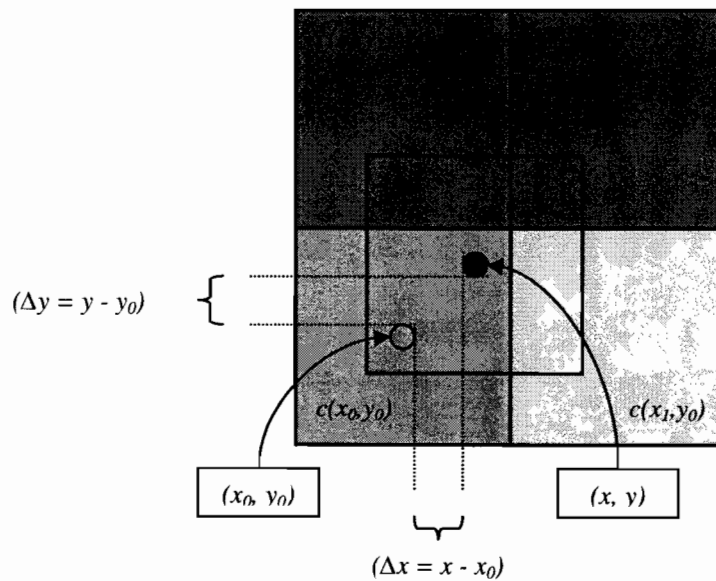


Figure 3.17: Derivation of bilinear interpolation

This concludes the theory behind warping. A flowchart summarizing warping is depicted in Figure 3.18.

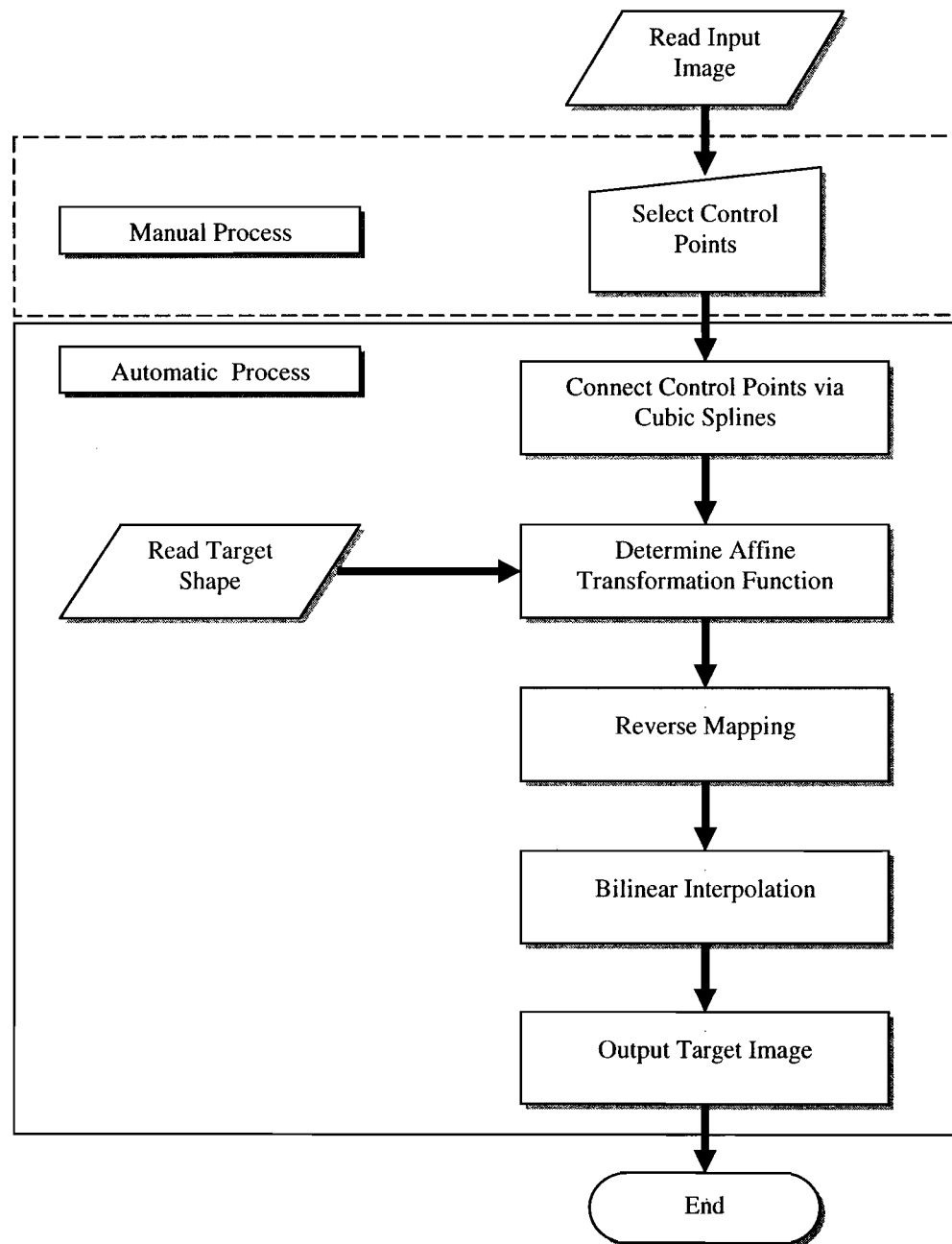


Figure 3.18: Warping program flowchart

3.2.4 Results of image warping

A few examples of the image warping process are illustrated in this section. The original images with manually selected landmarks are shown in Figure 3.19.

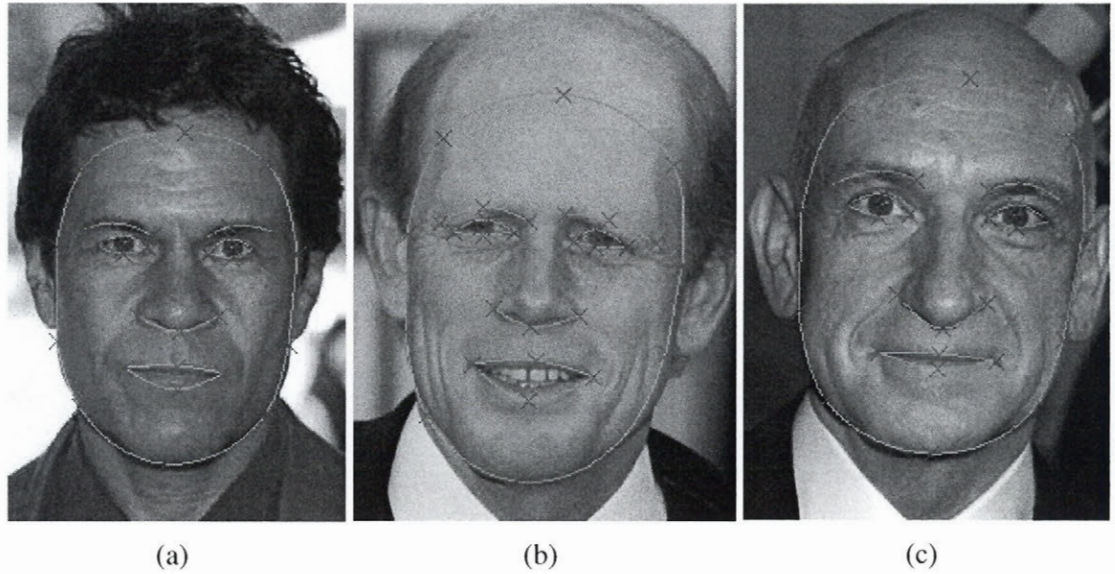


Figure 3.19: The original three test images with landmark points (red crosses) connected by smooth cubic splines (green curves).

All three faces have different shapes. To demonstrate the power of warping, an average image is produced by warping the input images. This image is compared to the average image computed by standard eye position alignment.

First, the three faces are warped to an ‘average’ shape. The latter is simply determined by finding the mean of landmarks, then connecting these average landmarks by cubic splines. Equation 3.14 summarises this process.

$$q_i = \bar{p}_i = \frac{1}{n} \sum_{j=1}^n (p_i)_j, \quad 3.14$$

where $(p_i)_j$ are the i^{th} landmark coordinates in image j . The average shape is defined by the landmark points, $(\bar{p}_1, \bar{p}_2, \bar{p}_3, \dots, \bar{p}_{33})$.

Figure 3.20 shows the average shape and input images warped to this average shape. The warped images have all facial features located around the same area, as expected from the warping. The average of these three images is illustrated in Figure 3.21.

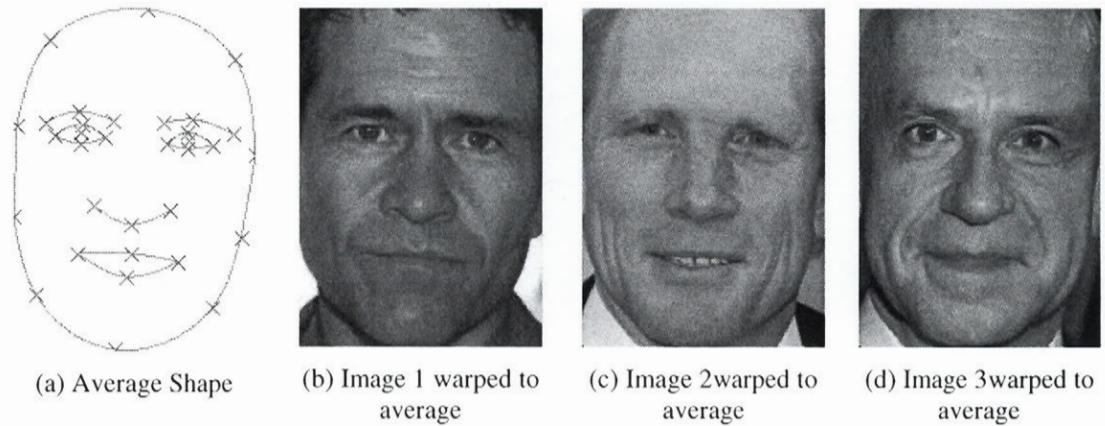


Figure 3.20: Original images warped to average shape

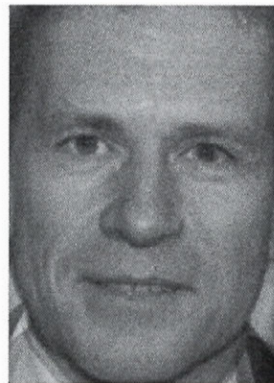


Figure 3.21: Average image produced by warped face images

During the alignment of eye positions, the face image is simply rotated and scaled so that the eye positions in all the images are at the same location. This is a widely used simple pre-processing step before face recognition. The alignment of the original three images from Figure 3.19 leads to the resulting images in Figure 3.22.

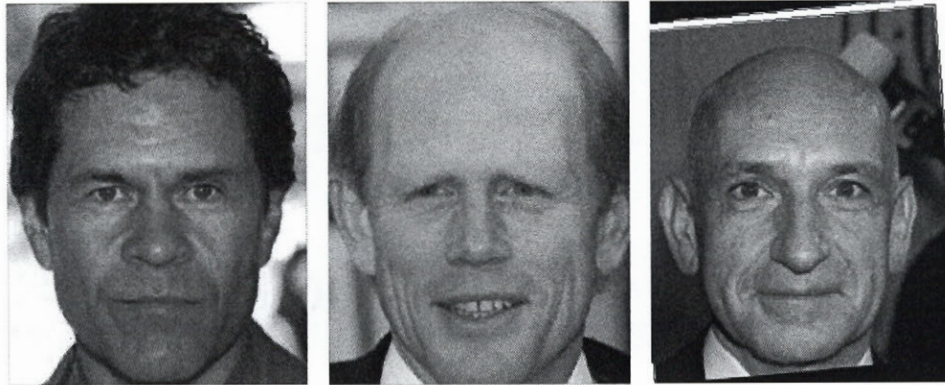


Figure 3.22: Original images scaled and rotated to align eye positions

Finally, the average of these three images is computed and shown in Figure 3.23.



Figure 3.23: Average image obtained by aligning eye positions

Comparison of Figure 3.21 and Figure 3.23 unleashes the power of warping for aligning facial features. The mean image produced by warping the input images to the mean shape looks realistic, like a combination of all the three faces to produce a new identity. The alignment performed is satisfactory. On the other hand, simply aligning eye positions does not guarantee alignment of other facial features as seen in the average image in Figure 3.23. The image appears blurry and is not a good average image.

3.3 Summary

We have developed a powerful and complete image normalization system that can be used as a pre-processing step for any face recognition task. Our illumination correction method works very well in eliminating shadows from the image, obtaining almost 100% recognition accuracy when trained with 2 random images per subject. The method would prove to be very useful when applied to images collected from a wide range of sources that need to be normalized for illumination. On the other hand, the warping technique we implemented also works exceptionally well in localizing facial features, and eliminating expressions (caused by smiling, frowning, etc). However, it is important to note that only minor out of plane rotation can be eliminated, since there is a loss of information due to occluded facial parts.

Chapter 4

AGE PREDICTION

Humans can easily categorize a person's age group and are often precise in this estimation. This task can be considerably intricate for a machine. In Section 2.1, we presented a review on studies that have attempted to derive an age prediction function. We concluded that researchers haven't been successful in finding a global aging function mainly due to a lack of good training databases. As a result, we collected our own training images. The McGill CBIR Aging database consists of 818 frontal face images labelled by the age and gender of the subjects in the image. Currently, it is the largest database to date for studying human aging.

This chapter is based on using Support Vector Machines [Vapnik, 1995] to derive a function that can predict the age of a given frontal face image. Support Vector Machines have recently gained a great deal of popularity in the computer vision community, proving to be very successful on several classical pattern recognition problems [Burges, 1999]. For our age estimation task, we use SVMs because of their numerous appealing features summarized in Section 4.2. We are the first to apply SVMs successfully to age prediction.

Section 4.1 describes the McGill CBIR Aging Database while Section 4.2 presents a brief theory on SVMs. Experiments on the database are outlined and discussed in Section 4.3.

4.1 Source and description of training database

The primary obstacle to researchers studying the aging problem is the lack of appropriate facial databases. Several databases are available for the study of illumination, pose and facial expression; the other three bugaboos that hinder accurate face recognition. These databases were created by taking photos of several subjects under controlled conditions with only one variable changing at a time, for example scene illumination. Unfortunately, this approach is impractical when it comes to building a facial aging database. We would have to gather subject's images over a number of years under the same conditions.

There are numerous frontal images available on the Internet. However, it is fairly difficult to find the age of the subject in the image. Scanned images from albums of family and friends are not a reliable source, due to the wide variation in the facial appearance of the pictures. Lanitis et al. [Lanitis et al., 2002] used this method, noting significant variation in 3D pose, lighting, expression, image quality and resolution. A suitable source of images with minimum variations would be criminal databases from the police. These images are taken by the same or similar cameras under controlled lighting, pose and facial expression. Also, the age of the subjects would be readily available. Unfortunately, due to security reasons, we could not get access to this information.

Considering our stringent requirements for a controlled source of images, we gathered images of *celebrities* from the Internet. Advantages of collecting celebrity images include:

1. Abundant sources of celebrity images are readily available on the Internet.
2. It is easy to determine the age of the celebrity in the image, since the images are generally dated and their birthdates are simple to find.
3. Since most photographers are professionals, the image quality and resolution is satisfactory for our purposes.

The images we collected formed the *McGill CBIR Aging database* – a database of 818 frontal face images labelled by age and gender. Our main source was the Internet Movie Database [Imdb, 2004], which contains a dated list of thousands of celebrity photographs with their birthdates. The most significant uncontrolled factor was facial expression, but we simply discarded face images with 'extreme' expressions. We looked for more 'neutral'

expressions while collecting the training data. In addition, the presence of makeup, particularly on older female celebrities, hid the majority of the wrinkles on the face. Consequently, the perceived age was a lot lower than the actual age, which could hinder the learning process. For this reason, mainly images of men were collected.

Initially, our goal was to collect several images of celebrities at different ages. Due to the difficulty in obtaining good quality images taken before the year 2000, our database consisted of images taken between 2000 and 2004. The images spanned over 500 different celebrities with diverse ethnicity to cover a broad range of ages. The ages range from 15 years up to 99 years. Hence, our database is good for training a global aging function covering this range.

In total, we collected 818 frontal images, of which 585 were male and 233 were female. A histogram of the number of images per age group (grouped by 5 year intervals) is shown in Figure 4.1. The mean age of males in the database is 46 years while the mean age of females is 40 years.

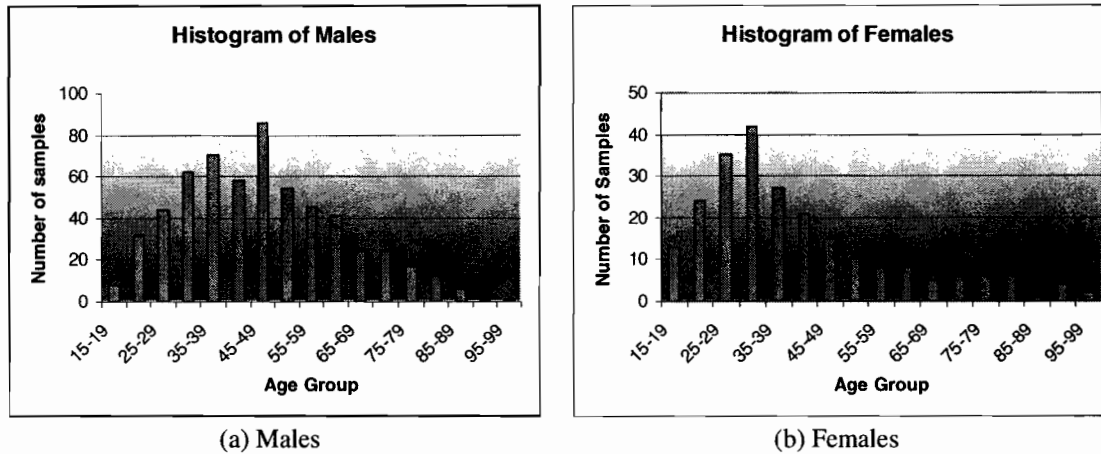


Figure 4.1: Histogram of male and female images in the database

4.1.1 Database normalization

Our image normalization scheme from Chapter 3 is applied to the aging database to isolate aging effects in the images. First, the feature points described in Section 3.2.1 are manually located in all the images. Then, the mean face shape is determined by averaging these feature points. Next, the warping procedure from Section 3.2 is applied to align all the faces to the

mean face shape. The mean face size was 116 x 160 pixels – sufficient resolution to retain enough wrinkle detail for learning purposes. Finally, the images are brought to the same dynamic range by histogram fitting (see Section 3.1.2).

Our images in the database contained very little shadows and specularities, since photographs were taken by professionals under good lighting conditions. Therefore, we did not pre-process the images with the Single Scale Retinex (Section 3.3.1) before applying histogram fitting. Simple histogram fitting was sufficient to normalize for illumination.

Figure 4.2 illustrates the database normalizing process in a block diagram.

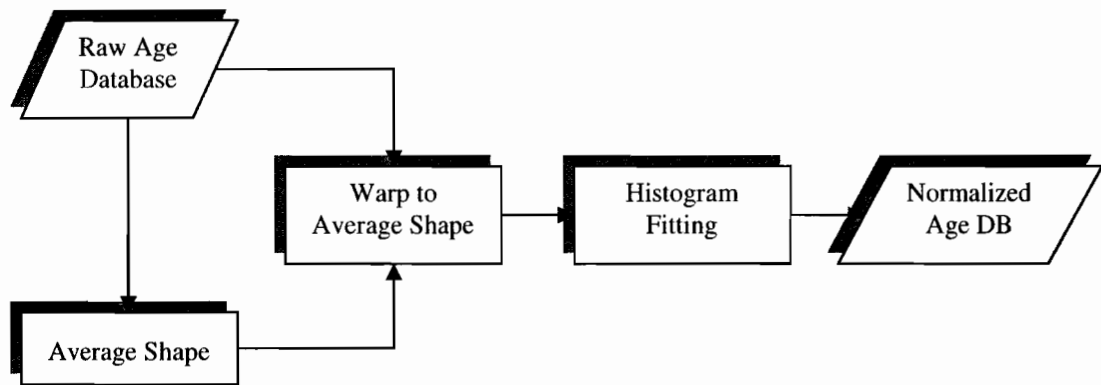


Figure 4.2: Block diagram of the normalization process

Some sample images and respective ages from our database are shown in Figure 4.3. As seen from the figure, the images we collected demonstrate good quality at a high resolution. The images have been resized to fit the page, but actual sizes were larger than those shown in Figure 4.3.

Subsequent sections describe SVMs and experiments carried out on the McGill CBIR Aging database.

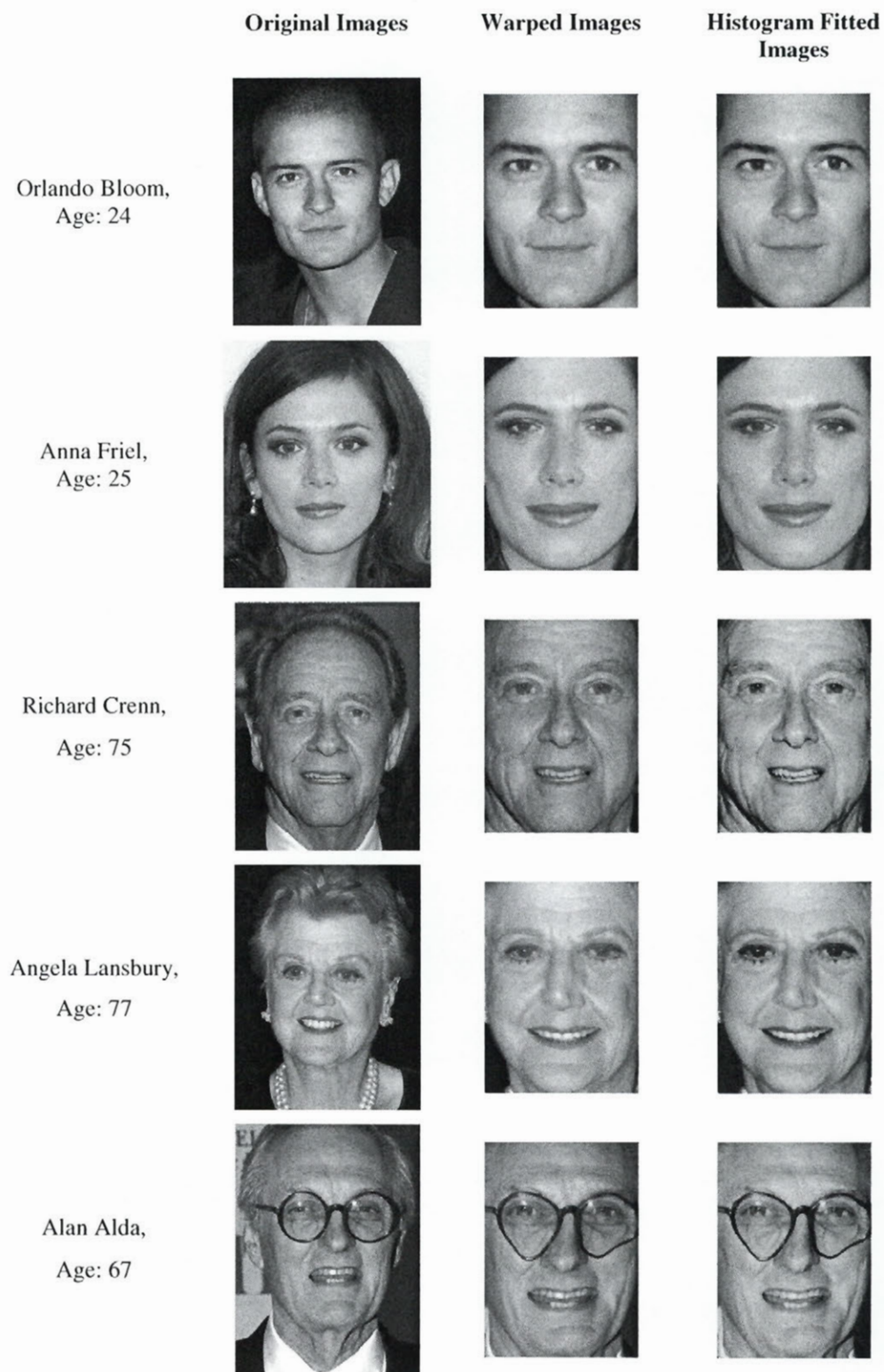


Figure 4.3: Sample cropped images and their age from the CBIR Aging Database

4.2 Support Vector Machines

A Support Vector Machine is a learning technique that learns the decision surface through a process of discrimination and has good generalization properties [Vapnik, 1995]. SVMs have been successfully applied to a number of applications ranging from particle identification, face identification, text categorization, engine knock detection, bioinformatics and database marketing [Bennett and Campbell, 2000].

4.2.1 Optimal separating hyperplanes

SVMs calculate *optimal separating hyperplanes*: hyperplanes that maximize the margin between the classes. Assuming that the data is linearly separable, Figure 4.4 illustrates an optimal separating hyperplane. Figure 4.4(a) demonstrates various hyperplanes that could separate the training data. SVMs choose the hyperplane h^* in Figure 4.4(b) with a margin of separation δ , that best separates data.

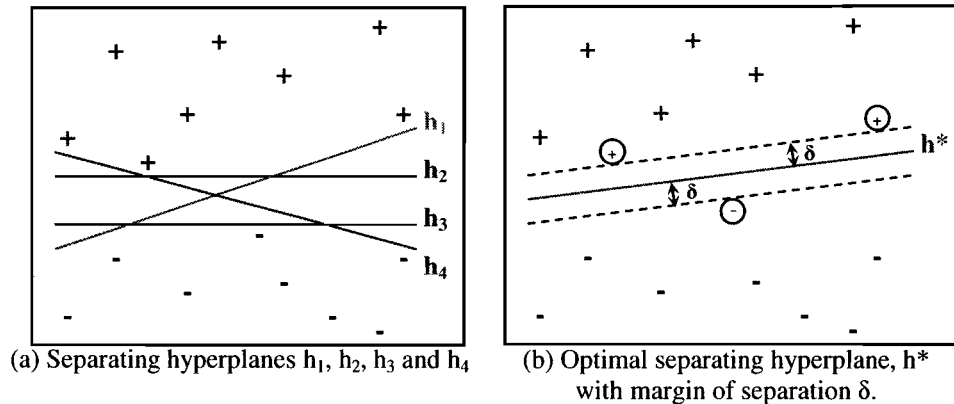


Figure 4.4: Notion of the optimal separating hyperplane

Maximizing the margin of separation, δ , reduces the function complexity. Thus by explicitly maximizing the margin, we are minimizing bounds on generalization error and can expect better generalization with high probability. Moreover, since the size of the margin is not directly dependable on the dimensionality of the data, we can expect good performance even for very high-dimensional data. Hence, problems caused by overfitting of high-dimensional data are greatly reduced, thereby resolving the so-called *curse of*

*dimensionality*²². In our study, the image vectors span a very high-dimension (close to 20,000), compared to the 800 training images. SVMs are therefore ideal learners for the aging function, as they are independent of the dimensionality of the training data.

A certain number of training patterns that lie on the margin of the separating hyperplane are retained, known as *Support Vectors*. They carry all the relevant information about the classification or regression problem. For example in Figure 4.4, the three circled training examples are retained as Support Vectors.

4.2.2 Support Vector Regression and the ‘ ϵ -insensitive zone’

Face recognition involves classification to categorize an unseen image into one of several predefined groups. However, age is a real valued number. Our aim is to learn the relationship between the coded representation of the image and a real number representing the age. As a result, we use Support Vector Regression (SVR) Machines. Equation 4.1 summarizes our objective.

$$A_p = F(I), \quad 4.1$$

where I is the coded representation of the face image and F is the regression function we want to train. A_p is the real number representing the predicted age of the subject in image I .

SVR has been formulated and motivated from statistical learning theory. We refer the reader to the large volumes of literature on this topic for more technical discussions e.g. [Cristianini and Shawe-Taylor, 2000] [Vapnik, 1995] and [Vapnik, 1998].

Support Vector Regression uses the ϵ -insensitive loss function shown in Figure 4.5. If the deviation between the actual and predicted value is less than ϵ ²³, then the regression function is not considered to be in error. Geometrically, we can visualize this as a band or tube of size 2ϵ around the hypothesis function $f(x)$, and any points outside the tube are seen as training errors. The ‘ ϵ -insensitive zone’, or width of the band, is this region which trades off accuracy

²² The “curse of dimensionality” is the requirement that the number of samples per variable increase exponentially with the number of variables to maintain a given level of accuracy.

²³ ϵ is in terms of years for the aging function

versus complexity of the function. The point enclosed in the dotted box in Figure 4.5 deviates by ζ and is seen as a training error. This point is penalized in a linear fashion by the SVR.

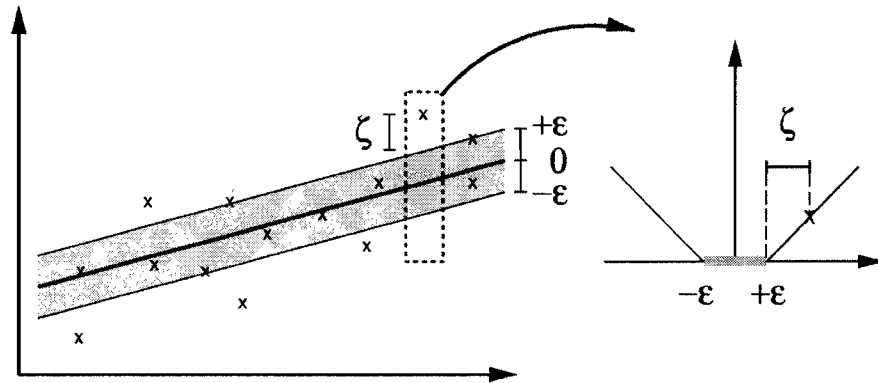


Figure 4.5: The ‘ ϵ -insensitive zone’ in linear Support Vector Regression

4.2.3 Kernel functions for non-linearity

In Figure 4.5, a linear function is depicted. Nonlinear functions can simply be generated by pre-processing the training vectors into another feature space I' , and then applying the standard SVR algorithm. However, it is not computationally practical to perform this transformation. SVMs have a neat feature known as *kernels* that can make the SVR nonlinear. By *Mercer's theorem* [Mercer, 1909], it is known that for certain mappings θ and any two points u and v , the inner product of the mapped points can be evaluated using the kernel function K without ever explicitly knowing the mapping, e.g. $\theta(u) \cdot \theta(v) \equiv K(u, v)$. Some popular kernels are given in Table 4.1.

Function $\theta(u)$	Kernel $K(u, v)$
Polynomial of degree d	$(u \cdot v + 1)^d$
Radial Basis Function Machine with standard deviation σ	$e^{\left(-\frac{\ u-v\ ^2}{2\sigma}\right)}$
Two Layer Neural Network with weight η	$\text{sigmoid}(\eta(u \cdot v) + c)$

Table 4.1: Examples of Kernel functions

Therefore, to change from a linear to a nonlinear classifier, we merely substitute a kernel evaluation instead of the original dot product. Thus, by changing kernels we can get different highly nonlinear classifiers. No algorithmic changes are required from the linear case other than substitution of a kernel evaluation for the simple dot product, maintaining all the benefits of the original linear SVM.

4.2.4 Convergence tolerance (ϵ)

SVM learning is an optimization algorithm, with its termination criterion given by the convergence tolerance. If the change in predicted age of training data is less than the convergence tolerance, ϵ , then the algorithm terminates and reports a solution. Convergence tolerance ϵ is defined relative to the feasible range of ages.

4.2.5 Benefits of SVMs

SVM provide several desirable features for learning the global aging function. Unlike other learners such as neural networks and decision trees, SVMs are independent of the dimensionality of the input data. When the training data contains images, the dimension of the input data is very high and generally would require many images to train the learner. However, with limited training data, SVMs are a better choice since they have no problems with getting stuck at *local minima*.

SVMs can produce non-linear functions based on the kernels selected. We have a fairly wide choice of SVM parameters, kernels and kernel parameters to determine a function that best represents the input data.

For these reasons, and the success of SVMs in various face recognition tasks, we decided to utilize them to learn the aging function.

4.3 Experiments on the CBIR Aging Database

We utilized Chang and Lin's software implementation of Support Vector Machines [Chang and Lin, 2001] for all our experimentation. The goal was to find the best set of parameters to learn a global aging function, F , from our database. Equation 4.2 summarizes this idea.

$$F = \text{SVR}(T, K, P, \varepsilon, e), \quad 4.2$$

where F is the function to be trained, T is a matrix of training vectors labelled by age, K is the kernel function (See Table 4.1), P is a vector containing the kernel parameters (depending on kernel), ε is the epsilon in the loss function of SVR (Figure 4.5) and e is the convergence criterion.

The first set of experiments determined the best feature space to use for learning, that is, the best T for training the SVM. The rest of the experiments involved fine-tuning SVR parameters (K, P, ε, e) to obtain a value of the function F that would minimize the error in the predicted age over unseen data. A 4-fold cross validation²⁴ was used for all the experiments with the following statistics reported for each experiment:

1. *Cross-validation Absolute Error (E)*: The average absolute error over unseen data in the validation sets. E is in terms of years and we intend to minimize this value.
2. *Average Number of Support Vectors (NSV)*: The average number of training samples retained for the cross-validation sets. It is desirable to have fewer support vectors so as to maintain a simple function. However, if there are too few Support Vectors retained, then the function may not be able to accurately predict the age.
3. *Cross validation squared correlation coefficient (r^2)*: The coefficient r^2 suggests the strength of a relationship between the variables in a learnt function. A value of 1 for r^2 indicates the strongest relationship between the variables and is desirable. It is inversely proportional to the absolute error, E .

In some cases, different combinations of parameters yield the same mean squared error. We selected a set of parameters with the fewest *Average Number of Support Vectors* as the best choice as it indicates a simpler function.

²⁴ In k -fold cross-validation, the data is divided into k subsets of (approximately) equal size. The SVM is trained k times, each time leaving out one of the subsets from training, but using only the omitted subset to compute the mean-squared error.

4.3.1 Optimum feature space

The feature space is a coded representation of the training data. Our image normalization scheme from Chapter 2 is first utilized to compensate for face illumination, pose and minor expressions to generate the normalized age database described in Section 4.1.1. Color information is discarded by converting the normalized age database into greyscale images. The image code could contain raw pixel values in the spatial domain or other representations of an image such as the frequency space.

Spatial Vs Frequency Domain Processing

First we needed to determine whether to use the raw pixel data or convert the images to frequency domain. It is possible that the frequency representation generates a better feature space than its spatial counterpart, since some measurements are easier made in the frequency domain.

The first experiment tested the spatial and frequency feature spaces. For the frequency representation, we applied a standard wavelet transform to the image. A key benefit of wavelets is that they provide a multi-resolution analysis of the image in the form of coefficients. The image can be decomposed into several layers of low and high frequency information separating the ‘edge’ or wrinkle information in the underlying face at various resolutions. We used the Daubechies-4 wavelet transform for our experiments. More information on wavelet processing and the various wavelet transforms can be found in [Daubechies, 1990].

A third degree polynomial kernel was selected for SVM experiments with margin size, $\varepsilon = 1$, and convergence tolerance, $e = 1$. Table 4.2 summarizes our results on the various feature spaces.

From the age prediction results in Table 4.2, the minimum error was realized by the raw pixel data in the spatial domain retaining fewer support vectors. The wavelet transformation did not help improve accuracy, although error rates were close to the raw image representation. Also we observe a slight increase in the error as we decompose the face image into more levels; however we did not pursue wavelet transformations further because error rates were lower with the raw normalized face images.



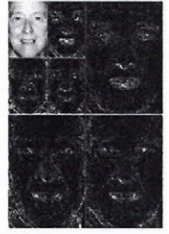
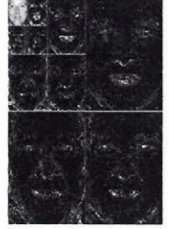
Feature Space		Average Number of Support Vectors	Absolute Error in Years	Squared Correlation Coefficient
Raw Image		562	9.83	0.62
Wavelet Processed (1 level decomposition)		571	9.85	0.61
Wavelet Processed (2 level decomposition)		573	9.88	0.61
Wavelet Processed (3 level decomposition)		576	9.98	0.60

Table 4.2: Spatial Vs Frequency Domain as a feature space

Masking the face region

The second issue was whether to ‘mask’ the face region or not. Lanitis [Lanitis, 2002] claims that introduction of the hairline has a negative effect on training for age prediction. The increased variability of the hair region distracts the age estimation task. We tested his theory empirically by comparing results obtained by the whole face image to the cropped face image.

Once again, a third degree polynomial kernel was selected for experiments with margin size, $\epsilon = 1$, and convergence tolerance, $e = 1$. Table 4.3 summarizes our results.



Feature Space		Average Number of Support Vectors	Absolute Error in Years	Squared Correlation Coefficient
Normalized Images (without cropping)		562	9.83	0.62
Normalized Images (face region cropped)		542	9.41	0.69

Table 4.3: Masking Vs Not Masking on the feature spaces

The normalized images with the face region cropped yielded a lower absolute error retaining fewer support vectors. The claim made by Lanitis [Lanitis, 2002] with regard to the variability of the hair region hindering the age estimation task is confirmed by our experiment.

Image Size

The size of the face image used for training was also evaluated. Intuitively, a larger image size will contain more wrinkle detail and age information and would be better for training. However, the dimensionality of the feature space explodes with increasing image size, requiring more training samples to train the SVM. Therefore we need to determine the smallest image size that captures all the aging information.

We used images of various sizes to observe how SVRs performed with increasing size. The images were resized by a bilinear interpolation. All the training images were normalized

and cropped before inputting them to the SVM. Experiments on the image sizes shown in Table 4.4 were carried out.

Relative Size	Actual Image Size	Dimensionality (Number of Pixels)
1.0	116 x 160	18560
0.95	110 x 152	16750
0.9	104 x 144	15034
0.85	99 x 136	13410
0.8	93 x 128	11878
0.75	87 x 120	10440
0.7	81 x 112	9094
0.65	75 x 104	7842
0.6	70 x 96	6682
0.55	64 x 88	5614
0.5	58 x 80	4640
0.45	52 x 72	3758
0.4	46 x 64	2970
0.35	41 x 56	2274
0.3	35 x 48	1670
0.25	29 x 40	1160
0.2	23 x 32	742
0.15	17 x 24	418

Table 4.4: Relative and actual image sizes

Graphs of the size versus all three measurements are shown in Figure 4.6. The graphs follow our intuition. As the image size increases, the error decreases almost stabilizing at an image size of 0.6, as seen in Figure 4.6(a). Plotting the same graph on a log scale, Figure 4.6(b) reveals a linear falloff in the error rate with increasing image size until a size of 0.4. After that, the rate of decrease is much slower. However, the lowest error is achieved by using the full image size. Inspection of the graph in Figure 4.6(c) also reveals an expected trend. In general, the number of support vectors decreases with increasing image size, indicating that the complexity of the learnt function decreases together with a decrease in the error rates. This is what we would expect when learning algorithms such as SVMs perform

well in high-dimension spaces. Finally, in Figure 4.6(d) the graph of squared cross correlation versus the size is seen to be exactly the inverse of the ‘Absolute error Vs size’ graph. The highest cross-correlation is attained at the full image size.

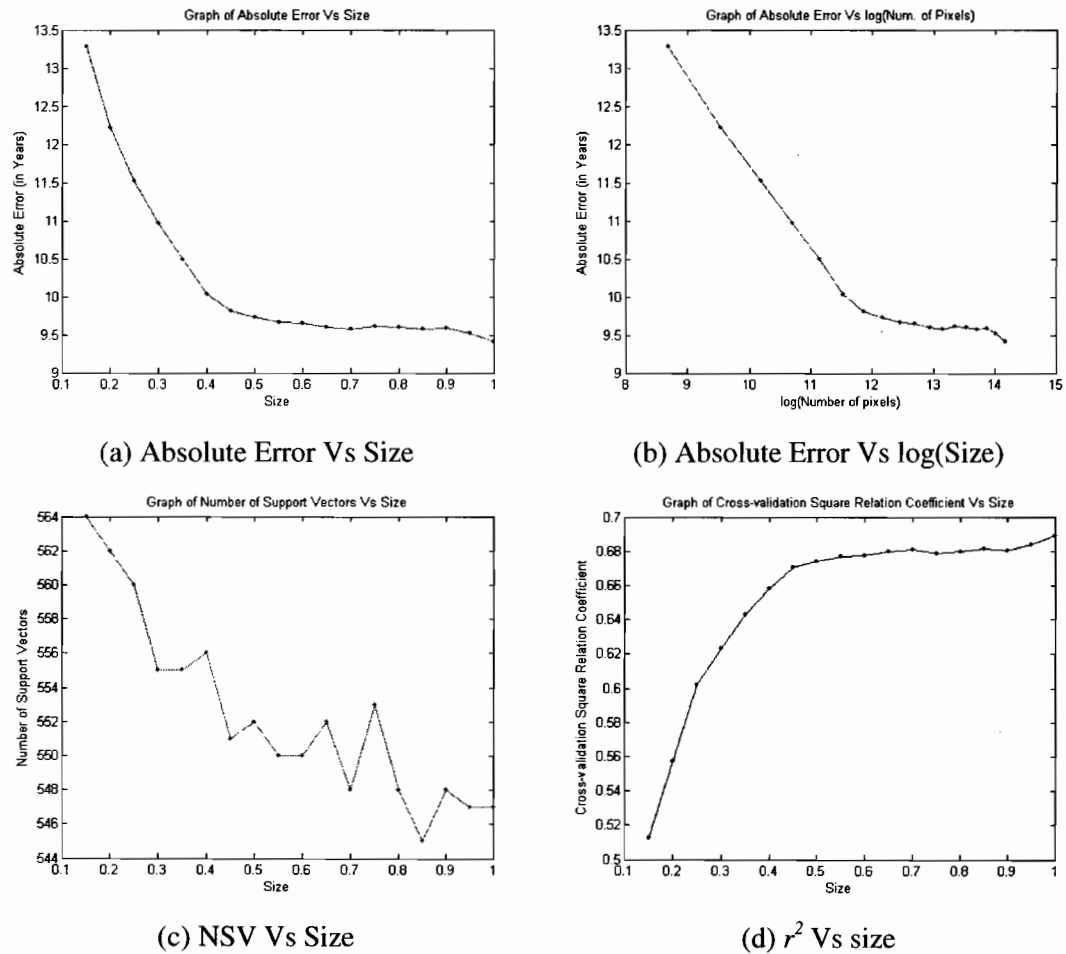


Figure 4.6: Results of experiments on various image sizes

From the first set of experiments, we observe the proficiency of SVMs when applied to very high dimensional data. With close to 20,000 dimensions, the SVM still performs the age estimation task very well even though only 818 images were available. In conclusion, the best feature space was covered by the full size normalized images with the face region cropped.

4.3.2 Best kernel and parameters (K , P)

With the feature space established, this next set of experiments determine the best kernel (K) and the corresponding parameters (P). As we recall from Section 4.1, the use of kernel functions can allow the SVM to produce nonlinear functions. The three standard kernels from Table 4.1 were tested and results are reported in Table 4.5. Each experiment was carried out with four different values of ε (margin size) and e (convergence tolerance) – with the average values reported.

Kernel		Average Number of Support Vectors	Absolute Error (in Years)	Squared Correlation Coefficient
K	P			
Linear	-	483	9.59	0.68
Polynomial	Degree = 2	477	9.41	0.69
<i>Polynomial</i>	<i>Degree = 3</i>	<i>474</i>	<i>9.31</i>	<i>0.69</i>
Polynomial	Degree = 4	476	9.43	0.69
Polynomial	Degree = 5	478	9.6	0.68
Polynomial	Degree = 6	478	9.65	0.67
Polynomial	Degree = 7	479	9.72	0.66
Polynomial	Degree = 8	480	9.73	0.66
Polynomial	Degree = 9	484	9.74	0.65
Polynomial	Degree = 10	477	9.80	0.65
Polynomial	Degree = 11	480	9.88	0.64
Sigmoid	Weight $\eta = 1$	520	15.23	0.004
Sigmoid	Weight $\eta = 3$	523	15.94	0.003
Sigmoid	Weight $\eta = 5$	527	16.21	0.003
Sigmoid	Weight $\eta = 10$	530	16.55	0.003
Sigmoid	Weight $\eta = 20$	532	16.97	0.002
Radial Basis	Width $\sigma = 1$	510	13.11	0.01
Radial Basis	Width $\sigma = 3$	523	13.35	0.01
Radial Basis	Width $\sigma = 5$	524	13.68	0.009
Radial Basis	Width $\sigma = 10$	526	13.56	0.009
Radial Basis	Width $\sigma = 20$	527	14.21	0.008

Table 4.5: Results of applying different kernels for age prediction

From the results in Table 4.5, we can conclude that the aging function is polynomial in nature since the least error was achieved with the polynomial kernels, while retaining relatively fewer Support Vectors. The linear, radial basis function and sigmoid kernels did not perform as well as the third polynomial kernel on the different combinations of ε and e . A third degree polynomial ($K = \text{Polynomial}$, $P = 3$) achieved the minimum absolute error over unseen data and was selected as the optimum kernel.

4.3.3 Best regression parameters (ε , e)

The parameters ε and e are the final two parameters that we needed to optimize. In this set of experiments, we used the best parameters determined thus far, which is a third order polynomial kernel is trained on normalized face images whose face region has been cropped.

The values of ε (margin size) and e (convergence tolerance) range from 1 to 10 years each, and experiments were performed at intervals of 1. Thus, a total of 100 experiments were carried out so as to determine the trend in our measurements. Figure 4.7 illustrates a surface plot of the absolute error with different combinations of ε and e . The blue regions indicate low error values while the red regions indicate high error rates.

As seen in Figure 4.7, there is a general increase in the error as the convergence tolerance increases. This trend is expected because intuitively, if we stop optimization when the error in the estimated age of the training images is still large (over 5 years), the function will be too simple and we will notice high errors over unseen images. However, if we terminate optimization at very low values of e (less than 1), we over-fit the training data, and note poorer performance over unseen data. Therefore, a value that lies between 1 and 10 would be ideal. The empirical results demonstrate this phenomenon and a global minimum in the mean squared error is found at a value of $e = 2$.

Similarly, as we tolerate more errors in the training data (with larger values of ε), the function will not be able to isolate aging effects. However, if we tolerate very little noise (with small values of ε such as 0 or 1) then we overfit the training data. Hence, we expect to find a value that lies in between 1 and 10. The surface in Figure 4.7 has a minimum that occurs at a value of $\varepsilon = 3$.

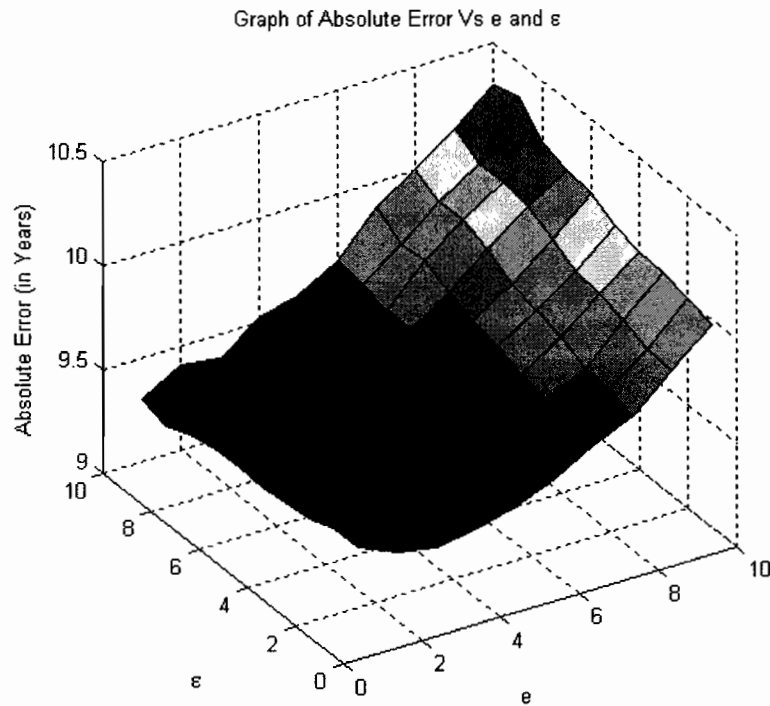


Figure 4.7: Graph of Absolute Error Vs e and ϵ

The optimum values obtained for ϵ and e are sensible in that we would tolerate errors in the training data of ± 3 years, terminating convergence when the difference in the predicted age over training data is 2 years. The minimum absolute error over unseen data attained was 9.2 years. This figure indicates that the error in the predicted age of the unseen data was approximately ± 9 years.

Next, the graph of Number of Support Vectors versus ϵ and e is shown in Figure 4.8. The trend is exactly what we would expect. The highest number of support vectors is retained when we can tolerate a minimal error on the training data (narrow margin) and when the convergence tolerance very small. This region is indicated by the dark red area of the surface plot in Figure 4.8. On the other hand, the least number of support vectors is retained when we tolerate more error on the training data (wide margin) and convergence tolerance is large.

The surface is a plane with the maximum number of Support Vectors retained at $\varepsilon = 1$ and $e = 1$, and the least number of Support Vectors retained at $\varepsilon = 10$ and $e = 10$.

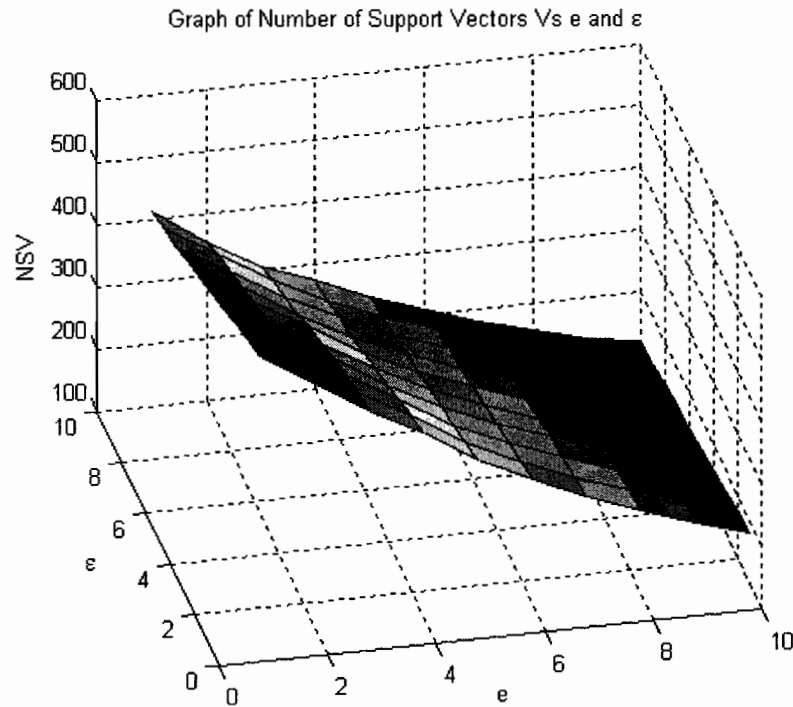


Figure 4.8: Number of Support Vectors versus e and ε

Finally, as we recall, the squared cross-correlation coefficient is simply the inverse of the graph in Figure 4.7. The graph is shown in Figure 4.9. The highest correlation is achieved when the error is a minimum, at $\varepsilon = 3$ and $e = 2$.

At this point, we have established all of the parameter values that generate the best function to can predict the age of an unseen individual, based on our data.

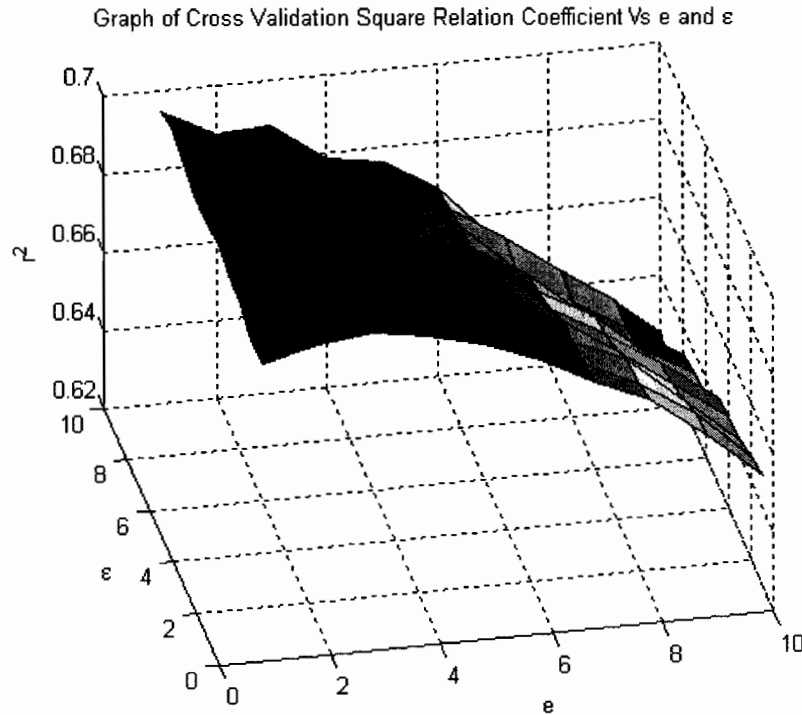


Figure 4.9: Squared Cross Correlation Coefficient versus ϵ and ϵ

4.4 Summary

In this chapter, we have discussed the training of a Support Vector Regression Machine to perform age prediction on labelled frontal images. The images were collected from the Internet and compensated for pose, illumination and expression by our normalizing scheme in Chapter 3.

We experimented with several different parameters to determine the best set that minimized the age estimation error over unseen images. The aging function was best represented by a third order polynomial regression function. Images whose face regions have been cropped performed better than the whole image itself, using the full size as the feature space. In addition, an error of ± 3 years on the training data is tolerated with a termination criterion of 2 years to gain best results over unseen data.

In the next chapter, we use this aging function to help synthesize aged portraits.

Chapter 5

AGE SIMULATION

In this chapter, we describe how to synthesize aged portraits with the aid of our global aging function derived in the previous chapter. We combine the aging function and a method to transfer wrinkle information between subjects to make an individual in an image appear older or younger.

We adapted Tiddeman et al.'s [Tiddeman et al., 2001] concept of creating texture enhanced age prototypes to create 'signature' images of various age groups that are 5 years apart. However, our approach is much simpler and more elegant than Tiddeman et al., producing similar results. We also propose a novel extension to the Image Based Surface Detail Transfer Method [Shan et al., 2001] to create these prototypes in Section 5.2.

Finally, we present a new automatic aging simulation process that utilizes the aging function, age prototypes and a method to transfer textures. The Image Based Surface Detail transfer method is used to transfer textures also. The age simulation process along with experiments and results are discussed in Section 5.3.

5.1 Image Based Surface Detail Transfer (IBSDT)

The IBSDT method was formulated by Shan et al. [Shan et. al., 2001] to transfer geometric details from one surface to another with simple image operations. Geometrically, the difference between an old person's face and a young person's face is that the older face's skin surface has more bumps (or wrinkles) on the face²⁵. If we transfer just the wrinkles from the older face to the younger face, the younger face will look older. A naïve approach would be to compute the difference of the two images, warping this difference image to make the target image look older. Burson and Nancy [Burson and Nancy, 1981] used this method to simulate aging effects. The problem with the latter was that the difference image contains skin colour information of the original two faces; therefore, the skin colour of the new face may be modified by the aging process.

To overcome the skin colour problem, Shan et al. [Shan et. al., 2001] formulated IBSDT, which captures the geometrical details of an object in a way that is independent of the surface reflectance property. The captured geometrical details can then be transferred to another surface to produce the appearance of the new surface with added geometrical details while its reflectance property is preserved. The key difference between IBSDT and other methods [Burson and Nancy, 1981], [Tiddeman et al., 2001] is that only one source image is required by IBSDT.

The IBSDT method is described next. Under a Lambertian model, the recorded intensity of a point P in the image I is given as:

$$I(\mathbf{p}) = \rho(P) l \mathbf{n}(P) \cdot \mathbf{l}(P), \quad 5.1$$

where l is the intensity of the light source (only one light source assumed) and $\mathbf{l}(P)$ denotes the light direction from P to the light source. The reflectance coefficient at P is denoted by $\rho(P)$. Finally, \mathbf{p} is the 2D projection of P onto the image where $\mathbf{p} = C(P)$ and $C(\cdot)$ is the camera projection function.

The problem is stated as follows: Given images I_1 and I_2 of two aligned surfaces S_1 and S_2 , respectively, what is the new image \hat{I}_2 of S_2 if we modify its surface normal such that:

²⁵ Aging also causes changes in face shape, which is beyond the scope of this thesis.

$$\hat{n}_2(P_2) = n_1(P_1), \quad 5.2$$

where P_1 and P_2 are the corresponding points on the images.

Because the distance between P_1 and P_2 is small (since the surfaces are aligned) and the light surface is far away, the approximation $I(P_1) \approx I(P_2)$ is made. The following relation is obtained from Equation 5.1 and Equation 5.2:

$$\hat{I}_2(p_2) \approx \frac{I_1(p_1)\rho(P_2)}{\rho(P_1)} \quad 5.3$$

In order to compute the ratio of $\rho(P_1)$ to $\rho(P_2)$, a smoothed image of I is defined as:

$$\bar{I}(p) \equiv \sum_{q \in \Omega(p)} w(q)I(q), \quad 5.4$$

where $\Omega(p)$ is the neighbourhood of p , and $w(\cdot)$ is the kernel function of a Gaussian smoothing filter²⁶.

Shan et al. prove that Equation 5.5 holds after making further assumptions. The formulation demonstrates that the transfer of the surface normal can be approximated by simple operations on the images of surfaces.

$$\hat{I}_2(p_2) \approx \frac{I_1(p_1)}{\bar{I}_1(p_1)} \bar{I}_2(p_2) \quad 5.5$$

Finally, IBSDT from Equation 5.5 is written as Equation 5.6 for simplicity.

$$\hat{I}_2 = \frac{I_1}{\bar{I}_1} \bar{I}_2 \quad 5.6$$

²⁶ The Gaussian smoothing filter is a 2-D convolution filter that is used to ‘blur’ images and remove detail and noise. It uses a kernel that represents the shape of a Gaussian (‘bell-shaped’) hump. Details can be found in [Gonzalez and Wintz, 1987].

From a signal processing viewpoint, Equation 5.6 merely substitutes the high frequency components of I_2 with those from I_1 . Once alignment of feature points is completed, a Gaussian filter with a user supplied σ is run on I_1 and I_2 , to obtain \bar{I}_1 and \bar{I}_2 . Finally, Equation 5.6 is applied to obtain the new image \hat{I}_2 .

The width of the Gaussian function, σ , controls how much geometrical smoothing is performed on the surfaces of I_1 and I_2 . Hence, it determines the scale of the surface details to be transferred. A small σ allows the fine geometrical details to be transferred, while a large σ allows only large-scale geometrical details to be transferred. It is inherent that the two images should be warped to the same shape before transferring textures, otherwise the resulting image would be meaningless.

The method is best illustrated via examples. We picked two synthetic images from Waldron²⁷ [Waldron, 2003] of the same person at different ages. The images are shown in Figure 5.1.

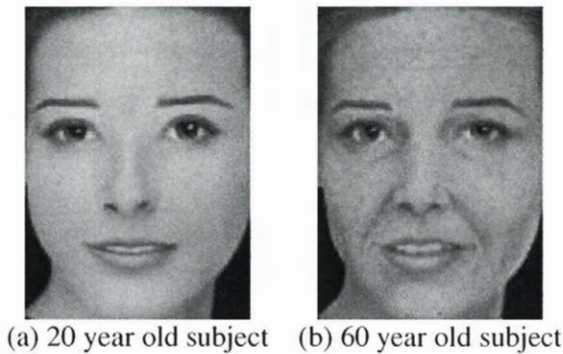


Figure 5.1: Two synthetic images of the same person at different ages

On application of different values of σ , we obtain the reconstructed images \hat{I}_2 shown in Figure 5.2 for an *age progression*. In Equation 5.6, the image I_2 is Figure 5.1(a) while image I_1 is Figure 5.1(b).

²⁷ Images are the original artwork of D'Lynn Waldron [Waldron, 2003] showing a woman at different ages.

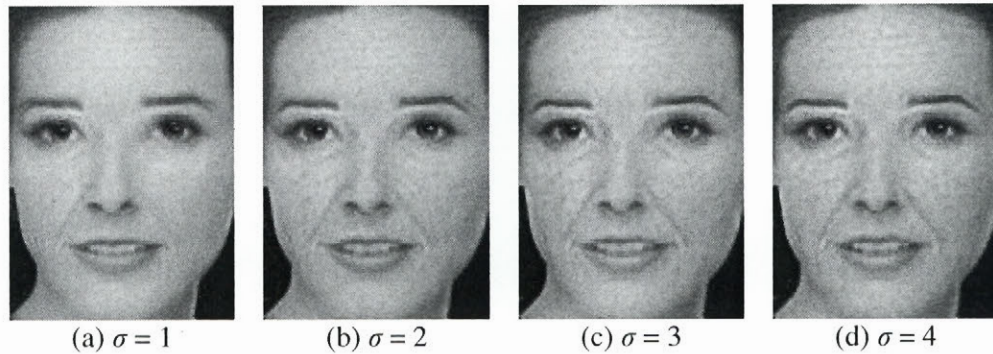


Figure 5.2: Age progression with various values of σ

Resulting images from Figure 5.2 demonstrate an increase in wrinkles captured as we increase the value of σ . In addition, with higher values of σ , some stubble is also captured in the reconstructed images. The subject retains her skin colour in reconstructions, even with higher values of σ , addressing the fact that geometric details are captured independent of its surface reflectance property.

Next, we performed *age regression* by switching I_1 and I_2 in Equation 5.5 i.e. using Figure 5.1(a) as I_1 and Figure 5.1(b) as I_2 . The resulting images are shown in Figure 5.3.

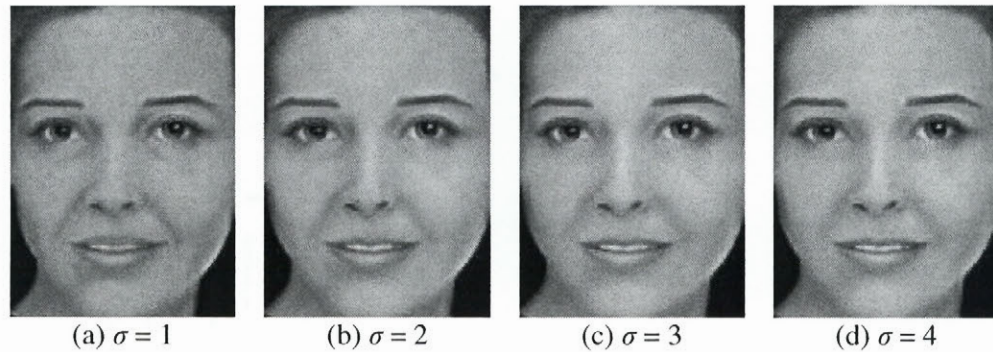


Figure 5.3: Age regression with various values of σ

Once again, it is evident that the images retained their skin colour while losing the wrinkles and stubble to appear younger. Indeed, the reconstructed images do simulate an age progression pattern by simply using different values of the width σ .

A limitation reported by Shan et al. was the requirement that the lighting conditions between the two images be similar. We easily resolved this constraint by applying our

illumination invariant signature obtained by histogram fitting the Retinex processed images (see Section 3.1). However, for our aging database, we simply needed to perform histogram fitting due to the good lighting conditions of these images.

We conclude that the IBSDT method works very well in transferring textures from one image to another. It is important to note that areas such as the eyes, nose and mouth have sharp changes in intensity, hence are regarded as areas of high frequency together with the wrinkles and stubble. Therefore, these textures will also be transferred during the reconstruction process.

5.2 Age Prototypes

The notion of creating age prototypes has been adapted from Burt and Perrett [Burt and Perrett, 1995] and Tiddeman et al. [Tiddeman et al., 2001]. Prototype images are used to define salient features of a particular age group. In our study, we create adult prototype images from our database for every age group 5 years apart. We use a scale of 5 years because a wider range would provide less accuracy while a narrower range would require more data to characterise the group. The age prototypes are used along with the IBSDT method to define an age transformation on a given image. By warping geometrical details (wrinkles and stubble) from the prototypes to the novel image, we can perceptually change the age of the subject in the image. The latter is discussed in Section 5.3.

5.2.1 Extension to IBSDT

In order to create an age prototype, we can simply warp all the images in that age group to the mean shape, and compute the average image. Burt and Perrett [Burt and Perrett, 1995] applied this method to generate the prototypes. For example, given images I_1, I_2, \dots, I_N in an age group, the age prototype is defined as:

$$\bar{I} = \frac{1}{N} \sum_{i=1}^N I'_i, \quad 5.7$$

where I'_i is the corresponding warped image of I_i

The perceived age ratings of male and female prototypes showed an age deficit compared to the original faces used to construct the prototypes. It was clear when looking at the prototype images of older faces that wrinkles weren't captured in the prototypes. The facial textures appear smoother than they do in real faces. This is because the warping stage does not align the wrinkles precisely. Aligning the wrinkles is difficult, if not impossible, due to the wide variability of their fine structure.

To improve the textures of the facial prototypes, Tiddeman et al. [Tiddeman et al., 2001] proposed a wavelet-based method. Results from their study are illustrated in Section 2.2. They created prototypes that looked realistic and captured wrinkle details well. However their method was quite involved. We propose a novel approach to perform the same task, but with simple 2D image operations.

Intuitively, if we warped all the high frequency information (including wrinkle and stubble) from the sample images into the average image, we would get a 'texture-enhanced' image that would be a better representative of that characteristic age group. Essentially, Tiddeman et al. did the same by manipulating the edge strength in the wavelet coefficients of the images. We extended the Image Based Surface Detail Transfer in Equation 5.6 to handle multiple images as shown in Equation 5.8.

$$\hat{I} = \frac{I_1}{\bar{I}_1} \cdot \frac{I_2}{\bar{I}_2} \dots \frac{I_N}{\bar{I}_N} \bar{I}, \quad 5.8$$

where \hat{I} is the 'texture-enhanced' prototype, \bar{I} is the average of images I_1, I_2, \dots, I_N from Equation 5.7 and \bar{I}_i is the Gaussian smoothed image of I_i . Gaussian smoothing is performed by a kernel of width σ .

The high frequency information from the images in the set is captured and transferred to the average image, resulting in a texture enhanced age prototype. As mentioned in the IBSDT method, the value of σ determines the level of textures to transfer. A small value will capture the fine geometrical details (fine lines and wrinkles) from all the images, while a large value will capture larger scale geometrical details.

5.2.2 Texture enhancement results

We generate a male age prototype for the 75–79 age group to demonstrate the effect of texture enhancement and σ . Our aging database contains 15 images that belong to this age group. Some sample images and the average image are shown in Figure 5.4.

As we expected, the average image \bar{I} appears to be smoothed out and is not a good representative of the age group due to the absence of wrinkles. The perceived age of \bar{I} demonstrates an age deficit. Next, we applied Equation 5.8 to generate the texture-enhanced prototypes with different values of σ . The resulting images are shown in Figure 5.5.

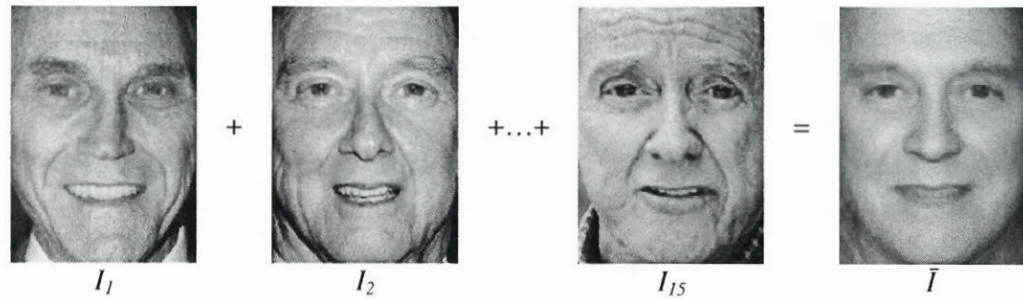


Figure 5.4: Sample images from the 40 - 44 year age group and their average

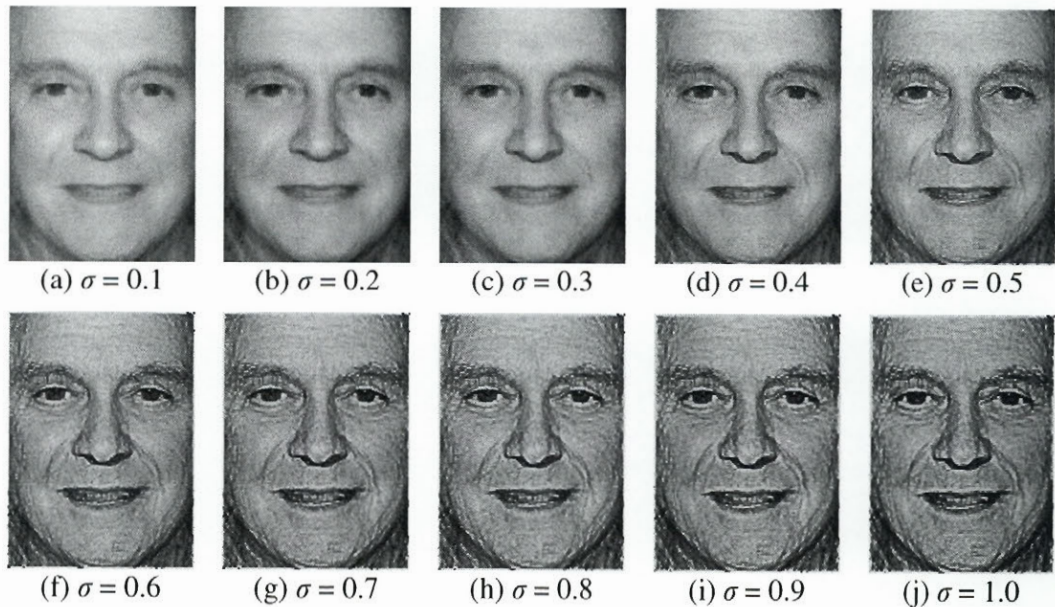


Figure 5.5: Texture enhanced male age prototype with different values of σ

We observe from Figure 5.5 that the prototype captures more wrinkles with increasing σ . With small values of σ , only the fine lines and wrinkles have been captured, while thicker wrinkles have been captured with larger values of σ . The images generated with σ between 0.4 and 0.7 look like they would be a good perceptual representative of the 75–79 age group. However, for different age groups, we would need to transfer different scales of wrinkles such that the prototype accurately represents its age group. Therefore, we used the aging function to determine the best value of σ for generating age prototypes.

5.2.3 Age prototyping

This algorithm to create age prototypes is summarized as follows:

1. Generate an age prototype I_T using Equation 5.8 with $\sigma = 0.1$, using all the images in the database that belong to this age group. i.e.

$$I_T = \frac{I_1}{\bar{I}_1} \cdot \frac{I_2}{\bar{I}_2} \dots \frac{I_N}{\bar{I}_N} \bar{I}$$

2. Predict the age of the prototype I_T using the aging function, F .

$$A_T = F(I_T)$$

3. Find the difference between the target age, A_T , and predicted age, A_N . Note it as τ .

$$\tau = A_N - A_T$$

4. Repeat steps 1-3 with different values of σ , in increments of 0.05 until $\sigma = 1.0$.
5. The value of σ used to generate I_T that minimizes τ is the ideal. Therefore, retain I_T as the prototype with this ideal value of σ (denoted as σ_{min}).

A program flowchart that summarizes this process is illustrated in Figure 5.6. It is important to note that, due to the number of pixel divisions and multiplications, the resulting image spanned a wide dynamic range. As a result, we required double precision to represent the pixels. In addition, we had to clip the intensity histogram at the edges and stretch the contrast of the images to fit between 0 and 256 for display purposes.

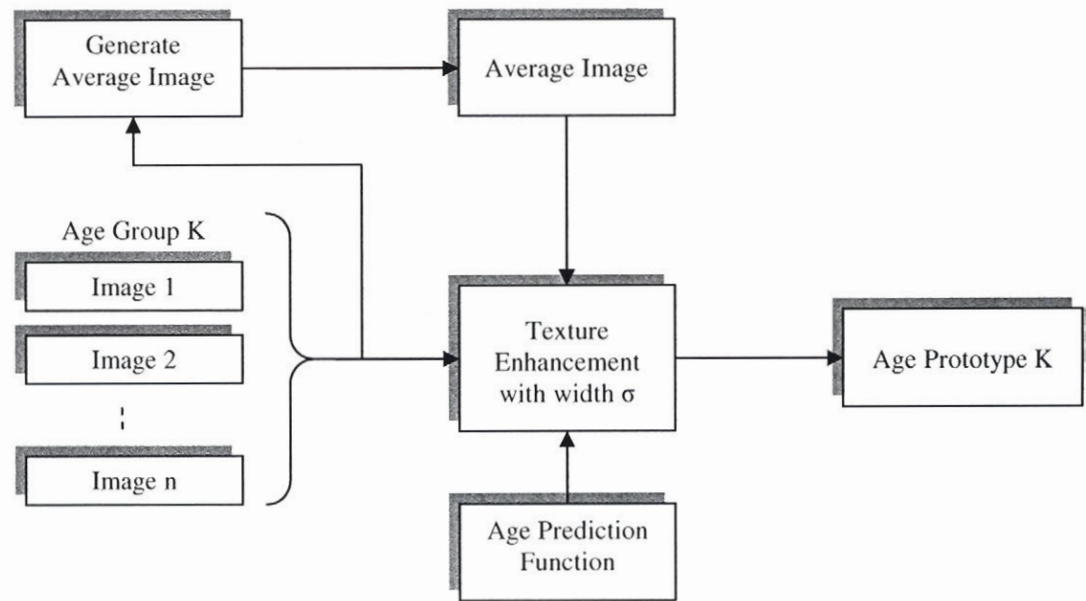


Figure 5.6: Flowchart of age prototype creation

We generated the age prototypes based on this algorithm, and the results were exceptional. The age prototypes, corresponding values of σ_{min} used to generate them and the predicted age, A_p are reported in Figure 5.7 and Figure 5.8. Typical graphs demonstrating the effect of σ on the age prototypes are shown in Figure 5.9.

First and foremost, from the results we conclude that the aging function correctly predicts the age of images presented to it. Perceptually, we can easily see that the age ratings of the prototypes correspond to the predicted age. The values of σ_{min} also fall in the range that we expected earlier, between 0.3 and 0.7. Beyond this range, the results would not correctly represent the age group – this is confirmed by our aging function, since the function did not accurately predict the ages of the prototypes beyond this range. It is also interesting to note that the female age prototypes did not exhibit many wrinkles, which could be due to various reasons such as application of makeup, plastic surgery, etc.

Finally, we observe from Figure 5.9(a) that the value of σ is directly related to the predicted and perceived age of the prototypes. As we add more texture information to the face images, in effect we increase the age rating of the prototype. Thus, our intuition is matched with experimental results.

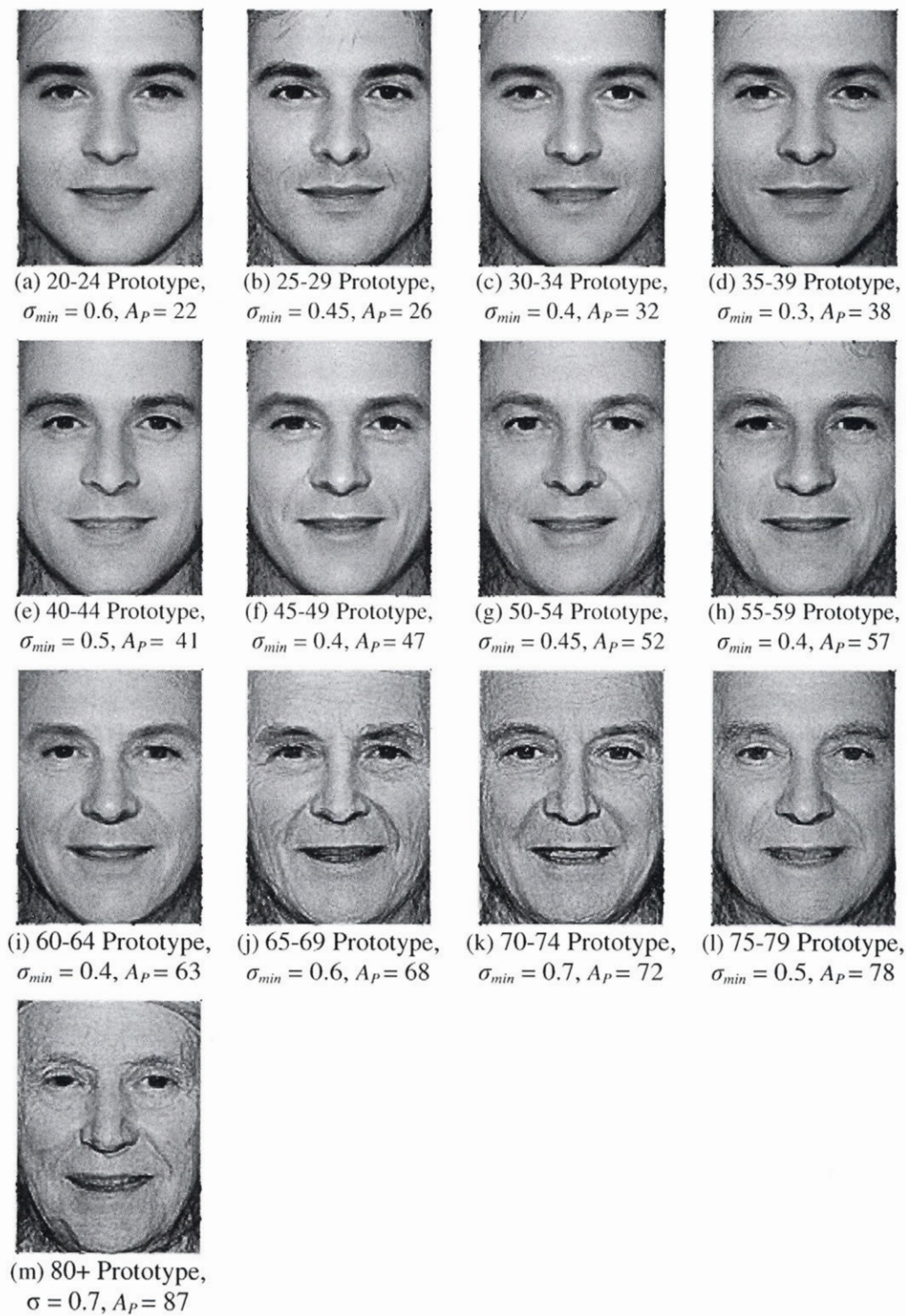


Figure 5.7: The male age prototypes

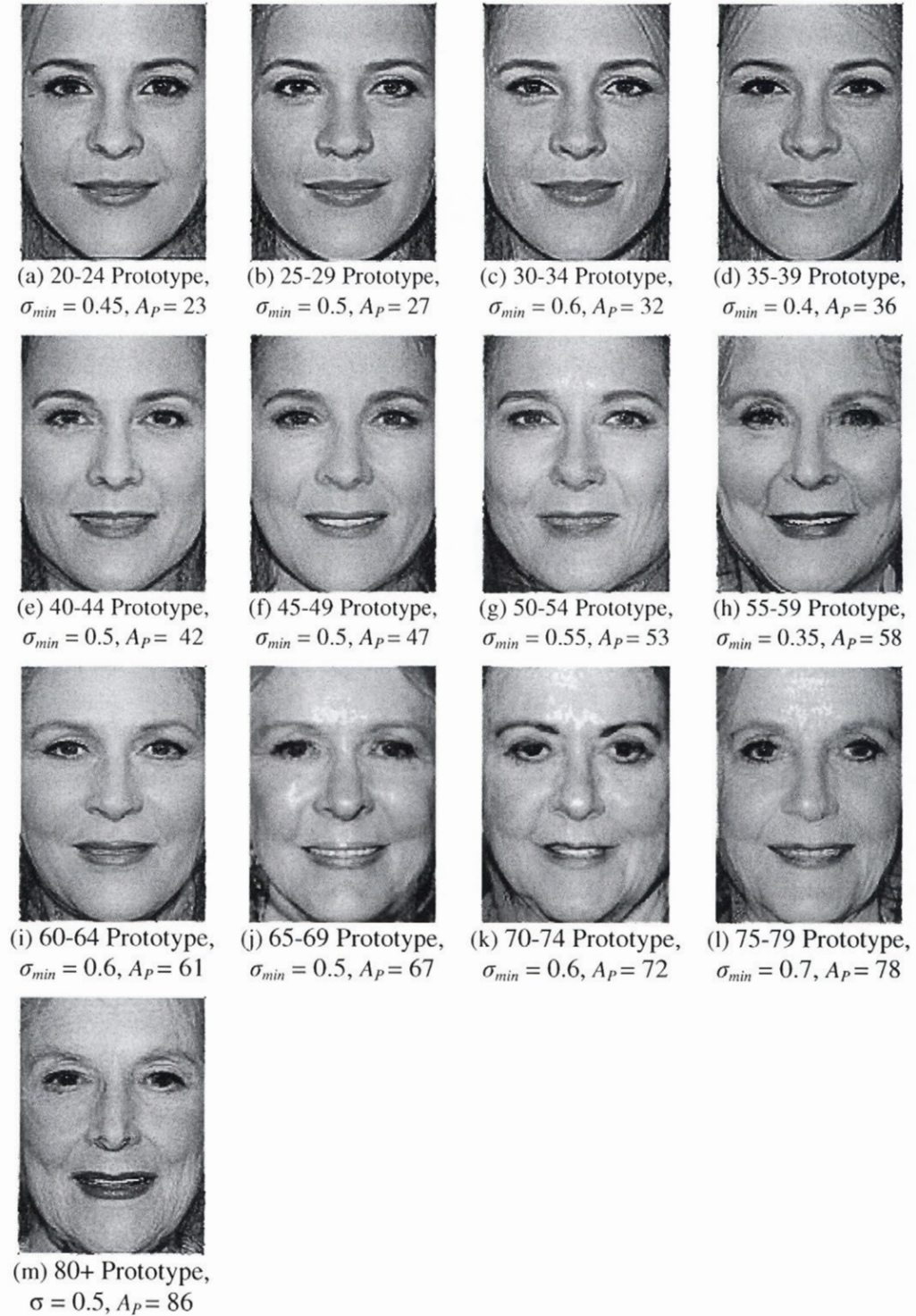
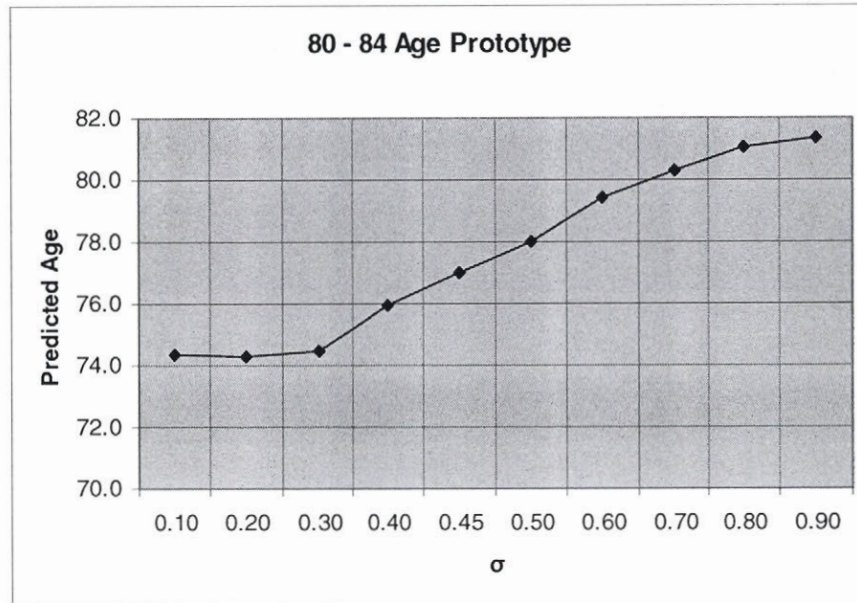
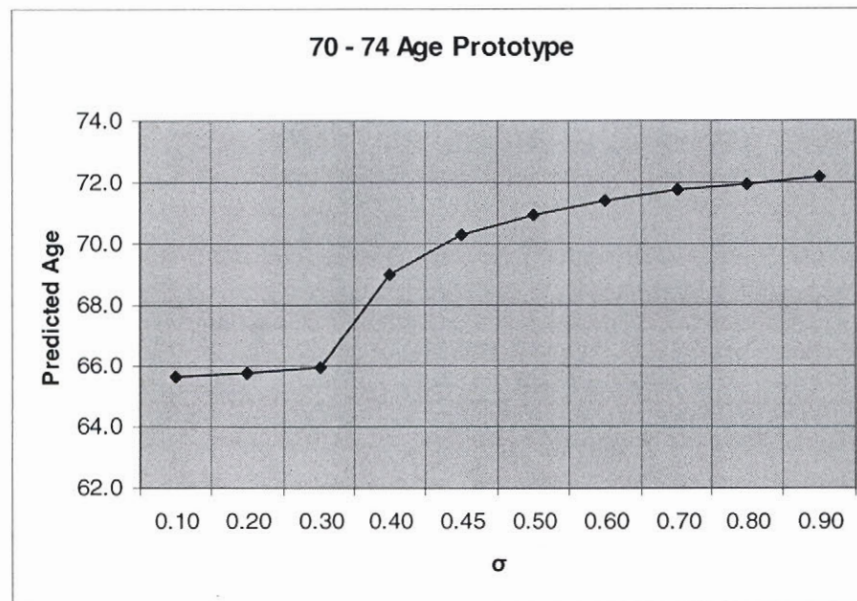


Figure 5.8: The female age prototypes



(a) Predicted age versus σ for the 80-84 male age prototype



(b) Predicted age versus σ for the 70 - 74 female age prototype

Figure 5.9: Typical graphs from the age prototypes creation process

A significant point to note is that we limited the prototypes from 20 years to 80 years in age. The primary reason for this limitation was the lack of images of subjects below 20 years and over 80 years. Therefore, our age simulation is restricted to this range of ages. It is also important to note that the face undergoes major changes in shape before the age of 20 years. Therefore we would have to take into account the face shape as well as texture when performing age simulation below 20 years. As mentioned earlier, this is beyond the scope of this thesis.

5.3 Image synthesis

The actual age simulation of a novel face image is described in this section. Thus far, we have described a method to transfer wrinkle information from one face to another and generated age prototypes that capture significant properties of an age group. In this section, we use the Image Based Surface Detail Transfer method to age a novel image to a target age that is represented by one of the age prototypes. We use the aging function to determine the precise amount of wrinkle information to transfer from the age prototypes to the novel face image for accurate aging.

5.3.1 Approach

Our age simulation approach is summarized as follows:

1. Given a novel image I_N , and the target age, A_T , find a corresponding prototype image, I_T , that represents A_T .
2. Use Equation 5.6 with a value of 0.5 for σ to synthesize an age portrait.

$$\hat{I}_N = \frac{I_T}{\bar{I}_T} \bar{I}_N \quad 5.9$$

3. Predict age of the synthesized image, A_N using the aging function.

$$A_N = F(\hat{I}_N) \quad 5.10$$

4. Find the difference between the target age, A_T , and predicted age, A_N . Denote it as τ .

$$\tau = A_N - A_T \quad 5.11$$

5. If $\tau > 3$:
 - a. If $\sigma > 0.1$, then reduce σ by 0.1 ($\sigma = \sigma - 0.1$). Go to step 2 with new value of σ
 - b. If $\sigma \leq 0.1$, then select new younger age prototype and go to step 2 with $\sigma=0.5$ and I_{T-J} .
6. If $\tau < 3$:
 - a. If $\sigma < 1.0$, then increase σ by 0.1 ($\sigma = \sigma + 0.1$). Go to step 2 with new value of σ
 - b. If $\sigma \geq 0.1$, then select new older age prototype and go to step 2 with $\sigma=0.5$ and I_{T+J} .
7. If $-3 < \tau < 3$, then the predicted age of the synthesized image is close enough to the target age and synthesis is complete.

A program flowchart is demonstrated in Figure 5.10 for a sample image. In this scenario, the novel image belongs to a male subject. For a female subject, the female prototypes would be used instead of the male.

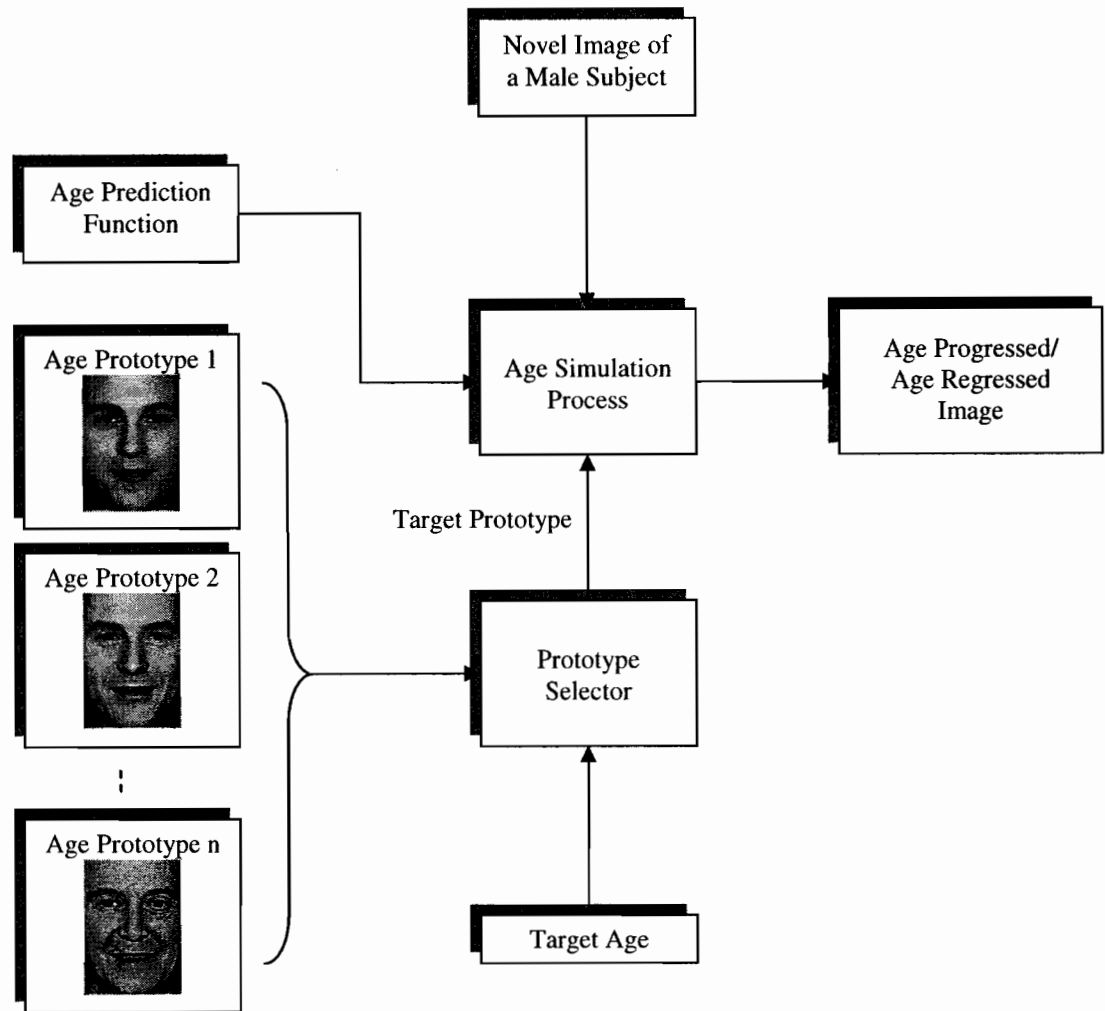
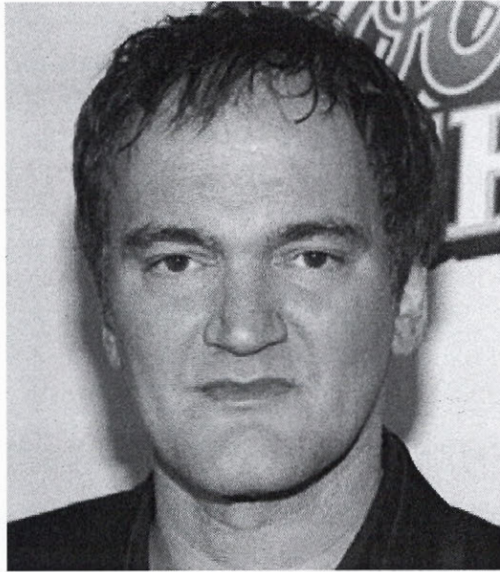


Figure 5.10: Flowchart of age simulation

5.3.2 Experimental Results

We present several age progressed and age regressed portraits generated by our software in this section. These test images were not present in the database while training the age prediction function and are previously unseen by the software.

Figure 5.11 illustrates age progression results on subject 1. Our method was used to generate the aged faces of subject 1 at various target ages. The synthesis demonstrates realistic examples of the aged portraits, preserving the identity of the subject and precisely aging the image.



Subject Name: Quentin Terentino

Actual Age: 41 years

Predicted Age: 36 years

Age estimation error: 5 years

Age Simulation: Age progression to 50, 60, 70 and 80 years respectively



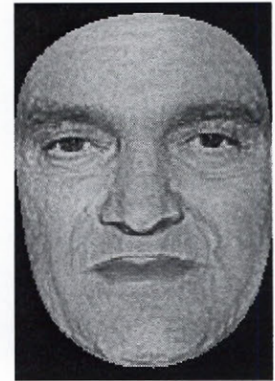
(a) Target Age: 50
years



(b) Target Age: 60
years



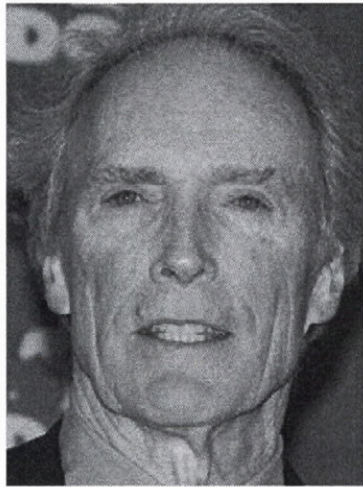
(c) Target Age: 70
years



(d) Target Age: 80
years

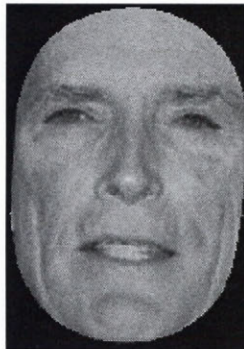
Figure 5.11: Age simulation results on subject 1

Next, we demonstrate age regression simulation on an older subject in Figure 5.12. Once again, we note the realistic simulations of the images while preserving the identity of the subject in the image. As the subject is ‘reverse-aged’ to much younger target ages, we notice the absence of wrinkles in the synthesized images that were present earlier. Therefore, the age regression is indeed realistic.

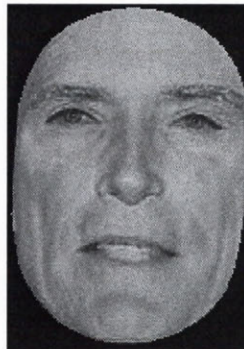


Subject Name: Clint Eastwood
Actual Age: 74 years
Predicted Age: 71 years
Age estimation error: 3 years

Age Simulation: Age regression to 60, 50, 40 and 30 years respectively



(a) Target Age: 60 years



(b) Target Age: 50 years



(c) Target Age: 40 years



(d) Target Age: 30 years

Figure 5.12: Age simulation results on subject 2

In Figure 5.13, we show age progression as well as age regression on a middle-aged man. Once again, the reconstructions tend to preserve the identity of the subjects while simulating age changes on them.



Subject Name: Tom Hanks
Actual Age: 44 years
Predicted Age: 40 years
Age estimation error: 4 years

Age Simulation: Age progression to 50, 60, and 70 years respectively



Target Age: 50 years



Target Age: 60 years



Target Age: 70 years

Age Simulation: Age regression to 30, 25, and 20 years respectively



Target Age: 30 years



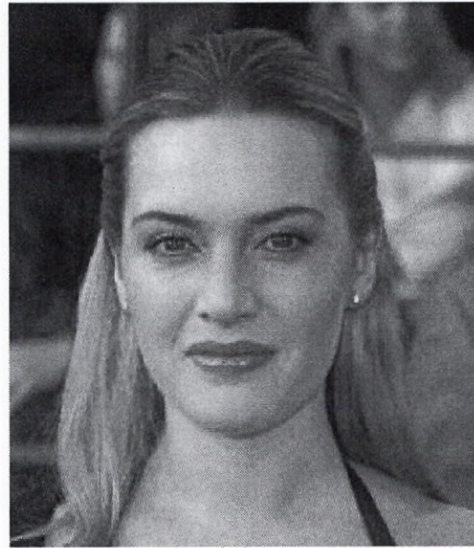
Target Age: 25 years



Target Age: 20 years

Figure 5.13: Age simulation results on subject 3

In Figure 5.14, we demonstrate age progression on a female subject. The female prototypes did not capture significant wrinkle information (seen in Figure 5.8), reflected in the age simulations of the subject in Figure 5.14.



Subject Name: Kate Winslet
Actual Age: 27 years
Predicted Age: 31 years
Age estimation error: 4 years

Age Simulation: Age progression to 40, 50, 60 and 80 years respectively



Target Age: 40 years



Target Age: 50 years



Target Age: 60 years



Target Age: 70 years

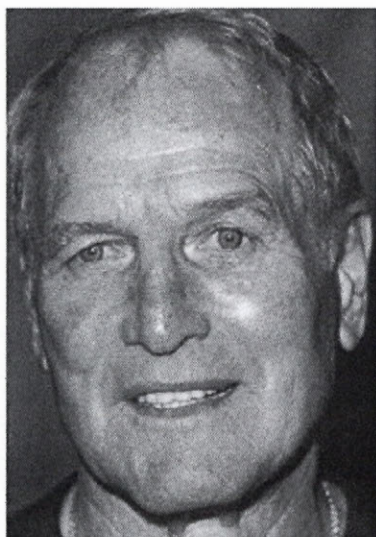
Figure 5.14: Age simulation results on subject 4

From the simulations that we have generated thus far, it is difficult to predict the accuracy of the age progression or regression. Due to the unavailability of actual images of the subjects at the target age, we cannot compare our simulations to ‘real’ images. In Figure 5.15, we demonstrate this phenomenon. We used two real images of a subject at two different ages, 31 years and 72 years respectively. In Figure 5.15(a), we perform age regression of the subject from 72 years down to 31 years. The image produced by our

software and the actual image are placed side by side for comparison. Then in Figure 5.15(b), we age progressed the 31 year old to 72 years and compared the output of our system to the actual image.

We observe that the age regression was quite close to the actual image from the results in Figure 5.15(a). On the other hand, the synthesized image in the age progression (seen in Figure 5.15(b)) had more facial wrinkles than the actual image. This factor can be explained by the error in age prediction. Our prediction function estimated the actual age of the 72 year old image to be 65 years. Hence, the subject appears younger for his age. Therefore, during the automatic age progression, more wrinkles are transferred while aging the subject to 72 years.

Finally, we present an example of age progression as well as age regression on subject 6 in Figure 5.6. We downloaded images of this subject at 3 different ages, 27, 37 and 71 years respectively from the Internet. Using the facial image at 37 years old, we age progressed subject 6 to 71 years old. The age progressed image and the actual image at 71 years are depicted in Figure 5.6(a) and Figure 5.6(b). The similarity between the synthesized image and the actual image is astounding. Next, using the same image at 37 years old, we age regressed subject 6 to look as if he were 27 years old. Once again the age regressed image and the actual subject image are shown in Figure 5.6(c) and Figure 5.6(d) respectively.



Subject Name: Paul Newman
Actual Age: 72 years
Predicted Age: 65 years
Age estimation error: 7 years



Subject Name: Paul Newman
Actual Age: 31 years
Predicted Age: 32 years
Age estimation error: 1 years

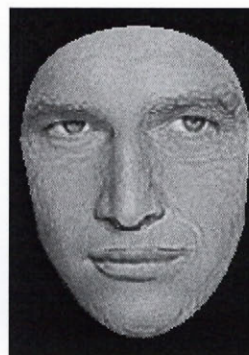


Synthesized Image



Actual Image

(a) Age Simulation: Age regression to 31



Synthesized Image



Actual Image

(b) Age Simulation: Age progression to 71

Figure 5.15: Comparison of age simulations to actual images from Subject 5

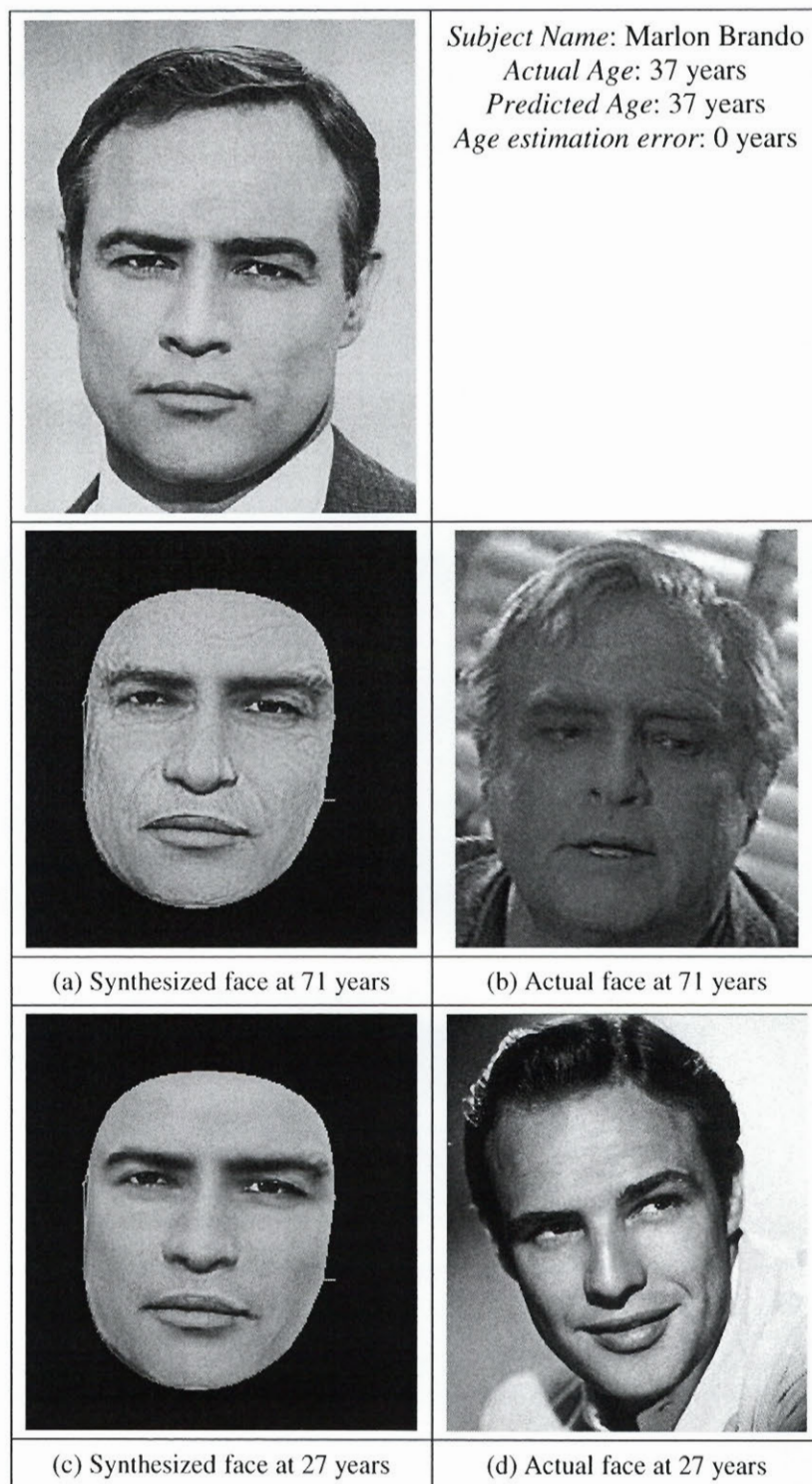


Figure 5.16: Comparison of age simulations to actual images from Subject 6

5.4 Summary

In this chapter we presented a novel method to generate age prototypes that perceptually reflect the age of the age group. Our simple extension to the IBSDT method made it possible to combine high frequency information from many images into a single image producing realistic age prototypes.

As well, we proposed a new framework for aging individuals based on the aging function and the age prototypes generated from our database. Our method is different from previous methods since we incorporated the age prediction function to the age simulation process.

The age simulation results on various images previously unseen by our software demonstrate realistic simulations. Also, the predicted ages of these individuals were very close to the actual ages, indicating the presence of a good age prediction function. The results demonstrate that we have a simple and powerful means for performing automatic age simulation in face images. Possibly due to the application of makeup or plastic surgery, we couldn't capture major wrinkle information from the female celebrities in the database. As a result, the age simulations on facial images of female subjects were not very accurate.

Chapter 6

CONCLUSION AND FUTURE DIRECTIONS

We have developed a fully automatic aging system in this thesis. A complete image normalization scheme was also developed that works better than the standard methods available in the literature.

We collected the *McGill CBIR Aging Database*, containing over 800 high quality images of subjects labelled by age and gender, as part of our research. Most of the images were collected from the Internet Movie Database [Imdb, 2004]. Due to application of makeup and possibly plastic surgery, we found it difficult to label the facial images of female subjects. There was a significant different in the actual age and perceptual age of these facial images, therefore mainly images of men were collected.

After image normalization, an age prediction function was established from our database. A Support Vector Regression Machine was trained with the data. Our age predictor was accurate to ± 9 years on unseen images.

The same database was used to generate signature images that captured the salient features of a particular age group. We used the aging function and a novel extension to a texture

enhancement procedure in order to create these signature images known as age prototypes. The age prototypes derived were realistic in characterising the age group they belong to.

Finally, using all the subsystems, we created a simple yet powerful automatic aging system that accurately ages an unseen individual, based on our database. The age simulations demonstrate that the method works well and is practical when applied to real world images.

We propose several improvements and additions to our system:

1. Incorporating the race and lifestyle of the individuals to the learning and age simulation process to increase the accuracy of the age progressions. It has been established that various social factors affect the aging process. Incorporating these into our system will increase accuracy of the simulations.
2. Integrating shape information from medical anthropometrical data to the age simulations. Although not very significant, the face undergoes minor shape changes as it ages. Face anthropometrical data are available [Pitanguy et al., 1996; Pitanguy et al., 1998 and Leta et al., 2000] to determine how the face changes. Warping the face to reflect these changes in shape would provide precise aging. Since we already have the feature points on the face image, the measurements would simply change the locations of these points.
3. In this thesis, we have demonstrated only age simulations from 20 to 80 years. Below this age range, we would require much more data as well as face shape information from improvement #2, since the face undergoes major changes during 0-20 years.
4. It is possible to automatically attach the hair information to the age progressed or age regressed portraits. Due to our geometric normalization, this task is trivial to implement. However, attaching the original hair information is deceptive in the age simulations. For instance, attaching the hair information from a 40 year old subject to the 80 year old age progressed image is deceptive, as we would expect balding or whitening of the hair to occur and vice versa. A future improvement would be to change the hair information as well in order to reflect the aging process. The example in Figure 6.1 illustrates this phenomenon.

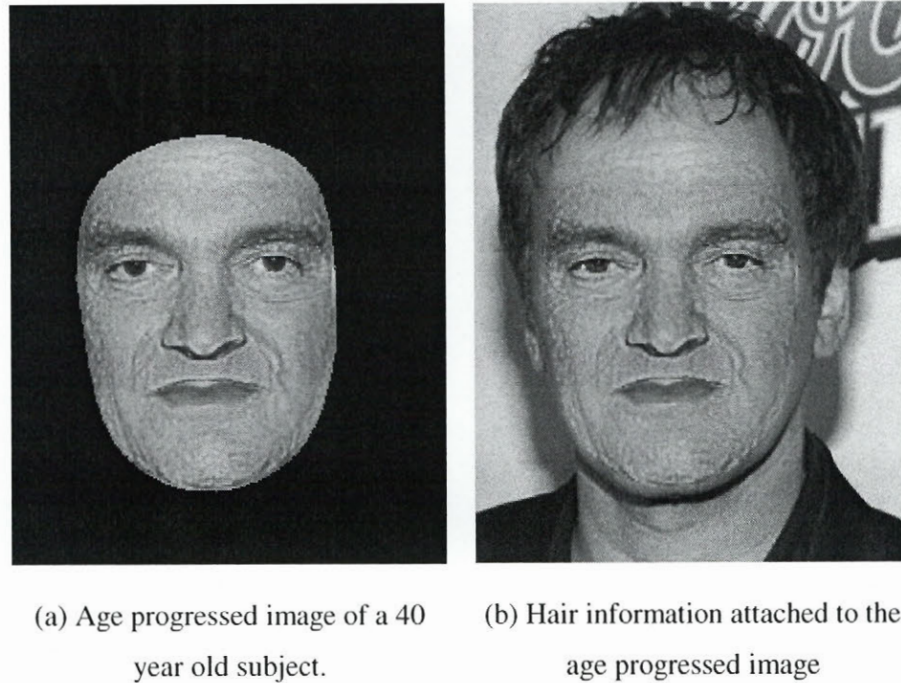


Figure 6.1: Attaching hair information to the age progression

5. We have established that Support Vector Machines can learn the aging function. In order to reduce the errors from our age prediction function, we need to collect more images, especially of females.
6. During age simulations, we pick an age prototype that is closest to the target age, and transfer wrinkle information from it to age a novel face image. However, it is also possible to take an older prototype and use smaller values of the width σ to transfer textures arriving at the same target age. Hence there are several ways of arriving at this target age, and we need to investigate the best method.
7. For our age simulations, we assumed a 'global' aging trend. This is only true to a certain extent. For more 'person-specific' aging, we need to devise other mechanisms to simulate age changes.

The proposed modifications can seamlessly be integrated into our software for improved results. In conclusion, our approach is an excellent stepping-stone for further research in the areas of age prediction and simulating new face images to reflect age changes.

Appendix A

Cubic Spline Polynomials

This appendix has been adapted mostly from Mathworld [Weisstein, 1999]. A cubic spline is a curve constructed of piecewise third-order polynomials which pass through a set of m control points. Figure A.1 illustrates some sample points connected via a cubic spline. The second derivative of each polynomial is commonly set to zero at the endpoints, since this provides a boundary condition that completes the system of $m - 2$ equations. This produces a so-called "natural" cubic spline and leads to a simple tridiagonal system which can be solved easily to give the coefficients of the polynomials.

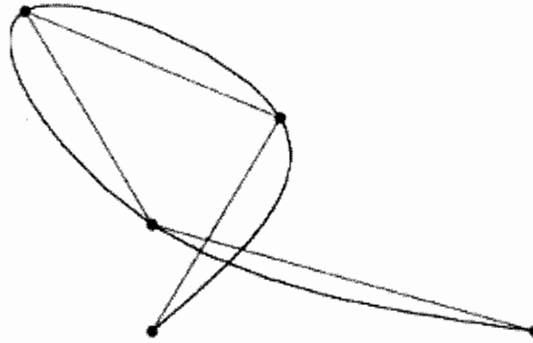


Figure A.1: 5 points connected via a cubic spline

Consider 1-dimensional spline for a set of $n + 1$ points $(y_0, y_1, y_2, \dots, y_n)$. Following Bartels et al. [Bartels et al., 1998], let the i^{th} piece of the spline be represented by Equation A.1.

$$Y_i(t) = a_i + b_i(t) + c_i(t^2) + d_i(t^3), \quad \text{A.1}$$

where t is a parameter $t \in [0,1]$ and $i = 0, \dots, n-1$. Then

$$Y_i(0) = y_i = a_i \quad \text{A.2}$$

$$Y_i(1) = y_{i+1} = a_i + b_i + c_i + d_i \quad \text{A.3}$$

Taking the derivative of $y_i(t)$ in each interval then gives

$$Y'_i(0) = D_i = b_i \quad \text{A.4}$$

$$Y'_i(1) = D_{i+1} = b_i + 2c_i + 3d_i \quad \text{A.5}$$

Solving Equations A.2 – A.5 for a_i , b_i , c_i and d_i then gives

$$a_i = y_i \quad \text{A.6}$$

$$b_i = D_i \quad \text{A.7}$$

$$c_i = 3(y_{i+1} - y_i) - 2D_i - D_{i+1} \quad \text{A.8}$$

$$d_i = 2(y_i - y_{i+1}) + D_i + D_{i+1} \quad \text{A.9}$$

Now require that the second derivatives also match at the points, so

$$Y_{i-1}(1) = y_i \quad \text{A.10}$$

$$Y'_{i-1}(1) = Y'_i(0) \quad \text{A.11}$$

$$Y_i(0) = y_i \quad \text{A.12}$$

$$Y''_{i-1}(1) = Y''_i(0) \quad \text{A.13}$$

since the interior points, as well as the end points satisfy

$$Y_0(0) = y_0 \quad \text{A.14}$$

$$Y''_{n-1}(1) = y_n \quad \text{A.15}$$

This gives a total of $4(n-1) + 2 = 4n - 2$ equations for the $4n$ unknowns. To obtain two more conditions, require that the second derivatives at the end points are zero, so

$$Y''_0(0) = y_0 \quad \text{A.16}$$

$$Y''_{n-1}(1) = 0 \quad \text{A.17}$$

Rearranging all the equations gives rise to the following tridiagonal system

[Bartels et al., 1998].

$$\begin{bmatrix} 2 & 1 & & & & \\ 1 & 4 & 1 & & & \\ & 1 & 4 & 1 & & \\ & & 1 & 4 & 1 & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & 1 & 4 & 1 \\ & & & & 1 & 2 \end{bmatrix} \begin{bmatrix} D_0 \\ D_1 \\ D_2 \\ D_3 \\ \vdots \\ D_{n-1} \\ D_n \end{bmatrix} = \begin{bmatrix} 3(y_1 - y_0) \\ 3(y_2 - y_0) \\ 3(y_3 - y_1) \\ \vdots \\ 3(y_{n-1} - y_{n-3}) \\ 3(y_n - y_{n-2}) \\ 3(y_n - y_{n-1}) \end{bmatrix} \quad \text{A.18}$$

If the curve is instead closed, the system becomes

$$\begin{bmatrix} 4 & 1 & & & 1 \\ 1 & 4 & 1 & & \\ & 1 & 4 & 1 & \\ & & 1 & 4 & 1 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ & & & 1 & 4 & 1 \\ 1 & & & & 1 & 4 \end{bmatrix} \begin{bmatrix} D_0 \\ D_1 \\ D_2 \\ D_3 \\ \vdots \\ D_{n-1} \\ D_n \end{bmatrix} = \begin{bmatrix} 3(y_1 - y_n) \\ 3(y_2 - y_0) \\ 3(y_3 - y_1) \\ \vdots \\ 3(y_{n-1} - y_{n-3}) \\ 3(y_n - y_{n-2}) \\ 3(y_0 - y_{n-1}) \end{bmatrix} \quad \text{A.19}$$

References

- [Aprilage, 2002] Aprilage, 2002. Description of a Smokers Face, Data Provided by Roswell Park Cancer Institute, Buffalo, N.Y, http://www.aprilage.com/smoking_effects.html
- [Bartels et al., 1998] R. H. Bartels, J. C. Beatty and A. B. Barsky, "Hermite and Cubic Spline Interpolation," *Ch. 3 in An Introduction to Splines for Use in Computer Graphics and Geometric Modelling*. San Francisco, CA: Morgan Kaufmann, Page(s): 9-17, 1998.
- [Bennett and Campbell, 2000] K. Bennett and C. Campbell, "Support Vector Machines: Hype or Hallelujah?," *SIGKDD Explorations*, Volume: 2, No:2, Page(s): 1-13, 2000.
- [Bhattacharyya, 2004] J. Bhattacharyya, "Detecting and Removing Specularities and Shadows in Images," Masters Thesis, Department of Electrical and Computer Engineering, McGill University, June 2004.
- [Blanz and Vetter, 1999] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," *Computer Graphics (Proc. Siggraph 99)*, ACM Press, New York, 1999, Page(s): 187-194.
- [Botox, 2004] Botulinum Toxin Type A, *Official website*. <http://www.botox.com/site/>
- [Burges, 1999] C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," <http://svm.research.bell-labs.com/SVMdoc.html>, *AT&T Bell Labs*, November 1999.
- [Burt and Perrett, 1995], D.M. Burt and D.I. Perrett, "Perception of Age in Adult Caucasian Male Faces: Computer Graphic Manipulation of Shape and Colour Information," *In Proc. Royal Society London, Series B*, Volume: 259, Page(s): 137-143, 1995.
- [Burson and Schneider, 1981] N. Burson and T. D. Schneider. "Method and apparatus for producing an image of a person's face at a different age," *United States Patent 4276570*, 1981.
- [Chan and Lau, 1998] K.H. Chan and R. Lau, "Contour-Based Warping," *CVGIP: Graphical Models and Image Processing*, Academic Press, Volume: 60, No: 5, Page(s): 331-348, September 1998.
- [Chang and Lin, 2001] C.C. Chang and C.J. Lin, "*LIBSVM : a library for support vector machines*," 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [Cootes et al., 1995] T.F Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active Shape Models – Their training and application," *Computer Vision and Image Understanding*, Volume:61, No: 1, 1995, Page(s) 38-59.
- [CrimeLib, 2002] Court TV's Crime Library, 2002. Criminal Minds and Methods, Age Progression, http://www.crimelibrary.com/criminal_mind/forensics/art/6.html
- [Cristianini and Shawe-Taylor , 2000] N. Cristianini, and J. Shawe-Taylor, "An Introduction to Support Vector Machines and Other Kernel Based Learning Methods," *Cambridge University Press*, 2000.
- [Daubechies , 1990] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inf. Theory*, Volume: 36, No: 5, Page(s):961-1005, 1990.

- [Eckhoff, 2000] N. D. Eckhoff, "Principles of interpolation," *ME 400 Computer Applications in Mechanical Engineering*. Course Notes [<http://athena.mne.ksu.edu/~nde/me400s02/Me40024.pdf>].
- [Edwards et al., 1998] G.J. Edwards, A. Lanitis, C.J. Taylor and T.F. Cootes, "Statistical Face Models: Improving specificity," *Image and Vision computing*, Volume: 16. No: 3, 1998, Page(s) 203 - 211.
- [Flores, 2002] Flores, J. R. "National Estimates of Missing Children: An Overview, National Incidence Studies of Missing, Abducted, Runaway, and Thrownaway Children," *U.S. Department of Justice*, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention, October 2002.
- [Farkas, 1987] Farkas, L., "Anthropometric facial proportions in medicine", Thomas Brooks, 1987.
- [Farkas, 1994] Farkas, L., "Anthropometry of the Head and Face", Raven Press, 1994.
- [Georghiades et al., 2001] Georghiades, A.S. and Belhumeur, P.N. and Kriegman, D.J., "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Volume: 23, No: 6, Page(s): 643-660, 2001.
- [Hayashi et al., 2002] J. Hayashi, M. Yasumoto, H. Ito, Y. Niwa, and H. Koshimizu, "Age and Gender Estimation from Facial Image Processing," *Proceedings of the 41st SICE Annual Conference*, Volume: 1, Page(s): 13 -18, Aug. 2002.
- [Holcomb, 2002] Holcomb J. D. 2002. Face Lift Surgery (Rhytidectomy) Pictures, Case 1432, <http://www.locateadoc.com/gallery.cfm/Action/Gallery/GalleryID/1432>
- [Hussein, 2002] Hussein H. K., "Towards Realistic Facial Modeling and Re-Rendering of Human Skin Aging Animation," *Proc. Of the Shape Modeling International 2002 (SMI'02)*.
- [Imdb, 2004] The Internet Movie Database, <http://www.imdb.com/>
- [Jähne, 1997] B. Jähne, "Geometric Transformations," Ch. 8.5 in *Digital Image Processing*, 4th Edition. La Jolla, CA: Springer, pp 244 – 261, 1997.
- [Jobson et al., 1997] D. J. Jobson , Z. Rahman, G. A. Woodell, "A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes," *IEEE Transactions on Image Processing*, Volume: 6, No: 3, Page(s): 965-976, July 1997.
- [Jones and Poggio, 1996] M. Jones and T. Poggio, "Model-based matching by linear combinations of prototypes," *A.I Memo No. 1583, C.B.C.L. Paper No. 139*, MIT, Nov. 1996.
- [Koch et al., 1996] Koch R. M., Gross M. H., Carls F. R., Buren D.F., Frankhauser G., and Parish Y. I. H., "Simulating facial surgery using finite element methods," *In SIGGRAPH*, Page(s) 421-428, August 1996
- [Kwon and Lobo, 1993] Y.H. Kwon and N. da Vitoria Lobo, "Locating facial features for age classification," *In proceedings of SPIE – the International Society for Optical Engineering*, Volume: 2055, Page(s): 62-72, 1993.

- [Kwon and Lobo, 1999] Y.H. Kwon and N. da Vitoria Lobo, "Age Classification from facial images," *Computer Vision and Image Understanding*, Volume: 74, Page(s): 1-21, April 1999
- [Land, 1977] E. Land, "The Retinex Theory of Color Vision," *Scientific American*, Page(s): 108-129, Dec. 1977.
- [Lanitis et al., 1997] A. Lanitis, C.J. Taylor and T.F. Cootes, "Automatic Identification and Coding of Human Faces using Flexible Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume: 19 No: 7, Page(s): 743 – 756, July 1997.
- [Lanitis and Taylor, 2000] A. Lanitis and C.J. Taylor, "Towards automatic face identification robust to ageing variation," *In Proc. Of 4th IEEE International Conference on Automatic Face and Gesture Recognition*, Page(s): 391-396, 2000.
- [Lanitis et al., 2002] 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*, Volume: 24 Issue: 4, Page(s): 442 - 455, April 2002.
- [Lanitis, 2002] A. Lanitis, "On the significance of different facial parts for automatic age estimation," *14th International Conference on Digital Signal Processing*, Volume: 2, Page(s): 1027 -1030, July 2002.
- [Lee et al., 1996] S. Lee, G. Wolberg, K.-Y. Chwa and S. Y. Shin, "Image Metamorphosis with Scattered Feature Constraints," *IEEE Trans. On Visual. & Comp. Graph.*, Volume: 2, No: 4, Page(s): 337-354, 1996.
- [Leta et al., 2000] F. R. Leta, D. Pamplona, H. I. Weber, A. Conci, and I. Pitaguy, "Study of the facial aging – a multidisciplinary approach," *Revista Brasileira CienCias Mecanicas*, Volume: 22, No: 3, Page(s): 489-501, 2000.
- [LSU, 2001] LSU Faces Laboratory 2001. Law Enforcement, Age Progression, <http://www.faceslab.lsu.edu/law/ageprogression.htm>
- [Mercer, 1909] J. Mercer, "Functions of positive and negative type and their connection with the theory of integral equations," *Philos. Trans. Roy. Soc. London*, A 209:415-446, 1909.
- [Moses et al, 1991] Y. Moses, Y. Adini and S. Ullman, "Face Recognition: The problem of compensating for changes in Illumination Direction," *European Conf. Computer Vision*, Page(s): 286 – 296, 1991.
- [NCMEC, 2001] NCMEC, National Center For Missing and Exploited Children, 2001. Age Progression and Facial Reconstruction Services, <http://www.missingkids.com/missingkids>
- [Osuna, 1997] E. E Osuna, R. Freund and F. Girosi, "Support Vector Machines: Training and Applications," *MIT AI Memo No. 1602*, March 1997.
- [Pessa et al., 1999] J. E. Pessa, V. P. Zadoo, C. Yuan, J. D Ayedelotte, F. J. Cuellar, C. S. Cochran, K. L. Mutimer, and J. R. Garza., "Concertina effect and facial aging: Nonlinear aspects of youthfulness and skeletal remodelling, and why, perhaps, infants have jowls," *Plastic and Reconstructive Surgery*, Volume: 103, No: 2, Page(s): 635-644, February 1999.

- [Pittanguy et al., 1996] I. Pitanguy, F. Leta, D. Pamplona, and H. I. Weber, "Defining and measuring aging parameters," In *Applied Mathematics and Computation*, Volume: 78, Page(s): 217-227, September 1996.
- [Pittanguy et al., 1998] I. Pitanguy, D. Pamplona, H. I. Weber, F. Leta, F. Salgado, and H. N. Radwanski, "Numerical modeling of facial aging," *Plastic and Reconstructive Surgery*, Volume: 102, No: 1, Page(s): 200-204, July 1998.
- [Ruprect and Muller, 1995] D. Ruprecht and H. Muller, "Image Warping with Scattered Data Interpolation," *IEEE Computer Graphics and Applications*, Vol. 3, Page(s): 37-43. 1995
- [Tiddeman et al., 2001] B. Tiddeman, M. Burt, and D. Perrett, "Protopyping and transforming facial textures for perception research," *Computer Graphics and Applications, IEEE*, Volume: 21 Issue: 5, Page(s): 42 -50, July-Aug. 2001.
- [Tiddeman et al., 2001] B. Tiddeman, N. Duffy and G. Rabey, "A general method for overlap control in image warping," *Computers and Graphics*, Volume: 25, No: 1, February 2001.
- [Shan et al., 2001] Y. Shan, Z. Liu and Z. Zhang, "Image-Based Surface Detail Transfer," *CVPR 2001*, Hawaii, Vol. II, Page(s): 794-799, Dec. 2001.
- [Shashua and Riklin-Raviv, 2001] A. Shashua and T. Riklin-Raviv, "The Quotient Image: Class-Based Re-Rendering and Recognition with Varying Illuminations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Volume: 23, No:2, Page(s): 129-139, 2001
- [Smith, 1997] J. Smith, "Sun Exposure: Precautions and Protection," *Ohio State University Extension Fact Sheet*, HYG-5550-97, Consumer And Textile Science, 1997 <http://ohioline.osu.edu/hyg-fact/5000/5550.html>
- [Smola and Schoelkopf, 1998] A. Smola and B. Schoelkopf, "A tutorial on support vector regression", *NeuroCOLT2 Technical Report NC2-TR-1998-030*, 1998.
- [Waldron, 2003] Waldron, D'L. 2003. Age-Progressed and Age-Regressed Portraits and Animations, <http://www.dlynnwaldron.com/ageprogression.html>
- [Weisstein, 1999] E.W. Weisstein. "Cubic Spline." *From MathWorld--A Wolfram Web Resource*. <http://mathworld.wolfram.com/CubicSpline.html>

Bibliography

- [Gonzalez and Wintz, 1987] R.C. Gonzales and P. Wintz. *Digital Image Processing*. Addison Wesley, Reading, MA, 2nd Edition, 1987.
- [Gonzalez and Woods, 1992] R.C. Gonzalez and R.E. Woods, “*Digital Image Processing*,” Addison-Wesley Publishing Company (New York), 1992.
- [Vapnik, 1995] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
- [Vapnik, 1998] V. Vapnik. *Statistical Learning Theory*. Springer, Wiley, 1998.
- [Wolberg 1990] G. Wolberg. *Digital Image Warping*. IEEE Computer Society Press, 1990.