FACE AGE ESTIMATION USING WRINKLE PATTERNS

CHOON-CHING NG

2015

SUPERVISED BY DR. MOI HOON YAP AND DR. NICHOLAS COSTEN

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

SCHOOL OF COMPUTING, MATHEMATICS AND DIGITAL TECHNOLOGY
MANCHESTER METROPOLITAN UNIVERSITY

Abstract

Face age estimation is a challenging problem due to the variation of craniofacial growth, skin texture, gender and race. With recent growth in face age estimation research, wrinkles received attention from a number of research, as it is generally perceived as aging feature and soft biometric for person identification. In a face image, wrinkle is a discontinuous and arbitrary line pattern that varies in different face regions and subjects. Existing wrinkle detection algorithms and wrinkle-based features are not robust for face age estimation. They are either weakly represented or not validated against the ground truth. The primary aim of this thesis is to develop a robust wrinkle detection method and construct novel wrinkle-based methods for face age estimation. First, Hybrid Hessian Filter (HHF) is proposed to segment the wrinkles using the directional gradient and a ridge-valley Gaussian kernel. Second, Hessian Line Tracking (HLT) is proposed for wrinkle detection by exploring the wrinkle connectivity of surrounding pixels using a cross-sectional profile. Experimental results showed that HLT outperforms other wrinkle detection algorithms with an accuracy of 84% and 79% on the datasets of FORERUS and FORERET while HHF achieves 77% and 49%, respectively. Third, Multi-scale Wrinkle Patterns (MWP) is proposed as a novel feature representation for face age estimation using the wrinkle location, intensity and density. Fourth, Hybrid Aging Patterns (HAP) is proposed as a hybrid pattern for face age estimation using Facial Appearance Model (FAM) and MWP. Fifth, Multi-layer Age Regression (MAR) is proposed as a hierarchical model in complementary of FAM and MWP for face age estimation. For performance assessment of age estimation, four datasets namely FGNET, MORPH, FERET and PAL with different age ranges and sample sizes are used as benchmarks. Results showed that MAR achieves the lowest Mean Absolute Error (MAE) of 3.00 ($\pm$4.14) on FERET and HAP scores a comparable MAE of 3.02 ($\pm$2.92) as state of the art. In conclusion, wrinkles are important features and the uniqueness of this pattern should be considered in developing a robust model for face age estimation.

Acknowledgements

In preparation of this thesis, there are many people have contributed towards my understanding and thoughts. I would like to acknowledge, and also I wish to express my deep gratitude to my supervisors, Dr. Moi Hoon Yap and Dr. Nicholas Costen for guidance, valuable time, technical and friendly dealing through out my study. Without their continued support, guidance and interest, thesis would not have been the same as presented here.

I have received direct or indirect help and support from many personalities that motivated and enabled me to conduct this research. Thanks are due to Dr. Baihua Li, Dr. Kevin Tan, Dr. John Darby, Dr. Muhammad Salman Haleem, Dr. Xia Han, Dr. Jamie McPhee, Dr. Liangxiu Han, Dr. Yannis Goulermas, Prof. Tim Cootes, Mr. Adrian Keith Davison, Mr. Daniel Leightley, Mr. Brett Hewitt, Mr. Ezak Fadzrin, Ms. Gemma Stringer, Ms. Ruth Shepherd and Taiwan collaborators - Dr. Gee-Sern Hsu and Mr. Yi-Tseng Cheng for their help and valuable comments. In addition, I would like to present my sincere appreciation to SciEng Research Degrees and IT Services teams for their friendly support during my study.

I would like to thank the coders for their valuable time in manual annotation, Phillips et al. (2000) for FERET, Ricanek and Tesafaye (2006) for MORPH, Minear and Park (2004) for PAL, Zhou et al. (2013) for FACE++ detector, Savran et al. (2008) for Bosphorus, Stegmann (2003) for AAM-API application, Cootes et al. (2001) for *am_tools* application, and gratefully acknowledge the advise and code of KLBP provided by Ylioinas et al. (2013).

A great gratitude also goes to Manchester Metropolitan University, United Kingdom, in providing a PhD studentship for this work. I am thankful to Dr. Simeon Gill and Prof. Neil D. Reeves for the part-time jobs which relieved my financial burden.

Finally, I thank my family and friends for their constant encouragement and my wife for her patience and support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2D	-	2-Dimensional Space
3D	-	3-Dimensional Space
AAM	-	Active Appearance Model
AGES	-	Aging Pattern Subspace
BIF	-	Bio-inspired Features
CLM	-	Cula Method
CS	-	Cumulative Score
CNN	-	Convolutional Neural Networks
FACS	-	Facial Action Coding System
FAM	-	Facial Appearance Model
FERET	-	Face Recognition Technology Dataset
FGNET	-	Face and Gesture Recognition Research Network Dataset
FORERET	-	Forehead Images of FERET
FORERUS	-	Forehead Images of BosphoRUS
FRF	-	Frangi Filter
GPA	-	Generalised Procrustes Analysis
HAP	-	Hybrid Aging Patterns
HHF	-	Hybrid Hessian Filter

HLT	-	Hessian Line Tracking
JSI	-	Jaccard Similarity Index
LBP	-	Local Binary Patterns
KLBP	-	Kernel-based Local Binary Patterns
LOPO	-	Leave One Person Out
LOWEX	-	Local Wrinkle-based Extractor
MAE	-	Mean Absolute Error
MAR	-	Multi-layer Age Regression
MLP	-	Multi-layer Perception
MORPH	-	Craniofacial Longitudinal Morphological Face Dataset
MWP	-	Multi-scale Wrinkle Patterns
PAL	-	Park Aging Mind Laboratory Dataset
PAW	-	Piece-wise Affine Warping
PCA	-	Principal Component Analysis
ROI	-	Region of Interest
SMO	-	Sequential Minimal Optimization
STD	-	Standard Deviation
SVM	-	Support Vector Machine
SVR	-	Support Vector Regression

List of Symbols

(x, y)	-	a pixel coordinates in a 2D image
(x', y')	-	a transformed coordinates of (x, y)
(x_c, y_c)	-	a center pixel coordinates
$\hat{a}$	-	a specified age of a range
A	-	a predictive age of a face image
ACC	-	an accuracy of wrinkle detection
$\mathbb{A}$	-	a 2D logical wrinkle mask of the first coder annotation
$\hat{b}$	-	a bias of a function
$\mathbf{b}$	-	a set of scores or parameters of principal component
B	-	a ground truth age of a face image
$\mathbb{B}$	-	a 2D logical wrinkle mask of the second coder annotation
$\mathcal{B}$	-	a 2D binary mask
$\tilde{c}$	-	a center point of a region or several landmarks
$\hat{c}$	-	a predetermined threshold of LBP
$\mathbf{c}$	-	a set of combined shape and texture vectors
$C1$	-	a complex layer of BIF
$\mathcal{C}$	-	a 2D confidence array matrix of HLT
$\hat{d}$	-	a degree of polynomial kernel in SVR

$\tilde{d}$	-	a Hamming distance
D	-	a dimension level
$\mathcal{D}$	-	a distance error of GPA
$\mathbf{D}$	-	a set of images
$\mathbf{e}$	-	a set of eigenvectors of PCA
E	-	a squared error of gradient descent
$\mathcal{E}$	-	a 2D matrix of curvelinear likeliness measure
f	-	a function
$\mathbf{f}$	-	a set of features or input patterns
$\hat{f}$	-	a kernel density estimator of KLBP
$\mathfrak{f}$	-	a wrinkle intensity value of a region
$\bar{F}$	-	a mean value of the Gabor filtered values
$\mathbf{F}$	-	a continuous vector valued mapping function
FN	-	false negative
FP	-	false positive
$\mathbf{g}$	-	a set of wrinkle lengths
g_c	-	a grey value of the center pixel, (x_c, y_c)
g_p	-	a set of the neighbourhood pixels in grey-scale
$\mathfrak{g}$	-	a wrinkle density value of a region
G	-	a 2D matrix of image gradient
G_x	-	a 2D matrix of image gradient in x-direction
G_y	-	a 2D matrix of image gradient in y-direction
$\mathbf{G}$	-	a set of input patterns of a specified age group

$\mathbb{G}$	-	a D-dimension matrix of Gaussian kernel
$\hat{h}$	-	a bandwidth parameter of KLBP kernel
$\mathbf{h}$	-	a set of candidate pixels
$\mathcal{H}$	-	a Hessian filter
H	-	a LBP histogram
I	-	a 2D matrix of a directional gradient image
$\mathbf{I}$	-	a 2D matrix of an image
$\mathcal{J}$	-	a Jaccard index
$\hat{k}$	-	a momentum rate of MLP
$\tilde{k}$	-	a constant variable
$\hat{l}$	-	a labeled pixel of KLBP
L	-	a set of overlapping patches of KLBP
$\hat{m}_p$	-	a magnitude of local pixel difference
M	-	a vector function
$\mathcal{M}$	-	a 2D matrix of vertical direction
$\mathcal{N}$	-	a 2D matrix of horizontal direction
N	-	a total number of images
$\hat{N}$	-	a set of neighbourhoods of S1 units
O	-	a set of output units of MLP layer
p	-	a total number of HLT sampling points
$\mathbf{p}$	-	a background pixel located $\hat{s}$ pixel(s) away from the candidate pixel
$\check{p}$	-	an input layer of MLP
$\mathbb{P}$	-	a total landmarks of a shape

P	-	a total number of LBP sampling points
q	-	a level
$\check{q}$	-	a hidden layer of MLP
$\mathcal{Q}$	-	a square region
r	-	a radius
$\hat{r}$	-	an image ratio
$\tilde{r}$	-	an intensity value
$\check{r}$	-	an output layer of MLP
R	-	a circle radius of LBP
$\mathcal{R}$	-	a 2D matrix of division of the Hessian eigenvalues
s	-	a feature dimension
$\hat{s}$	-	a scale of filter
$\tilde{s}$	-	a scale factor of a 2D transformation
$\mathbf{s}$	-	a shape vector
S	-	a scale band of $S1$ units
$S1$	-	a simple layer of BIF
$\mathbf{S}$	-	a 2D matrix of PCA covariance
$\mathcal{S}$	-	a 2D matrix of addition of the Hessian eigenvalues
$\mathbb{S}$	-	a standard deviation
sds	-	a set of HLT seeds
t	-	a specified threshold
$\mathbf{t}$	-	a texture vector
$\mathbf{T}$	-	a 2D matrix with dimension of $s \times s$

$\mathbb{T}$	-	a thresholding function
$\tilde{T}$	-	a translation factor of a 2D transformation
TP	-	true positive
$\mathbf{u}$	-	a set of wrinkle amounts
U	-	a total number of possible intensity levels in an image
$\mathcal{V}$	-	a cross sectional profile
w	-	a weight of a function
$\hat{w}$	-	an Otsu weight
$\mathbf{w}$	-	a set of weights
W	-	a set of weights of MLP layer
$\mathbb{W}$	-	a set of wrinkle coordinates
$\mathbf{W}$	-	a set of eigenvectors of the covariance matrix of $\mathbf{x}$
$\mathcal{W}$	-	a 2D logical wrinkle mask of HHF
$\mathbf{x}$	-	an input vector
$\mathcal{X}$	-	a 2D logical wrinkle mask of HLT
$\mathbf{y}$	-	a projection vector of ρ in a subspace
$\mathcal{Y}$	-	a 2D matrix of wrinkles mask
$\mathbf{z}$	-	a winning pixel coordinate
Z	-	a set of pixels
$\mathcal{Z}$	-	a 2D logical mask of Region of Interest (ROI)
α	-	a direction at the maximum rate of change of f
β	-	a sensitivity parameter of the Hessian filter, $\mathcal{H}$
χ	-	a wavelength of Gabor filter in BIF

Δ	-	a Dirac delta function
ℓ	-	an iteration number of MLP
η	-	a learning rate of MLP
ε	-	an epsilon value of loss function in SVR
γ	-	an aspect ratio of Gabor filter in BIF
κ	-	a kernel function
λ	-	a set of eigenvalues of the Hessian filter, $\mathcal{H}$
$\nabla^2 f$	-	a 2D Laplacian function
μ	-	a mean value
∇f	-	a gradient function
Λ	-	a magnitude function
ν	-	a total number of regions in wrinkle template
ω	-	a constant of kernel function
Ω	-	a bias of MLP layer
ψ	-	a total number of features
Ψ	-	a normalisation factor
ϕ	-	a reconstruction vector of AGES
Φ	-	a 2D matrix of column eigenvectors
ρ	-	an aging pattern vector
σ	-	a standard deviation
τ	-	a total number of ratios
θ	-	an orientation or angle
υ	-	a normalised eigenvector

ϖ	-	a generalised error of MLP layer
φ	-	a set of PCA eigenvalues
ρ	-	a set of Lagrange multipliers
ς	-	a set of coefficient of polynomial kernel in SVR
$\bar{\varepsilon}$	-	a mean reconstruction error
ϑ	-	a scale of Hessian filter
ξ	-	a set of slack variables
Ξ	-	a total number of scales
ζ	-	an effective width of Gabor filter in BIF

List of Publications

This thesis is based on material from the following publications:

Journals

PUB1 - C.-C. Ng, M.H. Yap, N. Costen and B. Li, “Wrinkle Detection using Hessian Line Tracking,” *IEEE Access*, vol. 3, pp. 1079–1088, 2015.

PUB2 - C.-C. Ng, M.H. Yap, B. Li, Y.-T. Cheng, G.-S. Hsu and T. Cootes, “Multi-scale Wrinkle Patterns for Face Age Estimation,” *IEEE Transactions on Multimedia*, 2015, under preparation.

PUB3 - C.-C. Ng, M.H. Yap, Y.-T. Cheng and G.-S. Hsu, “Multi-layer Age Regression for Face Age Estimation,” *IEEE Transactions on Cybernetics*, 2015, under preparation.

Conference Papers

PUB4 - C.-C. Ng, M.H. Yap, N. Costen and B. Li, “An Investigation on Local Wrinkle-based Extractor of Age Estimation,” *The 9th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, vol. 1, pp. 675–681, 2014.

PUB5 - C.-C. Ng, M.H. Yap, N. Costen and B. Li, “Automatic Wrinkle Detection using Hybrid Hessian Filter, in *Computer Vision ACCV*, vol. 9005, pp. 609-622, 2015.

PUB6 - C.-C. Ng, M.H. Yap, N. Costen and B. Li, “Will Wrinkle Estimate the Face Age?,” *IEEE International Conference on Systems, Man, and Cybernetics (SMC2015)*, 2015, in press.

Poster

PUB7 - C.-C. Ng, M.H. Yap, N. Costen and B. Li, “Automated Facial Aging Estimation,” Research Student Development Day, Poster, 2013, Faculty of Science and Engineering, Manchester Metropolitan University.

Others

PUB8 - M.H. Yap, C.-C. Ng, G.-S. Hsu and Y.-T. Cheng, “Performance Comparison of Human and Machine in Face-based Age Estimation,” International Exchange Scheme which funded by The Royal Society, £11k±, 2015-2017.

PUB9 - M.H. Yap and C.-C. Ng, “Image Processing Tool for Wrinkles and Pores Detection,” Investment by Croda Europe Ltd, 2015, under negotiation.

PUB10 - C.-C. Ng, “Skills Development Award for MPhil/PhD Students,” awarded by School of Research, Enterprise and Innovation, Manchester Metropolitan University, £960, 2015.

PUB11 - C.-C. Ng, “Student Travel Grant,” awarded by The SMC Society for SMC2015, $500, 2015.

Part I: Introductory Chapters

Chapter 1

Introduction

This chapter introduces the research and terminology for face age estimation. A number of important terms are defined. It explains the problem statement, highlights the thesis contributions, and outlines the thesis structure.

1.1 Background

There is a huge number of research on the facial image analysis over last few decades, driven by applications in diverse areas such as biometric, instructional technology, marketing, mental health, and entertainment (Cohn and De la Torre, 2014). Face age estimation is a broad topic that has grown considerably in recent years (Ramanathan et al., 2009; Fu et al., 2010) and many methods have been proposed. The taxonomy of Fu et al. (2010) is adopted to define the area in which this thesis attempts to contribute. In general, *facial image analysis* interprets face images in terms of facial attributes such as identity, expression, gender, age, ethnicity and pose. These can be done in an efficient and automated way through a number of *computer vision* methods such as *face recognition*, *facial expression recognition* and *face gender identification*. In this thesis, face age is the research of interest. *Face age estimation* labels a face image automatically with the exact age (year) or the age group (year range) of the individual face. From literature review, there are two main approaches: *global* approach tends to describe an image as a whole and *local* approach represents *aging features* in a series of image patches. In this work novel methods based on local approach are developed, using a fixed template of face regions. In order to extract wrinkles accurately from a face image, two novel wrinkle detection methods are proposed, using a ridge valley pattern

and wrinkle connectivity in a local neighbourhood. In addition, a *hierarchical* approach of both global and local features is developed where the performance of age estimation is further improved by combining both features.

Automatic facial age estimation is an important, yet largely unsolved, challenging problem. This challenge can be attributed to (i) large intra-subject variations and (ii) large inter-subject similarity. Major intra-subject variations include craniofacial growth and skin texture, meanwhile, the inter-subject similarities are gender and race. In addition, 2-Dimensional Space (2D) face images contain huge variations in expression, pose and illumination in the uncontrolled environment. From these variations, local aging features such as wrinkles, pores and spots are now beginning to receive increasing attention in the facial aging community (Batool and Chellappa, 2014).

Conventionally, methods in age estimation are based on appearance features. Appearance features are computed or modelled for an entire face which comprises both shape and texture (Cootes et al., 2001). For facial aging research, many algorithms (Geng et al., 2007; Chen et al., 2010; Chang et al., 2011; Chao et al., 2013) deal with appearance model parameters in FAM (Cootes et al., 2001) and age manifolds (Fu and Huang, 2008; Guo et al., 2008a). In training stage, a FAM is constructed from a set of training images with manually annotated landmarks. It consists of a shape model, a texture model and a combined appearance model. FAM is a training model that belongs to Active Appearance Model (AAM) where AAM is a generative model that constructs the target subject's face by a set of hidden parameters (Cootes et al., 2001; Gao et al., 2010). Different faces at different ages can be generated under a similar subspace with varied parameters for controlling the shape and texture of the individual face. However, age progression modelling is highly complex due to large intra-subject variation and inter-subject similarity (Li et al., 2011). In addition, dimensional reduction using Principal Component Analysis (PCA) in FAM decreases aging representation such as wrinkles because the least important variance could be aging information rather than noise (Patterson et al., 2007; Choi et al., 2011).

This thesis investigates the potential use of wrinkle for face age estimation. In order to detect the facial wrinkle, accurate wrinkle detection is an important task in face analysis (Batool and Chellappa, 2014). Most of the latest works are based on clinical perspective (subjective assessment) instead of computer vision (objective assessment). Judgements are typically made on neutral-expression images and the well-known shortcomings of subjective assessment limit the scientific study of treatment and

environmental effects on skin aging. Clinician perspective focuses on the level of wrinkle severity which is assessed using either descriptive or photographically calibrated scales (Tsukahara et al., 2000), but in computer vision, the concern is on how a wrinkle is located correctly relative to the ground truth (Batool and Chellappa, 2014). Therefore, the wrinkle detection method may become an enhanced tool for age estimation. In addition, extracted wrinkles can be transformed into a robust representation where wrinkle location, intensity and density are taken into consideration. Based on wrinkle statistics, an efficient age estimation method can be developed with a lower feature dimension if compared to the state of the art.

The field of face age estimation and wrinkle detection has recently benefited from the introduction of freely available datasets that include ground truth and allow for quantitative evaluation and comparison of techniques. In particular, the datasets such as Face and Gesture Recognition Research Network Dataset (FGNET) (FGNET aging dataset, accessed on September 2012; Panis et al., 2015), Craniofacial Longitudinal Morphological Face Dataset (MORPH) (Ricanek and Tesafaye, 2006), Face Recognition Technology Dataset (FERET) (Phillips et al., 2000) and Park Aging Mind Laboratory Dataset (PAL) (Minear and Park, 2004), provided the actual age of each subject in all face images. Moreover, two datasets of wrinkle detection, Forehead Images of BosphoRUS (FORERUS) and Forehead Images of FERET (FORERET), are proposed, where they were created from Bosphorus (Savran et al., 2012a) and FERET, respectively. These datasets have converged the result presentation within the field, making possible the quantitative cross-comparison of a range of existing techniques. In the remainder of this thesis each contribution is thoroughly tested on either the wrinkle datasets or other freely available face age datasets, quantitative results are presented and comparisons drawn with existing state-of-the-art approaches.

This chapter is organised as follows. Section 1.1 and 1.2 describes the background and motivation of this work; Section 1.3 states the aim and objectives of this work; Section 1.4 lists the contributions made by the thesis compared to the state of the art; Finally, Section 1.5 provides an insight to the structure and presentation of the thesis.

1.2 Motivation

There are many popular real-world applications related to facial aging. Age estimation by machine is useful in applications where we do not need to specifically identify the user, but want to know his or her age when accessing restricted content (Yacenda et al.,

2013). When humans age, it would be interesting to characterise the progressive but subtle variations in facial appearance because it has many significant implications such as the following:

i. Age estimation is a type of soft biometrics that provides ancillary information of the users identity information (Jain et al., 2004). It can be used to complement existing biometric features, such as fingerprint and iris, to improve the performance of primary (hard) biometrics system (Fu et al., 2010).

ii. With growing needs to regulate the content viewed by minors on the Internet, age-specific human computer interaction systems have found greater relevance in recent years. Hence, methods that perform automatic age estimation are very critical in developing such applications (Lanitis et al., 2004; Ramanathan et al., 2009). Further, age-based image retrieval and video retrieval systems are bound to benefit from automatic age estimation systems (Ramanathan et al., 2009). For example, age-based indexing of face images in e-photo albums, where users could have the ability to retrieve their photographs by specifying a required age-range (Lanitis et al., 2004).

iii. In marketing, the most challenging part is to obtain and analyse enough personal information from all customer groups, which needs companies to establish long-term customer relationships and sustain a large amount of cost input (Lanitis et al., 2002). For example, retail shops might want to identify the age demographic of people who spend more time viewing certain advertisements; a mobile phone company wants to know which age group is more interested in their new product models showing in a public kiosk; a store display might show a business suit as an adult walks by or jeans as a teenager walks by (Fu et al., 2010).

1.3 Objectives

The primary aim of this work is to propose novel wrinkle-based features for face age estimation. In order to achieve the primary aim, the following objectives have been established:

i. To explore novel methods for wrinkle detection where the pattern is an arbitrary curve and randomly grow.

ii. To investigate the use of discriminative features such as shapes, wrinkles and appearances, for face age estimation.

iii. To evaluate the proposed methods using the benchmarks and human judgement.

1.4 Contributions

The fundamental contributions of this thesis are:

i. A novel wrinkle detection method, Hybrid Hessian Filter (HHF), is proposed to segment the face wrinkles using the directional gradient and a ridge-valley Gaussian kernel, as shown in Section 4.2 [PUB5, PUB9].

ii. A novel wrinkle detection method, Hessian Line Tracking (HLT), is proposed to explore the wrinkle connectivity using a cross-sectional profile in addition to the ridge and valley pattern, as presented in Section 4.3 [PUB1].

iii. A novel feature representation method, Multi-scale Wrinkle Patterns (MWP), is proposed for face age estimation which takes into consideration the wrinkle location, intensity and density, as described in Chapter 5 [PUB2, PUB3, PUB6, PUB8, PUB10, PUB11].

iv. A novel feature representation method, Hybrid Aging Patterns (HAP), and a novel age estimation method, Multi-layer Age Regression (MAR), are proposed for face age estimation in complementary of both FAM parameters and wrinkle patterns, as discussed in Chapter 6 [PUB2, PUB3, PUB8].

These contributions have resulted in three journals (one is accepted), three conference papers, two collaborations and two awards (see List of Publications).

1.5 Thesis Organisation

This thesis is organised into two parts as shown in Figure 1.1. The first part includes introductory chapters providing fundamental and background knowledge of the subject area and the state of the art for wrinkle detection and age estimation. The second part of this thesis includes three contributory chapters where two novel wrinkle detection methods, two novel feature represention and one age estimation method are

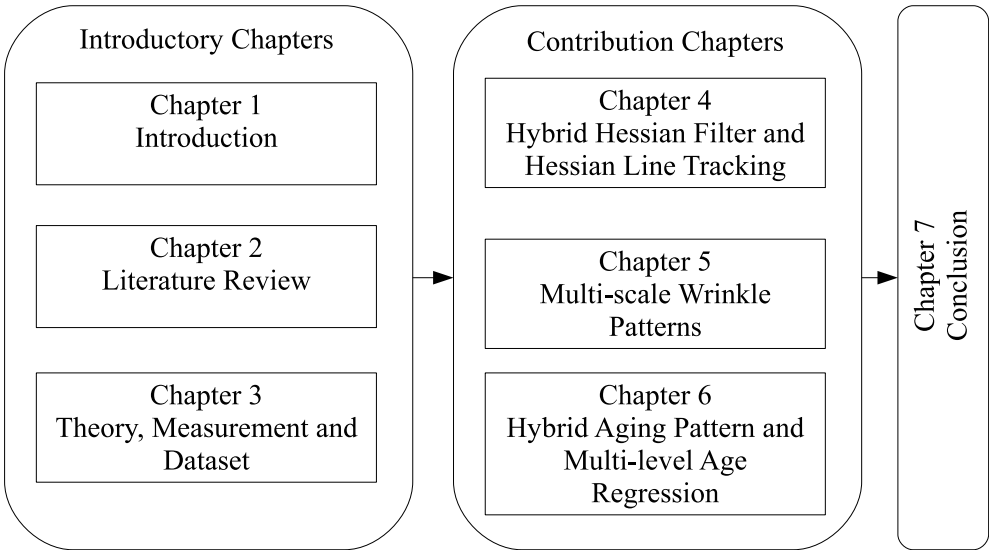


Figure 1.1: Thesis organisation. Novel methods for wrinkle detection are presented in chapter 4 and the main contributions for face age estimation are given in chapter 5 and 6.

discussed. Finally, this thesis concludes the works and an insight to future directions of research and development.

Part I: Introductory Chapters

Chapter 1 provides an overview of this thesis. It defines the problem domain, states the research motivation and specifies the thesis aims and objectives. It further summarises the proposed approaches and highlights the contributions made by the thesis to the state of the art. Finally it outlines the thesis organisation.

Chapter 2 presents fundamental knowledge about the conventional ways of wrinkle detection, the relationship between wrinkle and aging, the challenges in face age images and different age estimation methods. The advantages and disadvantages of each approach are compared and discussed.

Chapter 3 provides an overview of the research methodology. It further highlights methods being applied in both wrinkle detection and face age estimation.

Part II: Contribution Chapters

Chapter 4 proposes two novel wrinkle detection techniques. The first approach uses the directional gradient and a ridge-valley Gaussian kernel to determine the ridge and valley

pattern of wrinkle. The second approach extends the wrinkle exploration by looking into the wrinkle connectivity in between neighbourhood pixels using cross-sectional profile. Results of both approaches are compared with two benchmark datasets: FORERUS (Savran et al., 2008) and FORERET (Phillips et al., 2000).

Chapter 5 presents novel solution for effective face age estimation. This method is a local feature representation which comprises wrinkle location, intensity and density. Results are compared with the state-of-the-art methods such as FACE++ (Zhou et al., 2013), Bio-inspired Features (BIF) (Guo et al., 2009a) and Kernel-based Local Binary Patterns (KLBP) (Ylioinas et al., 2013) on FERET (Phillips et al., 2000).

Chapter 6 presents a detailed investigation on the use of global and local features for face age estimation. In particular the use of FAM and wrinkle patterns in a hybrid pattern and a hierarchical model. Four benchmark datasets: FGNET, MORPH, FERET and PAL are used for performance evaluation with the state-of-the-art methods. Results are critically analysed and conclusions made.

Finally, Chapter 7 concludes the thesis with a summary of contributions made by the thesis, limitations of the present context of research and an insight into future directions of research.

Chapter 2

Literature Review

This chapter presents a survey and general discussions that cover the state-of-the-art techniques for face age estimation. On top of that, it introduces the existing age image representation models and age estimation techniques. A performance comparison of existing techniques is also discussed.

2.1 Introduction

In this chapter an overview of face age estimation is presented. Due to the large volume of works in this area, the overview is not intended to be exhaustive but rather to define the areas in literature where the thesis attempts to contribute, and their wider context. A comprehensive review can be found in a number of review papers (Ramanathan et al., 2009; Fu et al., 2010; Panis et al., 2015).

In Section 2.2, popular applications of face age estimation are discussed. These include security control, information retrieval and marketing. It inspires a huge number of studies in facial aging analysis. Section 2.3 summarises the facial aging from birth to adulthood. It discusses the most significance changes of human faces in two stages: childhood and adulthood. Section 2.4 explains human perception and interpretation on facial aging. It covers the aesthetic views and related human experiments for face age estimation. An elaboration of the face wrinkle and its relationship with age is detailed in Section 2.5. Various methods of wrinkle detection and its performance are also included. To further understand the tasks, three concepts (Fu et al., 2010) about human age are defined as,

i. *Actual age*: The real age of an individual where the age is cumulated years

after birth.

ii. *Perceived age*: The individual age gauged by human subjects from a given face image.

iii. *Estimated age*: The individual age recognised by a computer machine from a given face image.

Following the taxonomy of Fu et al. (2010), the literature typically consists of two concatenated modules: feature representation and age estimation techniques. Feature representation attempts to construct the aging patterns from a set of images and it is described in Section 2.6. Age estimation technique uses a model to predict the exact age or age group of an unseen image as discussed in Section 2.7. Finally, this chapter concludes by highlighting the challenges of face age estimation and followed by the summary.

2.2 Real World Applications

As humans age, myriad changes occur chronically within the craniofacial complex. Notable soft tissue modifications may be seen across each decade of adult life that passes. As well, subtle hard tissue or bony changes slightly alter an overall shape of the human face, mainly in the dentoalveolar region (portion of the alveolar bone immediately about teeth). These age-related changes affect the accuracy and efficacy of face-related applications. Characterising the progressive, but subtle changes in facial appearances as human age has many significant implications. They are depicted as follows.

2.2.1 Security Control

Security control and surveillance monitoring issues are more and more crucial in our everyday life, especially when advanced technologies and explosive information become common to access and possess (Kloeppel, 2010). With the input of a monitoring camera, an age estimation system can warn or stop under-age drinkers from entering bars or off-license shops; prevent an under-age person from purchasing tobacco products from vending machines (Rhodes, 2009); refuse the youngster when he / she wants to try a roller coaster in an amusement park; and deny children access to adult web sites or restricted films (Guo et al., 2008a; Lanitis et al., 2004). In Japan, police found that a particular age group is more apt to money transfer fraud on ATM, in

which age estimation from surveillance monitoring can play an important role (Fu et al., 2010). Age estimation is a type of soft biometrics that provides ancillary information of the users (Jain et al., 2004). It can be used to complement the primary biometric features, such as face, fingerprint, iris, and hand geometry, to improve the performance of a primary (hard) biometric system (Fu et al., 2010). In a homeland security, face-based authentication systems that typically compare age-separated face images, are bound to benefit from facial aging models that extract age-invariant signatures (Lanitis et al., 2002; Ramanathan et al., 2009). Further, in the absence of such systems, authentication systems face the cumbersome task of periodically updating large face datasets with more recent face images (Ramanathan and Chellappa, 2006; Ramanathan et al., 2009).

2.2.2 Information Retrieval

With growing needs to regulate the content viewed by under-age on the Internet, age-specific Human Computer Interaction (HCI) systems have found greater relevance in recent years (Guo et al., 2008a). Hence, methods that perform automatic age estimation are very critical for developing such applications. Further, age-based image retrieval and video retrieval systems are bound to benefit from automatic age estimation systems (Ramanathan et al., 2009). For example, automatic album management of consumer photographs (Das and Loui, 2003) and age-based indexing of face images in e-photo albums, where users could have the ability to retrieve their photographs by specifying a required age-range (Lanitis et al., 2004; Gallagher and Chen, 2009). In HCI (Ricanek Jr et al., 2009), computer interface may use a face image to dynamically setup a workspace according to the user's age group; some companies are developing a software to determine Closed-Caption TeleVision (CCTV) feeds from stores to compile demographic information on customer purchasing behaviour. Age estimation software can also be used in health care systems, such as robotic nurse and intelligent intensive care unit, for customised services. For example, a personalised avatar will be selected automatically from the custom-built avatar dataset to interact with patients from different age groups with particular preferences (Fu et al., 2010). In cosmetology, dermatologists are interested in locating and removing the face wrinkles in order to achieve skin rejuvenation and look younger (Cula et al., 2013; Batool and Chellappa, 2014). Currently, most practices require expert intervention to manually locate and annotate the wrinkles before and after of a treatment. Such process is time-consuming and prone to human error. Therefore, an automatic wrinkle quantification system will

aid to human decision in cosmetology.

2.2.3 Marketing

Customer Relationship Management (CRM) is a management strategy to use information technology and multimedia interaction tools for effectively managing differentiated relationships with all customers and communicating with them individually (Kloeppel, 2010). Since different groups of customers have very different consuming habits, preferences, responsiveness, and expectation to marketing, companies can gain more profits by acknowledging this fact, responding directly to all customers' specific needs, and providing customised products or services. The most challenging part hereby is to obtain and analyse enough personal information from all customer groups, which needs companies to establish long-term customer relationships and sustain a large amount of cost input (Lanitis et al., 2002). For example, a fast food shop owner might want to know what percentage of each age group prefers and purchases what kind of sandwiches; determining the age demographic of people who spend more time viewing certain advertisements; a mobile phone company wants to know which age group is more interested in their new product models showing in a public kiosk; a store display might show a business suit as an adult walks by or jeans as a teenager walks by. Obviously, it is almost impossible to realise those due to privacy issues. However, with the help of a computer-based automatic age estimation system, a camera snapping photos of customers could collect demographic data by capturing face images of customers, automatically labelling age groups and deleting the processed images immediately. All of these can be done by a machine without violating anyone's privacy (Fu et al., 2010).

2.3 Natural Aging Process

Human face aging is generally a slow and irreversible process, even though some retinoic acid (e.g., tretinoin treatment[1]) may slightly reverse minor photo-aging effects (Zimbler et al., 2001). Although people are aging differently and aging shows different forms in different ages, there are still some general changes and resemblances can be described (Gilchrest, 1996; Albert et al., 2007; Farkas et al., 2013). For instance, craniofacial shape and skin texture. Figure 2.1 shows images of a male subject facial aging. It is noticed that the shape of the older face is slightly larger than the younger face and

[1] Tretinoin treatment modifies fine wrinkles and certain other features of human skin damaged by exposure to the sun or so-called photo-damage (Griffiths et al., 1993).

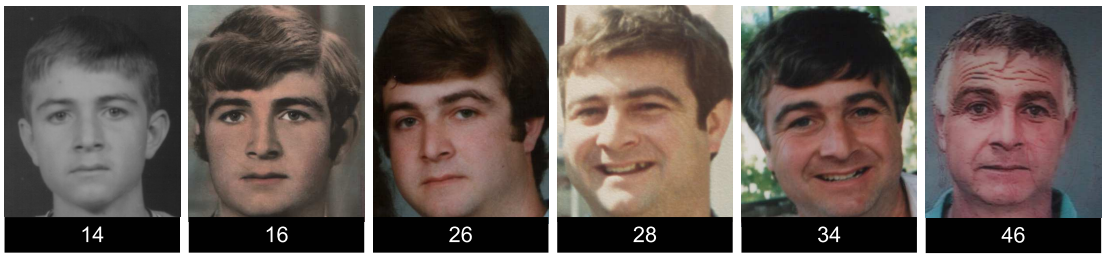


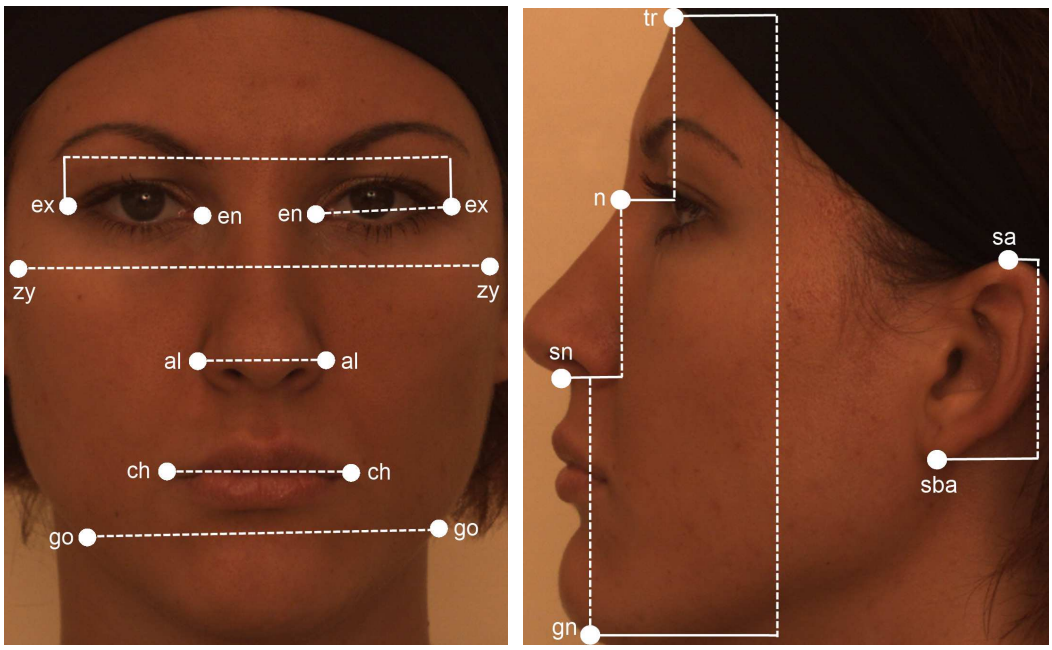
Figure 2.1: Facial aging in a male subject. The number indicates the age when the image was taken. Note that this figure was redrawn from FGNET.

skin patterns around forehead, eyes and nose are different as well. From a biological or anthropometric point of view, there are roughly two stages of the human life that are quite different in face growth, development, and aging forms (Mark et al., 1980). The following sections detail the two stages: childhood aging and adulthood aging.

2.3.1 Childhood Aging

During the early growth and development of the face, from birth to adulthood, the greatest change is the craniofacial growth (shape change) (Zebrowitz and Montepare, 1992; Corel Corp, 2012). Overall, the face size is getting larger gradually during the craniofacial growth (Meyer-Marcotty et al., 2013). With the growth of the cranium, forehead slopes back, shrinks, and releases spaces on the surface of the cranium, while the facial features, such as eyes, nose, ears, and mouth, expand their areas and tend to cover these interstitial spaces. Cheeks extend to larger areas and the chin becomes more protrusive. The facial skin relatively does not change too much compared with the craniofacial growth. But facial hairs, such as a moustache, may become dense and even bushy. Skin colour may change a bit. Pores may open and enlarge (Corel Corp, 2012).

Age-based anthropometric measurements have been shown to play a critical role in measuring sizes and proportions on human faces during formative years. Similar anthropometric data collected on adult faces are not available easily (Fu et al., 2010). These measurements are crucial in developing a facial aging model. Such studies provide a quantitative description of the craniofacial complex at different ages and hence, provide a plethora of options for learning-based approaches to be adopted to characterise facial aging. Face anthropometric studies give dense measurements taken between key landmarks on human faces across different ages and have played a critical role in surgical procedure implemented on the child face (Ramanathan et al., 2009).



(a) Frontal face (b) Profile face

Figure 2.2: Face ratio measurements. For frontal face, measurements of orbits: en-en (intercanthal width), ex-ex (biocular width), en-ex (eye fissure length); face: zy-zy (face width), go-go (mandible width); nose: al-al (morphological nose width); labio-oral region: ch-ch (mouth width). For profile face, head: tr-n (forehead height); face: tr-gn (physiognomical face height), n-gn (morphological face height), sn-gn (lower face height); nose: n-sn (nose height), in: (inclination of the nasal bridge); ear: sa-sba (length of the auricle) (Farkas et al., 2005). Note that this figure is drawn from Bosphorus (Savran et al., 2008).

craniofacial aging data are increasingly important where automated face recognition technology and computer 3-Dimensional Space (3D) modelling of adult faces for use in adult facial age progression are concerned (Albert and Ricanek Jr, 2008). Earlier efforts, such as FERET attest to the potential benefits of a cross-disciplinary interface between forensic anthropology and computer science with forensic science in general (Phillips et al., 2000).

Farkas (1994) provides a comprehensive overview of face anthropometry and its significant applications. He defines face anthropometry in terms of measurements taken from 57 carefully selected landmarks on human faces spread across six regions in the craniofacial complex (head, face, orbits, nose, lips and mouth, ear). The facial measurements are of three kinds: (i) projective measurements (shortest distance between two landmarks); (ii) tangential measurements (distance between two landmarks

measured along the skin surface); (iii) angular measurement (Gallagher et al., 2005; Ramanathan et al., 2009). Farkas et al. (2005) investigated the international anthropometric of facial morphology in various ethnic groups or races. They have selected 14 anthropometric measurements to determine the morphologic characteristics of the craniofacial complex. Ten of them were the neoclassical facial canons of the Renaissance used by by classic facial artists Leonardo da Vinci and Albrecht Dürer[2] (tr-n, zy-zy, sn-gn, en-en, en-ex, n-sn, al-al, ch-ch, sa-sba and the inclination of the nasal bridge), complemented by three facial (tr-gn, n-gn and go-go) and one orbital measurement (ex-ex), as shown in Figure 2.2. They attempt to establish the main facial characteristics of 25 national groups belonging to four races and thereby differentiate the morphological complex between one group and another.

Facial anthropometry plays an important role in ergonomic applications (Fang and Fang, 2011). Most ergonomically designed products depend on stable and accurate human body measurement data. Head and facial anthropometric dimensions provide detailed information on head and facial surfaces to develop well-fitting, comfortable and functionally effective facial masks, helmets or customised products. Fang and Fang (2011) developed an automated geometric-based facial feature extraction method to identify head and facial features. They managed to locate 17 digital length measurements and 5 digital tape measurements on the head and face. They claimed that nearly all length measurement data and tape measurement data meet the 5mm measuring error standard.

Kwon and da Vitoria Lobo (1999) presented a theory and practical computations for visual age classification from facial images. The computations are based on craniofacial development theory and skin wrinkle analysis. For primary features, ratios that distinguish babies from young adults and seniors are computed around the eyes, nose, mouth, chin, virtual-top of the head and the sides of face. In total, six ratios were used to build a classifier. They stated that the first ratio, the distance between eyes over the distance between eye and nose, is the most promising feature because it is not affected by the facial expressions or facial motions. However, imprecise localisation of landmarks might affect the results. Moreover, ratio 6, the distance between eye and chin over the

[2]The human sculptures produced in ancient Greece, notably in the 4^{th} to 5^{th} centuries Before Common Era (BCE), were derived from proportions that followed established rules or 'canons'. These rules were incorporated to the 'neoclassical canons' for the human face by Renaissance artists that included Leonardo da Vinci and Albrecht Dürer. These canons were based on the assumption that certain fixed ratios existed between different parameters of a harmonious face. Subsequently, these canons were adapted by medical artists, anatomists and aesthetic surgeons and continue to be used to this day (Le et al., 2002; Jayaraman and Kulanthaivel, 2012).

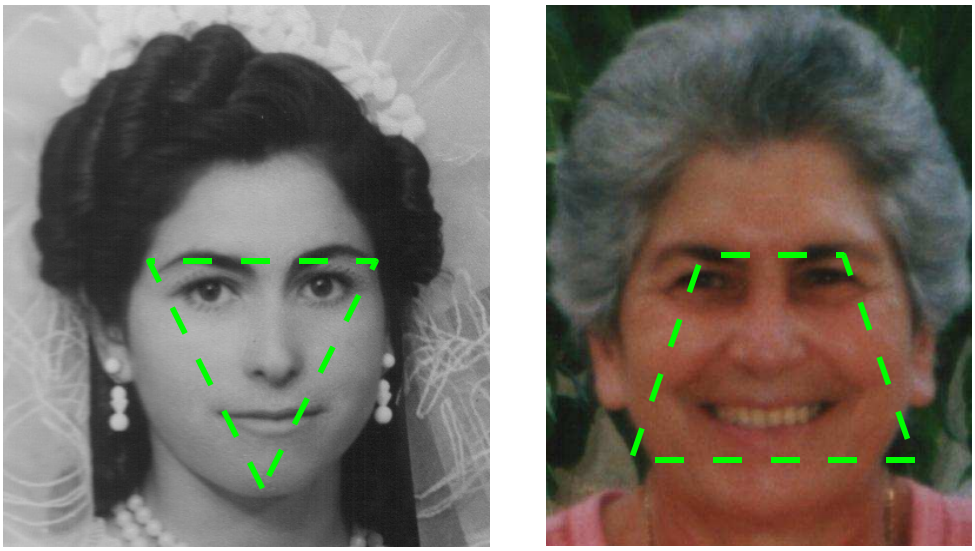


Figure 2.3: Face shape changes. Left is female subject of 18 years old and her face shape looks like a upside-down triangle. Right is the same subject of 61 years old and her face shape looks like to a trapezoid. Note that this figure was redrawn from FGNET.

distance between virtual-top of the head and chin, is the most reliable feature but it suffers from errors in estimating the virtual-top of the head. In secondary feature analysis, a wrinkle geography map is used to guide the detection and measurement of wrinkles. A combination rule for the ratios and the wrinkle index thus permits categorisation of a face into one of three classes: babies, young adults and senior adults. However, there was no validation against the extracted wrinkles where it could be noise like shading or lines of interest. Overall, they have demonstrated an interesting pioneer work for face age estimation where computing ratios and detecting the presence of wrinkles can yield age categorisation.

2.3.2 Adulthood Aging

During adult aging, from adulthood to old age, the most perceptible change becomes skin aging (texture change). The shape change still continues, but less dramatically, mostly due to typical patterns in skin and tissue. Originally shown in Gonzalez-Ulloa and FLORES (1965) and also in Stone (2012), as the face matures and ages with loss of collagen beneath skin as well as gravity effects, skin becomes thinner, darker, less elastic, and more leathery (Corel Corp, 2012). A set of wrinkles and blemishes due to biologic aging gradually appear. Dynamic wrinkles and folds due to muscle motion

Table 2.1: Probable soft tissue or facial appearance effect. This table is derived from article of Albert et al. (2007).

Age range	Description
20-30	Upper eyelid drooping begins. Eyes appear smaller. Nasolabial lines begin to form. Lateral orbital lines begin to form. Upper lip retraction begins in women.
30-40	Circumoral striae begin to form. Lines begin to form from lateral edges of nose to lateral edges of mouth. Upper lip thickness decreasing.
40-50	Facial lines and folds continue to increase in depth. Nose and chin positioning affected as dental arch lengths decrease. Most profound morphological changes of the head, face, and neck are evident.
50-60	Facial lines and folds continue to increase in depth. Protuberance of nose and ears due to greater craniofacial convexity.
$>$ 60	Protuberance of nose and ears continues. Concave appearance in cheek hollows. Diminished jaws.

become more distinct. In the areas of deeper attachment, such as cheeks, eyelids, chin, and nose, elasticity of muscles and soft tissues get weak and fat continues depositing. In other areas, fat may atrophy or be absorbed. These changes lead to the downward descent or sagging of skin, such as double chin, dropping cheek, and lower eyelid bags. Ezure et al. (2012) stated that sagging is correlated with various age-related facial morphological changes, and its severity is influenced by subcutaneous tissues such as subcutaneous adipose tissue and mimetic tissue. Although the craniofacial growth is not dramatic during this aging period, the facial geometry change is still evident from 30 to 80 years, especially on the female faces. Faces change from a U-shaped or upside-down triangle shape to a trapezoid or rectangle, as shown in Figure 2.3 (Gonzalez-Ulloa and FLORES, 1965; Stone, 2012; Michaud et al., 2015). The bony framework underneath skin may also deteriorate to accelerate the development of skin aging, such as wrinkles, creases, and droops. In addition, face aging during this age period may cause the loss of flexible control of facial muscles and consequently the facial movements and behaviours may also change unintentionally (Corel Corp, 2012).

Table 2.1 shows the adult hard and soft tissue age-related changes (Albert et al., 2007). Generally, normal aging of the facial soft tissues begins in the 20s with fine facial lines appearing horizontally across forehead, vertical lines emerging between the eyebrows, and faint lines developing around the outer corners of the eyes (i.e., crows' feet). These faint lines continue to deepen in the 30s and beyond. In the 30s, horizontal lines appear at the top of nose between the eyes, and nasolabial lines begin to form

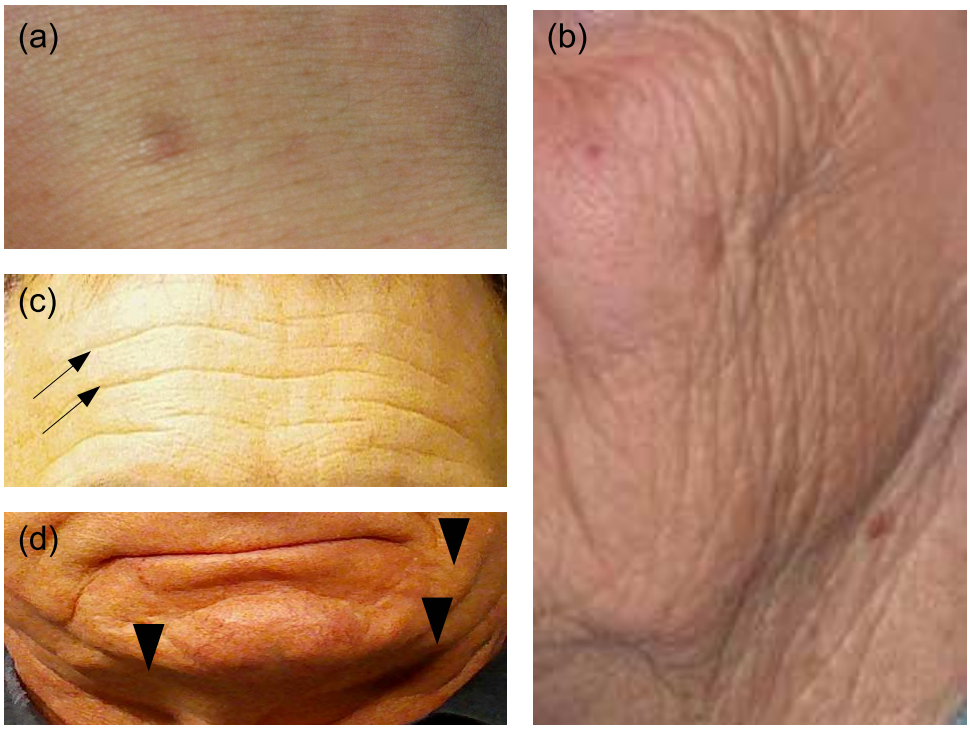


Figure 2.4: Wrinkle types. (a) Type 1 wrinkles due to atrophy. (b) Type 2 wrinkles due to solar elastosis. (c) Type 3 wrinkles due to expression lines on forehead. (d) Type 4 wrinkles due to gravity near the lower mandible.

(lines or creases from the outer corners of the base of nose diagonally down to the sides of mouth). During the 40s, many soft tissue age changes occur. The eyebrows may drop, and the upper eyelid may droop and obscure the upper eyelid crease around the lateral margins. The lower jaw becomes less firm, circumoral striae develop (lines around mouth), and the lips can thin out. In the 50s, more prominent soft tissue age changes become evident. The groove or crease below the eyes may evolve into a pouch of skin. More tissue may develop on the upper eyelid where drooping further obscures the crease of the upper eyelid. Existing creases, lines, and grooves continue to deepen. The lips become thinner. If tooth loss occurs, the cheeks can appear hollow. Jowls and a double chin may develop. The neck wrinkles can become apparent. In the 60s, the aging features of the previous decade become more pronounced. During this decade, nose and ears appear longer, jaw has lost its firmness, and the soft tissues of neck sag. By the 70s and beyond, creases, lines, and grooves are all exaggerated. There is a loss of skin elasticity, and significant skin sagging. In summary, key facial age changes of adult are noted to appear in the upper face before the lower face (Taister et al., 2000).

Table 2.2: Terms that describe the skin marks. This table is derived from the book of Ekman et al. (2002).

Terms	Description
Line	A surface line with no depth, usually quite fine in terms of width. Some faces may show permanent surface lines; these may deepen to a wrinkle when certain actions occur.
Wrinkle	A line which has some depth and often has more width than a surface line; some faces may show no permanent wrinkles, but they will appear with certain actions. Other faces may show permanent wrinkles but they will deepen with certain actions.
Furrow	This term describes a place on face where certain wrinkles may appear: lower eyelid furrow; infraorbital furrow; or nasolabial furrow. In some faces there is no wrinkle or line in such a facial location until there is action. Some faces show a line permanently, but it will deepen to a wrinkle with certain actions, e.g., some faces show a permanent line in the lower eyelid furrow, infraorbital furrow or nasolabial furrow, which will deepen with an action.

2.3.2.1 Wrinkle

Face is part of the body most visible to others, and its wrinkles are a hallmark of aging (Cula et al., 2013; Hayashi et al., 2003; Li et al., 2006). Wrinkle has been widely investigated for face analysis especially for face age estimation (Kwon and da Vitoria Lobo, 1999; Takimoto et al., 2007; Dehshibi and Bastanfard, 2010; Choi et al., 2010; Jeong et al., 2014; Jana et al., 2015). Usually, wrinkles are surface skin features which appear in images as line segments rather than continuous lines, reasonably similar with those in fingerprints and palm prints (Kong et al., 2009). Table 2.2 presents the terms that describe the skin marks, as defined by Ekman et al. (2002). A surface line with no depth, usually quite fine in terms of width; A wrinkle which has some depth and often has more width than a surface line; furrow describes a place on face where certain wrinkles may appear.

According to Piérard et al. (2003), there are four basic types of wrinkles can be distinguished according to their histological aspects and pathogenesis (as shown in Figure 2.4). Type 1 wrinkles (atrophic) are fine, almost parallel each other, and they vanish when skin is put under transversal tension. They are due to the atrophy of the collagen bundles both in the reticular dermis and the hypodermal connective tissue strands. Type 2 wrinkles (elastotic) become progressively permanent lines, developed on certain sun-exposed skin areas where solar elastosis is hypertrophic, compact and

Figure 2.5: Facial hyper-pigmentation.

largely overcompensating in volume the collagen atrophy. Type 3 wrinkles (expressional) become permanent lines, always orientated in a stereotyped pattern according to the forces imposed by facial muscles. Type 4 wrinkles (gravitational) result from gravitational forces inducing folding and sagging of skin which has lost its turgidity.

Ramanathan and Chellappa (2008) proposed a two-step approach towards modelling facial aging in adults. It comprises a shape and a texture variation models. The texture variation model was designed specifically to characterise facial wrinkles in predesignated facial regions. They identified four facial regions which tend to have a high propensity towards developing wrinkles, namely forehead region, the eye-burrow region, the nasal region and the lower chin region. Then, they categorised the region-based facial wrinkle variations across age transformation into one of the three classes: subtle wrinkle change, moderate wrinkle change and strong wrinkle change. This classification is performed by studying the pixel-based differences in gradient magnitudes and orientations. They found that the rates at which facial wrinkles are manifested on individuals across different ages is often subjective. They claimed that the proposed texture variation model can be used to predict the different wrinkle patterns that could have been observed on the individuals.

2.3.2.2 Pigmentation

Hyper-pigmentation is the darkening of an area of skin or nails caused by increased melanin, often due to sun damage, as shown in Figure 2.5 (Hakozaki et al., 2010; Perez-Bernal et al., 2000). Pigmentary changes are a hallmark of aging in all races and ethnicities, and are of growing concern in a world with an aging demographic. Solar lentigines or age spots remain a major concern amongst individuals globally, but are of particular concern in Asian populations (Nip et al., 2010; Chang et al., 2013). The continually increasing understanding of the pigmentation process and the underlying problems in hyper-pigmentation conditions provide basis for establishing targets against which to screen new compounds to identify those that may be effective pigmentation

control agents (Hakozaki et al., 2010).

2.3.3 Facial Aging Factor

The real-world applications are very rich and attractive, existing facts and attitudes from the perception field reveal the difficulties and challenges of automatic age synthesis and estimation by computer (Guo et al., 2008a,b). It is imperative to understand that patterns, rates and characteristics of aging may be different due to culture and lifestyle (environment, gender) and biology (sex, ancestry or genetics, trauma and disease), as well as idiosyncratic features (unique to an individual). Moreover, it varies among people at any given age, or change over time in any one person, are not completely understood (Farkas and Schendel, 1995; Alley, 1988; Albert et al., 2007). In particular, Stone (2012) stated that aging can be accelerated by smoking, genetic predisposition, emotional stress, disease processes, dramatic changes in weight, and exposure to extreme climates. Following are the discussion of two major factors of aging: intrinsic and extrinsic. Intrinsic aging is caused by internal biological factors whereas extrinsic aging is caused by environmental influences (Albert et al., 2011).

2.3.3.1 Intrinsic Factor

As investigated by Berry and McArthur (1986) using an ecological approach, age-related variations in craniofacial growth play an important role in social perception. Facial characteristics may affect impressions if they typically reveal psychological attributes. According to Albert et al. (2007), innate adult facial aging results from: (i) changes in the bony support structure of face and subsequent changes in musculature; (ii) gravity; (iii) hyper-dynamic facial expressions. Other changes related to advancing age include loss of tissue elasticity and facial volume and alterations in skin texture (Scali and Iorio, 2014). Although the manner of aging can be highly unpredictable, there is a sequence of changes that appears to adhere to a basic progressive pattern across time. The rate of aging, however, varies across adult ages. There is a strong likelihood that certain adult age spans yield a greater magnitude of age-related changes than other time (Albert et al., 2007). For example, there may be minimal changes seen from ages 20-30, while by contrast there may be noticeable changes from ages 40-50. Attempts have been made to effectively express the sequence or stages of aging (Taister et al., 2000). The general sequence of aging is somewhat predictable, with an understanding that the pace, exact timing, and extent of any one aging feature may be

unpredictable or unnoticeable due to the normal human variation. The appearance of facial lines and wrinkling (Taylor, 2000), and their appearance in response to loss of muscle, bony changes and the natural passage of time have been well documented by Zimbler et al. (2001).

2.3.3.2 Extrinsic Factor

Suppa et al. (2011) have proven that smoking, obesity and excessive sun exposure increase the appearance of aging of the skin, especially in the periorbital region. They also produced some evidence that sunscreen use is protective for age-related cutaneous damage. The effects of Ultraviolet Radiation (UVR), usually through exposure to sunlight, may cause solar aging, which is another strong cause for advanced signs of face aging. Jdid et al. (2012) found that environmental conditions especially high ambient UVR can aggravate age-related skin features, even in women with dark skin. Aside from photo-aging which mainly affects the skin and exaggerates normally occurring age changes, general exposure to the elements such as wind and arid air can influence facial aging as well (Taister et al., 2000). Smoking disrupts the microvasculature of skin, which in turn affects the integrity of elastin and collagen. Dehydration can result in a greyish cast to a complexion. Lines around the mouth, circumoral striae, may become evident. In addition, exposure to the elements such as drug use and stress-related activities decrease in sleep are deemed influential in facial aging (Taister et al., 2000). Despite these possible effects, it is thought that photo-aging results in more severe wrinkling than smoking.

Boissieux et al. (2000) also pointed out the challenge to accurately modelling aging skin because of the different aesthetic values and focuses of attention people have. Thus, face aging is uncontrollable and personalised, which makes the task of accurately capturing aging variations more difficult (Ramanathan and Chellappa, 2006; Fu and Huang, 2008; Geng et al., 2007). Further, facial appearances get altered with increase in age due to factors such as changes in hairstyle. Hence, the proposed characterisation of facial aging effects should ideally account for multiple factors that induce aging effects (Ramanathan et al., 2009). From a technical point of view, males and females may have different face aging patterns displayed in images due to the different extent in using makeups and accessories (Fu et al., 2007; Guo et al., 2008b). Many female face images may potentially show younger appearances.

Rexbye et al. (2006) investigated the influence of environmental factors on facial aging. They conducted a population-based survey where participants (aging Danish

twins) provided information on a wide range of exposures and health indicators. They were also asked to have a face photograph taken. A total of 1826 elderly (aged 70+) twins who had a high-quality face photograph taken. Ten nurses assessed the visual age of each twin from the face photographs. The mean of the nurses' age estimates for each twin was used as the twin's perceived age. Multivariate linear regression and intra-pair comparison (for intact pairs) were used for analyses. Results confirmed that an influence of sun exposures, smoking and Body Mass Index (BMI) on facial aging. Their study also indicated that high social status, low depression score and being married are associated with a younger look. However, the strength of association varies between genders.

2.4 Human Perception

Changes in skin condition throughout a person's life are indicators of the irreversible aging processes that occur in the human body. The texture and appearance of skin gradually shifts from a smooth, translucent and uniformity youth to a rougher and more unevenly pigmented skin. Facial wrinkling, sagging, and pore enlargement, all generally irreversible without therapeutic intervention, mark the passage of time or, perhaps more accurately, the accumulation of photo-damage. Though these aging phenomena are important social and self-esteem determinants, only recently have the biochemical and physiological changes associated with them begun to be investigated (Gartstein and Shaya, 1996).

From the aesthetic views, face aging can be explained in a different way. Assuming that faces are points in a high-dimensional space (Valentine, 1991), the distinctiveness indicates the distance between any two faces. The attractiveness is negatively correlated with distinctiveness because less distinctive and more typical is seen as more attractive, presented by Rhodes and Tremewan (1996). Following this theory and concepts, Deffenbacher et al. (1998) found that face aging is actually related to attractiveness and distinctiveness. Their experiments are based on a 3D face model built on European-descent male and female young adults. When the facial caricature technique is performed, they found that young faces are more attractive but less distinctive than old faces. When the degree of facial caricature increases, the distinctiveness increases while the attractiveness decreases linearly and, at the same time, faces are looking older. Langlois and Roggman (1990) supported the 'Averageness Hypothesis' of attractiveness, which indicates that an average face is the most attractive face. Rexbye

and Povlsen (2007) explored visual age and how signs of age are read and interpreted in consumer culture. The main indicators of age are biological: skin, eyes and hair colour, but supplemented by vigour, style and grooming. From the in-depth interviews, they found that activity and dress code are key themes in the reading and interpretation of visual aging signs. Braun et al. (2001) believed that the merging of several faces can remove unpleasant asymmetries, irregularities, wrinkles, and pimples so that the skin looks perfectly smooth, clear, and younger. Moreover, baby-face makes faces look younger and can enhance attractiveness. This indicates that facial components such as wrinkles and pigments would affect the face age. Alley (1981) also found that the morphological changes of face aging using shape transformations can induce the decrease of perceived facial cuteness. Based on the facial aging perception, O'Toole et al. (1997) applied a facial caricature algorithm to a 3D face model for face aging synthesis. They support the conclusion that an average face is younger and more attractive than an individual face. They suggest that the distinctiveness of an individual face, defined as its distance from the average face in a 3D space, embodies the information highly related to face aging. So, in perception and aesthetics, facial aging is often explained and implemented by the concepts of attractiveness and distinctiveness with explicit caricatures or transformations from the empirical knowledge of biology anthropometry (face growth, development, and aging forms).

Mark et al. (1980) investigated the changes in the shape of a human head and the development of facial wrinkles as potential sources of information about age level. In the first experiment, participants estimated the ages of faces that had been produced by systematically manipulating characteristic head shapes and levels of wrinkles associated with ages 15, 30, 50 and 70 years. The results indicated that participants used both sources of craniofacial change in making age estimates, but the effects of either source of change on perceived age depended upon the level of the other source of change. In the second experiment, the participant ratings of apparent conflict between levels of head shape and wrinkles further substantiated the conclusion that the participants are sensitive to the coordination of products of the two sources of change. These findings suggest that the information specifying perceived age level is a complex relationship among different types of craniofacial change.

Tsukahara et al. (2000) reported a photographic scale for the assessment of human wrinkles. A five-grade photo scale of wrinkles at eye corners was developed using photos obtained from 411 female participants aged 17 to 83 years. Based on this photo scale, scorings of all photos were performed by two specialists, and a standard photo

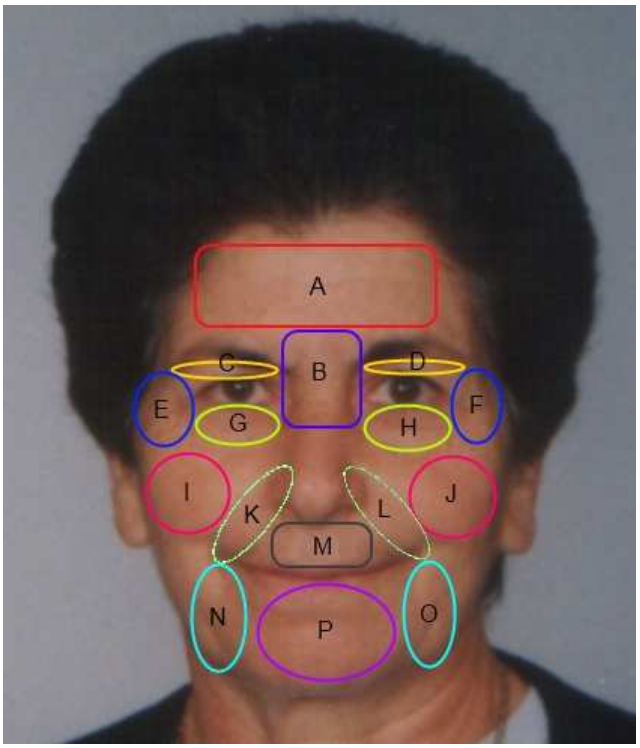


Figure 2.6: Wrinkle geography. Each region is represented by different colours and letters. Note that this image was redrawn from FGNET.

for each wrinkle grade was obtained. In addition, another scoring was performed by five observers in order to evaluate the influence of inter-observer differences in grading criteria. The agreement between the grade standard and the score given by the general observers was evaluated by calculating the chance-corrected coefficient, i.e., the kappa value. The degree of agreement for the five-grade scale was more consistent than that for the nine-grade scale system, in which the kappa values were 0.499 and 0.396, respectively. When scoring was performed using the five-grade photo scale for the eye corners after 15 participants used a wrinkle-improving agent, a significant reduction of the wrinkle scores was confirmed. They also found that wrinkles more rapidly develop in the order of: eye corners $>$ forehead $>$ lower eyelids $>$ upper eyelids $>$ mouth angles $>$ nasolabial grooves $>$ cheeks $>$ glabella. Wrinkles around the eyes that are highly correlated with age also showed a high wrinkle development rate, same finding was confirmed by Akazaki et al. (2002); Merinville et al. (2015). Taken together, their findings indicated that the the five-grade wrinkle photo scale is valid and useful to assess the degree of facial wrinkles. Figure 2.6 illustrates the wrinkle geography map which has been used in facial aging analysis (Kwon and da Vitoria Lobo, 1994; Kwon and da Vitoria Lobo, 1999; Takimoto et al., 2007; Bae et al., 2008; Suo et al., 2010,

2012; Choi et al., 2011). The regions are forehead (A), frown / glabella (B), upper eyelid (C and D), crows' feet (E and F), lower eyelid (G and H), cheeks (I and J), nasolabial segment (K and L), upper lid (M), Marionette (N and O), and lower lid (P) [3]. Upper and lower lid are less common in research but wrinkles in these areas are very obvious especially on those who smoke a lot.

Aznar-Casanova et al. (2010) studied the influence of wrinkles on facial age judgements. The qualitative (type of wrinkle) and quantitative (density of wrinkles and depth of furrows) contributions of wrinkles were analysed. Their results indicated that the greater the number of wrinkles and the depth of furrows, the older a face was rated. In addition, the quantitative component (density of wrinkles) had a stronger effect on facial age judgements than the qualitative element (type of wrinkle). Their findings are similar to Mark et al. (1980) where faces with deep furrows were considered older than faces with shallow furrows. Moreover, their experiments showed that the perceived facial age was more strongly influenced by the gender of face; the youngest groups of participants (undergraduate students) attributed lower ages to faces than senior adults; wrinkles contributed to making male faces appear older than female faces.

Recently, the work of face aging has begun to consider diurnal changes in individual faces (Albert et al., 2011). Diurnal changes are those alterations in a person's face that occur over the course of a single day, from morning until evening. In a study by Tsukahara et al. (2004), image analysis of 38 Japanese females and males with a mean age of 34 years showed significant intensification of wrinkling in all areas of face evaluated in the afternoon when compare with the morning. The areas include forehead, corners of the eyes and nasolabial grooves. Wrinkles were likely exacerbated later in the day due to a decrease in the face swelling thought to occur in the morning since during sleep gravity is not pulling fluids downward; wrinkles may be plumped out in the morning. As the day progresses and a person is vertical, gravity may pull fluids downward from face and into the appendages. As well, a person's repeated facial movements or expressions may reinforce already existing lines, where lines become intensified and deepen into wrinkles. If the intensification in wrinkling occuring from morning to late afternoon is discernible, the degree to which the change alters the appearance of an individual's face is meaningful.

Geng et al. (2006) proposed Aging Pattern Subspace (AGES) for automatic age estimation. They compare the performance of AGES with the human-based age estimation. 51 face images are randomly selected and presented to 29 human observers.

[3] Face anatomy is available at Taylor (2000) and Zimbler et al. (2001).

There are two stages in the experiment. In each stage, the 51 face images are randomly shown to the observers, and the observers are asked to choose an age from 0 to 69 for each image. In the first stage, only the grey-scale face regions are shown, while in the second stage the whole colour images are shown. Multiple cues including face, hair, skin colour, clothes and background are shown in the second stage due to the colour information. They claimed that the proposed method outperforms the human observers in the ability of facial age estimation although human observers are provided with more information than the input into the algorithms.

Age and identity information of a facial image is encoded in both texture and shapes. Facial shape variations due to aging are often manifested as subtle drifts in facial features and progressive variations in the shape of facial contours. Although facial shape could be affected by many factors, such as expression, pose and age, it still conveys much information about the identity and age of the subject (Wu et al., 2012). O'Toole et al. (1999) examined the relationship between facial attractiveness, age, and 'averageness', using laser scans of faces that were put into complete correspondence with the average face. Their results indicate clearly that normalising the faces with respect both to their 3D structure and their 2D texture increased the attractiveness of faces. They also found that the shape-normalised faces which retained their image-based imperfections were considered more attractive than the texture-normalised faces. Turaga et al. (2010) investigated the role of geometry for face age estimation. They tested on a standard FGNET and a passport dataset which illustrate the effectiveness of their approach. Their results showed that exploiting geometric cues in a principled manner provides comparable performance to several systems that utilise both geometric and textural cues. In summary, facial geometry has a strong influence on age perception.

Han et al. (2015) proposed a generic framework for automatic demographic (age, gender and race) estimation for a given face image. They extracted demographic informative features from the commonly used BIF, and predicted the demographic attributes of a face image using a hierarchical classifier. Quality assessment was used to identify low-quality face images, which allows possible reacquisition of new face images in cooperative scenarios, or rejection of the input face image otherwise. Their results were compared with human estimation using crowd-sourced data obtained via the Amazon Mechanical Turk (MTurk) service. In a correlation analysis, they claimed that human tends to overestimate the ages of individual subjects while machine provides relatively unbiased estimates. For FGNET estimation, human outperformed machine of without quality assessment. They argued that FGNET is significantly biased age distribution

Table 2.3: Contributions in the topics of facial aging from psychophysics, human perception and physiology. Partial of this table is derived from the article of Ramanathan et al. (2009).

Reference	Summary of Contributions
Pittenger et al. (1979)	Interestingly, observed that the cardioidal strain transformation was effective in altering the perceived age of cartoon drawings of monkeys, birds, dogs and inanimate objects such as the 'Volkswagen' beetle.
Mark and Todd (1983)	Extended the 2D cardioidal strain transformation model into 3D and demonstrated its effectiveness in characterising facial growth in 3D.
Bruce et al. (1989)	Observed that a subject's sensitivity to cardioidal strain related changes in 3D faces were comparable, when viewed in pairs of face profiles or pairs of 3/4 faces or pairs of mixed profiles
O'Toole et al. (1997)	Applied a standard facial caricaturing algorithm on 3D faces. Noted that an exaggeration or a de-emphasis of facial creases, increased or decreased the perceived age of faces, respectively.

where a majority of images belong to subjects less than 18 years old. As a result, the machine does not contain a sufficient number of images in the age range of 30-69. In contrast, both MORPH and Pinellas County Sheriffs Office (PCSO) (Han et al., 2015) datasets have subjects with more uniform age distributions, thus the machine estimation achieved low MAEs across all age ranges. Their comparisons suggest that the age estimation performance of human is relatively independent of the age distribution in the dataset and the dataset size because of their prior knowledge. Table 2.3 shows the related contributions topics of facial aging from psychophysics, human perception and physiology.

2.5 Wrinkle Detection Techniques

Although the relationship between age and physical and physiology skin attributes has been studied for the last century, there was no attempt before the advent of image processing to accurately measure visual characteristics of skin aging (Gartstein and Shaya, 1996). The visible signs of damage to skin, such as wrinkling, sagging, irregular

pigmentation, telangiectasia[4], and hyperkeratosis[5], are frequently clinically assessed using either descriptive scales or photographically calibrated ones. The well-known shortcomings of subjective assessment limit the scientific study of treatment and environmental effects on skin aging. Due to facial wrinkle is very similar to blood vessel in the retinal image, different segmentation techniques in facial and retinal images are reviewed here.

Gartstein and Shaya (1996) showed the use of image processing technique to the analysis of profilometric traces of skin replicas. It is to isolate individual wrinkles and measure their depth. There are four steps in the processing: masking, shading correction, enhancement and segmentation, and object identification and morphological measurement. They claimed that the proposed method complements the measurement of wrinkle width and length as extracted from conventional optical images. In comparing the image-analysis measurements of subjects' pictures to their perceived ages, a high correlation of perceived-age with the total length of wrinkles. Thus, they stated that visual perception technique should be useful for validation of age-related measurements and for evaluation of therapies which modify the facial appearance. They also demonstrated that the proposed technique able to distinguish the old from the young and have sufficient sensitivity to detect small skin aging related effects.

Cula et al. (2013) developed an algorithm for automatic detection of facial wrinkles. It is based on estimating the orientation and the frequency of elongated spatial features, capture via digital image filtering. They used a similar idea of fingerprint detection where noise is removed and the contrast of image is improved in order to capture the irregularities of linear surface. They stated that wrinkle detection is more difficult than fingerprint detection because the appearance of wrinkles is more subtle, more varied and significantly noisier, whereas fingerprint features appear as continuous lines with almost constant thickness and frequency. They defined wrinkle index (WI) as a product between wrinkle depth and wrinkle length. They claimed that the proposed computational assessment correlates well with the corresponding clinical scores. However, the correctness of wrinkle detection is not validated where the found wrinkles could be noises or lines of interest. Details of this algorithm are depicted in Section 3.1.2.1.

[4]According to Oxford English dictionary online (Dictionary, 2015), telangiectasia is a condition characterised by dilatation of the capillaries causing them to appear as small red or purple clusters, often spidery in appearance, on the skin or the surface of an organ, illustration can be found in (Negishi et al., 2002).

[5]According to Oxford English dictionary online (Dictionary, 2015), hyperkeratosis is a condition characterised by abnormal thickening of the outer layer of the skin.

Batool and Chellappa (2014) introduced a novel method for facial wrinkle detection. It is based on Gabor features using image morphology and geometric constraints. Gabor filter bank is to highlight the subtle curvilinear discontinuities in skin texture based on wrinkles. Then, image morphology is used to incorporate geometric constraints to localize curvilinear shapes of wrinkles at image sites of large Gabor filter responses. They pointed out few image characteristics of wrinkles as follows.

- Wrinkles cause intensity gradients which are highlighted by Gabor filter bank.
- Wrinkles appear as curvilinear features instead of blob like features.
- Wrinkles are continuous and not a series of disconnected curvilinear segments.
- Wrinkles are less probable to intersect each other or to be congested in small areas.

Experiments are conducted on two sets of low and high resolution images and results showed that the proposed algorithm is significantly faster than conventional approach but also provides visually better results. However, their work is lack of intra- and inter-reliability measurements where the annotation of single coder could be bias. In addition, their method is time consuming due to a maximum of 30 iterations is required to verify the neighbouring sites of connected segments. There are two limitations during the evaluation setup. First, high detection rate did not ensure results of visually high quality where the term 'quality' can be described as the continuity of a wrinkle curve as well as its smooth curvature looking more similar to real wrinkles to a user. Second, the hand-drawn wrinkles can be subjective depending on the user perception of wrinkles. This introduced subjectivity in the quantitative results calculated from the experiments. On the other hand, they also highlighted factors contributed to the degradation of localisation results as follows.

- Very light wrinkles. The most common reason for low detection rates was wrinkles' being very light. The intensity gradients caused by the granular skin texture are comparable to those caused by wrinkles. Consequently, the algorithm picks non-wrinkle sites as wrinkles which increasing false alarm rate.
- Skin discolourations. This problem was more pronounced in subjects of lighter skin colour which is more prone to having moles and brown / dark spots.

- Illumination and bright spots. Due to the publicly available images being under uncontrolled acquisition settings, most of the images had some illumination variation. The variation caused 'bright spots' in images where light is reflected from skin surface due to greasy skin.

Chaudhuri et al. (1989) presented an operator for detecting blood vessels in retinal images. It is based on the optical and spatial properties of objects to be recognised. The grey-level profile of the cross section of a blood vessel is approximated by a Gaussian shaped curve. The concept of matched filter detection of signals is used to detect the piecewise linear segments of blood vessel in these images. They claimed that the proposed scheme retains the computational simplicity of the enhancement / thresholding type of edge operators, and at the same time incorporates the advantages of using model-based edge detectors. However, they stated that it may take a long time to run this algorithm due to the large size of convolutional kernel. They also observed three interesting properties of the blood vessels in retinal images as follows:

- Since the blood vessels usually have small curvatures, the anti-parallel pairs may be approximated by the piecewise linear segments.

- Since the vessels have lower reflectance compared to other retinal surfaces, they appear darker relative to the background.

- Although the width of a vessel decreases as it travels radially outward from the optic disk, such a change in blood vessel is a gradual one.

Frangi (2001) proposed a vessel enhancement filter for examining the multi-scale second order structure of an image. A measure of vessel-likeliness is obtained by all eigenvalues of Hessian. They tested on 2D Digital Subtraction Angiography (DSA) and 3D aortoiliac and cerebral Magnetic Resonance Angiographic (MRA) data. Since vessels appear in different sizes, they introduced a measurement scale which varies within a certain range. They conceived vessel enhancement as a filtering process that searches for geometrical structures which can be regarded as tubular. Results showed that the proposed method performs an excellent noise and background suppression in a two clinical imaging modalities. Details of the algorithm are described in Section 3.1.2.2.

2.6 Feature Representation

Feature representation is very important for age estimation, since the extracted features greatly affect the classification performance. These features can be categorised into local, global and hybrid features (Mark et al., 1980; Kwon and da Vitoria Lobo, 1994; Kwon and da Vitoria Lobo, 1999; Günay and Nabiyev, 2008; Aznar-Casanova et al., 2010). As stated in the first chapter, local features represent aging features in a series of image patches, global features tend to describe an image as a whole and hybrid features are a combination of both. Local features have been commonly used to classify people into age groups (e.g. babies, young adults and senior adults) as they possess unique characteristics that distinguish specific age group. Consequently, these features are better suited to applications requiring an age group classification (e.g. the class of less than 20 years old, the class of 20-39 years old, etc.) rather than a detailed age estimation (e.g. 17, 23 years old, etc.). Conversely, unlike local features, global features tend to be used for a detailed age estimation and contain not only aging characteristics, but other individual characteristics such as identity, expression, gender, ethnicity, among others. Therefore, individual characteristics are reflected better in global features than aging clues. For instance, FAM (Cootes et al., 2001) is mainly used to estimate an age as global features, because they offer greater amounts of information concerning the appearance and shape of a face than local features (Lanitis et al., 2002; Geng et al., 2007; Yan et al., 2008a). However, FAM features do not include detailed wrinkle and skin informations, due to the dimensional reduction made by PCA. This defect can be resolved by hybrid features through a combination with local features. Suo et al. (2008) proposed an age estimation method based on hybrid features and the defect found in each feature can then be compensated by combining both global and local features. Therefore, hybrid features are desirable for an accurate estimation of face age (Chao et al., 2013).

2.6.1 Global Feature

As mentioned above, Geng et al. (2006) proposed AGES for learning the facial aging patterns for automatic age estimation. The basic idea is to model the aging pattern, which is defined as a sequence of personal aging face images, by learning a representative subspace. A proper aging pattern for an unseen face image is then determined by the projection in the subspace that can best reconstruct the face image, while the position of the face image in that aging pattern will indicate its age. They claimed that AGES

treated each aging pattern as a sample, which naturally integrates the personal identity and the time concept. This changes the learning task from an image-class problem based on complete data into an (image sequence)-(class sequence) problem based on incomplete data. An iterative learning algorithm is specially developed to learn a representative subspace from the highly incomplete aging patterns. Their results showed that AGES outperforms the existing algorithms and also exceeds the human ability for age estimation. However, AGES has three main problems. First, the subspace of one particular age requires at least few images during the training stage, otherwise no results can be predicted at that age. Second, the reconstruction of missing face images are highly redundant which may increase the prediction error. Third, multiple aging subspaces might appear as the lowest distance of an unseen image and it may enlarge the margin of prediction error.

Fu and Huang (2008) construct a low-dimensional manifold from a set of age-separated face images using different manifold learning approaches such as Locality Preserving Projections (LPP), Orthogonal Locality Preserving Projections (OLPP), Conformal Embedding Analysis (CEA). They used linear and quadratic regression functions on the low-dimensional feature vectors from the respective manifolds, in estimating the age of a face image. Along very similar lines, Guo et al. (2008a) adopted the manifold learning approach and the use of Support Vector Regression (SVR) to estimate the age from low-dimensional representation of faces.

Chen et al. (2013) introduced a novel cumulative attribute space based framework for solving a number of computer vision problems invoking the need for regression estimation. Noisy and sparse low level visual features are mapped onto a cumulative attribute space where each dimension is designed specifically to give a clear semantic meaning that captures how the scalar output (e.g., age, people count) changes continuously. They claimed that the proposed framework requires no additional human annotation to assign attributes and the face age can be estimated efficiently given sparse and imbalanced training data.

Luu et al. (2011) proposed a novel Contourlet Appearance Model (CAM) to localize facial landmarks and extract texture features from the convex hull bounded by them. CAM is a combination between features of Modified Active Shape Model (MASM) and Nonsubsampled Contourlet Transform. MASM decouples the fitting of facial landmarks from the extraction of texture features, while Contourlet features represent local features such as wrinkles in the adult period. They claimed that CAM is more accurate at reconstructing an unseen texture compared to a conventional AAM. Results showed

that the proposed method achieves a MAE of 4.12 on FGNET.

Yang et al. (2011) proposed a feature fusion method which ensembles global features from FAM and local features from a Gabor wavelet transformation. First, they performed FAM modelling on a face image to retrieve global features and the shape-free patch. Second, they used a Gabor wavelet transformation on the shape-free patch to get the Gabor feature representation and then PCA is applied to reduce the dimensionality of local features. Third, they utilised SVR to perform regression analysis for estimating the age. Results showed that a MAE of 5.88 is achieved on PAL.

Chao et al. (2013) proposed a new age estimation approach considering the intrinsic factors of human ages. In order to avoid over-fitting and explore the connections between facial features and age labels, LPP (He and Niyogi, 2004) is utilised to reduce the dimensionality of features drastically and preserve the most important information. They introduced a label-sensitive concept to better exploit the ordinal relationship among age labels and presented an age-oriented local regression algorithm. Ordinal relationship among age labels is one of the factors that causes age estimation more difficult learning task than the traditional multi-group classification. For example, age of 30 years is closer to age of 25 years than age of 10 years. Therefore, the ordinal relationship among age labels are considered in distance metric adjustment and dimensionality reduction. They also proposed several imbalance treatments to alleviate the imbalance problem. They achieved a MAE of 3.06 on FGNET with ages smaller than 30 and a MAE of 4.11 with a 4-fold cross validation on whole FGNET.

2.6.2 Local Feature

For local descriptor, it can be divided into two categories: synthetic and natural. Synthetic features are a set of artificial patterns created by numerous texture-based descriptors such as Local Binary Patterns (LBP) (Ojala et al., 1994; Ylioinas et al., 2013), Histogram Of Gradient (HOG) (Dalal and Triggs, 2005; Fernández et al., 2015) and BIF (Serre et al., 2005; Guo et al., 2009a). Natural features are pixels derived originally from an image without any preprocessing (i.e., wrinkle, moustache, hair, etc.). Different methods of synthetic and natural descriptor are discussed in the following section.

2.6.2.1 Synthetic Feature

Yang and Ai (2007) applied Local Binary Pattern Histograms (LBPH) as aging descriptors. Given a restricted local patch, a Chi square distance between the extracted LBPH

and a reference histogram is used as a measure of confidence belonging to the reference class. Adopting the AdaBoost technique, they identified a sequence of local features which when combined into a strong classifier performs the task of age classification successfully. They claimed that the error rate of age classification is as low as 7.88% on FERET. However, LBPH is weakened by the sparse nature of LBP representation Ylioinas et al. (2013).

Guo et al. (2009a) proposed BIF for age estimation via faces. The basic idea is derived from a feedforward model of the primate visual object recognition pathway so-called the "HMAX" model. It contains alternating layers called Simple (S) and Complex (C) cell units creating increasing complexity as the layers progress from the primary Visual cortex (V1) to Inferior Temporal (IT) cortex (Serre et al., 2005). The first layer S1 is created with a Gabor filtering on the input image. The second layer C1 is obtained with a "MAX" operation on S1. This kind of feature has the advantage of handling small translations, rotations, and scale changes, which, if effective, captures aging patterns. The BIF feature combined with Support Vector Machine (SVM) can achieve a MAE of 4.77 years on FGNET, and MAEs of 3.91 and 3.47 years for female and male on the dataset of Yamaha Gender and Age (YGA) (Fu and Huang, 2008), respectively. These results demonstrate the superior performance of BIF for the task of age estimation. Details of this algorithm is presented in Section 3.2.1.3.

Ruiz-Hernandez et al. (2010) proposed a tensor representation from Gaussian receptive field. A tensorial representation is used to encode the shape changes in the aging process and it conserves the 3D structure of the scale space. LBP is applied over each Gaussian receptive field and this operator encodes a set of micro-patterns from the neighbourhood appearance. Multi-linear PCA is used to reduce the dimensionality of the tensorial space. Results showed that the best MAEs of FGNET and MORPH are 4.96 and 6.19, respectively. They claimed that a tensorial representation conserves the natural structure information contained in the binary maps.

Ylioinas et al. (2013) proposed KLBP for face age estimation. This method includes a pose correction method and a comparison of two local binary pattern based facial representations, namely a spatially enhanced histogram and a novel kernel density estimate. It generates sign and magnitude features through face patches. They claimed that the sparse nature of LBP representation is improved using the proposed kernel estimator. They achieved a MAE of 5.20 on original images of FGNET and a MAE of 5.09 on pose corrected images. However, this method is not yet tested on FERET. Details of this algorithm is given in Section 3.2.1.4.

Fernández et al. (2015) presented a comparative evaluation of four regression learning algorithms for facial age estimation. They are kernel techniques, ensemble learning, neural networks and projection algorithms. They used simple HOG descriptors as feature representation of each image. They claimed that advantages of using HOG features are no face alignment needed, no modelling of facial aging and no additional demographic annotations. In a 5-fold cross validation, they achieved a MAE of 4.83 on MORPH and a MAE of 2.88 on the dataset of Face Recognition Grand Challenge v2.0 (FRGC) (Phillips et al., 2005).

2.6.2.2 Natural Feature

As mentioned above, Kwon and da Vitoria Lobo (1999) proposed a combination rule of the facial ratios and wrinkle index for age group classification. These criteria were suggested by craniofacial research and the observation that aging skin develops wrinkles. Initially, different ratios are measured from a number of facial landmarks such as oval, chin, sides of face, virtual top of head, eyes, mouth and nose. Next, a population of snakelets are used to find wrinkles. A snakelet is a small snake segment designed to find a small individual curve segment in an image. A wrinkle geography map drops multiple snakelets in polygonal regions, where wrinkles may typically be found. Their focus is concentrated on the wrinkles on forehead, next to the eyes and near the cheek bones. If several curves are found in a particular region, they claimed that wrinkles exist in that region. The different locales of evidence for the presence of wrinkles are then weighted appropriately for classifying into one of three groups, babies, young adults and seniors. Their experiments showed that computing ratios and detecting the presence of wrinkles can yield age categorization. However, there are at least two problems need to be addressed. First, there was no objective validation against the extracted wrinkles. Lines found by the snakelet could be wrinkles or noises. Second, facial alignment was not done automatically. Wrinkle geography map was placed manually for dropping the snakelets. It is worth to mention that they failed to locate any wrinkles with an image resolution of 256×256. A zoom in process is required to further capture the details and they stated that a high resolution image (1280×1024) is a potential area to be explored for wrinkle analysis.

Takimoto et al. (2007) investigated the appearance-age features for age estimation. The features are wrinkle, freckle, shape, hair and colour where these features are distributed at different parts of face such as forehead, eye corners, eye bags, nasalobial

region ends of mouth and chin. They used different measurements for different features, i.e., edge of wrinkle, strength of freckle, coordinates of shape, brightness and saturation. A genetic algorithm (GA) is used to select the optimum features and a Multi-layer Perception (MLP) is implemented to estimate the age from 12 to 68 years old. Their results suggested that appearance-age features are influenced by different age groups.

Choi et al. (2010) studied the local feature extraction for age estimation. They compared the performances of Sobel filter, difference image between original and smoothed image, Ideal High Pass Filter (IHPF), Gaussian High Pass Filter (GHPF), Haar features and Discrete Wavelet Transform (DWT). They claimed that the facial wrinkles and skin texture are very important information for age estimation. With age increasing, facial wrinkles are increased and spots or blotches are appeared on skin texture. These aging features are generally shown as high frequency components on images. They also found that local features are largely influenced by the image resolution and it can be used to estimate detailed age, not limited to age group classification. However, they only achieved MAE of 6.85 ($\pm$5.34) on their dataset of Biometrics Engineering Research Center (BERC) and 8.44 ($\pm$7.16) on PAL. This implies that age estimation using local features still a challenging problem.

2.6.3 Hybrid Feature

Chen et al. (2011) introduced a facial feature fusion which combines global features extracted from FAM and local features extracted from LBP. They also proposed an advanced age estimation system combining feature fusion and model selection schemes such as Least Angle Regression and sequential approaches. Due to the fact that different facial feature representations may come with various types of measurement scales, normalisation scheme is applied for both facial features. Their results showed that feature fusion achieves better results over single feature. A min-Max normalisation works better than other normalisation methods.

Choi et al. (2011) proposed a hierarchical classifier based on hybrid features. Hybrid features are a combination of global and local features. In the global feature extraction step, FAM appearance and shape parameters are extracted as global features. In the local feature extraction step, the skin wrinkle is extracted using a set of region specific Gabor filters, each of which is designed based on the regional direction of the wrinkles, while the skin texture is extracted using a LBP which is capable of extracting the detailed textures of skin. Figure 2.7 demonstrates an example of Gabor feature extraction with 8

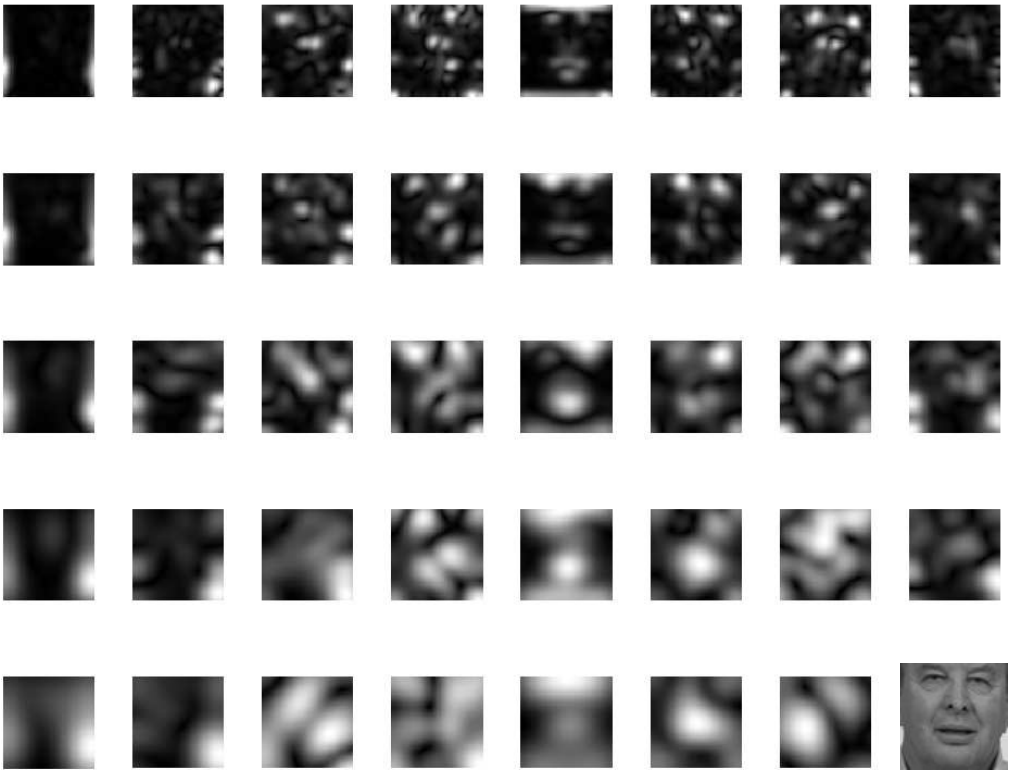


Figure 2.7: Feature extraction using Gabor filter. Each column shows a different scale and each row presents a different orientation. Bottom right corner is a cropped image from FERET.

different orientations and 5 different scales. It is noticed that a perceptible face outline can be seen on small size of filter especially around the eyes and mouth. They claimed that this combination can compensate the defects found in individual global and local features. Table 2.4 presents the pros and cons of each descriptors. It is noticed that a combination of features would increase the computation time due to a large size of feature dimension. Moreover, it might influence the estimation performance because a mixture patterns with different ranges of values would affect the modelling process of a classifier, unless a preprocessing step of feature normalisation is applied.

Wu et al. (2012) proposed a model based on the facial shapes and texture patterns. Facial shapes are distributed as points on a Grassmann manifold. A Grassmann manifold is the space whose points are k-planes where each k-plane is associated with an equivalence class. It is not a vector space, however points on a Grassmann manifold can be projected onto the tangent space at a mean-point and standard vector-space methods can then be applied on the tangent space as aging signatures. Using these tangent-vectors as shape features, they applied a standard Euclidean space regression

Table 2.4: A comparison between local, global and hybrid features. Note that partial of this table was derived from the article of Choi et al. (2011).

Category of Features	Feature Description	Strength	Weakness
Local	Skin texture, wrinkle, hair	Ability to extract aging features without an influence of personal characteristics	Better suited to age group classification
Global	FAM-based parameters, age manifolds, age subspaces	Provides a detailed representation of face age	Does not include enough information on wrinkles and aged skin
Hybrid	A combination of both global and local features	Compensates weaknesses of each features	Time consuming and feature reliability is doubtful

methods for the age estimation problem. For texture features, they used a Histogram of Gabor Phase Patterns (HGPP) (Zhang et al., 2007). A Partial-Least Squares (PLS) regression method is implemented for texture features due to its ability in handling a very high dimension features. In addition, they proposed a fusion which is based on the intermediate scores obtained from the shape and textural cues. Experimental results showed that a fusion between HGPP and geometric features outperform others for age estimation on FGNET. They argued that facial geometry has a strong influence on age perception.

2.7 Age Estimation Techniques

Typically, the existing facial aging estimation methods consist of two concatenated modules: feature representation and feature classification, as shown in Figure 2.8. In the first step, facial appearance changes are extracted from the face images for a compact representation; in the second step, an age estimation model is built to determine the face age based on the represented features (Choi et al., 2011; Chao et al., 2013). In general, age estimation is divided into three categories: age group classification, single-layer age estimation and hierarchical age estimation (Lanitis et al., 2004; Günay and Nabiyev, 2008). Age group classification is an approach that predicts an age group,

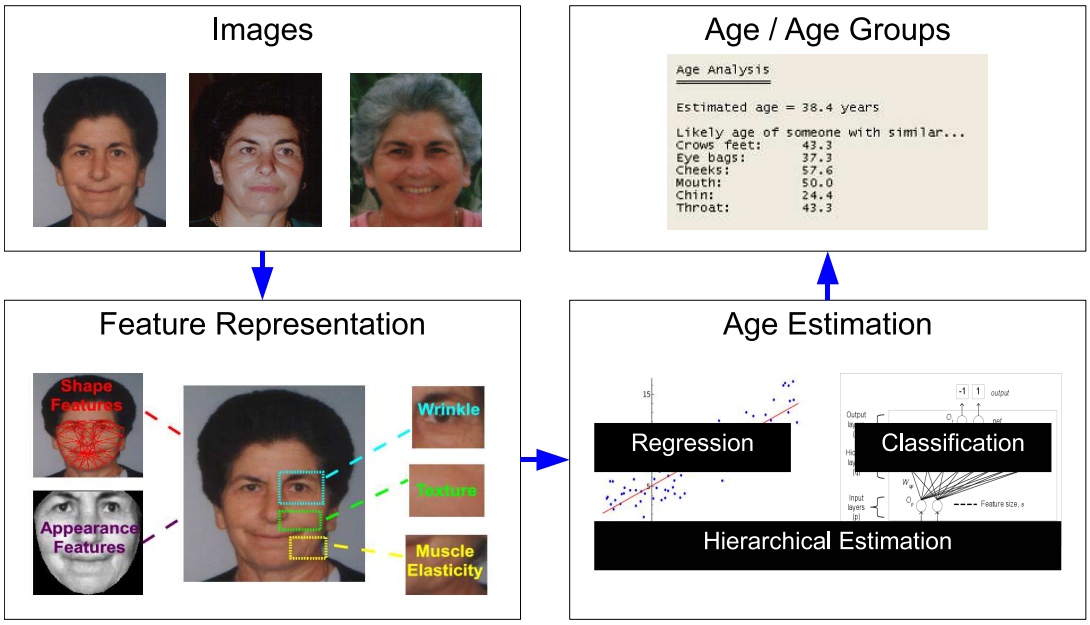


Figure 2.8: Flow of face age estimation. Note that this figure was redrawn from FGNET.

whereas single-layer and hierarchical age estimations are focused on detailed age prediction. Moreover, hierarchical age estimation is a coarse-to-fine method that finds the age label in a pre-defined group. Of these methods, hierarchical age estimation provides the most improved performance (Luu et al., 2009; Choi et al., 2011). As facial aging is perceived differently in various age groups, age estimation in a specific age group provides a more accurate result. Moreover age estimation on a smaller age group simplifies the computational load (Guo et al., 2009b; Choi et al., 2011).

2.7.1 Age Group Classification

Ramanathan and Chellappa (2006) proposed a method to estimate the age separation between a pair of face images of an individual. Considering four age separation categories namely, 1-2 years, 3-4 years, 5-7 years and 8-9 years, they created intra-personal subspaces for each category using the difference of images obtained from pairs of age-separated face images from the respective categories. Their approach was based on the premise that facial appearance variations are induced as a result of facial aging would be more pronounced. As a result, the 'image difference' features would get progressively more exaggerated with an increase in age separation. They used a Bayesian framework to classify pairs of face images based on age separation and further extended the approach to perform face verification across age progression.

Suo et al. (2007) adopted the multi-resolution grammatical face model (Xu et al., 2005) and augmented the model with age and hair features. Facial aging was modeled by means of a dynamic Markov process on the And-Or graph representation. They created a dictionary of different facial components and regions across five age groups: 20-30 years, 30-40 years, 40-50 years, 50-60 years and 60-70 years. The transitional probabilities comprised in the dynamic model are computed by means of 'geometric distances' and 'photometric distances' computed between facial components from different age groups. They claimed the following attributes related to facial aging are accounted for, in three different resolutions of the face image as,

- Global appearance variations in facial shape, hair style and skin colour are addressed at the lowest resolution.
- Deformations in facial components are addressed in the next resolution.
- Finally, finer aspects such as wrinkles and skin pigments are accounted for in the highest resolution.

Luu et al. (2010) introduced an advanced age determination technique that combines a feature set derived from a face image. It is using multi-factored PCA on the face shape and skin in order to produce a 30×1 linear encoding of the face image. The linearly encoded features are combined with Spectral Regression (SR) to further reduce the dimensionality of the encoded features such that inter-class distances are minimized while maximizing intra-class distances. The encoded features are used to recognise a face into one of two age groups, pre-adult (0-20 years old) and adult (> 20). Then, an age determination function is constructed for each age group in accordance to physiological growth periods for humans. Results showed that the proposed method achieved a MAE of 4.25 on FGNET.

2.7.2 Single-layer Age Estimation

Ricanek Jr et al. (2009) proposed a generalised multi-ethnic age estimation technique. They used FGNET and MORPH as the benchmark datasets where MORPH is divided into five race groups: Asian, African-American, Hispanic, Indian and Caucasian. They claimed that an overall performance of the proposed method is comparable to the existing algorithms that are tuned for a specific ethnic group. They highlighted three interesting points when developing a general model for face age estimation: feasibility of reduced features, selectivity of optimized features and size of features.

Weng et al. (2013) proposed a multi-feature ordinal ranking method for age estimation. It is based on learning weighted classifiers on multiple feature sets. Complementary information between different feature sets are explored by assigning weights to their correspondent classifiers through joint learning. Moreover, age estimation problem is divided into a group of $K-1$ sub-problem of binary classification according to the ordinal property of age labels where K is the number of age labels in the dataset. Their results demonstrated that the proposed ranking method outperforms other multi-view data fusion approaches.

Guo and Mu (2014) presented a new framework for joint estimation of age, gender and ethnicity. Under the proposed framework, they explored different methods to accomplish a one-step procedure for the joint estimation problem. They studied the feature dimensionality problem using the rank theory methods, Canonical Correlation Analysis (CCA) and PLS. CCA describes a linear relation between two multi-dimensional variables as the problem of finding basis vectors for each set, while PLS uses latent variables to learn a new space to make the data correlates to each other. BIF is used to characterise the facial patterns for age, gender and ethnicity together. They found that CCA based methods can derive an extremely low dimensionality where only three dimensions of features are needed to estimate age, gender, and ethnicity together.

2.7.3 Hierarchical Age Estimation

Thukral et al. (2012) proposed a hierarchical approach for age estimation where face images are divided into different age groups and a separate regression method is learned for each group. The groups are 0-15, 15-30 and above 30. They claimed that this division is based on the perceived homogeneity of the geometric growth of human faces with age and also the number of images available at different age groups. They used geometric feature vectors (Turaga et al., 2010) to represent each face in the dataset. Results showed that age estimation in the age group of above 30 is worse than other groups due to less variation of the facial geometry. They claimed that the proposed method hits a MAE of 3.40 on FGNET. However, the boundary of age group is a challenging problem. Such a hard boundary might not applicable to other datasets because different ethnicity or gender might have different rates of growing.

Han et al. (2015) presented a generic framework for automatic demographic estimation that includes age, gender and race. Given a face image, they first extracted

Table 2.5: Summary of NIST findings (Ngan and Grother, 2014).

Impact	Description
Core accuracy and speed	Age difference of 5 years on 6 million images is 67%, with a MAE of 4.3 years. All algorithms $<$ 0.15 seconds.
Demographic data on accuracy	In 240 thousand images, MAE of males $<$ females; MAE of an ethnically-heterogeneous population $>$ a homogeneous population; MAE of age group 56-99 $>$ 18-55.
Image quality on accuracy	A 17 year-old has a 29% chance of passing for 21. MAE of high quality mugshot images $<$ poor quality webcam photos.
Number of image samples on accuracy	Number of image samples $\uparrow$ age estimation times $\uparrow$.

demographic informative features via a boosting algorithm, and then employed a hierarchical approach consisting of between-group classification and within-group regression. An automated quality assessment model is also proposed to identify low-quality face images that are difficult to obtain reliable demographic features. Prior to select the demographic features, they used BIF for feature extraction. Then, multi-class AdaBoost is used to select the demographic features. The most informative features for age estimation are located in the regions where wrinkles typically appear such as eyes and mouth corners, nasolabial folds and cheeks. In the classification stage, three binary SVM classifiers are used to build a two-level binary decision tree and a test image is classified into one of the four groups. In the regression stage, a separate SVM regression is trained within each group to make an accurate age prediction. They claimed that the proposed hierarchical classifier with a quality assessment model can closely match human performance in demographic estimation with a MAE as low as 3.8 on FGNET.

2.8 Research Direction

In 2014, National Institute of Standards and Technology (NIST), United States (U.S.) Department of Commerce, performed a large scale empirical evaluation of face age estimation algorithms, with participation from five commercial providers and one university, using three large operational datasets comprised of facial images from visas and law enforcement mugshots, leveraging a combined corpus of over 7 million images (Ngan and Grother, 2014). They employed a black-box testing methodology designed to model operational reality where software is shipped and used 'as-is' without algorithmic training. Core age estimation accuracy was baselined over a large homogeneous

population, then assessed demographic by age group, gender and ethnicity. The impact and results of this study are summarised in Table 2.5. For core accuracy and speed, it highly depends on the provider of the core technology. Using the most accurate age estimation algorithm, the chance of accurately estimating the age of a person within five years of their actual age over 6 million images is 67%, with a MAE of 4.3 years. All algorithms can perform age estimation on a single image in less than 0.15 seconds with one server-class processor. For impact of demographic data on accuracy of 240,000 images, age is more accurately estimated in males than females, with the tendency for adult females to be underestimated. A majority of the algorithms demonstrated lower accuracy and higher MAE on an ethnically heterogeneous population than a homogeneous population, which suggests that ethnicity has an impact on age estimation. A majority of the algorithms estimate age more accurately for the most operationally relevant age group, 18-55, while the majority of algorithms exhibit the highest MAE in the senior age group, 56-99. For age verification accuracy that a person is at least 21 years old, a 17 year-old has a 29% chance of passing for 21, as achieved by the most accurate algorithm. This false verification percentage increases as a person gets closer to age 21. For impact of image quality on accuracy, high quality mugshot images are more accurate than poor quality web-cam photos due to the overestimation is seen in web-cam photos. For the impact of number of samples on accuracy, age estimation computation times increase linearly with respect to the number of image samples. For contemporaneous mugshot images of the same subject collected within a one year period, the results showed a MAE monotonically decreasing as the number of images provided increased for all algorithms. Appendices A and B show the state-of-the-art results of detailed age estimation and age group classification on different datasets.

The process of age estimation attempts to label a face image automatically with the exact age (year) or the age group (year range) of the individual face. By deriving significant features from faces of known ages, the age of an individual face can be estimated by solving the inverse problem using the same feature-extraction technique. Although many algorithms have been proposed since 1994 (Kwon and da Vitoria Lobo, 1994; Ramanathan et al., 2009; Fu et al., 2010), age estimation is still a challenging problem due to three reasons (Geng et al., 2006, 2007):

i. The aging process is uncontrollable. The procedure of aging is slow and irreversible. Thus the collection of sufficient training data for age estimation is extremely laborious.

ii. Personalised aging patterns. Different persons age in different ways. The aging

Table 2.6: Potential research areas for face age estimation.

Feature Types	Age Group Classification (Ramanathan and Chellappa, 2006; Suo et al., 2007; Luu et al., 2010)	Single-layer Age Estimation (Ricanek Jr et al., 2009; Weng et al., 2013; Guo and Mu, 2014)	Hierarchical Age Estimation (Choi et al., 2011; Thukral et al., 2012; Han et al., 2015)
Local Features (Kwon and da Vitoria Lobo, 1999; Takimoto et al., 2007; Guo et al., 2009a; Choi et al., 2010; Li et al., 2011; Ylioinas et al., 2013)	Wrinkle, spots, freckle, etc. were explored, but the features are not robust and no objective quantification against the extracted features.	BIF and LBP were well studied, additional methods which could preserve the complex aging features are worth to be studied.	No research yet.
Global Features (Geng et al., 2006; Fu and Huang, 2008; Chen et al., 2013; Luu et al., 2011; Yang et al., 2011; Chao et al., 2013)	Anthropometric model and appearance model were thoroughly investigated, but an extension to detailed age estimation is preferable.	FAM-based parameters and age manifolds were widely investigated, but feature dimension and selection remains a challenging issue.	Limited research.
Hybrid Features (Chen et al., 2011; Choi et al., 2011; Wu et al., 2012)	A combination of natural features were proposed in literature, results were not promising.	A number of hybrid features were introduced for detailed age estimation, but feature normalisation is an issue.	The hierarchical design and age range selection are challenging issues.

pattern of each person is determined by his/her gene as well as many external factors such as health, living style and weather conditions.

iii. The aging patterns are temporal data. The aging progress must obey the order of time. The face status at a particular age will affect all older faces, but will not affect those younger ones.

First, facial age progression is uncontrolled and personalised. Characteristics of aging variation cannot be captured easily due to the large variations conveyed by human faces. Facial aging effects are manifested in different forms during different ages. While facial aging effects are predominantly manifested in the form of facial shape variation during formative years, textural variations in the form of wrinkles and other skin artefacts take precedence over shape variations during later stages of adulthood. Hence, facial aging can be described as a problem of characterising facial shape and facial texture as functions of time. Since aging affects induce progressive variations in facial appearances, models characterising the same need to account for the temporal nature of the induced variations (Ramanathan et al., 2009). Second, there is no complete facial aging dataset with chronological ages. Developing facial growth models or building characterisations of facial aging begins with identifying the appropriate form of data that provide a fair description of the event. The data could be individual-specific or population-specific. Fiducial features (2D or 3D) extracted from age-separated faces, 2D facial imagery or 3D facial scans extracted from individuals across different ages, face anthropometric measurements extracted from a population, etc. are some forms of data that can help characterise facial growth (Ramanathan et al., 2009). It is hard to collect a large facial image set of people throughout their life which are sufficient to present detailed aging progression. Third, it is difficult to define an absolute aging pattern which can be used to quantify one particular age. For example, these are weak arguments if we say two wrinkles at the eye corner is classified as age 20, three wrinkles as 30 years old and so on. Such vague arguments will cause ambiguity if the person's age is between 20 and 30. Thus, a robust wrinkle-based representation is needed to cope with the identified problem.

Table 2.6 shows the potential research areas for face age estimation. It is noted that most of the works are focusing on global features, hybrid features, age group classification and single-layer age estimation. Local features such as wrinkles are particularly a research of interest as many works have confirmed that wrinkles are highly correlated with aging. Based on the reviews, here are few observations of facial wrinkles,

i. Facial wrinkles appear when aging.

ii. Facial Wrinkles grow randomly across different parts of face.

iii. Density of wrinkles grows when aging.

iv. Depth of wrinkles increases when aging.

In addition, hierarchical age estimation is also a potential area to be explored. As a result, few research questions are highlighted as follows and these will be answered in the following contribution chapters,

i. How to extract wrinkles from face accurately without human intervention.

ii. How to represent wrinkles as aging features in the context of face age estimation.

iii. How to make use of extracted wrinkles in hierarchical age estimation.

2.9 Summary

This chapter introduces the state-of-the-art methods for face age estimation. It includes questions like 'why is face age estimation is important?', 'what happens to the human face when aging?', 'how a human estimates people's age?', 'what are the related works of face age estimation?'. The advantages and disadvantages of each method are also highlighted. Chapter 3 will present a report of the theory, measurement and datasets being applied in this work. The idea is to systematically review the existing methods so that the performance of the proposed methods can be compared to the benchmark algorithms in terms of accuracy and functionality.

Chapter 3

Theory, Measurement and Datasets

This chapter describes each of the component techniques, measurement and datasets that were explored together to develop the novel methods described in later chapters. These consist of methods for solving wrinkle detection and face age estimation. A number of different scientific measurements and datasets are introduced. Most of them are applied in the following chapters as the benchmark methods and datasets.

3.1 Segmentation Algorithms

In this section, different methods for edge detection and line detection are described.

3.1.1 Edge Detection

In image segmentation, the most common approaches for detecting meaningful discontinuities in intensity values are convolution-based techniques such as the Sobel filter, Canny edge detector, Laplacian Of a Gaussian detector (LoG) typically followed by thresholding (Sobel and Feldman; Canny, 1986; Gonzalez et al., 2004). Such discontinuities are detected using the first and second order derivatives. The gradient of a 2D function, f, is defined as

$$\nabla f(x,y) = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \tag{3.1}$$

The magnitude, Λ, of this vector is

$$\Lambda = \text{mag}\left(\nabla f\right) \tag{3.2a}$$
$$= \left[G_x^2 + G_y^2\right]^{1/2} \tag{3.2b}$$
$$= \left[\left(\partial f/\partial x\right)^2 + \left(\partial f/\partial y\right)^2\right]^{1/2} \tag{3.2c}$$

The result of gradient approximations can be simplified by the square-root operation,

$$\Lambda \approx G_x^2 + G_y^2 \tag{3.3}$$

or using absolute values,

$$\Lambda \approx \left|G_x\right| + \left|G_y\right| \tag{3.4}$$

These approximations still behave as derivatives; that is, they are zero in areas of constant intensity and their values are proportional to the degree of intensity change in areas whose pixel values are variable. It is common practice to refer to the magnitude of the gradient or its approximations simply as 'the gradient'.

A fundamental property of the gradient vector is that it points in the direction of the maximum rate of chance of f at coordinates (x,y). The angle, α, at which this maximum rate of chance occurs is

$$\alpha\left(x,y\right) = \tan^{-1}\left(\frac{G_y}{G_x}\right) \tag{3.5}$$

One of the key issues is how to estimate the derivatives G_x and G_y digitally. The various approaches for edge detection are discussed later in the following sections.

The second order derivatives in image processing are generally computed using the Laplacian function. That is, the Laplacian of a 2D function f is formed from the second order derivatives, as follows:

$$\nabla^2 f\left(x,y\right) = \frac{\partial^2 f\left(x,y\right)}{\partial x^2} + \frac{\partial^2 f\left(x,y\right)}{\partial y^2} \tag{3.6}$$

The Laplacian function is seldom used by itself for edge detection because, it is unacceptably sensitive to noise; its magnitude produces double edges, and it is unable to encode edge direction. However, the Laplacian function can be a powerful component when used in combination with other edge-detection techniques. For example, although its double edges make it unsuitable for edge detection directly, this property

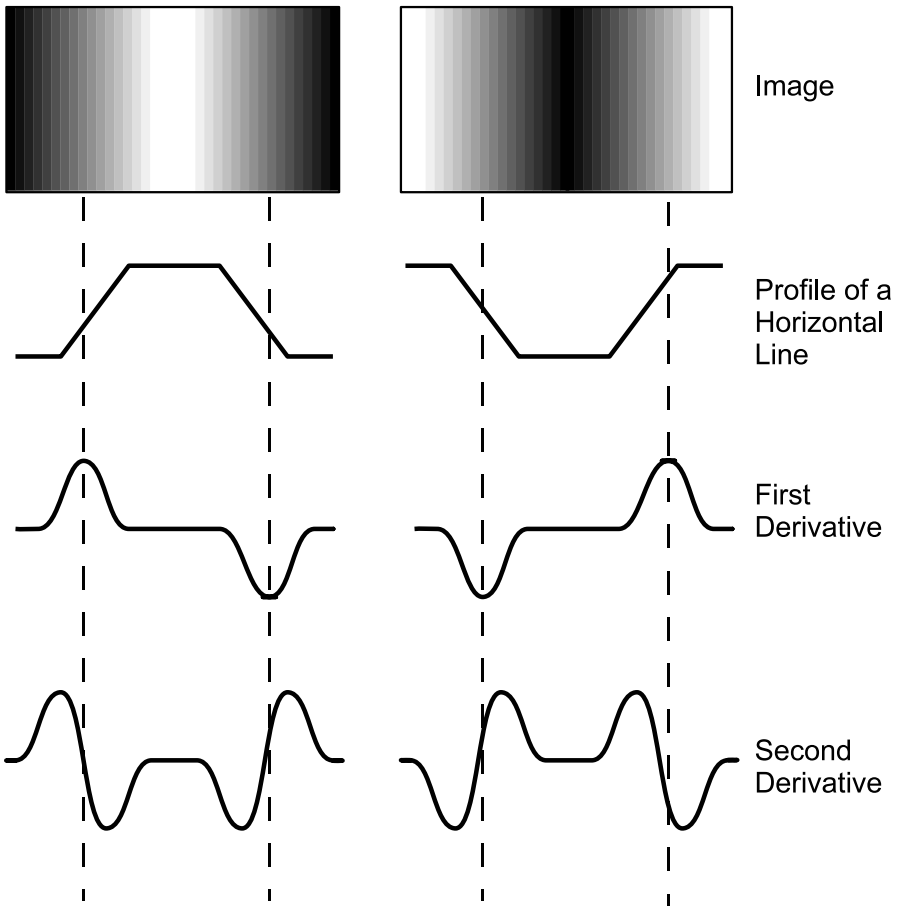


Figure 3.1: Profile of an edge. Top row shows an image with different edges. The second row is a profile of a horizontal line where y-axis is the grey-scale value. Third row presents profile of the first order derivative and the last row demonstrates the profile of the second order derivative. Dotted line means the location of edge of interest. Note that this figure was redrawn from the book of Sinha and Patel (2014, Chapter 6).

can be used for edge location.

With the preceding discussion as background, the basic idea behind edge detection is to find places in an image where the intensity changes rapidly, and significantly using one of two general criteria:

i) Find places where the first derivative of the intensity is greater in magnitude than a specified threshold.

ii) Find places where the second derivative of the intensity has a zero crossing.

Figure 3.1 shows an example of edge image on a dark background. For the left image, grey-scale values change from black to white and then to black again, while

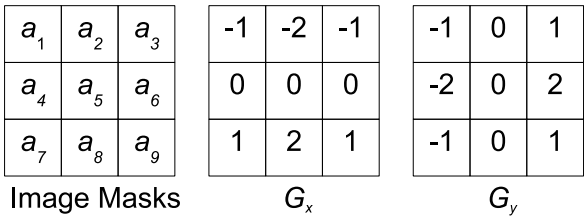


Figure 3.2: Sobel detector masks. Image mask shows the neighbourhood in a 3×3 filter. G_x and G_y are the first derivatives of Sobel detector. Note that this figure was redrawn from book of Gonzalez et al. (2004, Chapter 10).

the right image is in the opposite direction. The edge is between black and white area (dotted line in the figure). The first derivative is positive when the image changes from dark to light and zero where the image is constant. The second derivative is positive for the part of the transition associated with the dark side of the edge and negative for the transition associated with the light side of the edge. Thus the magnitude of the first derivative can be used to detect the presence of an edge in the image and the sign of the second derivative can be used to detect whether a pixel lies on the light or dark side of the edge (Sinha and Patel, 2014, Chapter 6).

3.1.1.1 Sobel Filter

The Sobel edge detector uses the masks in Figure 3.2 to approximate digitally the first derivatives G_x and G_y by convolution. In other words, the gradient at the center point in a neighbourhood is computed as

$$G = \left[G_x^2 + G_y^2\right]^{1/2} \tag{3.7a}$$

$$= \Big(\left[(a_7 + 2a_8 + a_9) - (a_1 + 2a_2 + a_3)\right]^2 + \left[(a_3 + 2a_6 + a_9) - (a_1 + 2a_4 + a_7)\right]^2\Big)^{1/2} \tag{3.7b}$$

Then, a pixel at location (x, y) is an edge pixel if $G \geq t$ at that location, where t is a specified threshold. G is a logical image containing 1s at locations where edges were detected and 0s elsewhere.

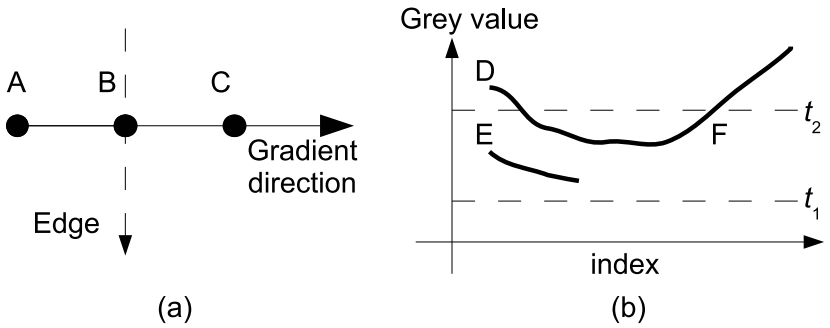


Figure 3.3: Properties of Canny edge detector. (a) Non-maximal suppression and (b) hysteresis thresholding.

3.1.1.2 Canny Edge Detector

The Canny detector (Canny, 1986) is the one of the most powerful edge detector for segmentation. It has been used for wrinkle detection and age group classification by Dehshibi and Bastanfard (2010). The method can be summarised as:

i. The image is smoothed using a Gaussian filter as Eq. (3.8) with a specified standard deviation, σ.

ii. The local gradient, $G = {G_x}^2 + {G_y}^2$, and edge direction, $\alpha = tan^{-1}(G_y/G_x)$, are computed at each point. Any of the edge detectors such as Sobel filter can be used to compute G_x and G_y. An edge point is defined to be a point whose strength is locally maximum in the direction of the gradient. Note that both smoothing (i) and edge detection (ii) functions can be combined into a single step by first order derivative of Gaussian filter.

iii. The detected edge points give rise to ridges in the gradient magnitude image. The algorithm then tracks along the top of these ridges and sets to zero all pixels that are not actually on the ridge top so as to give a thin line in the output, a process known as non-maximal suppression as shown in Figure 3.3(a). Point A is on the edge (in vertical direction). Gradient direction is normal to the edge. Point B and C are in gradient directions. So point A is checked with point B and C to see if it forms a local maximum. If so, it is considered for next stage, otherwise, it is suppressed (put to zero).

iv. The ridge pixels are then thresholded using hysteresis thresholding as shown in Figure 3.3(b). Let two threshold, t_1 and t_2, with $t_1 < t_2$. Ridge pixels with values

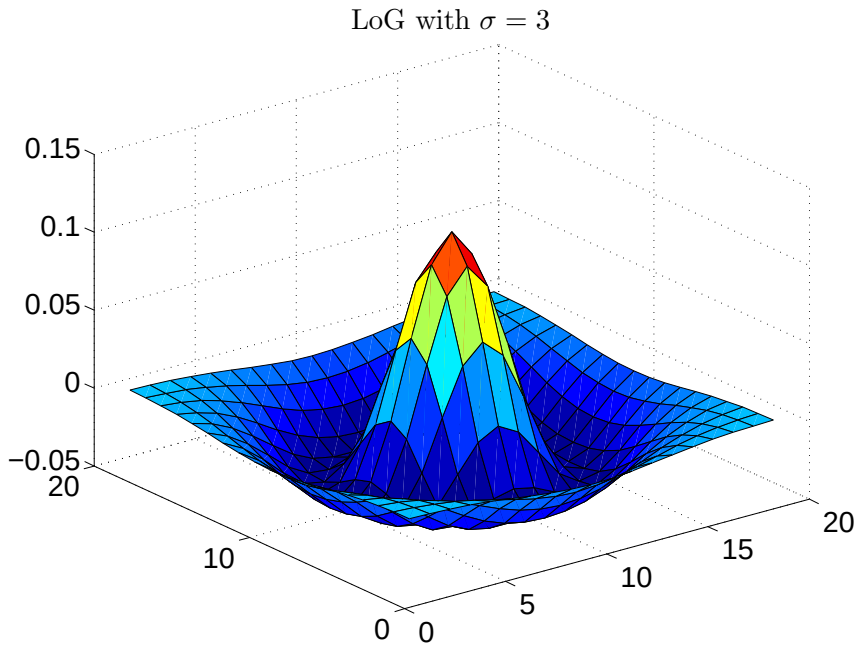


Figure 3.4: An illustration of the Laplacian of a Gaussian kernel.

greater than t_2 are said to be “strong” edge pixels. Ridge pixels with values between t_1 and t_2 are said to be “weak” edge pixels. The edge D is above the t_2, so considered as ”strong-edge”. Although edge F is below maxVal, it is connected to edge D, so that also considered as valid edge and we get that full curve. But edge E, although it is above t_1 and is in same region as that of edge C, it is not connected to any ”strong-edge”, so that is discarded. So it is very important that we have to select t_1 and t_2 accordingly to get the correct result.

3.1.1.3 Laplacian of a Gaussian Detector

Consider a Gaussian function as

$$f(x,y) = -e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.8}$$

where σ is the standard deviation (Gonzalez et al., 2004). This is a smoothing function which, if convolved with an image, will blur it. The degree of blurring is determined by the value of σ. The Laplacian of this function is defined as

$$\nabla^2 f(x,y) = -\left[\frac{x^2+y^2-\sigma^2}{\sigma^4}\right] e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.9}$$

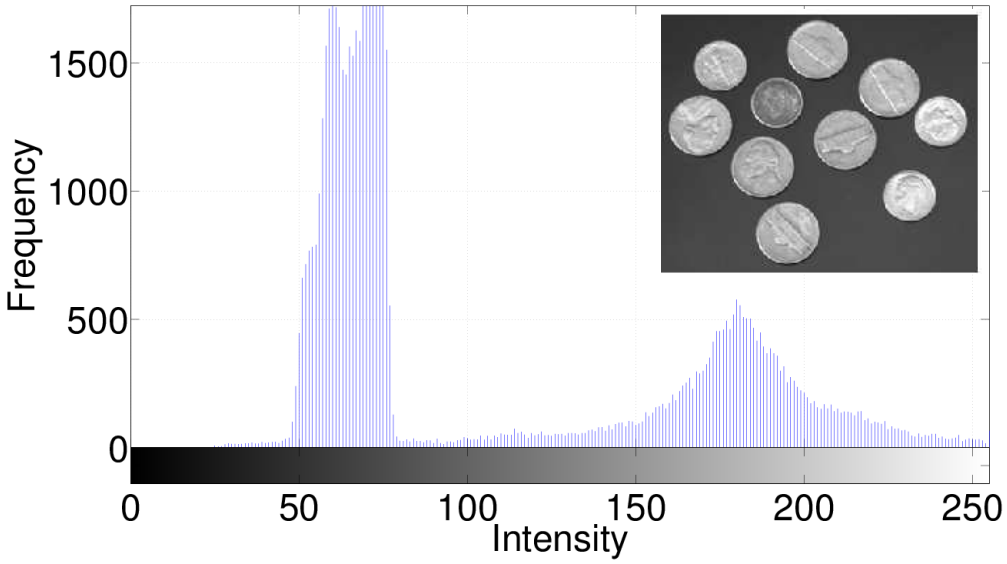


Figure 3.5: Selecting a threshold by visually analysing a image histogram. Top right corner shows an image of Matlab.

For obvious reasons, this function is called the Laplacian of a Gaussian (LoG) (Gonzalez et al., 2004). Because the second derivative is a linear operation, convolving (filtering) an image with $\nabla^2 f$ is the same as convolving the image with the smoothing function first and then computing the Laplacian of the result. This is the key concept underlying the LoG. We convolve the image with $\nabla^2 f$, knowing that it has two effects, it smoothes the image (thus reducing noise), and it computes the Laplacian, which yields a double-edge image. Figure 3.4 demonstrates an example of LoG where $\sigma = 3$.

3.1.1.4 Thresholding

Thresholding is the most common and easy way to undertake image segmentation. Suppose that the intensity histogram shown in Figure 3.5 corresponds to an image composed of light objects (coins) on a dark background, in such a way that object and background pixels have intensity levels grouped into two dominant modes. One obvious way to extract the objects from the background is to select a threshold t that separates these modes. Then any point (x,y) for which $\mathbf{I}(x,y) \geq t$ is called an object point; otherwise, the point is called a background point. In other words, the thresholded image $\mathbf{I}'$ is defined as

$$\mathbf{I}'(x,y) = \begin{cases} 1 & \text{if } \mathbf{I}(x,y) \geq t \\ 0 & \text{if } \mathbf{I}(x,y) < t \end{cases} \tag{3.10}$$

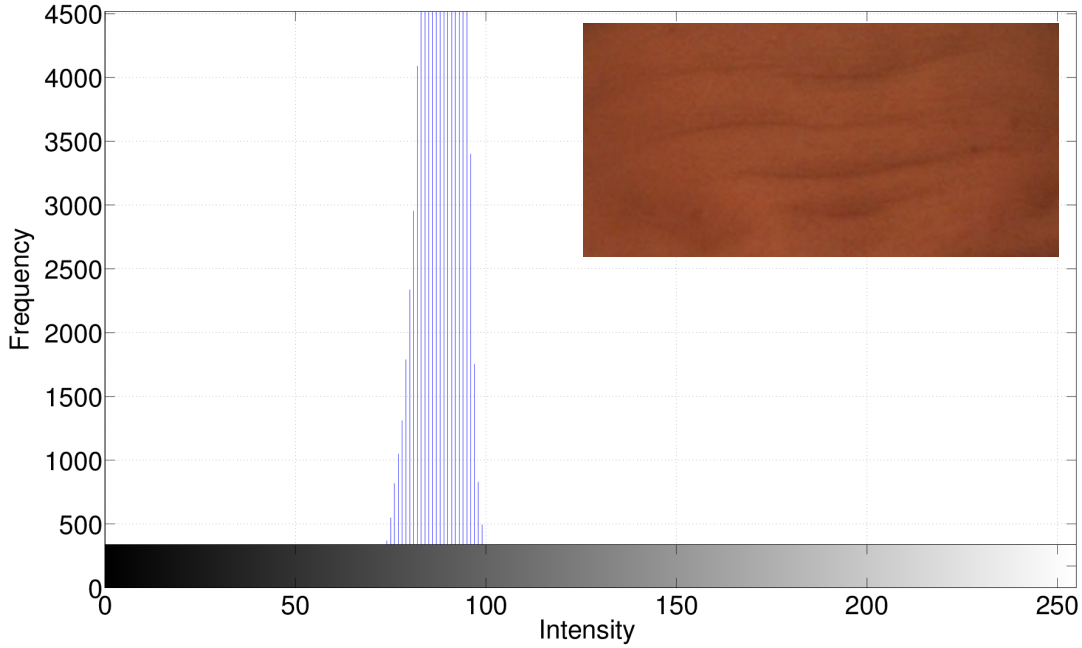


Figure 3.6: Histogram of forehead wrinkle image. Top right corner illustrates the original image used.

where pixels labeled 1 correspond to objects, whereas pixels labeled 0 correspond to the background. When t is a constant, this approach is called global thresholding. Figure 3.6 presents a forehead image with wrinkle and its histogram. As we noticed that thresholding is not an appropriate method for wrinkle segmentation because it is hard to set a threshold value between the wrinkles (objects) and background.

Otsu (1979) proposed an adaptive thresholding function where the threshold value t maximizes the between-class variance σ^2_{Otsu}. To examine this method, a normalised histogram as a discrete probability density function is defined as,

$$\hat{p}_{\tilde{r}}\left(\tilde{r}_q\right) = \frac{n_q}{n}, \quad q = 0,1,2,\ldots,U-1 \tag{3.11}$$

where n is the total number of pixels in the image, n_q is the number of pixels that have intensity level $\tilde{r}_q$, and U is the total number of possible intensity levels in the image. Now suppose that a threshold t is chosen such that Z_0 is the set of pixels with levels $[0,1,...,t-1]$ and Z_1 is the set of pixels with levels $[t,t+1,...,U-1]$. The between-class variance σ^2_{Otsu} is defined as

$$\sigma^2_{\text{Otsu}} = \hat{w}_0\left(\mu_0 - \mu_z\right)^2 + \hat{w}_1\left(\mu_1 - \mu_z\right)^2 \tag{3.12}$$

where $\hat{w}$ and μ are defined as

$$\hat{w}_0 = \sum_{q=0}^{t-1} \hat{p}_q\left(\tilde{r}_q\right) \tag{3.13a}$$

$$\hat{w}_1 = \sum_{q=t}^{U-1} \hat{p}_q\left(\tilde{r}_q\right) \tag{3.13b}$$

$$\mu_0 = \sum_{q=0}^{t-1} q\hat{p}_q\left(\tilde{r}_q\right)/\hat{w}_0 \tag{3.13c}$$

$$\mu_1 = \sum_{q=t}^{U-1} q\hat{p}_q\left(\tilde{r}_q\right)/\hat{w}_1 \tag{3.13d}$$

$$\mu_z = \sum_{q=0}^{U-1} q\hat{p}_q\left(\tilde{r}_q\right) \tag{3.13e}$$

3.1.2 Line Detection

In this section, two existing methods for line detection are discussed.

3.1.2.1 Cula Method

Cula et al. (2013) explored the use of the first order derivatives and Gabor filter for detecting wrinkle length and depth, respectively. The estimation is based on the orientation and frequency of elongated spatial features. They defined Wrinkle Index (WI) as the product of both wrinkle length and wrinkle depth. Wrinkle depth is derived from the Gabor filter responses. They claimed that WI achieves a high correlation between clinical scores (expert rating) and output of Cula Method (CLM). However, from a computer vision perspective, their work was lack of detailed quantification of line segment localisation.

Figure 3.7 shows the flow chart of CLM. A portion of the following description is derived from Cula et al. (2013). The computational process of CLM comprises several steps. First, the red channel of parallel-polarised face image is selected as the input image. They stated that it captures the most information about the skin surface structure and minimises information on skin colour variation which is modulated by hemoglobin and melanin. To better separate between the smooth skin areas and the wrinkles, the input contrast is enhanced via histogram equalisation. The enhanced image is further

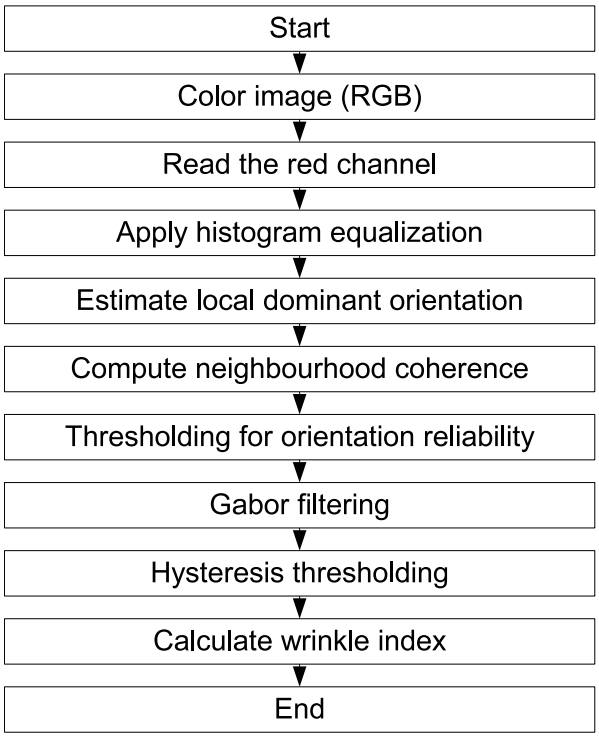
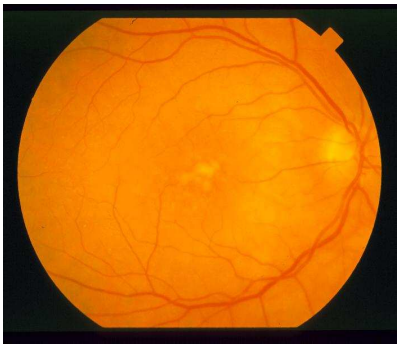
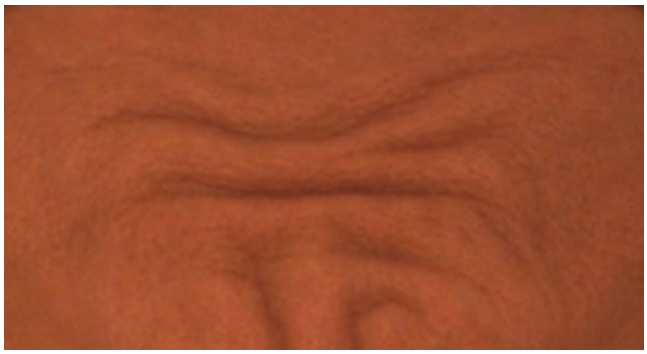


Figure 3.7: Flow chart of CLM method.

estimated via the corresponding oriented texture field, where for each pixel in the image the local dominant orientation (Chen et al., 2005) within a local window (centered around the current pixel) is estimated. Because, in general, skin has alternating smooth areas and wrinkled areas, a reliability image of the estimated orientation is calculated, based on the coherence of the image covariance within the neighbourhood of each pixel. The reliability image is adaptively thresholded such that areas in the image with high orientation reliability are preserved as skin areas with a high chance of presenting wrinkles. The final enhanced image is obtained by filtering the original normalised image with finely tuned Gabor filters, oriented according to the estimated local orientation. To detect the presence of the wrinkles, the enhanced image is binarised by globally thresholding its values using hysteresis thresholding, the result being a wrinkle mask, which is further used to select from the enhanced image only the areas where wrinkles appear. The resulting image is a wrinkle mask where the value is 0 indicates that there are no skin wrinkles, or Gabor filter responses which reflect the degree of shading in the image. They argues that Gabor responses are correlated to the depth of wrinkles. According to Gartstein and Shaya (1996), the length of individual wrinkles correlates well with the perceived age of the subjects, thus they defined the wrinkle measure, WI, as the product between wrinkle depth and length.



Retinal Image



Wrinkle Image

Figure 3.8: Retinal versus wrinkle image. Retinal image was obtained from Hoover and Goldbaum (2003) and wrinkle image was from FORERUS.

3.1.2.2 Frangi Method

Frangi et al. (1998) proposed a vessel enhancement filter, referred to here as Frangi Filter (FRF), for extracting blood vessels and interpreting the dataset of Magnetic Resonance Angiography (MRA) cerebral vasculature. It has been widely used in retinal vessel detection (Sofka and Stewart, 2006; Martinez-Perez et al., 2007), but not for wrinkle detection. FRF is based on the use of the second order partial derivatives for the ridge and valley detection.

Pertinent equation of FRF is derived from Frangi et al. (1998). A common approach to analyse the local behaviour of an image, $\mathbf{I}$, is to consider its Taylor expansion in the neighbourhood of a point $\mathbf{x}_0$,

$$\mathbf{I}(\mathbf{x}) \approx \mathbf{I}(\mathbf{x}_0) + (\mathbf{x}-\mathbf{x}_0)^T \nabla \mathbf{I}(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T \mathcal{H}\mathbf{I}(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0) \tag{3.14}$$

This expansion approximates the structure of the image up to the second order where ∇, $\mathcal{H}$, are the gradient vector and Hessian matrix of the image computed in $\mathbf{x}_0$ at scale ϑ. They defined the differentiation as a convolution with derivatives of Gaussian as,

$$\frac{d}{d\mathbf{x}}\mathbf{I}(\mathbf{x},\vartheta) = \vartheta^{\Psi}\mathbf{I}(\mathbf{x}) * \frac{d}{d\mathbf{x}}\mathbb{G}(\mathbf{x},\vartheta) \tag{3.15}$$

where Ψ is a normalised factor and $\mathbb{G}$ is a D-dimensional Gaussian defined as,

$$\mathbb{G}(\mathbf{x},\vartheta) = \frac{1}{\sqrt{2\pi\vartheta^2}^D}\exp -\frac{|\mathbf{x}|^2}{2\vartheta^2} \tag{3.16}$$

The parameter Ψ was introduced by Lindeberg (1998) to define a family of normalised

derivatives. This normalisation is particularly important for a fair comparison of the response of differential operators at multiple scales. When no scale is preferred Ψ should be set to unity.

Let $\lambda_{\vartheta,k}$ denote the eigenvalue corresponding to the k^{th} normalised eigenvector, $\upsilon_{\vartheta,k}$, of the $\mathcal{H}_{0,\vartheta}$, all computed at scale ϑ. From the definition of eigenvalues,

$$\mathcal{H}_{0,\vartheta}\upsilon_{\vartheta,k} = \lambda_{\vartheta,k}\upsilon_{\vartheta,k} \tag{3.17}$$

and it follows that

$$\upsilon_{\vartheta,k}^{T}\mathcal{H}_{0,\vartheta}\upsilon_{\vartheta,k} = \lambda_{\vartheta,k} \tag{3.18}$$

The eigenvalue decomposition extracts three orthonormal[1] directions which are invariant up to a scaling factor when mapped by the Hessian matrix. λ_k will be the eigenvalue with the k^{th} smallest magnitude. In particular, a pixel belonging to a vessel region will be signalled by λ_1 being small or ideally zero, and λ_2 and λ_3 of a large magnitude and equal sign. The sign is an indicator of brightness or darkness. In their experiment, vessels emerge as bright tubular structures in a darker environment. To summary, for an ideal tubular structure in a 3D image,

$$|\lambda_1| \approx 0 \tag{3.19a}$$

$$|\lambda_1| \ll |\lambda_2| \tag{3.19b}$$

$$\lambda_2 \approx \lambda_3 \tag{3.19c}$$

where the sign of λ_2 and λ_3 indicate its polarity. Since both wrinkle detection and vessel enhancement present similar line patterns as shown in Figure 3.8, it would be interesting to find out the performance of FRF on wrinkle detection. Details of FRF performance for wrinkle detection are given in Chapter 4.

3.2 Facial Age Descriptors

In this section, different facial feature representations are discussed.

[1] A set of vectors is orthonormal if every vector has magnitude 1 and the set of vectors are mutually orthogonal. 2 vectors are orthogonal if they are perpendicular to each other, i.e., the dot product of the two vectors is zero (Axler, 1997).

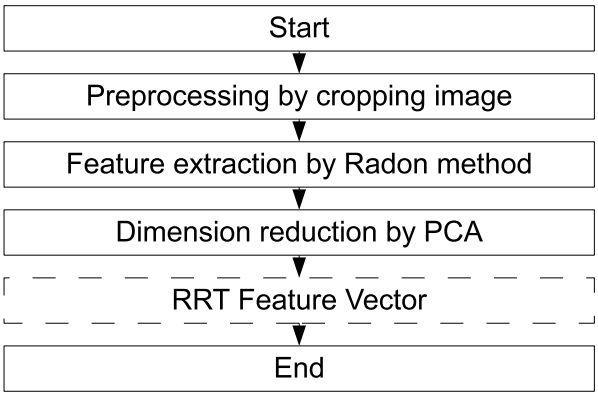


Figure 3.9: Flow chart of RRT method. Dotted rectangle shows the output vector of RRT.

3.2.1 Local Feature Representation

In this section, four different methods of aging representation are discussed. It includes the Radon method, Face++, BIF and KLBP.

3.2.1.1 Regional Radon Transform

Günay and Nabiyev (2013) proposed Regional Radon Transform (RRT) features for age estimation. It consists of three modules: preprocessing, feature extraction with RRT, and dimensionality reduction using PCA, as shown in Figure 3.9. In the preprocessing module, the facial images are cropped, scaled and transformed to the size of 88×88, based on the eye center locations. For feature extraction, RRT computes the projection of an image along specified direction. In third step, RRT features are projected into a lower dimension using PCA in order to improve the efficiency. A D-dimensional feature vector $\mathbf{i}$ is transformed into a d-dimensional vector $\mathbf{j}$ with $d < D$. Günay and Nabiyev (2013) claimed that RRT features convert pixels and geometric information into a low dimension representation. It brings an advantage of achieving global geometric affine invariance. Given a set of distances r and angles θ, the projection of RRT along specified direction is defined as

$$\mathrm{RRT}(r,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\Delta(r - x\cos\theta - y\sin\theta)\,d_x d_y \tag{3.20}$$

where RRT is the line integral of a 2D function f along a line from $-\infty$ to ∞, $\Delta(\cdot)$ is a Dirac delta function, r is a radius and $\theta = k\pi/6$ where $k = 0,1,2,3,4,5$. In its discrete form, a Radon transform is the summation of pixel intensities along line of different

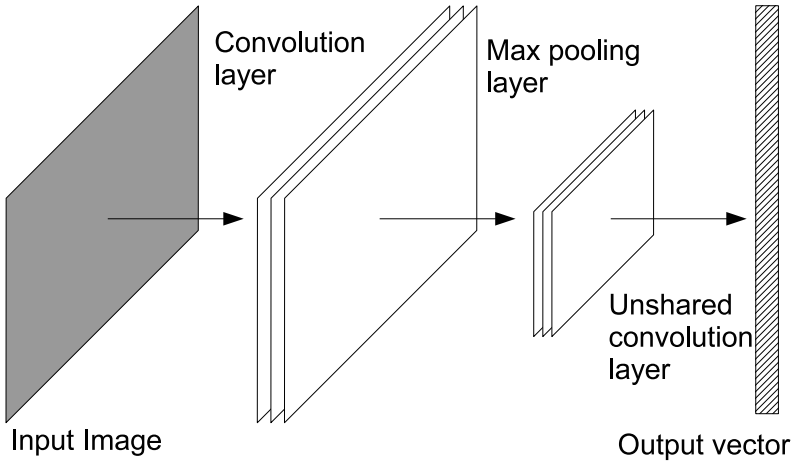


Figure 3.10: An illustration of CNN architecture. Convolution, max-pooling and non-linear operators are three key modules to form a CNN. Note that this figure was redrawn from Zhou et al. (2013).

directions and it is concatenated into a single feature vector. It produces a feature vector of 3920 units (Günay and Nabiyev, 2013). Finally, the age of vector **j** could be estimated by regression function such as linear regression (Günay and Nabiyev, 2013). Their results showed that a high resolution dataset, FERET achieves a higher MAE of 6.98 than a low resolution dataset, FGNET, with a MAE of 6.18. This indicates RRT features might not a reliable representation for age estimation.

3.2.1.2 Face++

Face++ is an online tool for facial landmark detection (Zhou et al., 2013; Inc., 2013). In addition, it provides additional prediction for facial analysis such as gender, age, race and expression. Face++ utilised a deep learning approach, Convolutional Neural Networks (CNN), with two advantages, geometric constraints among facial points are implicitly utilised and huge amount of training data can be leveraged. CNN design is a four-level convolutional network cascade, which tackles the problem in a coarse-to-fine manner. The network takes the raw pixels as input and performs regression on the coordinates of the desired points. CNN is a composition of multiple linear and non-linear operators. The first operator is the convolution layer which filters the multi-channel image signal. The second operator is max-pooling which reduces the size of the image. Third is a non-linear operator of unshared convolution layer where different positions are computed, so the layer is local-receptive rather than convolutional. CNN is illustrated in Figure 3.10. Convolution, max-pooling and non-linear operators are three

Figure 3.11: Flow chart of BIF descriptor.

key modules to form a CNN. It could be extended to a multi-layer CNN by repeating the operators needed. The output vector is represented by pixel values of all the image patches after unshared convolution layer.

According to Fan et al. (2014), FACE++ can be extended to the traditional tasks of face analysis such as gender analysis by a Look-Up-Table (LUT) approach. It identifies the gender on same bin that share common attributes in LUT table. They claimed that a classification accuracy of 96.8% is achieved on gender classification. In this work, FACE++ is applied as a benchmark algorithm for face age estimation. The performance of FACE++ is described in Chapter 5.

3.2.1.3 Bio-inspired Features

Recently, Bio-inspired Features (BIF) (Riesenhuber and Poggio, 1999) have shown good performance for age estimation (Guo et al., 2009a,b), as well as object category recognition (Serre et al., 2005; Mutch and Lowe, 2008) and face recognition (Meyers and Wolf, 2008). A specially-designed BIF with two layers: the simple layer $S1$ and complex layer $C1$ as shown in Figure 3.11.

The $S1$ layer contains a set of Gabor filters with parameters designed based on visual cortex models (Serre et al., 2005), and it is defined as

$$G(x,y) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\zeta^2}\right) \times \cos\left(\frac{2\pi}{\chi}x'\right) \tag{3.21}$$

where $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$ are the rotations of the Gabor filters for angle θ which varies between 0 and π. The aspect ratio is fixed as $\gamma = 0.3$, the effective width ζ, the wavelength χ as well as the filter sizes $\hat{s}$ were adjusted accordingly as Table 1 in Guo et al. (2009a). The orientation θ varies from 0 to π uniformly with different intervals, resulting in different numbers of total orientations, such as 4, 6, 8, 10, and 12.

The $C1$ layer contains some non-linear operations including the "MAX" (maximum) pooling and an "STD" (standard deviation) operation (Guo et al., 2009a), in order to have some invariance to translation, rotation, and scaling, as well as a characterisation

of the aging details. The "STD" is defined as

$$\text{STD} = \sqrt{\frac{1}{\hat{N}_S \times \hat{N}_S} \sum_{i=1}^{\hat{N}_S \times \hat{N}_S} (F_i - \bar{F})^2} \tag{3.22}$$

which reveals the variability in the data within a neighborhood $\hat{N}_S \times \hat{N}_S$ of $S1$ units, where F_i is the maximum value of two consecutive $S1$ units in the same scale band S (there are 8 bands in total but using different filters) at pixel index i,

$$F_i = \max\left(x_i^j, x_i^{j+1}\right) \tag{3.23}$$

where x_i^j and x_i^{j+1} are the filtered values with scales j and $j+1$ at position i. $\bar{F}$ is the mean value of the filtered values within the neighbourhood $\hat{N}_S \times \hat{N}_S$. The pooling of the "MAX" operation over two consecutive scales (i.e., in the same scale band) increases the tolerance to 2D transformations, such as scale changes with a small amount. The "MAX" operation merges two filtered images using filters of the same orientation but different scales into one, and then the "STD" operation is performed on the merged image within a neighbourhood $\hat{N}_S \times \hat{N}_S$. According to Guo et al. (2009a), the original model of the second "MAX" can tolerate more shift and size changes, but it cannot reveal the variability in the data. For age estimation, the description of local variability in data might be important for subtle age variation, such as wrinkles.

In order to estimate the face age, Guo et al. (2009a) applied the linear SVM for age classification and SVR for age regression. In this work, the BIF representation is used for age regression and it is defined as

$$\text{BIF} = \left[C1_1^{\text{STD}}, C1_2^{\text{STD}}, \cdots, C1_n^{\text{STD}}\right] \tag{3.24}$$

where n is the feature dimension of BIF representation.

3.2.1.4 Kernel-based Local Binary Pattern

The current state-of-the-art method for local-based age estimation is proposed by Ylioinas et al. (2013). The method is based on KLBP. It generates the sign and magnitude features through the face patches as shown in Figure 3.12. They claimed that the sparse nature of LBP representation is improved by the proposed kernel estimator. However, this method is not yet tested on FERET.

According to Ylioinas et al. (2013), the step before representing a face image is to

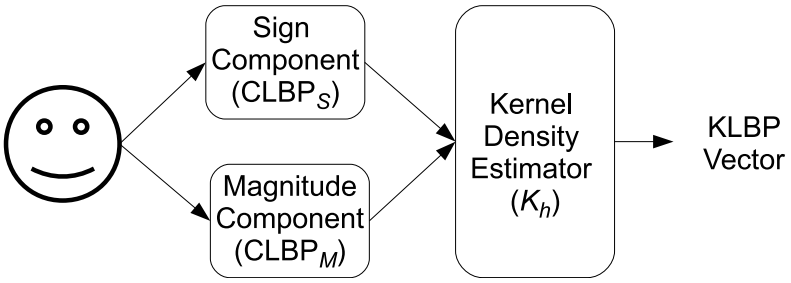


Figure 3.12: Flow chart of KLBP descriptor.

first align them to reduce the effect of scale, rotation, translation variations. Hence, to get rid of the most of the variation, clearly visible in the images of Figure 2.1, all face images are geometrically normalised by a similarity transformation with respect to both eyes and corners of the mouth. Once normalised, local features are extracted from uniformly distributed patches across the face. A face image is first divided into a set of L overlapping patches of a size 13×13 pixels, each patch overlapping its vertical and horizontal neighbours by 4 units. With a face image of size 76×76, this results $L = 64$ patches. In a particular patch, it is decomposed into two complementary components, sign and magnitude. The sign component is coded using the conventional LBP operator defined as

$$\text{LBP}^{S}_{P,R} = \sum_{p=0}^{P-1} \mathbb{T}\,(g_p - g_c)2^p \tag{3.25}$$

where g_c corresponds to the grey value of the center pixel (x_c, y_c), g_p refers to grey values of P equally spaced pixels on a circle of radius R, and $\mathbb{T}$ defines a thresholding function with $\mathbb{T}(x) = 1$ if $x \geq 0$ and $\mathbb{T}(x) = 0$ otherwise. The magitude component is defined as

$$\text{LBP}^{M}_{P,R} = \sum_{p=0}^{P-1} \mathbb{T}\,(\hat{m}_p, \hat{c})2^p \tag{3.26}$$

where $\hat{m}_p$ is the magnitude of local pixel difference and $\hat{c}$ a predetermined threshold value of LBP usually set as the mean value of local pixel differences in the whole image. As the magnitude operator encodes the difference in local pixel intensities, it gives a measure of contrast. The key idea of LBP is to gain more comprehensive image representation by combining these two complementary descriptions.

According to Ylioinas et al. (2013), kernel method proposed by Aitchison and Aitken (1976) is more suitable for estimating the probability distribution of LBP-like random

variables. The kernel is given by

$$K_{\hat{h}}\left(\hat{l}|\hat{l}'\right) = \hat{h}^{P-\tilde{d}\left(\hat{l},\hat{l}'\right)}\left(1-\hat{h}\right)^{\tilde{d}\left(\hat{l},\hat{l}'\right)} \tag{3.27}$$

where $\hat{l}$ and $\hat{l}'$ are both P-dimensional binary variables, $\tilde{d}$ is the Hamming distance between them, and $\hat{h}$ is a bandwidth parameter. Finally, using the given kernel instead of a histogram, KLBP pattern is a concatenation of both sign and magnitude components of LBP after multiplication with the kernel as

$$\text{KLBP} = \left[\text{LBP}^{S}_{1,P,R} * K_h, \text{LBP}^{M}_{1,P,R} * K_h \cdots, \text{LBP}^{S}_{n,P,R} * K_h, \text{LBP}^{M}_{n,P,R} * K_h\right] \tag{3.28}$$

where n is the total number of KLBP features. The age is estimated using the KLBP features and a SVR with a non-linear radial basis function.

3.2.2 Global Feature Representation

In this section, different global feature representations are discussed.

3.2.2.1 Principal Component Analysis

PCA is an orthogonal linear transformation that projects the data into a new coordinate system with greatest variance as principal component, the second greatest variance on the second coordinate, and so on (Jolliffe, 2002). According to Cootes and Taylor (2001), there is a quicker way of calculating PCA when there are fewer samples than dimensions. Suppose we wish to apply a PCA to vectors, $\mathbf{x}_i$ ($\mathbf{x}_i \in \mathbf{D}$), from $n\mathbf{D}$ to $s\mathbf{D}$, where $s < n$. The covariance matrix is $n \times n$, which may be very large. However, we can calculate its eigenvectors and eigenvalues from a smaller $s \times s$ matrix derived from the data. Because the time taken for an eigenvector decomposition goes as the cube of the size of the matrix, this can give considerable savings. Subract the mean from each data vector and put them into the matrix $\mathbf{D}$

$$\mathbf{D} = \left[(\mathbf{x}_1 - \bar{\mathbf{x}}) \,|\, \ldots \,|\, (\mathbf{x}_s - \bar{\mathbf{x}})\right] \tag{3.29}$$

The covariance matrix can be written

$$\mathbf{S} = \frac{1}{s}\mathbf{D}\mathbf{D}^T \tag{3.30}$$

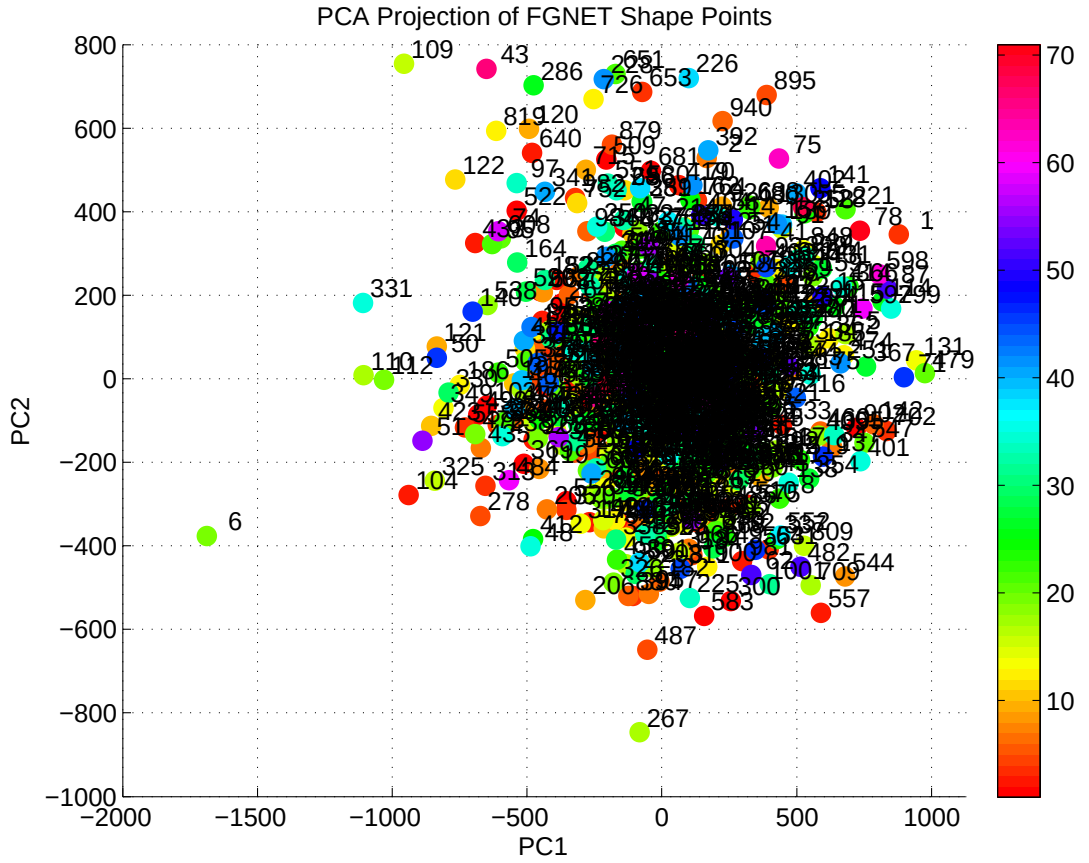


Figure 3.13: PCA projection of FGNET shape points. There are 1002 instances on FGNET where each instance consists of 68 landmarks. The landmarks in an instance are represented by (x,y) and it is sorted into a single vector with 136 units. Based on this vector, the mean, covariance, eigenvector and eigenvalues are calculated. Label shows the subject identity in FGNET and the legend presents the age of each subject. Axis-x and -y are first and second principal components. Instance of 6 is far away from the main cluster because it is the largest image compared to others.

Let $\mathbf{T}$ be the $s \times s$ matrix

$$\mathbf{T} = \frac{1}{s}\mathbf{D}^T\mathbf{D} \tag{3.31}$$

Let $\mathbf{e}_i$ be the s eigenvectors of $\mathbf{T}$ with corresponding eigenvalues φ_i, sorted into descending order. It can be shown that the s vectors $\mathbf{D} \cdot \mathbf{e}_i$ are all eigenvectors of $\mathbf{S}$ with corresponding eigenvalues φ_i, and that all remaining eigenvectors of $\mathbf{S}$ have zero eigenvalues. Note that $\mathbf{D} \cdot \mathbf{e}_i$ is not necessarily of unit length so may require normalising.

Figure 3.13 presents a PCA projection of FGNET shape points of 68 landmarks. It is noticed that the 6th object is far away from the main cluster and this object is the

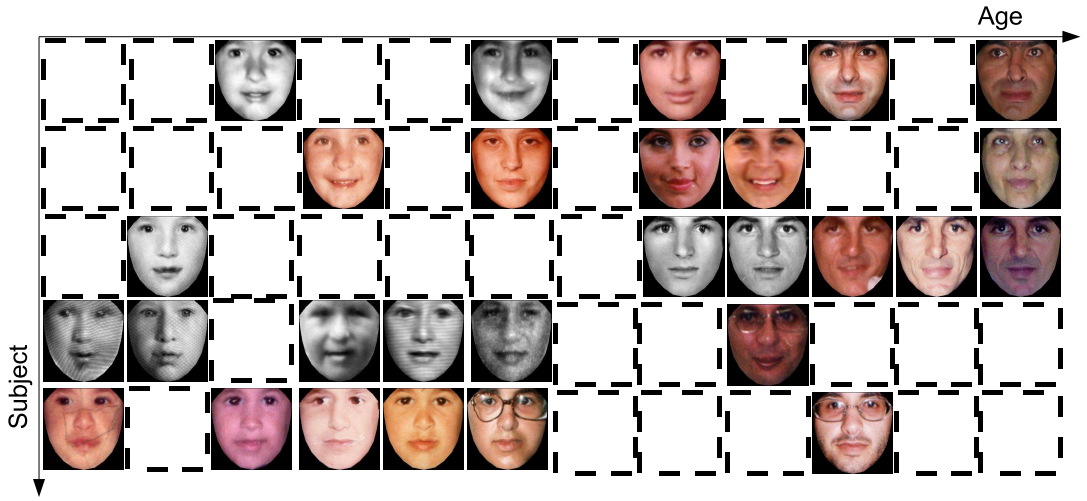


Figure 3.14: An illustration of AGES. Each row shows the same subject and each column represents a chronological age. The main idea of AGES is to reconstruct the empty space (dotted rectangle) by the known subspace (the images). Note that this figure was redrawn from FGNET and the images were sorted randomly for the illustration purpose.

largest image size in FGNET where the resolution is 639×772 compare to the rest where the resolution is around 400×400. It is one of the greatest benefit of PCA on analysing the data.

PCA has been widely used in facial analysis. Buchala et al. (2005) applied it to encode different properties of face such as gender, ethnicity, age and identity. Geng et al. (2007) used PCA to construct a subspace of missing images and then predict the age of unseen image based on PCA reconstruction. It is elaborated in the next section.

3.2.2.2 Aging Pattern Subspace

Geng et al. (2007) proposed an automatic age estimation method named AGES. A representative model for the aging patterns can be built up by the information theory approach of coding and decoding. One widely adopted technology is using PCA to construct a subspace that captures the main variation in the data set (Jolliffe, 2002). The basic idea of AGES is to model the aging pattern, which is defined as the sequence of individuals face images sorted in time order, by constructing a representative subspace as shown in Figure 3.14. The projection in the subspace is computed by

$$\mathbf{y} = \mathbf{W}^T (\rho - \mu) \tag{3.32}$$

where μ is the mean vector of ρ and $\mathbf{W}^T = \mathbf{W}^{-1}$ is the transpose of transformation matrix, $\mathbf{W}$, which is composed by the orthogonal eigenvectors of the covariance matrix of ρ. The difficulty is that the aging pattern vector ρ is highly incomplete. Note that, the vector ρ is initialised by filling in the known appearance model parameters according to age. Based on the characteristics of aging patterns, an Expectation Maximization (EM)-like algorithm is proposed here to learn a representative subspace.

Suppose the training set has N aging pattern vectors $\mathbf{D} = \{\rho_1, \cdots, \rho_N\}$. Any sample in this set can be written as

$$\rho_k = \{\rho_k^a, \rho_k^m\} \tag{3.33}$$

where ρ_k^a are the available features and ρ_k^m are the missing features of ρ_k. Suppose the transformation matrix is $\mathbf{W}$, the projection $\mathbf{y}_k$ of ρ_k in the subspace can be calculated by Eq. (3.32) and the reconstruction of ρ_k is calculated by

$$\phi_k = \mu + \mathbf{W}\mathbf{y}_k \tag{3.34}$$

ϕ_k can also be written as $\phi_k = \{\phi_k^a, \phi_k^m\}$, where ϕ_k^a is the reconstruction of ρ_k^a and ϕ_k^m is the reconstruction of ρ_k^m. It is well known that standard PCA can be derived by minimizing the mean reconstruction error (residuals) of the data set $\mathbf{D}$ in the subspace. With the presence of the missing features ρ_k^m, the goal is changed into finding a $\mathbf{W}$ that minimizes the mean reconstruction error ($\bar{\varepsilon}$) of the available features

$$\bar{\varepsilon}^a = \frac{1}{N}\sum_{k=1}^{N}(\rho_k^a - \phi_k^a)^T(\rho_k^a - \phi_k^a) \tag{3.35}$$

In case the number of missing features in different instances is highly uneven, Eq. (3.35) should be normalised by the dimensionality of the missing part. This is equivalent to a preprocess of dividing each instance by its missing dimensionality.

3.2.2.3 Facial Appearance Model

FAM establish a compact parameterisation of object variability, as learned from a training set by estimating a set of latent variables (Cootes and Taylor, 2001; Stegmann et al., 2003; Matthews and Baker, 2004). It decouples and models two parts of an object: shape and texture. The shape is a vector formed by concatenating the position elements of the labelled landmarks, while the texture means the measure of pixels, which is usually represented by intensities or colours. Figure 3.15 shows an example of

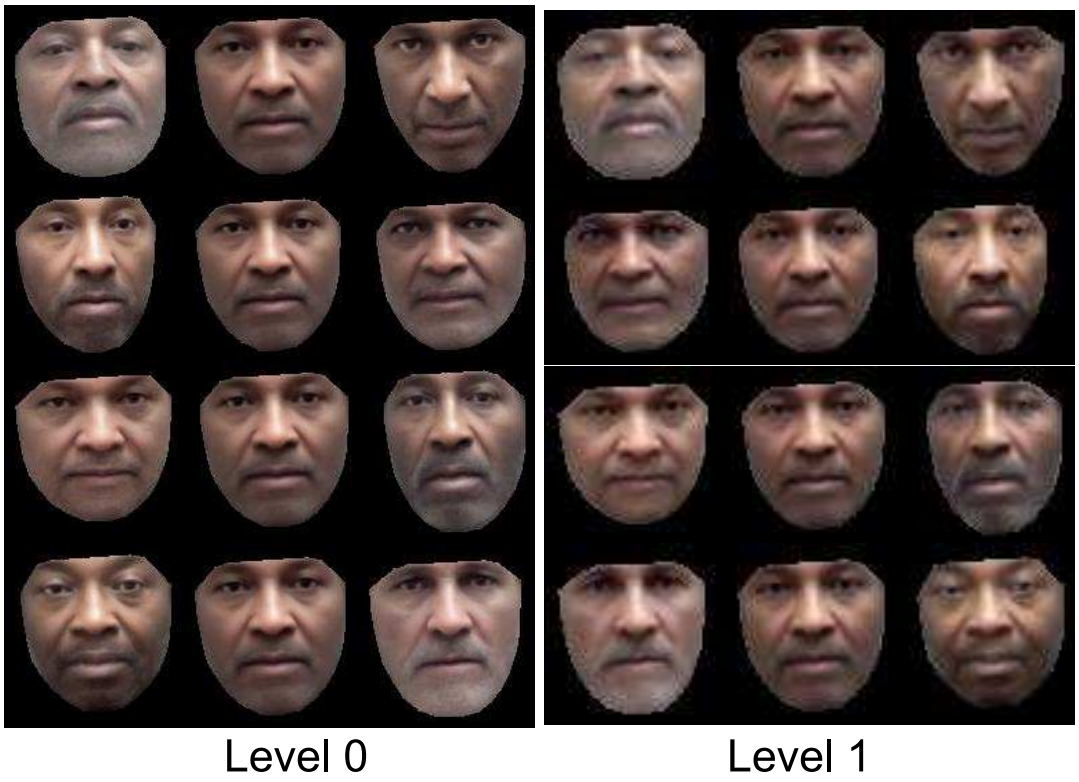
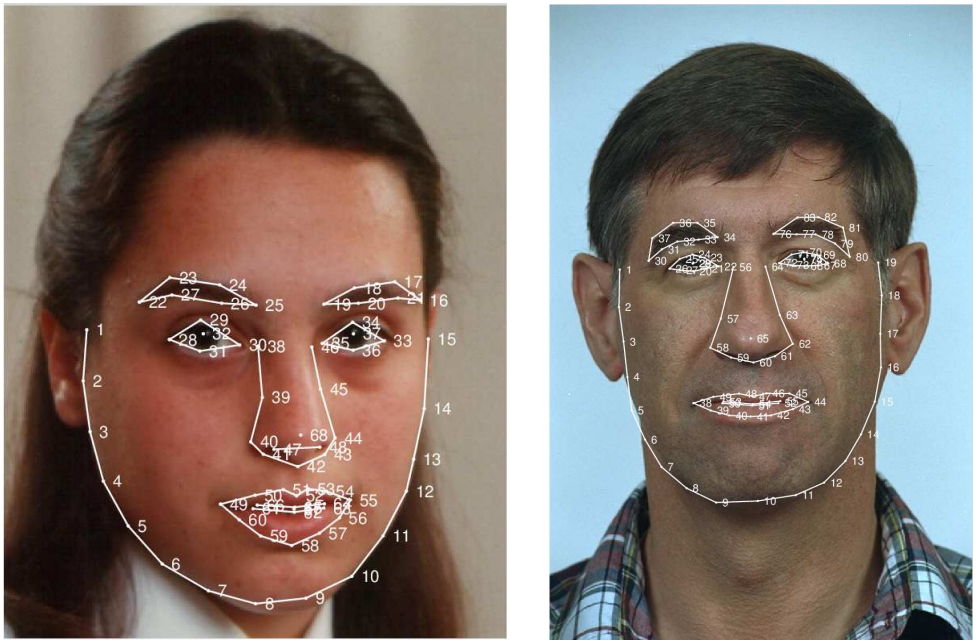


Figure 3.15: A combined appearance model. Each column shows different levels of image resolution and each row presents different modes of model. Note that this figure was redrawn from MORPH using *am_tools* of Cootes and Taylor (2001).

a combined appearance model where each column shows different levels of image resolution and each row presents different modes of model. Also, a training set of images with correspondingly labelled landmarks is required for model training. A face image with labelled landmarks is shown in Figure 3.16. Further, given a suitable warp function a dense (i.e. per-pixel) correspondence is established between training objects, thereby enabling a proper modelling of texture variability. By exploiting prior knowledge of the nature of the optimization space, these models of shape and texture can be rapidly fitted to unseen images, thus providing image interpretation through synthesis.

Variability is modelled by the means of a PCA where an eigen analysis of the dispersions of shape and texture. Given N training samples for an object class and let each sample be represented by a set of $\mathbb{P}$ landmarks. The shape samples are aligned to a common mean shape (or shape-free patch) using Generalised Procrustes Analysis (GPA) (Goodall, 1991; Cootes and Taylor, 2001) where all effects of translation, rotation and scaling are removed. The alignment of each shape is the sum of distances



68 points 83 points

Figure 3.16: Landmark points. The left image is based on 68 landmark points of FGNET and the right image is based on 83 landmark points of FACE++. Both images were redrawn from FGNET and FERET.

of each shape to the mean as

$$\mathcal{D} = \sum |\mathbf{x}_i - \bar{\mathbf{x}}|^2 \tag{3.36}$$

where $\mathcal{D}$ is the Procrustes distance and $\mathbf{x}$ is a set of n landmark points $\{(x_i, y_i)\}$ of a single example. It is poorly defined unless constraints are placed on the alignment of the mean (for instance, ensuring it is centred on the origin, has unit scale and some fixed but arbitrary orientation) (Cootes and Taylor, 2001).

In the next step, the texture samples are then warped into correspondence using a Piece-wise Affine Warping (PAW) (see Section 3.4.2) and subsequently sampled from this shape-free reference. Let $\mathbf{s}$ and $\mathbf{t}$ denote a synthesized shape and texture of an image in the reference frame and let $\bar{\mathbf{s}}$ and $\bar{\mathbf{t}}$ denote the corresponding sample means. New instances are now generated by adjusting the principal component scores, $\mathbf{b}_s$ and $\mathbf{b}_t$ in

$$\mathbf{s} = \bar{\mathbf{s}} + \Phi_s \mathbf{b}_s \tag{3.37}$$

$$\mathbf{t} = \bar{\mathbf{t}} + \Phi_t \mathbf{b}_t \tag{3.38}$$

where Φ_s and Φ_t are matrices of column eigenvectors of the shape and texture dispersions estimated from the training set. To obtain a combined shape and texture parameterization, $\mathbf{c}$, the values of $\mathbf{b}_s$ and $\mathbf{b}_t$ over the training set are combined into

$$\mathbf{b} = \begin{bmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{W}_s \Phi_s^T (\mathbf{s} - \bar{\mathbf{s}}) \\ \Phi_t^T (\mathbf{t} - \bar{\mathbf{t}}) \end{bmatrix} \tag{3.39}$$

A suitable weighting between pixel distances and pixel intensities is carried out through the diagonal matrix $\mathbf{W}_s$. To make the normalised measures of pixel distance and pixel intensities commensurate, the shape principal component scores are typically weighted by the square root of the ratio between the sums of the texture and shape eigenvalues.

To recover any correlation between shape and texture the two eigen-spaces are usually coupled through a third principal component transform as

$$\mathbf{b} = \Phi_c \mathbf{c} = \begin{bmatrix} \Phi_{c,s} \\ \Phi_{c,t} \end{bmatrix} \mathbf{c} \tag{3.40}$$

where $\mathbf{c}$ is a combined appearance model parameters that generates new object instances by

$$\mathbf{s} = \bar{\mathbf{s}} + \Phi_s \mathbf{W}_s^{-1} \Phi_{c,s} \mathbf{c} \tag{3.41}$$

$$\mathbf{t} = \bar{\mathbf{t}} + \Phi_t \Phi_{c,t} \mathbf{c} \tag{3.42}$$

To regularize the model and improve speed and compactness, Φ_s, Φ_t and Φ_c are truncated, usually such that a certain amount of variance in the training set is preserved. This eventually results in k ($k < N$) combined modes where N is the total images of training set. Details of FAM can be found in Cootes and Taylor (2001).

3.3 Estimation Algorithms

In this section, different methods of classification and regression are discussed.

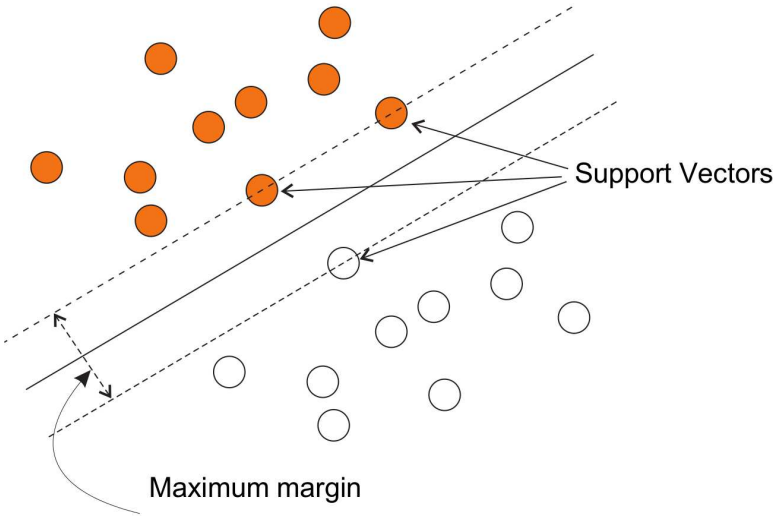


Figure 3.17: An illustration of SVM hyperplane. Different colors show the objects of different class. Support vectors contribute to a hyperplane separation where a maximum margin is determined. This figure was redrawn from the article of Selamat et al. (2009).

3.3.1 Classification

In this section, two classification methods, SVM and MLP, are described.

3.3.1.1 Support Vector Machine

A SVM is a statistical classification method proposed by Cortes and Vapnik (1995). Based on the principle of Structural Risk Minimization (SRM) principle, SVM tries to find a separating hyperplane with maximum margin to separate the positive examples and negative examples from the training data sets. It makes decisions based on the support vectors that are selected as the only effective elements from the training set. SVMs belong to the general category of kernel methods (Smola and Schölkopf, 1998; Shawe-Taylor and Cristianini, 2004; Ben-Hur and Weston, 2010). A kernel method is an algorithm that depends on the data only through dot-product. When this is the case, the dot-product can be replaced by a kernel function which computes a dot-product in some possibly high dimensional feature space. This has two advantages: (1) the ability to generate non-linear decision boundaries using methods designed for linear classifiers. Second, the use of kernel functions allows the use of a classifier to the data that have no obvious fixed-dimensional vector space representation. For example, the features of facial image analysis such as face recognition, face age estimation and facial expression recognition. Figure 3.17 shows SVM finds the hyperplane, which is

separated from the positive and negative training examples with a maximum margin. The examples that are close to the hyperplane are called support vectors, which are marked with arrows.

Given a training set of instance-label pairs, $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, where $(\mathbf{x}_i) \in \Re^d$ and $y_i \in \{-1, +1\}$. $\Re^d$ is a real number in d-dimension and y_i is a class label. SVM requires the solution of the following optimization problem,

$$\underset{\mathbf{w},b,\xi}{\text{minimize}} \qquad \frac{1}{2}\mathbf{w}^T\mathbf{w} + \tilde{k}\sum_{i=1}^{n}\xi_i \tag{3.43a}$$

$$\text{subject to: } y_i\left(\mathbf{w}^T M(\mathbf{x}_i) + \hat{b}\right) \geq 1 - \xi_i, \;\; \xi_i \geq 0 \tag{3.43b}$$

where training vector $(\mathbf{x}_i$ are mapped into a higher dimensional space by the vector function M, $\hat{b}$ is a bias, $\mathbf{w}$ is a linear combination of the training vector $\mathbf{x}_i$, $\xi \geq 0$ are slack variables that allow an example to be in a margin error ($0 \leq \xi_i \leq 1$) or to be misclassified ($\xi_i > 1$) and the constant $\tilde{k} > 0$ sets the relative importance of maximizing the margin and minimizing the amount of slack. This formulation is called the soft-margin SVM and it was introduced by Cortes and Vapnik (1995). SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. Furthermore, $\kappa\left(\mathbf{x}_i, \mathbf{x}_j\right) \equiv M(\mathbf{x}_i)^T M\left(\mathbf{x}_j\right)$ is called the kernel function. There are four basis kernels used in SVM: linear, polynomial, Radial Basis Function (RBF), sigmoid. Each of the kernel is defined as:

$$\kappa_{\text{linear}}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \mathbf{x}_i^T \cdot \mathbf{x}_j \tag{3.44a}$$

$$\kappa_{\text{polynomial}}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left(\omega \cdot \mathbf{x}_i^T \cdot \mathbf{x}_j + \varsigma\right)^{\hat{d}}, \omega > 0 \tag{3.44b}$$

$$\kappa_{\text{RBF}}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \exp^{\left(-\omega \cdot \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)}, \omega > 0 \tag{3.44c}$$

$$\kappa_{\text{sigmoid}}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \tanh\left(\omega \cdot \mathbf{x}_i^T \cdot \mathbf{x}_j + \varsigma\right) \tag{3.44d}$$

where ω is a constant to adjust the width of the Gaussian function, ς is the coefficient, $\hat{d}$ is the degree of polynomials. In testing, the classification is given by

$$f(\mathbf{x}) = sign\left(\mathbf{w} \cdot \mathbf{x} + \hat{b}\right) \tag{3.45}$$

for any new data point $\mathbf{x}$. If the training data are non-separable, slack variables ξ can be used.

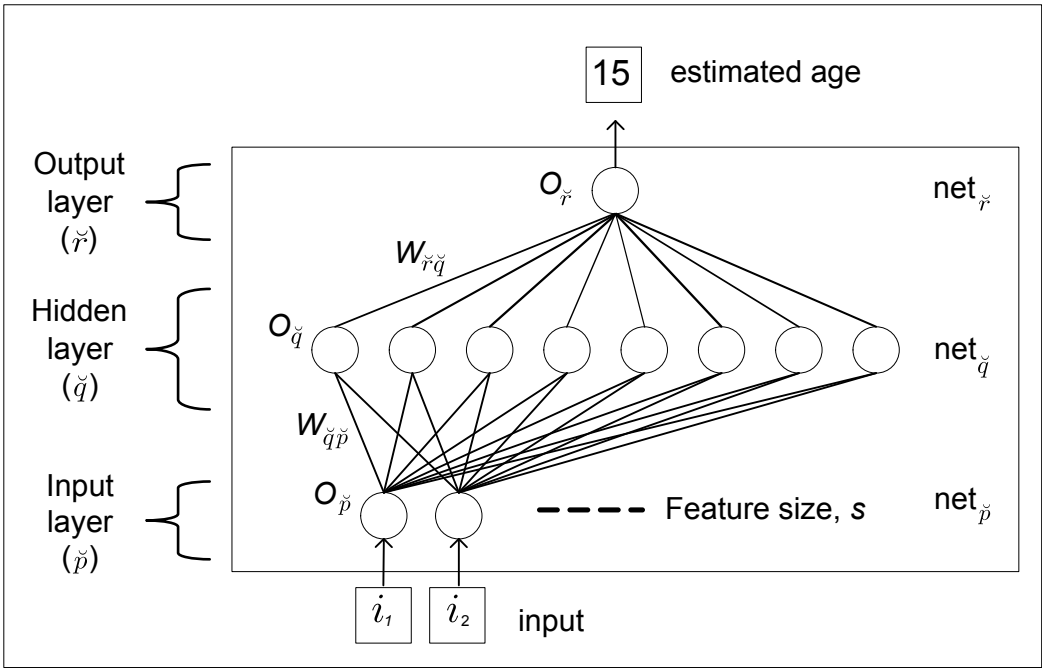


Figure 3.18: MLP architecture. Note that this figure was redrawn from the article of Selamat et al. (2009).

3.3.1.2 Multi-layer Perceptron

Figure 3.18 shows the example of architecture of MLP (Selamat et al., 2009). It has been applied by numerous studies as in Rumelhart et al. (2002); Lanitis et al. (2004); Takimoto et al. (2007); Dehshibi and Bastanfard (2010). According to Takimoto et al. (2007), MLP consists of at least three layers: one input layer $\breve{p}$, one hidden layer $\breve{q}$ (or more) and one output layer $\breve{r}$. Total nodes of an input layer is according to the number of selected facial feature, s. If the feature size, s, is 55 then the number of input layers will be set correspondingly. The output layer consists 1 unit only which is the age range from 12 to 68 years. The input values to the MLP are represented by $i_1, i_2, ..., i_s$, and normally it is scaled using a min-max normalisation. Adoption of the weight between hidden $\breve{q}$ and input $\breve{p}$ layers is given by,

$$W_{\breve{q}\breve{p}}(\ell+1) = W_{\breve{q}\breve{p}}(\ell) + \Delta W_{\breve{q}\breve{p}}(\ell+1) \tag{3.46}$$

$$\Delta W_{\breve{q}\breve{p}}(\ell+1) = \eta \varpi_{\breve{q}} O_{\breve{p}} + \hat{k} \Delta W_{\breve{q}\breve{p}}(\ell) \tag{3.47}$$

$$\varpi_{\breve{q}} = O_{\breve{q}}(1 - O_{\breve{q}}) \sum_{\breve{r}} \varpi_{\breve{r}} W_{\breve{r}\breve{q}} \tag{3.48}$$

where ℓ is the iteration number, η is a learning rate, $\hat{k}$ is a momentum rate, $\Omega_{\breve{p}}$ is the output units of $\breve{p}$, $\Omega_{\breve{q}}$ is the output units of $\breve{q}$, $\Omega_{\breve{r}}$ is the output units of $\breve{r}$, $W_{\breve{q}\breve{p}}$ is a $\breve{q}$ weight to the unit $\breve{p}$, $W_{\breve{r}\breve{q}}$ is a $\breve{r}$ weight to the unit $\breve{q}$, $\varpi_{\breve{q}}$ is the generalised error through a layer $\breve{q}$, and $\varpi_{\breve{r}}$ is the generalised error through a layer $\breve{q}$ and $\breve{r}$. Note that the first transfer function at hidden layer $\breve{q}$ is given by,

$$\text{net}_{\breve{q}} = \sum_{\breve{q}} W_{\breve{q}\breve{p}} O_{\breve{p}} + \Omega_{\breve{q}} \tag{3.49}$$

$$O_{\breve{q}} = f\left(\text{net}_{\breve{q}}\right) = 1/\left(1 + e^{-\text{net}_{\breve{q}}}\right) \tag{3.50}$$

where $\text{net}_{\breve{q}}$ is the first transfer function at hidden layer $\breve{q}$ and $\Omega_{\breve{q}}$ is a bias on hidden unit $\breve{q}$. Adaptation of the weights between output and hidden layers is given by,

$$W_{\breve{r}\breve{q}}(\ell + 1) = W_{\breve{r}\breve{q}}(\ell) + \Delta W_{\breve{r}\breve{q}}(\ell + 1) \tag{3.51}$$

$$\Delta W_{\breve{r}\breve{q}}(\ell + 1) = \eta \delta_{\breve{r}} O_{\breve{q}} + \hat{k} \Delta W_{\breve{r}\breve{q}}(\ell) \tag{3.52}$$

$$\delta_{\breve{r}} = O_{\breve{r}}(1 - O_{\breve{r}})(\ell_{\breve{r}} - O_{\breve{r}}) \tag{3.53}$$

Then the output function at the output layer $\breve{r}$ is given by,

$$\text{net}_{\breve{r}} = \sum_{\breve{r}} W_{\breve{r}\breve{q}} O_{\breve{q}} + \Omega_{\breve{r}} \tag{3.54}$$

$$O_{\breve{r}} = f\left(\text{net}_{\breve{r}}\right) = 1/\left(1 + e^{-\text{net}_{\breve{r}}}\right) \tag{3.55}$$

where $\text{net}_{\breve{r}}$ is the second transfer function at output layer $\breve{r}$ and $\Omega_{\breve{r}}$ is a bias on output unit $\breve{r}$. The output values to the MLP are represented by a real value which are corresponding to the age of a person. Table 3.1 shows the parameters of MLP given in Takimoto et al. (2007). The input node, hidden node and output node are 55, 30, and 1, respectively; learning cycle is 20000; and step size is between 0.1 and 0.3.

Table 3.1: Parameter settings of MLP derived from Takimoto et al. (2007).

Description	Value
Function	Backpropagation
Input Node	55
Hidden Node	30
Output Node	1
Epochs	20000
Step Size	0.1 - 0.3

3.3.2 Regression

In this section, two regression methods, quadratic regression and SVR with Sequential Minimal Optimization (SMO) optimization, are explained.

3.3.2.1 Quadratic Regression

Quadratic regression is the simplest algorithm to consider when value of the target function is numeric such as face age (Lanitis et al., 2002). The basic idea of quadratic function is to express the target as a linear combination of the attributes with predetermined weights. Given the extracted features for each face age, a regression function is often used to characterise the relationship between the extracted features $\mathbf{x}$ and the age labels A as

$$A(\mathbf{x}) = f(\mathbf{x}) \tag{3.56}$$

A typical choice of the regression function f is the Quadratic Model (QM). For example, Lanitis et al. (2002) proposed to use the QM function for age regression as

$$\hat{A}(\mathbf{x}) = w_0 + \mathbf{w}_1^T \mathbf{x} + \mathbf{w}_2^T \mathbf{x}^2 \tag{3.57}$$

where $\mathbf{x}$ and $\mathbf{x}^2$ are the extracted feature vector, $\hat{A}$ is the predicted age of input $\mathbf{x}$, w_0 is the offset, $\mathbf{w}_1$ and $\mathbf{w}_2$ are weight vectors.

The model parameters are optimized by minimizing the difference between the actual ages of the individuals, A, and the ages estimated using Eq. (3.57), i.e., $||A - \hat{A}||$. The loss function usually corresponds to a Least Square Estimation (LSE) criteria. However, there are some disadvantages for the QM method (Guo et al., 2008a): (1) the aging is a complex non-linear regression problem, especially for a large span of years, i.e., 0-90. The simple quadratic function may not model properly the complex aging

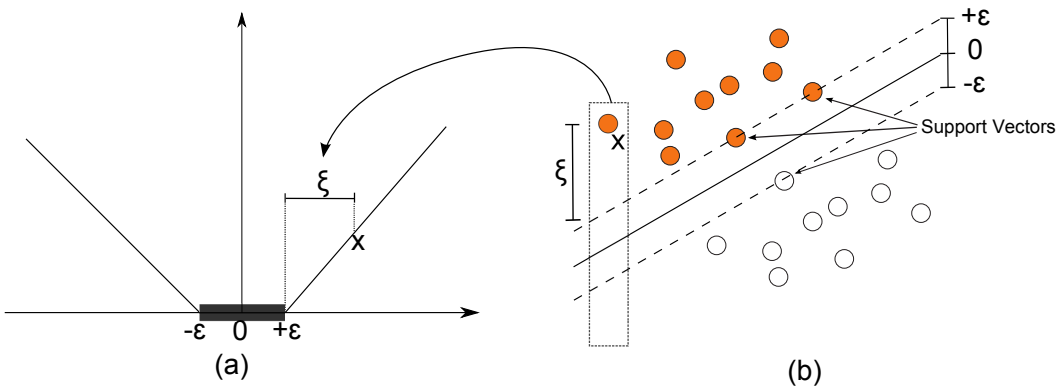


Figure 3.19: Linear SVR model. (a) ε-insensitive loss function where the horizontal and vertical axes are $\mathbf{x}$ and $f(\mathbf{x})$. Using a small ε is preferable when a large number of outliers are present. (b) An illustration on how support vectors lay in between $+\varepsilon$ and $-\varepsilon$ and the slack variable ξ from a particular instance $\mathbf{x}$. Note that this figure was redrawn from the article of Smola and Schölkopf (2004).

process; (2) the LSE is sensitive to outliers that come from incorrect labels in collecting a large image dataset; (3) the LSE may not generalize well for unseen examples which may increases the error for age prediction.

3.3.2.2 Support Vector Regression

The basic idea of SVR is to find a function f that has most ε deviation from the actually obtained target y_i for the training data $\mathbf{x}_i$. In other words, errors are less important as long as they are less than ε. Figure 3.19 depicts the situation graphically. Only the points outside the shaded region contribute to the regression function, as the deviations are penalised in a linear fashion. As stated by Guo et al. (2008a), this property determines SVR to be less sensitive to outliers than the quadratic function.

First, a notion of a linear SVR is defined. Consider the problem of approximating the set of data $\mathbf{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ where $\mathbf{x}_i \in \Re^d$ and $y_i \in \Re$ (y_i is a real number instead of class), a linear function (Guo et al., 2008a) is defined as

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + \hat{b} \tag{3.58}$$

where $\hat{b}$ is a bias, $\langle \cdot \rangle$ denotes the dot product. The optimal regression function (Vapnik, 1998) is given by

$$\underset{\mathbf{w},\xi}{\text{minimize}} \quad \frac{1}{2}\left\|\mathbf{w}^2\right\| + \tilde{k}\sum_{i=1}^{n}\left(\xi_i^+ + \xi_i^-\right) \tag{3.59a}$$

$$\text{subject to:} \quad \begin{aligned} & y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - \hat{b} \leq \varepsilon + \xi_i^+ \\ & \langle \mathbf{w}, \mathbf{x}_i \rangle + \hat{b} - y_i \leq \varepsilon + \xi_i^- \\ & \text{where } \xi_i^+, \xi_i^- \leq 0 \end{aligned} \tag{3.59b}$$

where constant $\tilde{k} > 0$ determines the tradeoff between the flatness of f and data deviations, and ξ_i^+, ξ_i^- are slack variables to cope with otherwise infeasible constraints on the optimization problem of Eq. (3.59). $\tilde{k}$ is a penalty factor where a high penalty for non-separable points may lead to over-fitting and vice versa (Alpaydin, 2014). Therefore, in practice, $\tilde{k}$ should be selected through a separate validation set or it can be determined through the technique of cross-validation. A ε determines the level of correctness of the approximated function (Cherkassky and Ma, 2003). It depends on the ground truth in the training set. If ε is larger than the range of the ground truth, a bad result would be expected. If ε is zero, the over-fitting would occur. Therefore, ε must be chosen to reflect the data in some way such as grid search.

According to Guo et al. (2008a), the linear regression cannot model the complex aging process. Therefore, a non-linear regression function may be required in practice to adequately model the data. It can be obtained using kernels such as polynomials and RBF functions as Eq. (3.44). Given the kernel mapping, the solution of the non-linear SVR is obtained by

$$\langle \mathbf{w}, \mathbf{x} \rangle = \sum_{i=1}^{n} \left(\rho_i - \rho_i^*\right) \kappa\left(\mathbf{x}_i, \mathbf{x}_j\right) \tag{3.60}$$

where ρ_i and ρ_i^* are Lagrange multipliers. The target is predicted as

$$f(\mathbf{x}) = \sum_{i=1}^{n} \left(\rho_i - \rho_i^*\right) \kappa\left(\mathbf{x}_i, \mathbf{x}_j\right) + \hat{b} \tag{3.61}$$

In the non-linear case, the optimization problem corresponds to finding the flattest or linear regression function in the higher dimensional feature space instead of input space (Guo et al., 2008a).

According to Platt et al. (1999), SMO does not require extra matrix storage except the amounts of memory required to store 2×2 matrices required by SMO. Traditional quadratic programming algorithms are not suitable for large size problems because of the following reasons (Shevade et al., 2000). First, they require that the kernel matrix

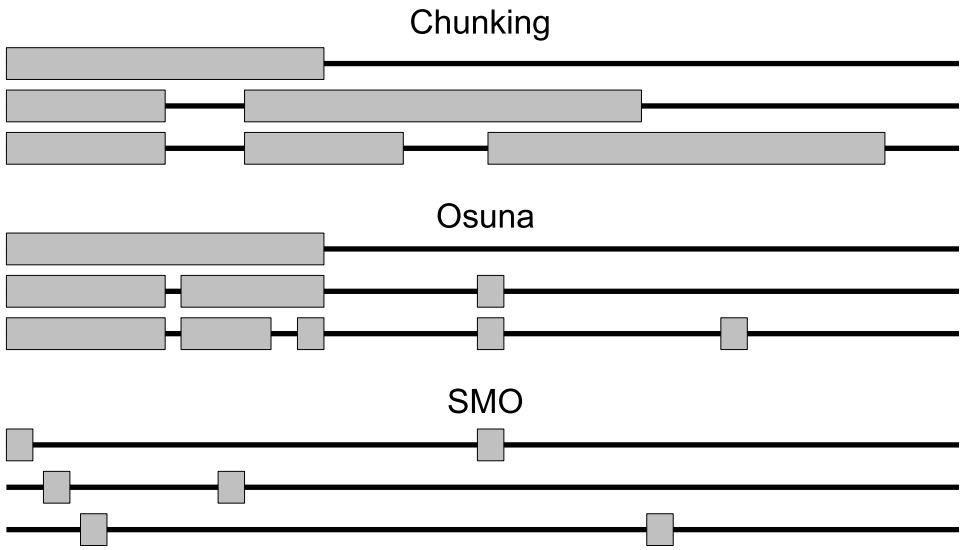


Figure 3.20: Different training methods of SVM: Chunking, Osuna's algorithm and SMO The horizontal thin line shows the training set, the thick boxes present the Lagrange multipliers being optimized at each iteration. A given group of three lines corresponds to three training iterations where the first iteration at the top. Note that this figure was redrawn from the article of Platt et al. (1999).

$\kappa\left(\mathbf{x}_i, \mathbf{x}_j\right)$ be computed and stored in memory. This requires extremely large memory. Second, these methods involve expensive matrix operations when decomposing a large sub-matrix of the kernel. Third, coding of these algorithms is difficult. Tsang et al. (2005) also claimed that SMO is very efficient in terms of both time and space, it often converges in a small number of iterations.

SMO is an iterative algorithm for solving the optimization problem by operating on a fixed size subset of the training set at a time. SMO breaks this problem into a series of smallest possible sub-problems, which are then solved analytically. Figure 3.20 shows different training methods of SVM such as Chunking (Cristianini and Shawe-Taylor, 2000), Osuna's algorithm (Osuna et al., 1997) and SMO (Shevade et al., 2000). For chunking, a fixed number of samples are added at each iteration, Thus, the number of samples of training per iteration tends to grow. For Osuna's algorithm, a fixed number of samples are optimized at each iteration and the same number of samples is added to and discarded from the problem at every step, and it is still not very efficient. SMO algorithm puts chunking to the extreme by iteratively selecting working sets of size two and optimizing the target function with respect to them. One advantage of using working sets of size two is that the optimization sub-problem can be solved analytically. The chunking process is repeated till all the training samples satisfy KKT

conditions. Because of the linear equality constraint involving the Lagrange multipliers ρ_i, the smallest possible problem involves two such multipliers. Then, for any two multipliers ρ_1 and ρ_2, the constraints are reduced to $0 \leq \rho_1, \rho_2 \leq \tilde{k}$, where $\tilde{k}$ is a constant. Details of SMO can be found at Platt et al. (1999); Shevade et al. (2000).

Several kernel functions are available for nonlinear transformation of the input space, such as linear, polynomial, RBF and universal Pearson VII function based kernels (PUK). A kernel method is an algorithm that depends on the data only through dot-products (Ben-Hur and Weston, 2010). It is replaced by the kernel functions which calculate the variations in a high dimensional space. Therefore, it extends the ability of a linear classifier to generate non-linear decision boundaries and apply a classifier to data that have no obvious fixed-dimensional vector space representation, for example, data in bioinformatics and facial wrinkle patterns. Let $\mathbf{x}_i, \mathbf{x}_j \in \Re^N$ denote input vectors of SVM, the PUK kernel, κ_{puk}, is defined as,

$$\kappa_{\text{puk}}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \frac{1}{\left[1 + \left(\left[2\sqrt{\left|\mathbf{x}_i - \mathbf{x}_j\right|^2 \sqrt{2^{(1/\omega)} - 1}}\right] \div \sigma\right)^2\right]^{\omega}} \tag{3.62}$$

where the parameters ω and σ control the half-width and the tailing factor of the peak of the Pearson VII function. In this way, the PUK kernel will lead to a symmetric matrix with ones on the diagonal and all other entries ranging between the values 0 and 1 for any arbitrary pair of $\left(\mathbf{x}_i, \mathbf{x}_j\right)$. The PUK kernel is robust and has an equal or even stronger mapping power as compared to the standard kernel functions, which leads to an equal or better generalisation performance of SMO (Zhang and Ge, 2013).

3.4 Face Alignment Techniques

In this section, the fundamental theory of geometric transformation is discussed.

3.4.1 Affine Transformation

In a 2D geometric transformation, any points (x, y) in the space could be transformed into a new set of coordinates (x', y') using translation, rotation and scaling as shown in Figure 3.21. Instead of applying a transformation to every point in every line that makes up an object (normally is a triangle), the transformation is applied only to the vertices of the object and then new lines are drawn between the resulting endpoints. Translation is

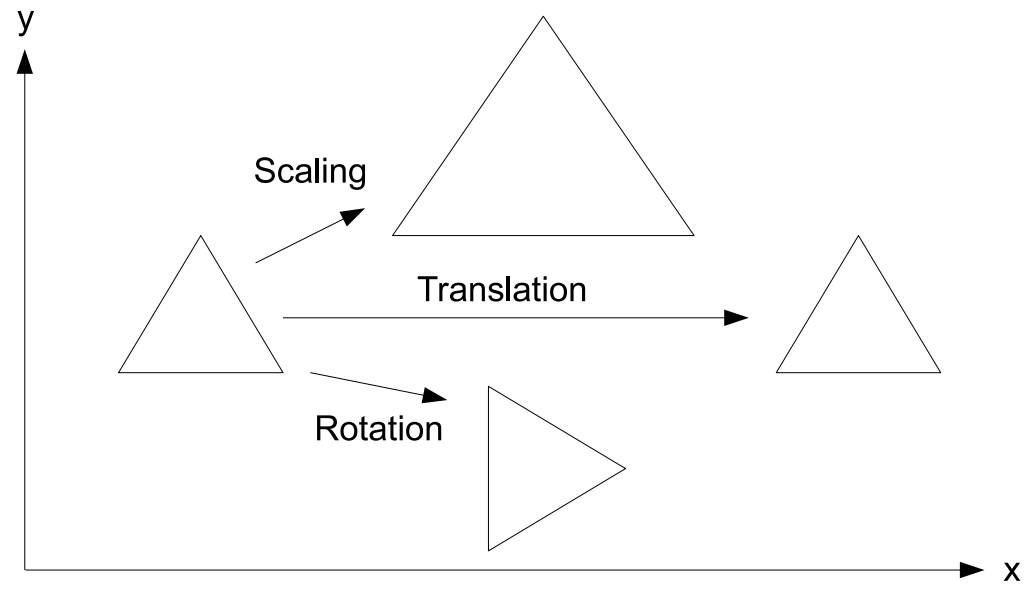


Figure 3.21: Affine transformation of a triangle.

moving an object from one location to another with a translation factor $\tilde{T}$ as

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \tilde{T}_x \\ 0 & 1 & \tilde{T}_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3.63}$$

where $\begin{bmatrix} x & y & 1 \end{bmatrix}^T$ is named as homogeneous coordinates and the $[3 \times 3]$ is a transformation matrix. Scaling is changing the size of an object with a scale factor $\tilde{s}$ as

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \tilde{s}_x & 0 & 0 \\ 0 & \tilde{s}_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3.64}$$

If the scaling factors are less than one, the object will appear smaller. If the scaling factors are greater than one, the object will appear larger. If $\tilde{s}_x = \tilde{s}_y = 1$, the object is unchanged. Rotation is a transformation that causes a point $\mathbf{x}$ to be moved relative to a central point, without changing the distance of $\mathbf{x}$ from that point. It is defined as

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3.65}$$

where θ indicates how many degree of rotation is desired and the range is between

$0 \leq \theta \leq 2\pi$. A series of transformations are feasible by multiplying the transformation matrices together as a composite transformation. The order of the multiplication sequence is important. Suppose we wish to first rotate an object by θ, translate it by $(\tilde{T}_x, \tilde{T}_y)$, and finally scale it by $(\tilde{s}_x, \tilde{s}_y)$, the composite transformation is defined as

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \tilde{s}_x & 0 & 0 \\ 0 & \tilde{s}_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & \tilde{T}_x \\ 0 & 1 & \tilde{T}_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3.65a}$$

$$= \begin{bmatrix} \tilde{s}_x\cos\theta & -\tilde{s}_x\sin\theta & \tilde{s}_x\tilde{T}_x \\ \tilde{s}_y\sin\theta & \tilde{s}_y\cos\theta & \tilde{s}_y\tilde{T}_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3.65b}$$

3.4.2 Piece-wise Affine Warping

PAW is a method to project each pixel of image $\mathbf{I}$ into a new image $\mathbf{I}'$ so that a set of n control points $\{\mathbf{x}_i\}$ are mapped to new positions, $\{\mathbf{x}'_i\}$ (Cootes and Taylor, 2001). A continuous vector valued mapping function $\mathbf{F}$ is defined as

$$\mathbf{F}(\mathbf{x}_i) = \mathbf{x}'_i \;\; \forall i = 1, \ldots, n \tag{3.66}$$

In practice, in order to avoid holes and interpolation problems, it is better to find the reserve map, $\mathbf{F}'$, taking $\{\mathbf{x}'_i\}$ into $\{\mathbf{x}_i\}$. For each pixel in the target warped image, $\mathbf{I}'$ we can determine where it came from in $\mathbf{I}$ and fill it in. In general $\mathbf{F}' \neq \mathbf{F}^{-1}$, but is a good approximation. Note that $\mathbf{F}$ can be decomposed as

$$\mathbf{F}(\mathbf{x}) = \sum_{i=1}^{n} f_i(\mathbf{x})\,\mathbf{x}'_i \tag{3.67}$$

where the n continuous scalar valued functions f_i each satisfy

$$f_i(\mathbf{x}_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{3.68}$$

and this ensures

$$\mathbf{F}(\mathbf{x}_i) = \mathbf{x}'_i \tag{3.69}$$

The simplest warping function is to assume each f_i is linear in a local region

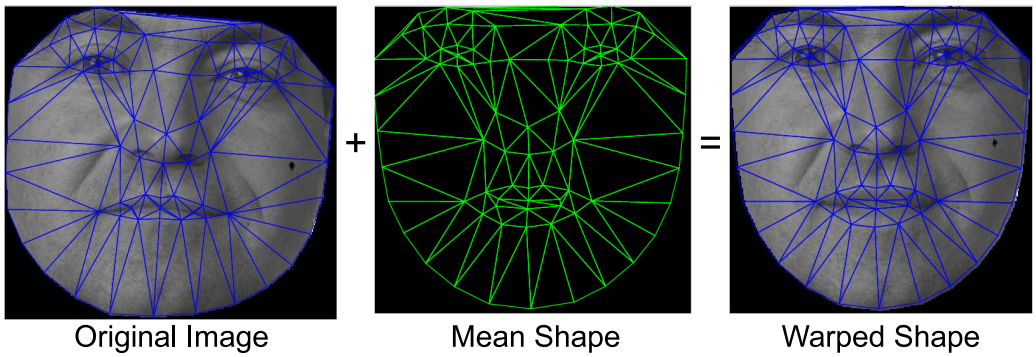


Figure 3.22: Warping between a shape and the mean shape using PAW. Note that this figure was redrawn from FERET.

and zero everywhere else. For instance, in the one dimensional case (where each $\mathbf{x}$ is a point on a line), suppose the control points are arranged in ascending order $(x_i < x_i + 1)$. If $\mathbf{F}$ maps a point $\mathbf{x}$ which is half way between x_i and x_{i+1} to a point halfway between x'_i and x'_{i+1}. This is achieved by setting

$$f_i(x) = \begin{cases} (x - x_i)/(x_{i+1} - x_i) & \text{if } x \in [x_i, x_{i+1}] \text{ and } i < n \\ (x - x_i)/(x_i - x_{i-1}) & \text{if } x \in [x_{i-1}, x_i] \text{ and } i > 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.70}$$

where the control points in the region is between x_1 and x_n.

In 2D dimensions, a triangulation (i.e., Delauney) can be used to partition the convex hull of the control points into a set of triangles. To the points within each triangle, an affine transformation which uniquely maps the corners of the triangle to their new positions in $\mathbf{I}'$ (as shown in Figure 3.22). Suppose $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ are three vectors of such a triangle. Any internal point can be written as

$$\mathbf{x} = \mathbf{x}_1 + b(\mathbf{x}_2 - \mathbf{x}_1) + c(\mathbf{x}_3 - \mathbf{x}_1) \tag{3.71a}$$

$$= a\mathbf{x}_1 + b\mathbf{x}_2 + c\mathbf{x}_3 \tag{3.71b}$$

where $a = 1 - (b + c)$ and so $a + b + c = 1$. For $\mathbf{x}$ to be inside the triangle, $0 < a, b, c < 1$. Under the affine transformation, this point simply maps to

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}) = a\mathbf{x}'_1 + b\mathbf{x}'_2 + c\mathbf{x}'_3 \tag{3.72}$$

To generate a warped image we take each pixel, $\mathbf{x}'$ in $\mathbf{I}'$, decide which triangle it belongs

to, compute the coefficients a, b, c giving its relative position in the triangle and use them to find the equivalent point in the original image, $\mathbf{I}$.

3.4.3 Shape-free Patch

According to Cootes and Taylor (2001), a simple iterative approach could be used for aligning the shapes as following steps:

i. Translate each example so that its centre of gravity is at the origin where the origin is $(0,0)$.

ii. Choose one example as an initial estimate of the mean shape and scale so that $|\bar{\mathbf{x}}| = 1$.

iii. Record the first estimate as $\mathbf{x}_0$ to define the default reference frame.

iv. Align all the shapes with the current estimate of the mean shape.

v. Re-estimate mean from aligned shapes.

vi. Apply constraints on the current estimate of the mean by aligning it with $\mathbf{x}_0$ and scaling so that $|\bar{\mathbf{x}}| = 1$.

vii. If not converged, return to step iv.

where $\mathbf{x}$ is a set of n landmark points, $\{(x_i, y_i)\}$ of a single example. Note that the convergence is declared if the estimate of the mean does not change significantly after an iteration. In this work, the tangent space is applied as constraints to transform each shape to the mean so as to minimize $\mathcal{D}$. The tangent space to $\mathbf{x}_t$ is the hyperplane of vectors normal to $\mathbf{x}_t$, passing through $\mathbf{x}_t$. All the vectors $\mathbf{x}$ such that $(\mathbf{x}_t - \mathbf{x}) \cdot \mathbf{x}_t = 0$, or $\mathbf{x} \cdot \mathbf{x}_t = 1$ if $\mathbf{x}_t = 1$. This leads to the corners of shape varying along a straight lines, orthogonal to the lines from the origin to the corners of the mean shape. This preserves the linear nature of the shape variation. The simplest way to achieve this is to align the shapes with the mean, allowing scaling and rotation, then project into the tangent space by scaling $\mathbf{x}$ by $1/(\mathbf{x} \cdot \bar{\mathbf{x}})$. Alternative approach can be found in Cootes and Taylor (2001) for reducing the non-linear variation in the point positions.

3.5 Scientific Measurements

For wrinkle detection, the objective quantification includes Jaccard Similarity Index (JSI) (Jaccard, 1901) and an accuracy of wrinkle detection (ACC). Since the subjective age estimation is biased and costs expensive manpower, the evaluations for age estimation are usually objective tests, evaluated quantitatively, and set up with either individual images or age groups (Miyamoto et al., 2005). This evaluation task often adopts two measures, Mean Absolute Error (MAE) and Cumulative Score (CS) , with the protocol of cross validation or Leave One Person Out (LOPO) approach (Geng et al., 2006; Guo et al., 2008a).

3.5.1 Jaccard Similarity Index

The segmentation results can be assessed using JSI to measure the overlap between the computerised method with manual annotation. In this context, JSI is used (Jaccard, 1901) to measure the reliability of wrinkle detection method. The Jaccard index $\mathcal{J}$ is calculated by the intersection of $\mathbb{A}$ and $\mathbb{B}$ divided by the union of $\mathbb{A}$ and $\mathbb{B}$. $\mathbb{A}$ and $\mathbb{B}$ are annotations of two different coders, respectively.

$$\mathcal{J}(\mathbb{A},\mathbb{B}) = \frac{|\mathbb{A} \cap \mathbb{B}|}{|\mathbb{A} \cup \mathbb{B}|} \tag{3.73}$$

3.5.2 Accuracy of Wrinkle Detection

In order to validate the correctness of the proposed method, ACC is defined as

$$\text{ACC} = \sum_{i=1}^{N} w_i, \quad w_i = \begin{cases} 1 & \text{if JSI}_i > 40\% \\ 0 & \text{otherwise} \end{cases} \tag{3.74}$$

where N is the total number of images in the experiment and w is true if JSI $>$ 40%. Any overlap between A and B larger or equal to 40% is considered as correct detection (Real, 1999; Drukker et al., 2002).

3.5.3 Mean Absolute Error

A MAE is defined as the average of the absolute errors between the estimated age labels and the ground truth age labels, that is,

$$\text{MAE} = \sum_{k=1}^{N} \left(|A_k - B_k| / N \right) \tag{3.75}$$

where B_k is the ground truth age for the k-th test image, A_k is the estimated age, and N is the total number of test images.

3.5.4 Cumulative Score

A CS, is defined as

$$\text{CS}(j) = N_{e \leq j} / N \times 100\% \tag{3.76}$$

where $N_{e \leq j}$ is the number of test images on which the age estimation makes an absolute error no higher than j years. As a classification problem, the age estimation performance can also be revealed by the classification accuracy.

CS(j) can be viewed as the classification accuracy, which, like in face recognition, might be a more important criterion than the average performance MAE for practical applications. Here, in the age estimation field, both CS and MAE are used as metrics since different methods, datasets, and systems may be biased or unbalanced for evaluation. For example, if the ordinal label is available in a large training data set with dense distribution of different ages, CS could be the best way to reflect the performance. But, if the training data are not dense, i.e., there are a lot of missing data or missing ages, MAE could be very useful to measure the performance.

3.5.5 Cross Validation

k-fold cross validation is the statistical practice of partitioning a sample of data into subsets such that an analysis is initially performed on a single subset, while the other subset(s) are retained for subsequent use in confirming and validating the initial analysis (Kohavi, 1995; Geisser, 1993). The first subset is used as a testing set and the other subset(s) are applied as training set. Then, the second subset is used as a testing set and the rest is implemented as training set, and so on. A collection of images - dataset ($\mathbf{D}$) is used in that method by randomly split it into k mutually exclusive subsets

(named as fold), $D_1, D_2, ..., D_k$ of approximately equal size. For example, 10-fold cross-validation is implemented for estimating a MAE as,

$$\text{MAE}_{10} = \frac{1}{k}\sum_{i=1}^{k}\text{MAE}_i \tag{3.77}$$

where $k = 10$ and MAE_i is the result of a particular D_i. LOPO is similar to cross validation but instead of dividing a dataset into number of subsets, it replaces k with a total number of images (N).

3.6 Benchmark Datasets

In this section, five datasets including Bosphorus (Savran et al., 2012a), FGNET (FGNET aging dataset, accessed on September 2012), MORPH (Ricanek and Tesafaye, 2006), FERET (Phillips et al., 2000), PAL (Minear and Park, 2004) are discussed.

3.6.1 Bosphorus

In order to evaluate the potential 3D modality for facial expression analysis, Savran et al. (2012a) prepared an extensive facial expression dataset, called Bosphorus, where 3D faces were acquired with a structured light system and the companion 2D face images with a normal light camera. The image acquisition was done under good illumination conditions and without any background clutter. The colour images have 1600×1200 resolution. Their dataset contains 105 subjects enacting a large repertoire of expressions, and displaying systematic head poses (13 fixed rotations including yaw, pitch and cross rotations) and occlusions (beard, moustache, glasses, hand, hair, etc.), resulting in a total of 4666 number of face scans. The facial expressions were instructed by the experimenter and the ground-truth Facial Action Coding System (FACS) (Ekman and Friesen, 1977; Ekman et al., 2002) annotations were obtained by one certified FACS coder. The limitation of using only one FACS coder does not allow for inter-coder reliability of their coding. Some of the characteristics of the dataset are as follows: The majority of the subjects are aged between 25 and 35, mostly Caucasian, and the cohort consists of 60 men and 45 women in total. They employed 29 professional actors and actresses for acting the expressions while the rest were recruited from students and staffs. They selected 25 action units, split between seven lower action units and 18 upper facial action units, and tested them over 2902 neutral and expression samples

Figure 3.23: Bosphorus samples (Savran et al., 2012a). Top and bottom row shows a female subject and a male subject of same person with different expressions and poses, respectively. The label of each image describes the expression (**N**eutral / **F**ear / **H**appy) and pose angle.

in the dataset. They also included the pitch, yaw and roll statistics using 22 landmark points.

3.6.2 FGNET

FGNET comprises 1002 images of 82 subjects (6-18 images per subject) in the age range 0-69 years (FGNET aging dataset, accessed on September 2012) . The dataset also provides 68 landmark features that were identified manually, on all the face images. In addition, the following meta-information is available for all the images in the dataset namely: image size, age, gender, spectacles, hat, mustache, beard, horizontal pose and vertical pose. Since the images were retrieved from real-life albums of different subjects, the face images involve all possible variations including illumination, pose, expression, beards, moustaches, spectacles, etc (Zhou et al., 2005). Figure 3.24 illustrates some samples of the type of variation seen in FGNET. Based on human observation, the first and second row are clear images but the last row is considered as blurred images which do not have sufficient texture information such as wrinkles.

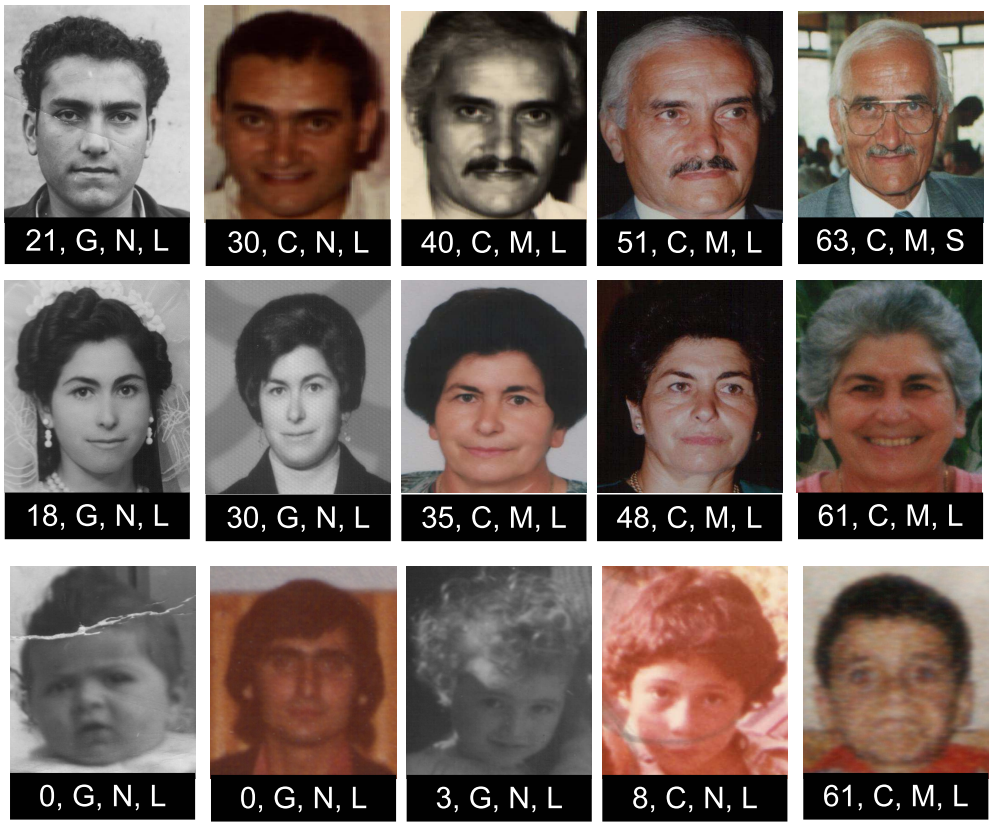


Figure 3.24: FGNET samples (FGNET aging dataset, accessed on September 2012). Top row shows a male subject of same person at different age, mid row also the same but the subject is a female and bottom row presents different subjects with poor image quality. The label of each image describes the age, colour property (**G**reyscale / **C**olor), **M**oustache or **N**on-moustache, **S**pectacles or Spectacle**L**ess.

3.6.3 FERET

FERET is a comprehensive dataset that addresses multiple problems related to face recognition. (Phillips et al., 2000). It consists of 2366 images of 994 subjects that presents multiple problems related to face recognition such as illumination variations, pose variations, and facial expressions. Moreover, it consists of a few hundred age-separated face images of subjects with the age difference of 18 months or more and the age range is between 10 and 70. Batool and Chellapa Batool and Chellappa (2015) claimed that FERET is a low resolution dataset like FGNET. However, due to wrinkle extraction requiring decent resolution, we assess the performance of proposed algorithm on FERET, the highest resolution amongst the benchmark datasets. Figure 3.25 illustrates the samples of FERET which involve the difference in age, gender and ethnicity.

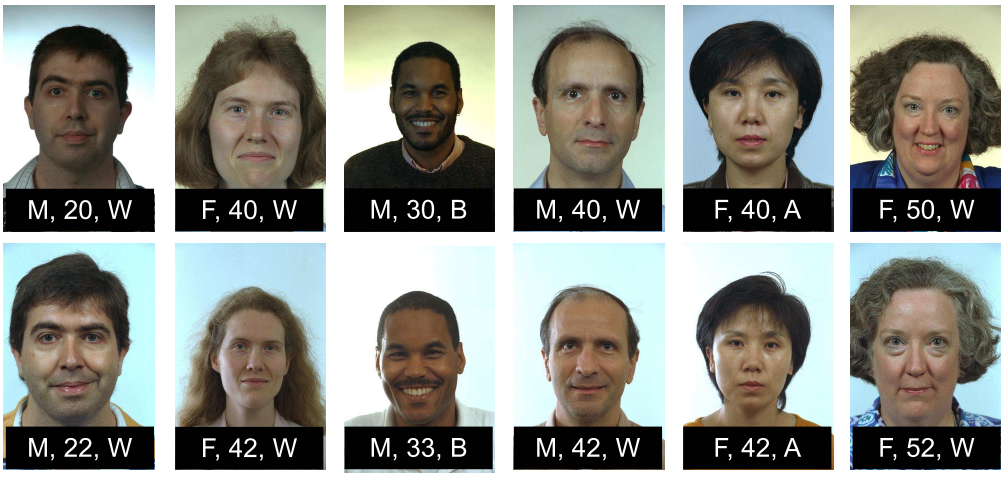


Figure 3.25: FERET samples (Phillips et al., 2000). Top and bottom of same column shows the face images of same subject at different ages. The label of each image describes the gender (**M**ale / **F**emale), approximate age in 2014 and race (**W**hite / **B**lack / **A**sian).

3.6.4 MORPH

The publicly available dataset of MORPH was collected by the Face Aging Group at the University of North Carolina at Wilmington for the purpose of face biometrics applications. This longitudinal dataset records individuals metadata, such as age, gender, ethnicity, height, weight, and ancestry, which is organized into two albums (Albert and Ricanek Jr, 2008; Ricanek and Tesafaye, 2006). In the latest version on 2015[2], it contains 55,000 unique images of more than 13,000 individuals, spanning from 2003 to late 2007. Ages range from 16 to 77 with a median age of 33. The average number of images per individual is 4 and the average time between photos is 164 days, with the minimum being 1 day and the maximum being 1681 days. The standard deviation of days between images is 180. Figure 3.26 shows the samples from MORPH.

3.6.5 PAL

Faces constitute a unique and widely used category of stimuli. In spite of their importance, there are few collections of faces for use in research, none of which adequately represent the different ages of faces across the lifespan. The lack of age range representation has limited the majority of researchers to using predominantly young faces as stimuli even when their hypotheses concern both young and old participants. Minear

[2]MORPH Non-Commercial Release Whitepaper, http://www.faceaginggroup.com

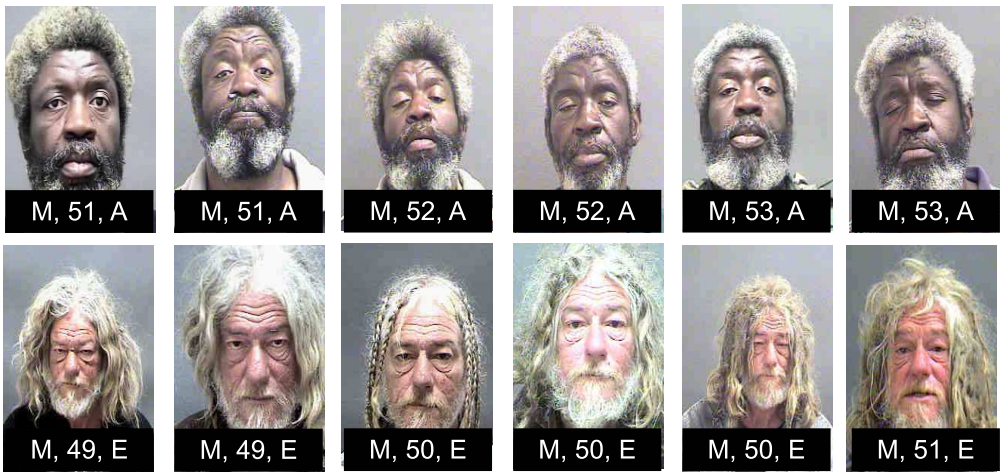


Figure 3.26: MORPH samples (Ricanek and Tesafaye, 2006). Top and bottom row shows the face images of two different subjects at different ages. The label of each image describes the gender (**M**ale / **F**emale), age and race (**A**frican / **E**uropean).

Figure 3.27: PAL samples (Minear and Park, 2004). The label of each image describes the gender (**M**ale / **F**emale), approximate age in 2014 and race (**A**frican-American / **C**aucasian / **O**ther).

and Park (2004) from PAL collected a dataset of 576 individual faces ranging from ages 18 to 93. It was developed to be more representative of age groups across the lifespan, with a special emphasis on recruiting older adults. The resulting dataset has faces of 218 adults age 18-29, 76 adults age 30-49, 123 adults age 50-69, and 158 adults age 70 and older. This will allow researchers interested in using facial stimuli access to a wider age range of adult faces than has previously been available. In this work, only neutral faces are used for the experiments due to profile faces might not have enough

wrinkle information.

3.7 Summary

The techniques, measurements and datasets described in this chapter provide the basis for contributions in later chapters. In the next chapter, two novel methods for wrinkle detection are described. Then, the subsequent chapters describe the novel feature representations and age estimation methods.

Part II: Contribution Chapters

Chapter 4

Novel Methods of Wrinkle Detection

This chapter proposes two novel methods of wrinkle detection. First, human performance for wrinkle detection is explained. Second, the proposed HHF which integrates the directional gradient and Hessian matrix in discovering the ridge and valley pattern is described. Third, the proposed HLT that discovers the wrinkle connectivity through the neighbourhood pixels is presented. Finally, it provides an insight of the strength and weakness of each method which could be explored in the future.

4.1 Human Performance for Wrinkle Detection

Aging as a natural phenomenon affects different parts of the human body under the influence of various biological and environmental factors. Based on literature review, humans rely on a number of facial features to predict the face age, i.e., face shape, hair, moustache, beard, mole, scar and wrinkle. The most pronounced changes that occur on the face is the appearance of wrinkles, which are the focus of this research. Accurate wrinkle detection automatically is an important task in face analysis. Some have been proposed in the literature, but the poor localisation limits the performance of wrinkle detection. It will lead to false wrinkle detection and consequently affect the processes such as age estimation and clinician score assessment. Prior to developing a robust algorithm, it is important to understand how human perform wrinkle detection.

In this section, three coders have been instructed to annotate the wrinkles of 2D forehead images manually. A dataset namely FORERUS was created for this experiment where 100 forehead images were randomly selected and cropped from Bosphorus (Savran et al., 2012b).

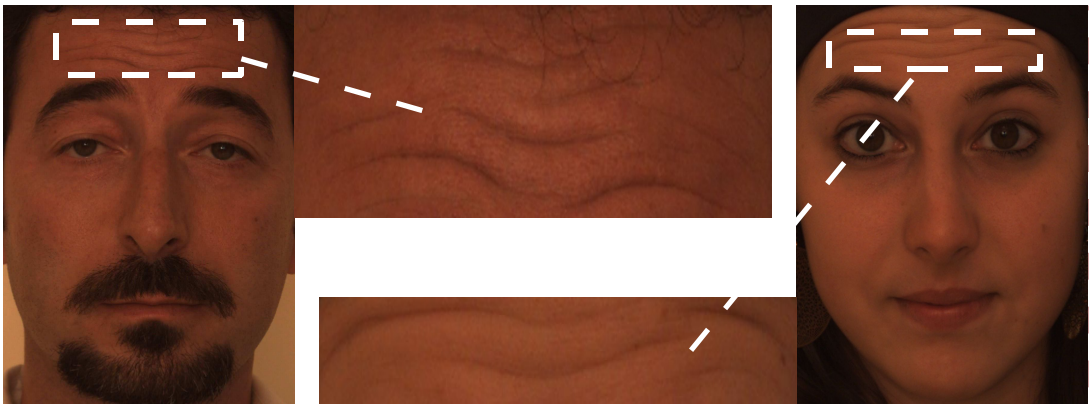


Figure 4.1: FORERUS samples. The dotted lines show the portion of forehead image cropped from the original facial image and it is used for wrinkle detection analysis. Both face images were from Bosphorus (Savran et al., 2012b).

4.1.1 Dataset Preparation

In order to assess the performance of the wrinkle detection algorithm, images were selected from Bosphorus (Savran et al., 2012b). This dataset consists of 106 subjects and the 2D face images were acquired under good illumination conditions with a normal light camera. According to Batool and Chellappa (2012), wrinkles are more obvious on forehead in most of the images. Therefore, the same procedure is repeated by manually cropping the forehead images from Bosphorus. For each image, forehead was manually cropped with a rectangle selector as the dotted lines shown in Figure 4.1. Rectangle size varied from one to another due to different face shapes exist in dataset. This experiment validates the performance of the proposed algorithms in detecting wrinkles compared to the human annotations on different images. Thus the image size normalisation is not included in the preprocessing step. In total, 100 random forehead images are cropped from Bosphorus and each forehead image is annotated with hand labelling as the ground truth. This dataset is named as FORERUS. Three coders have been instructed to do the annotation in Matlab under a controlled environment as illustrated in Figure 4.2(c) (Microsoft Windows 7 Enterprise 64-bit SP1, Intel Core i7-3770 CPU @ 3.40GHZ, 8.0GB RAM, NVIDIA Quadro 200, lab with similar lighting conditions). To prevent the bias from the use of software, the annotators were allowed to practise on the annotation software for 10 minutes. They were instructed to annotate the interior centre line of the wrinkle, which is defined as the deepest wrinkle area. Center line of a wrinkle means the deepest wrinkle area. The reason for this is to minimise the factors such as lighting, screen size and the noise interference while annotating the wrinkles.

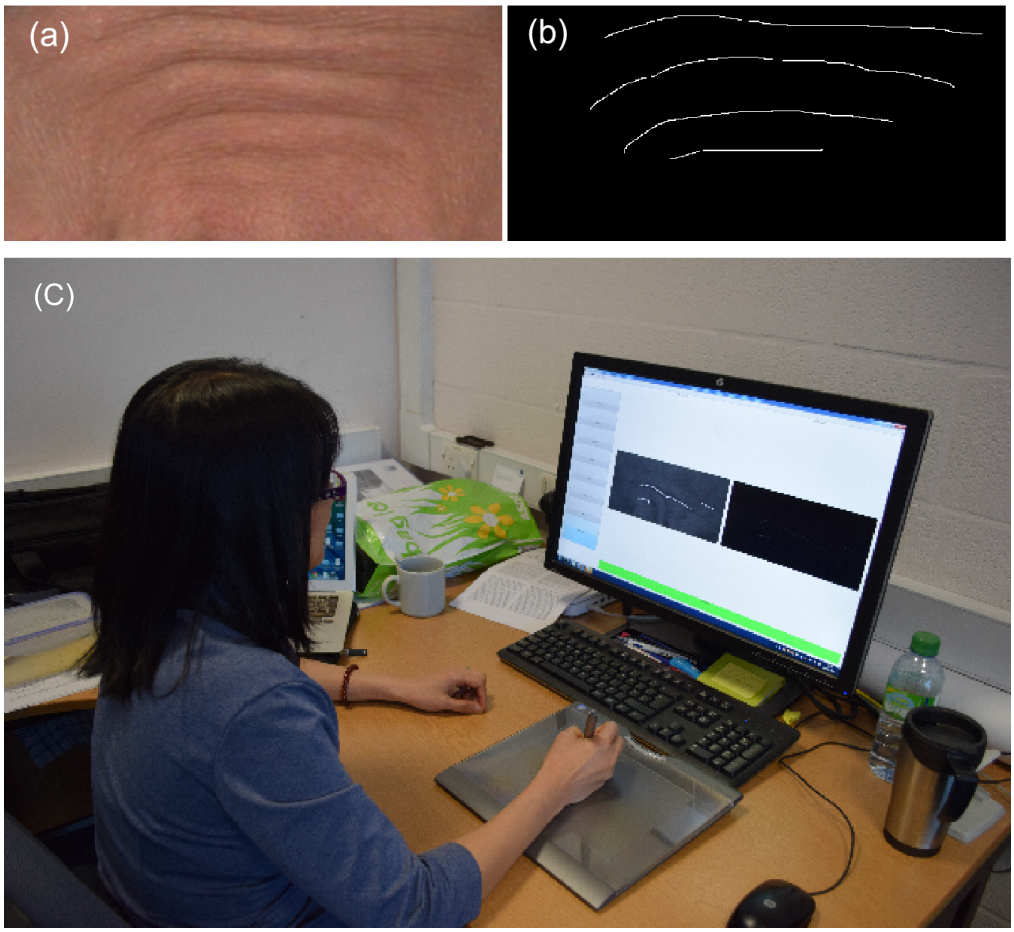


Figure 4.2: Illustration of the working environment settings for human annotation. (a) Original image. (b) Four lines drawn by human annotation in a binary mask. (c) An illustration of how the coder performs the wrinkle annotation in Matlab.

One pixel line was used for each human annotation. Figure 4.2(b) illustrates the mask of human annotation.

In this work, the expansion threshold of nine pixels is utilised when calculating a JSI. From the experiment, it is noticed that nine pixels produced the best overlapping in between the ground truth and estimated line. A small number of pixels will fail to hit the estimated line and large number of pixels will bias the result due to certain wrinkles being close to each other. Figure 4.3 illustrates an example of how the annotation is expanded with nine pixels to top and bottom from its original location. Assume coder A is the prediction and coder B is the ground truth. The white area in Figure 4.3(b) is an annotation expansion of coder A and it is represented by the blue and green lines in Figure 4.3(c). Note that, only prediction line is expanded when calculating a JSI to prevent over segmentation.

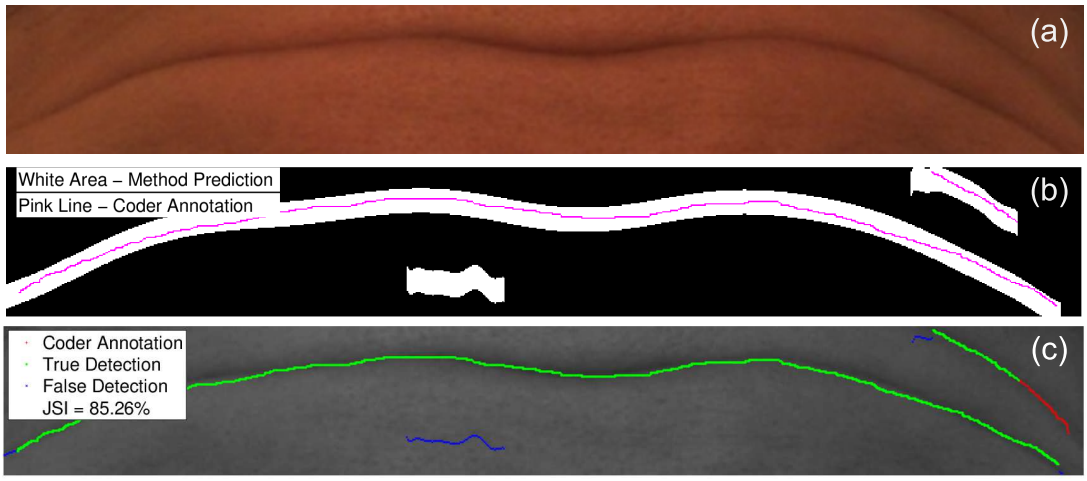


Figure 4.3: Pixel overlapping between coders. (a) Original image. (b) Annotation of coder A (or white area) has expanded nine pixels to top and bottom from its original location as shown in (c) as blue and green line. Pink line represents the annotation of coder B or the ground truth. (c) Blue line represents the false alarm and green line shows the true prediction. The image JSI is 85.26%.

4.1.2 Experimental Results of Human Annotation

Figure 4.4 shows the intra-coder reliability within coders. A round shape plot indicates the annotation is highly consistent. First, the coders annotated the complete set of images on the first day, and then they were asked to repeat the annotation on 30% of the images on the second day. The intra-coder reliabilities of coders A, B and C in terms of average JSI are 82.55%, 78.99% and 74.56%, while the Standard Deviation (STD)s are 8.58%, 11.28%, 10.45%, respectively. The intra-coder reliability of coder C is the lowest but the performance of coder B is the most inconsistent. From the human annotation result, wrinkle annotation is not only a laborious and time-consuming task, but it is also a difficult perceptual task with low consistency within the same annotator.

In addition to intra-coder reliability, inter-coder reliability is a measurement of consistency among coders. Table 4.1 illustrates the JSI statistics between coders A, B and C. Overall, the accuracy of inter-coder reliability between coders is 94%. These results showed that human annotation is often subjective. Moreover, it is time consuming if human annotation involves a large dataset. Therefore, an automatic wrinkle detection is needed to speed up and automate the annotation.

Figure 4.5 shows the samples of human annotation between three coders. It is observed that coder B did not annotate certain lines of all images as wrinkles but coder A did annotate it such as *img* i, ii and iv. Such contradictions resulted in the variations of

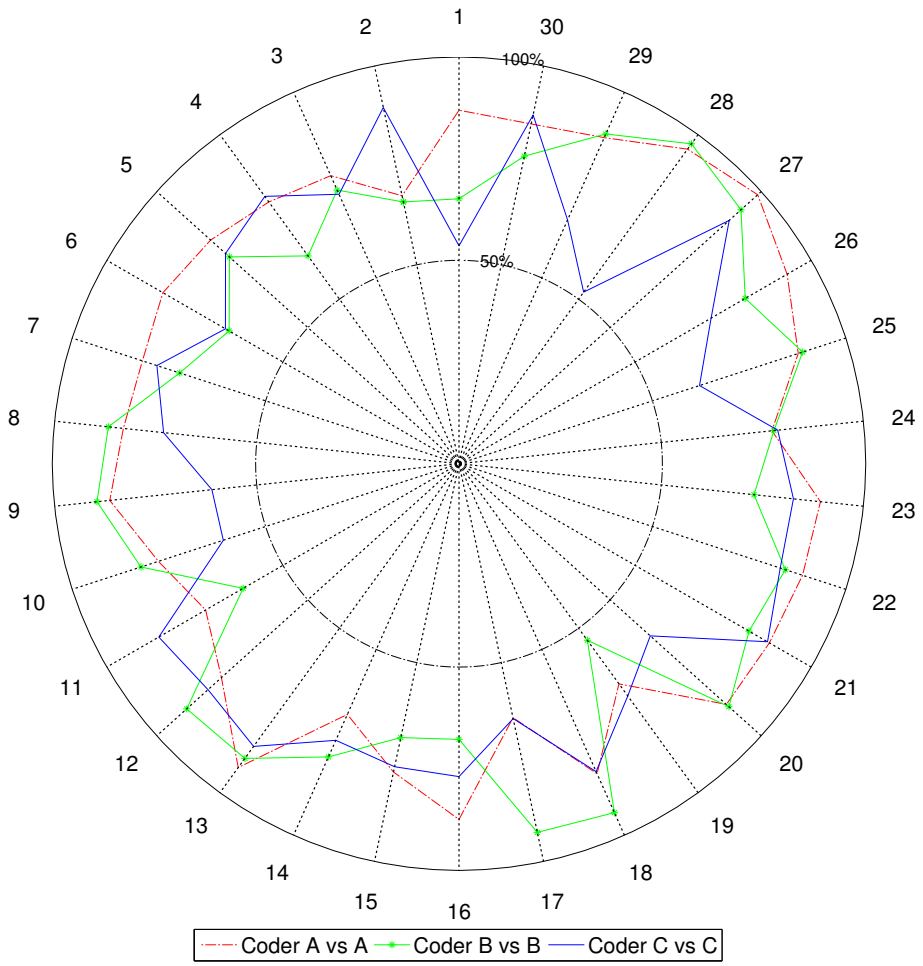


Figure 4.4: Intra-coder reliability of 30% dataset. Angle represents image number and radius illustrates the JSI percentage.

Table 4.1: Inter-coder reliability of human annotation.

Coders	Mean	STD	Min	Max	ACC
A vs B	67.28	14.10	18.41	94.88	99.00
A vs C	75.46	13.59	33.98	96.69	97.00
B vs C	61.46	16.92	23.05	98.34	86.00

annotation and hence yielded different JSI results. Moreover, this result demonstrates the difficulty of wrinkle localisation and the high technical challenge for automatic wrinkle detection. Therefore, two novel methods, HHF and HLT are proposed to automatically perform the wrinkle detection as follows.

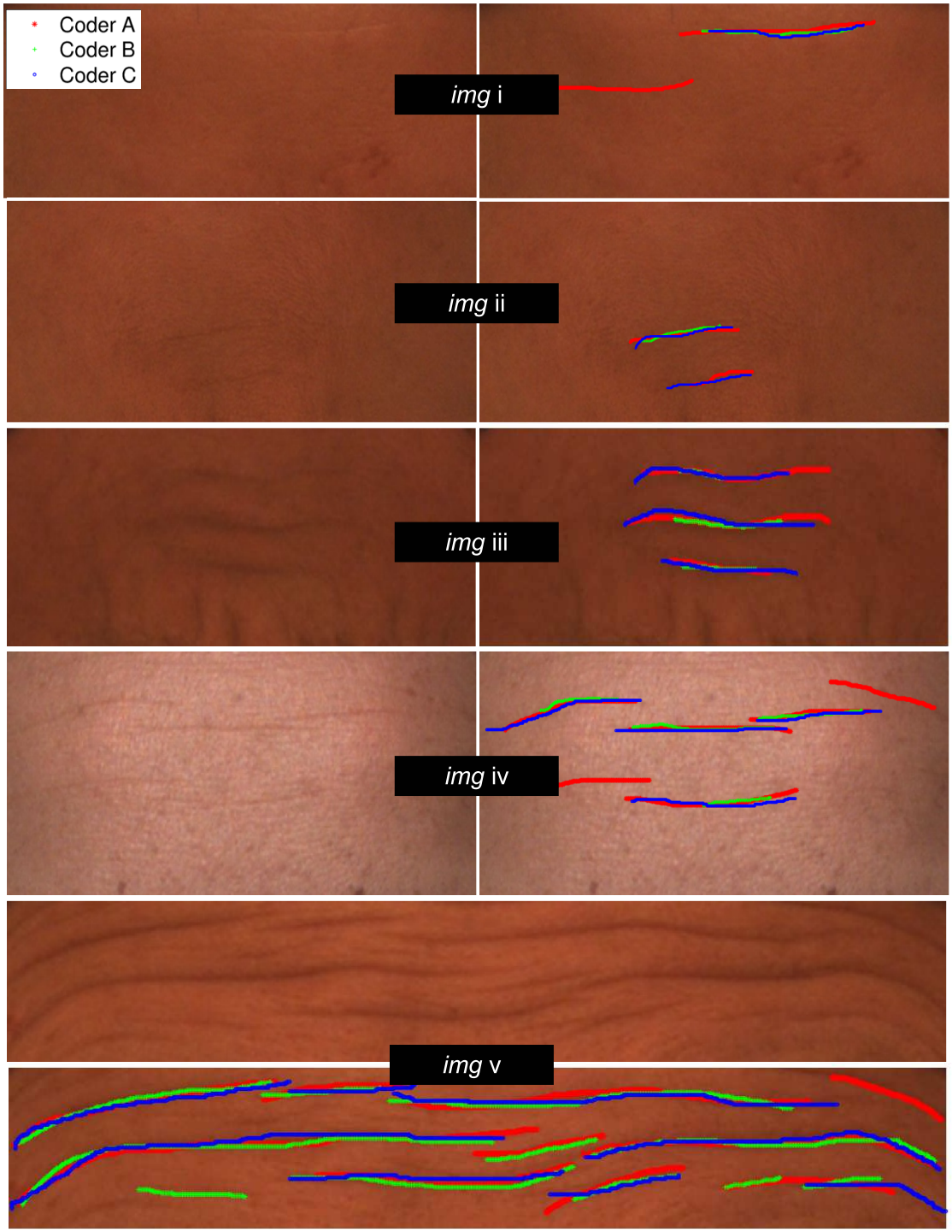


Figure 4.5: Human annotation on FORERUS images. Each row shows an original image (left) and annotated image of three coders (right). Red line represents coder A, green line is coder B and blue line is coder C. *img* i, ii, iii, iv, v mean different images.

4.2 Novel Algorithm of Hybrid Hessian Filter

A Hybrid Hessian Filter (HHF) is proposed to segment the wrinkles using the directional gradient and a ridge-valley Gaussian kernel. The proposed filter is conceptually simple, however, it significantly increases the true wrinkle localisation when compared with the conventional methods.

Wrinkles are considered as stochastic spatial arrangements of line segment sequences, reasonably similar with those in retinal blood vessels. A wrinkle should not be confused with edges. An edge is the border between two areas while a wrinkle is a line that is either darker or lighter than their neighbourhood. Therefore, edge detection methods such as Canny and Sobel are not suitable for wrinkle detection because it will produce wrinkle boundaries and not a wrinkle.

In this section, the multi-scale second order local structure of an image is explored. According to Frangi et al. (1998), a vesselness measure is obtained on the basis of all eigenvalues of the Hessian. The eigenvalues of the Hessian matrix evaluated at each point quantify the rate of change of the gradient field in various directions. The eigenvalues are independent vector measures by the components of the second derivatives of the field at each point (x, y). A small eigenvalue indicates a low change rate of the field in the corresponding eigen-direction, and vice versa (Frangi, 2001). Although both vessel and wrinkle present similar curve patterns, the underlying image quality is different. The vessel image was captured by the TopCon TRV-50 fundus camera at a 35° Field Of View (FOV), which were digitized with 24-bit grey-scale resolution and a spatial resolution of 700×605 pixels. In contrast, the original image of wrinkles were acquired using a standard camera with the image resolution of 1600×1200. Higher resolution images contain more properties but also noise. Therefore, the proposed method has to deal with the pepper noise and illumination without damaging the region of interest. In addition, a retinal scan involves using a low-intensity light source through an optical coupler to scan the unique patterns of the retina (Liu and Silverman, 2001). Retinal scanning can be quite accurate but does not require the user to look into a receptacle and focus on a given point. As a result, these images present constant intensity across the dataset. In contrast to retinal vessels images, skin images present different challenges. Skin surface consists of pores, hair and pigmentation signal, among these attributes wrinkles present significant variations such as curve pattern, length, thickness and orientation with varied waviness and roughness.

Figure 4.6 shows the proposed HHF for wrinkle detection. The main difference

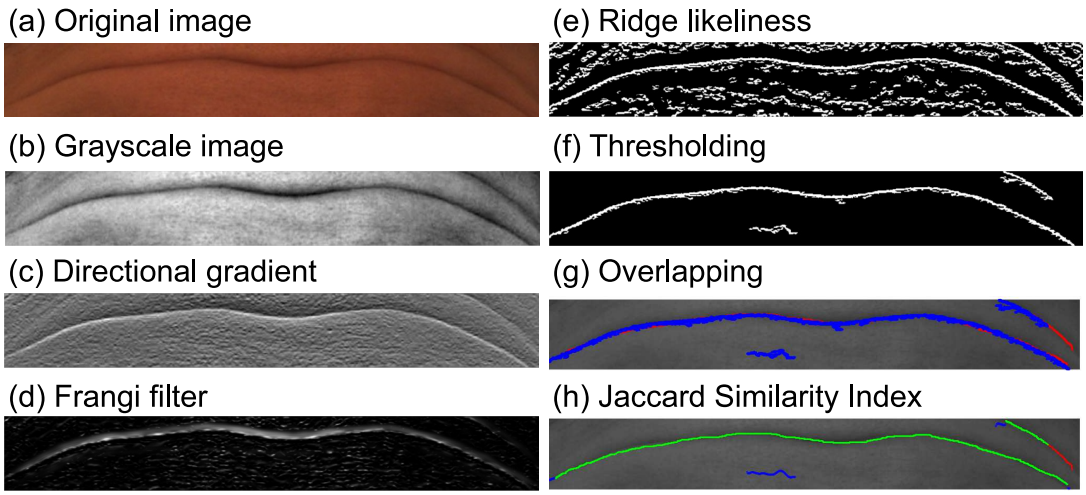


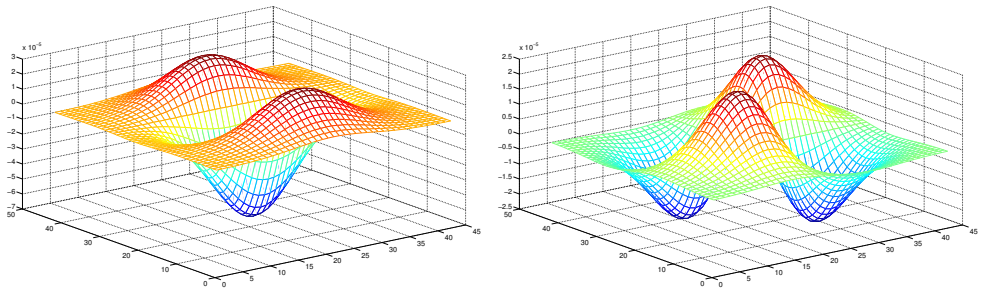
Figure 4.6: Process flow of HHF. (a) Original image was cropped from Bosphorus (Savran et al., 2012b). (b) Colour image was converted into grey-scale image. (c) Gaussian filter was used to derive the directional gradient from grey-scale image. (d) FRF was applied on directional gradient image to approximate the structure around each pixel at certain scale. (e) Image vectors less than zero was preserved as ridge-like pattern. (f) Ridge image was thresholded in a degree that only wrinkle-like patterns are extracted. (g) The overlapping between coder annotation (red line) and method estimation (blue line). (h) JSI was calculated based on the intersection area only (green line).

between HHF and FRF is the use of gradient image in HHF and a ridge-valley Gaussian kernel. This ensures the direction of interest is highlighted before the processing of Hessian filter. Given a 2D forehead image $\mathbf{I}$ as illustrated in Figure 4.6(a), it is converted into grey-scale as shown in Figure 4.6(b). The directional gradient (G_x, G_y) is computed from the grey-scale image and the G_y is illustrated as in Figure 4.6(c). Let G_y denoted as I, due to I emphasizes the y-variation (or horizontal line in the forehead image), it is used as the input for calculating the Hessian matrix $\mathcal{H}$. The Hessian matrix $\mathcal{H}$ at scale ϑ is defined as

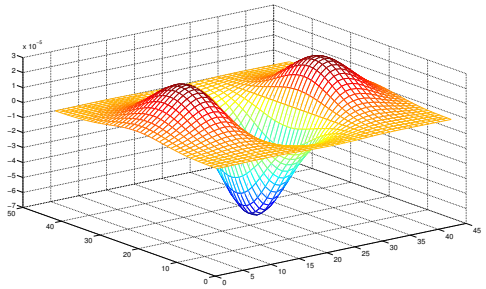
$$\mathcal{H}(x,y,\vartheta) = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \tag{4.1}$$

where a, b and c are the outputs of second derivative (Frangi, 2001). Each approximates the convolution of I by the Gaussian kernels $\mathbb{G}_1(\vartheta)$, $\mathbb{G}_2(\vartheta)$ as

$$\mathcal{H}_a(x,y,\vartheta) = I(x,y) * \mathbb{G}_1(i,j,\vartheta) \tag{4.2}$$



(a) $\mathbb{G}_1$ (b) $\mathbb{G}_2$



(c) $\mathbb{G}_1^T$

Figure 4.7: Gaussian kernels $\mathbb{G}_1$, $\mathbb{G}_2$ and $\mathbb{G}_1^T$ where $\vartheta = 7$. In this work, $\vartheta \in \{1,3,5,7\}$ for deriving different sizes of wrinkle.

$$\mathcal{H}_b(x,y,\vartheta) = I(x,y) * \mathbb{G}_2(i,j,\vartheta) \tag{4.3}$$

$$\mathcal{H}_c = [\mathcal{H}_a]^T \tag{4.4}$$

The second derivative of a Gaussian kernel at scale ϑ generates a probe kernel that measures the contrast at the selective scale in the direction of the derivative. They are given by

$$\mathbb{G}_1(i,j,\vartheta) = \frac{1}{2\pi\vartheta^4}\left[\frac{\mathcal{M}_{i,j}^2}{\vartheta^2} - 1\right] e^{\frac{-\mathcal{M}_{i,j}^2+\mathcal{N}_{i,j}^2}{2\vartheta^2}} \tag{4.5}$$

$$\mathbb{G}_2(i,j,\vartheta) = \frac{1}{2\pi\vartheta^6}\left[\mathcal{M}_{i,j}\mathcal{N}_{i,j}\right] e^{\frac{-\mathcal{M}_{i,j}^2+\mathcal{N}_{i,j}^2}{2\vartheta^2}} \tag{4.6}$$

where $\mathcal{M}$ and $\mathcal{N}$ are the kernels with vertical and horizontal directions as

$$\mathcal{M}_{i,j} = -3\vartheta + i - 1 \tag{4.7}$$

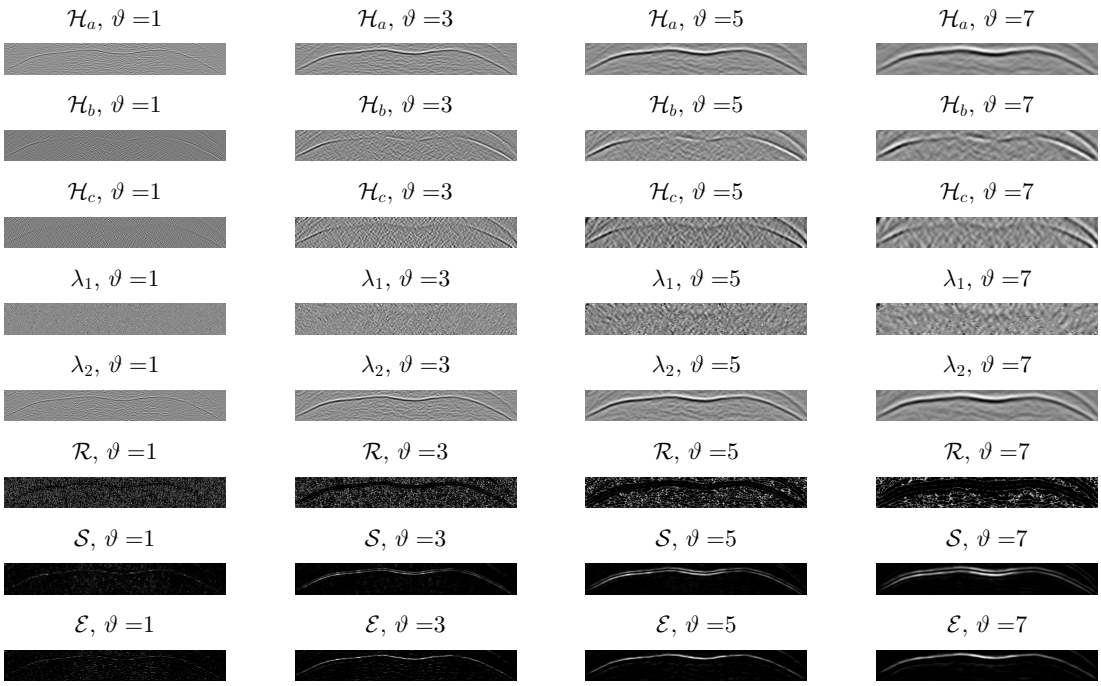


Figure 4.8: HHF variables at different scales. Each row shows different variables of HHF and each column is one of the scales, ϑ. It is observed that $\mathcal{H}_a$, λ_2 and $\mathcal{E}$ show the lines of interest. A multi-scale detection presents a different kind of lines. This implies different sizes of wrinkles can be detected by HHF.

$$\mathcal{N}_{i,j} = -3\vartheta + j - 1 \tag{4.8}$$

where the range of i and j is $[-3\vartheta, 3\vartheta]$. Figure 4.7 shows the Gaussian kernels $\mathbb{G}_1$, $\mathbb{G}_2$ and $\mathbb{G}_1^T$ where $\vartheta = 7$. Since wrinkles are similar to the patterns of ridge and valley, the Gaussian kernels are designed in the same way. To categorize the texture pattern, eigenvalues λ_1 and λ_2 of the Hessian at specific scale and coordinates are given by

$$\lambda_1 = \frac{1}{2}\left[\mathcal{H}_a + \mathcal{H}_c + \left(\sqrt{(\mathcal{H}_a - \mathcal{H}_c)^2 + 4\mathcal{H}_b^2}\right)\right] \tag{4.9}$$

$$\lambda_2 = \frac{1}{2}\left[\mathcal{H}_a + \mathcal{H}_c - \left(\sqrt{(\mathcal{H}_a - \mathcal{H}_c)^2 + 4\mathcal{H}_b^2}\right)\right] \tag{4.10}$$

The similarity measures $\mathcal{R}$ and $\mathcal{S}$ are given by

$$\mathcal{R} = \left(\frac{\lambda_1}{\lambda_2}\right)^2, \ \lambda_2 \neq 0 \tag{4.11}$$

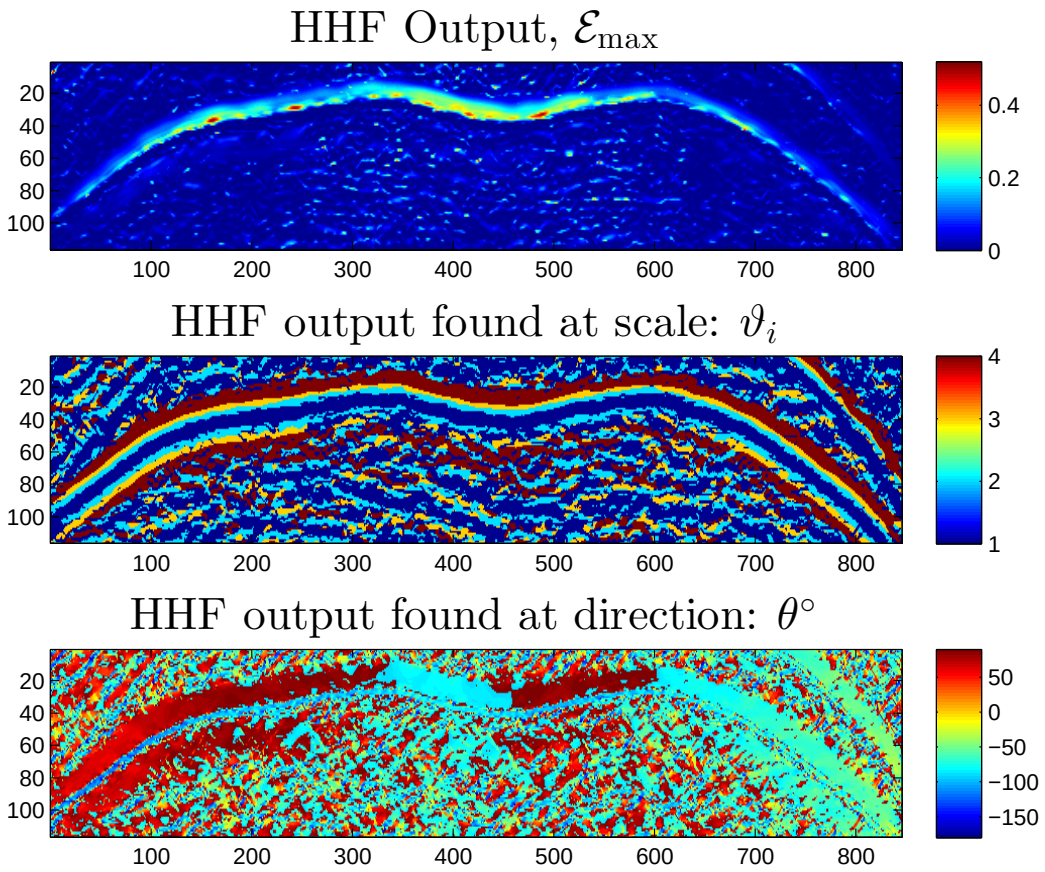


Figure 4.9: HHF values at different scales and orientations of Figure 4.6(a). Top row is the HHF optimum values selected from different scales. Mid-row is where the scale of optimum is selected. Bottom is where the orientation of optimum is selected.

$$\mathcal{S} = \lambda_1^2 + \lambda_2^2 \tag{4.12}$$

The curvilinear likeliness measure $\mathcal{E}$ is defined as

$$\mathcal{E}(x, y, \vartheta) = \begin{cases} 0 & \text{if } \lambda_2 < 0 \\ e^{-\frac{\mathcal{R}}{2\beta_1^2}} \left[1 - e^{-\frac{\mathcal{S}}{2\beta_2^2}}\right] & \text{otherwise} \end{cases} \tag{4.13}$$

In this work, the kernel scale, ϑ, is set as 1, 3, 5, 7; β_1 controls the sensitivity of the filter to the curve deviation $\mathcal{R}$ and the default value is 0.5; β_2 depends on the greyscale range of the ridge of interest and controls the sensitivity of the filter to the similarity measure $\mathcal{S}$ and the default value is 15 (refer to Frangi (2001)). If a ridge is analyzed at different scales ϑ, the response of the filter $\mathcal{E}_{\max}$ will be the maximum of all scales if it

approximately matches the filter scale and it is defined as,

$$\mathcal{E}_{\max}(x,y) = \max_{\vartheta_{\min} \leqslant \vartheta \leqslant \vartheta_{\max}} [\mathcal{E}(x,y,\vartheta)] \tag{4.14}$$

where $\vartheta_{\min}$ and $\vartheta_{\max}$ are the minimum and maximum scales at which relevant structure are expected to be found (as shown in Figure 4.6(d)). λ_1 and λ_2 highlights the data of interest and discards noisy patterns (Frangi, 2001). Figure 4.8 illustrates HHF variables at different scales. It is noticed that λ_2 presents a pattern of interest where wrinkles are highlighted in a ridge and valley curve. This indicates that λ_2 is a key indicator of wrinkle location. In all scales, if λ_2 appears to be negative, then the wrinkle is detected. Moreover, if $\mathcal{E}$ appears exactly as zero at all scales, then the wrinkle is detected as well. Figure 4.9 shows the HHF values at different scales and orientations. It is observed that the HHF output generates a high range of values where wrinkles lie. Most of them are found at similar scale but different angles. This implies a multi-scale filtering is very useful for different sizes of wrinkles. Once the similarity is derived from the Eq. (4.14), the wrinkle-like pattern is preserved. The initial wrinkle mask $\mathcal{W}$ is generated as shown in Figure 4.6(e) and it is defined as

$$\mathcal{W}(x,y) = \begin{cases} 0 & \text{if } \mathcal{E}_{\max}(x,y) > 0 \\ 1 & \text{otherwise} \end{cases} \tag{4.15}$$

Next, each region of interest (8-connected pixels) is filtered by an area threshold where regions less than 250 pixels are removed and the output is the estimated forehead wrinkle as shown in Figure 4.6(f). Note that the area threshold is based on the initial image resolution and it is empirically determined. A small image might need a lower threshold and vice versa.

Figure 4.6(g) illustrates the overlapping between human annotation and predicted wrinkle of HHF. A polynomial fitting is implemented on the predicted wrinkle to localize the center line of the wrinkle and it is used for calculating a JSI denoted as Eq. (3.73).

Figure 4.6(h) shows how a JSI is calculated, red represents the human annotation or ground truth, blue means the false detection and green is the true detection.

Figure 4.10 demonstrates the pseudo code of HHF. The novel idea is to segment wrinkles using a ridge-valley Gaussian kernel on the directional gradient (as step 3 and 5 in the Figure 4.10). Given a grey-scale image $\mathbf{I}$, the directional gradient is computed, giving (G_x, G_y). Since G_y represents changes in the y-direction, it is suitable for horizontal line detection, and it is used to calculate the Hessian matrix $\mathcal{H}$ at a specific scale

1: **BEGIN**
2: $\mathbf{I} = ImageAcquisition()$
3: $(G_x, G_y) = ImageGradient(\mathbf{I})$
4: **for** $\vartheta = FirstScale : step : LastScale$ **do**
5: $\mathcal{H}_{a,b,c} = GaussianConvolution(G_y, \vartheta)$
6: $\lambda_{1,2} = TextureOrientation(\mathcal{H}_{a,b,c})$
7: $\mathcal{R} = CurveDeviation(\lambda_{1,2})$
8: $\mathcal{S} = SimilarityMeasure(\lambda_{1,2})$
9: $\mathcal{E} = CurvelinearLikeliness(\mathcal{R}, \mathcal{S})$
10: **end for**
11: $\mathcal{E}_{\max} = Maximum(\mathcal{E}, \vartheta)$
12: $\mathcal{W} = Thresholding(\mathcal{E}_{\max})$
13: **return** $\mathcal{W}$
14: **END**

Figure 4.10: Pseudo code of HHF.

ϑ. Let $I = G_y$, each of the approximations $\mathcal{H}_a, \mathcal{H}_b, \mathcal{H}_c$ in a particular position, (x, y), is the convolution between $\mathbf{I}$ and a Gaussian kernel, $\mathbb{G}$. Then, the eigenvalues $\lambda_{1,2}$ of $\mathcal{H}$ are derived from each approximation in order to determine the texture orientation. After that, both curve deviation $\mathcal{R}$ and a similarity measure $\mathcal{S}$ are computed from $\lambda_{1,2}$. Then, the curvilinear likeliness $\mathcal{E}$ at specific ϑ is calculated from $\mathcal{R}$ and $\mathcal{S}$. Next, the approximation, $\mathcal{E}_{\max}$, is computed from the maximum response of the filter at all scales. Finally, the output, $\mathcal{W}$, is produced by assessing on each pixel where $\mathcal{W} = 1$ if $\mathcal{E}_{\max} \leq 0$, and vice versa.

4.3 Novel Algorithm of Hessian Line Tracking

In this section, Hessian Line Tracking (HLT) is proposed to overcome the weakness of HHF. The proposed HLT is composed of Hessian seeding and directional line tracking. It is an extension of a Hessian filter; however it significantly increases the accuracy of wrinkle localisation when compared with existing methods.

HHF is capable of estimating the presence of the ridge and valley patterns. However, the relationship between wrinkle pixels with those surrounding them has not previously been explored. Both current and adjacent pixels within the same line (wrinkle) should have the same value or less. Therefore, the main focus of this experiment is to introduce a new method, HLT, which has the capacity to track wrinkle lines based on neighbourhood properties. HLT was inspired by a multi-scale retinal vessel segmentation technique, which uses line tracking, proposed by Vlachos and Dermatas (2010).

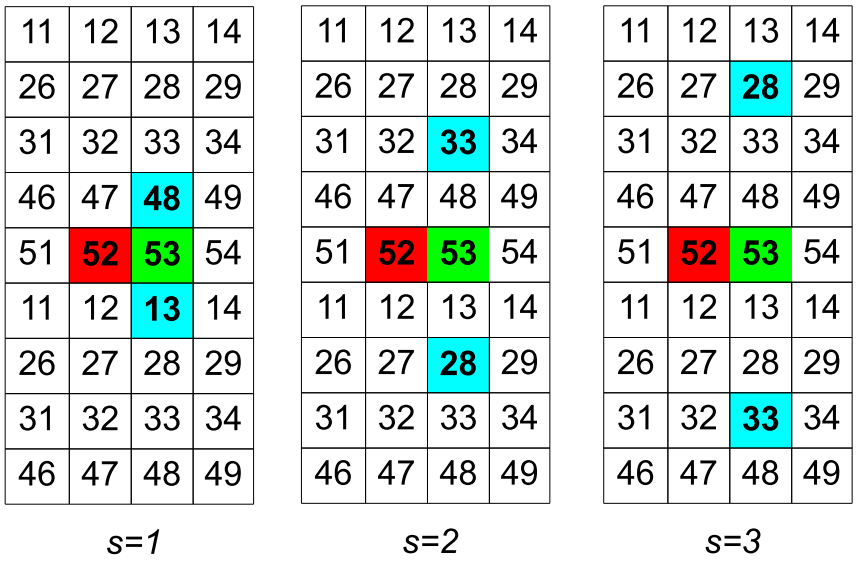


Figure 4.11: Multi-scale tracking. $\hat{s}$ is the scale of tracking. Red is the current pixel, green is the candidate pixel and blue is the background pixel.

However, they used the green channel as starting pixels or seeds for line tracking. This is not suitable for this work as the green channel is a weak representation of skin colour. The initialisation step is crucial in HLT because too many seeds will lead to over segmentation and too few seeds will cause under segmentation. Therefore, the ridge areas of HHF were selected as seeds for HLT.

Several image processing modules are implemented in HLT: seed extraction by HHF provides the most appropriate pixels as starting points, line tracking at multiple scales determines the confidence of each pixel belonging to a wrinkle line, and post processing removes outliers. It is worth mentioning that HLT is applied to all seeds at multiple scales. The total number of scales, Ξ, is selected based on the size of wrinkles to be detected and also the resolution of the original image. The scale will be increased if the wrinkle width is larger or the resolution is higher, and vice versa. Figure 4.11 presents the multi-scale tracking of the same pixel with scales of 1, 2 and 3. In this study, initial scale $\hat{s}$ is 1, in step of 1, and the last scale is 10.

First, a grey-scale image $\mathbf{I}$ is selected from the dataset and HHF is performed upon it. As described in the previous section, the HHF output is defined as $\mathcal{W}$ and it is divided to ridge and valley areas. Figure 4.12 illustrates an ideal wrinkle profile from non-wrinkle area to wrinkle area. The ridge is the area of interest as tracking always

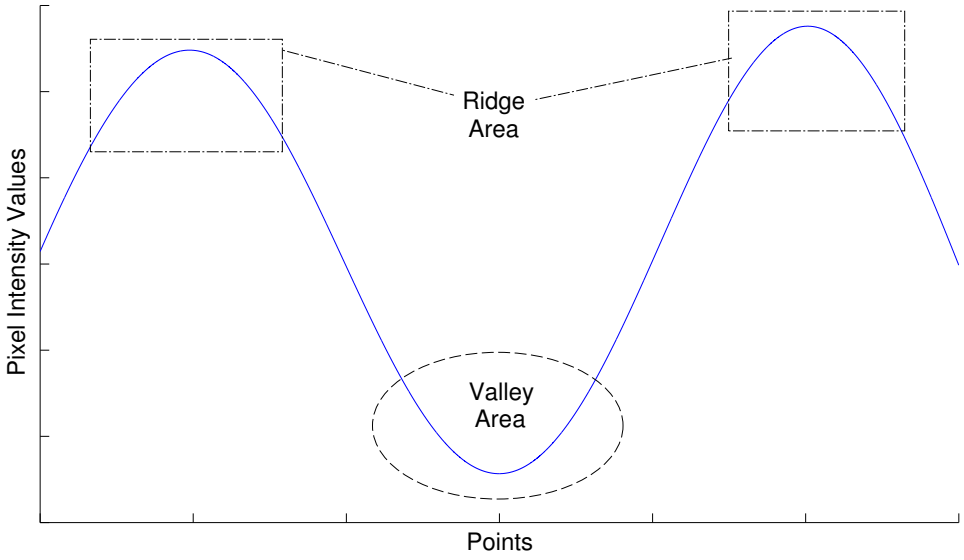


Figure 4.12: Ideal wrinkle profile. Valley area presents the wrinkle while ridge area shows the non-wrinkle area.

takes place from bright to dark areas. Therefore, the inverse of $\mathcal{W}$ is denoted as

$$\mathcal{W}'_{\text{inv}} = \begin{cases} 1 & \text{if } \mathcal{W} = 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.16}$$

.

Assume $\mathbf{I}(x,y)$ is thresholded by the $\mathcal{W}'_{\text{inv}}$ as

$$\mathbf{I}_{\text{new}}(x,y) = \begin{cases} \mathbf{I}_{\text{old}}(x,y) & \text{if } \mathcal{W}'_{\text{inv}} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.17}$$

where the pixel value of $\mathbf{I}$ is preserved if the logical value of $\mathcal{W}'_{\text{inv}}$ is true and vice versa. Let $\mathbf{h}_{p,\hat{r}}(x,y)$ be denoted as a set of candidates pixels to a particular center pixel or seed, sds ($sds \in \mathbf{I}_{\text{new}}$), where p represents the number of sampling points and $\hat{r}$ is the radius of the neighbourhood. These sampling points, $\mathbf{h}$, lie at Cartesian coordinates as,

$$(x_\theta, y_\theta) = (x + r\cos\theta, y + r\sin\theta) \tag{4.18}$$

where $\theta \in \{0°, 180°, 45°, 225°, 135°, 315°\}$ and $(x_\theta, y_\theta) \in \mathbf{h}_{p,\hat{r}}$ and $\hat{r}$ is unity by default. In this work, $(p,\hat{r}) = (6,1)$ where tracking of vertical directions are excluded. Therefore, each time a seed is derived, there are six elements in $\mathbf{h}$. Figure 4.13 illustrates the cross-sectional profile of different orientations. Let $\mathbf{p}_1$ denote the first background pixel located $\hat{s}$ pixel(s) away from the candidate pixel and let $\mathbf{p}_2$ as the second background

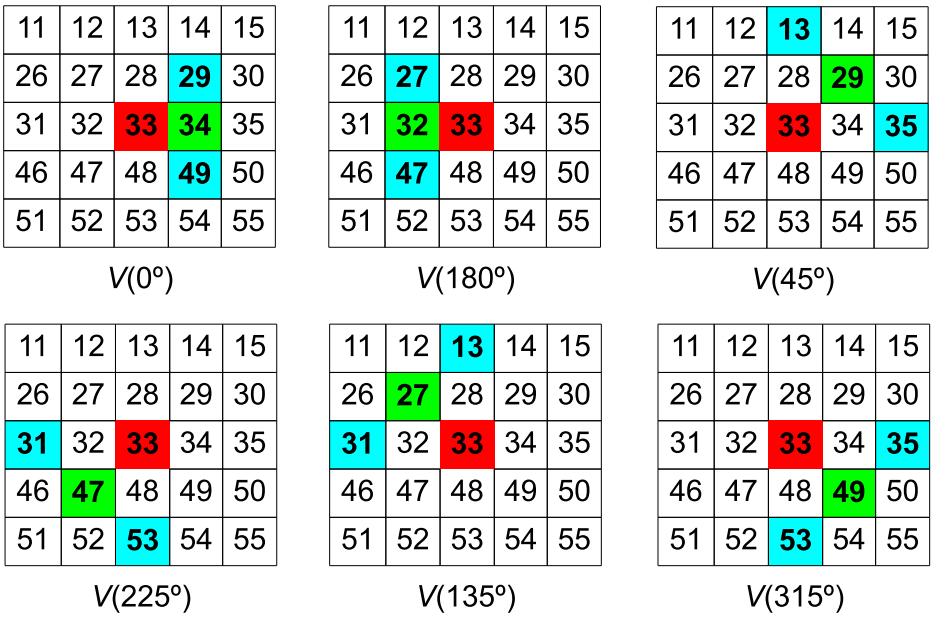


Figure 4.13: Cross-sectional profile. Red is the current pixel, green is the candidate pixel and blue is the background pixel.

pixel located $\hat{s}$ pixel(s) away from the candidate pixel but in the opposite direction. The cross-sectional profile $\mathcal{V}$ at a particular scale, $\hat{s}$, is estimated as

$$\mathcal{V}(x_\theta, y_\theta, \hat{s}) = \mathbf{I}(\mathbf{p}_1, \hat{s}) + \mathbf{I}(\mathbf{p}_2, \hat{s}) - \mathbf{I}(x, y, \hat{s}) - \mathbf{I}(x_\theta, y_\theta, \hat{s}) \tag{4.19}$$

where $\hat{s} \in \{1, 2, \ldots, \Xi\}$ and Ξ is the total number of scales, in this case $\Xi = 10$. Note that $\mathbf{I}(x, y, \hat{s})$ is the current pixel or seed while $\mathbf{I}(x_\theta, y_\theta, \hat{s})$ is the candidate pixel. If the current and candidate pixels belongs to the set of wrinkles, the cross-sectional profile $\mathcal{V}$ has a large positive value. If the current and candidate pixels belongs to the non-wrinkle pixels, $\mathcal{V}$ has a negative value or is near to zero. The winner pixel $\mathbf{z}$ with a maximum positive cross-sectional profile, exceeding a predefined positive threshold t, is defined as

$$\mathbf{z} = \arg\max_{\mathbf{h}} \left(\{ \mathcal{V}(x_\theta, y_\theta, \hat{s}) \} > t \right) \tag{4.20}$$

where $\mathbf{z}$ could be one of the elements of $\mathbf{h}$ or $\mathbf{z}$ could be null if none is larger than t. t is an important threshold in identifying the intensity difference between the current pixel and the candidate pixel relative to the background pixels. The default value of t is 9, which is the best threshold that has been empirically explored for FORERET. Note that $\mathbf{z}, \mathbf{p}_1, \mathbf{p}_2$ refer to a particular coordinates, (x, y). If $\mathbf{z}$ is not null, then the confidence

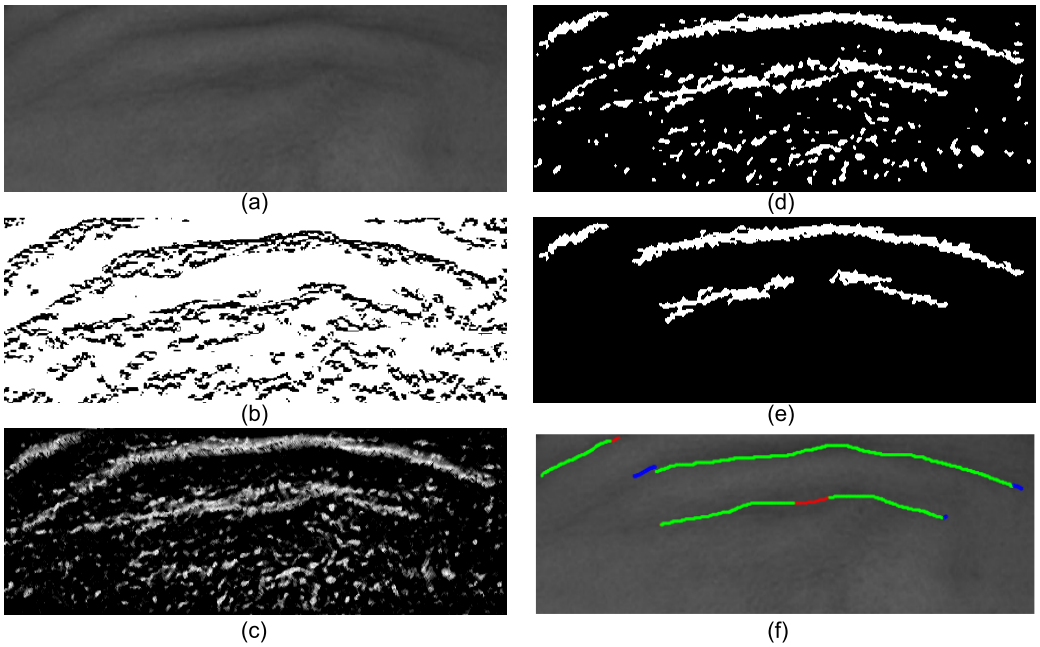


Figure 4.14: Hessian line tracking. (a) Original image $\mathbf{I}$, (b) Seeds sds, (c) confidence array $\mathcal{C}$, (d) initial wrinkle map $\mathcal{B}$, (e) wrinkle map after post processing of filtering, (f) inter-coder reliability results where red is ground truth, green is true positive and blue is false positive.

array, $\mathcal{C}$, is increased by one as

$$\mathcal{C}_{\text{new}}(x,y) = \begin{cases} \mathcal{C}_{\text{old}}(x,y)+1 & \text{if } \mathbf{z} \neq \text{null}, (x,y) = \mathbf{w}_0 \\ \mathcal{C}_{\text{old}}(x,y) & \text{otherwise} \end{cases} \tag{4.21}$$

where the next pixel to be tracked is $\mathbf{I}(\mathbf{z})$. If $\mathbf{z}$ is *null*, the next candidate pixel is drawn from the seeds. Note that the tracked pixel is unique from candidate pixels, otherwise the tracking will be redundant. Once multi-scale line tracking is completed for all scales $\hat{s}$, the initial wrinkle map $\mathcal{B}$ is generated by consulting the confidence array as

$$\mathcal{B}(x,y) = \begin{cases} 1 & \text{if } \mathcal{C}_{\text{new}}(x,y) \geq \Xi \\ 0 & \text{otherwise} \end{cases} \tag{4.22}$$

where Ξ is the total number of HLT scales and $\Xi = 10$. $\mathcal{B}$ is *true* if $\mathcal{C}_{\text{new}}$ is greater than or equal to amount of scales, Ξ. If the wrinkle width is of particular interest, it can be estimate from a predefined scale. In other words, Ξ can be defined according to the width preference. For example, large Ξ is preferred for the coarse wrinkles and a smaller one for the fine wrinkles.

1: **BEGIN**
2: $\mathbf{I} = ImageAcquisition()$
3: $\mathcal{W} = HybridHessianFilter(\mathbf{I})$
4: $sds = Inverse(\mathcal{W})$
5: **for** $\hat{s} = InitScale : step : FinScale$ **do**
6: **for** $k = 1 : length(sds)$ **do**
7: $(x,y) = sds(k)$
8: $[\mathbf{h}, \mathbf{p}_1, \mathbf{p}_2] = Neighborhood(x,y,\hat{s})$
9: $\mathcal{V}(\mathbf{h}) = CrossSectionalProfile(\mathbf{h}, \mathbf{p}_1, \mathbf{p}_2)$
10: $\theta = \arg\max_{\mathbf{h}} \{\mathcal{V}(\mathbf{h}_1) : \mathcal{V}(\mathbf{h}_6) > t\}$
11: **if** $\theta != null$ **then**
12: $\mathcal{C}_{\text{new}}(x,y) = \mathcal{C}_{\text{old}}(x,y) + 1$
13: $(x_\theta, y_\theta) = (x + r\cos\theta, y + r\sin\theta)$
14: **while** $(x_\theta, y_\theta) \notin sds$ **do**
15: Repeat from step 9
16: **end while**
17: **end if**
18: **end for**
19: **end for**
20: **for all** (i,j) in $\mathcal{C}_{\text{new}}$ **do**
21: **if** $\mathcal{C}_{\text{new}}(i,j) >= \Xi$ **then**
22: $\mathcal{B}(i,j) = 1$
23: **else**
24: $\mathcal{B}(i,j) = 0$
25: **end if**
26: **end for**
27: $\mathcal{B}_{\text{mf}} = MedianFiltering(\mathcal{B})$
28: $\mathcal{B}_{\text{df}} = DirectionalFiltering(\mathcal{B}_{\text{mf}})$
29: $\mathcal{X} = AreaThresholding(\mathcal{B}_{\text{df}})$
30: **return** $\mathcal{X}$
31: **END**

Figure 4.15: Pseudo code of HLT.

Figure 4.14 shows the outputs of HLT, while (a), (b) and (c) illustrate the original image $\mathbf{I}$, seeds sds and confidence array $\mathcal{C}$, respectively. Due to noise, and very rarely, some non-wrinkle pixels can be tracked in all scales and erroneously are considered as part of the wrinkle map. As a final step, post processing is required to reduce the presence of noisy pixels from the wrinkle map. This involves median filtering, directional filtering and area filtering. Median filtering is used to remove pepper noise pixels that do not belongs to any wrinkles. A 3×3 median filter is applied to $\mathcal{B}$ and the output is

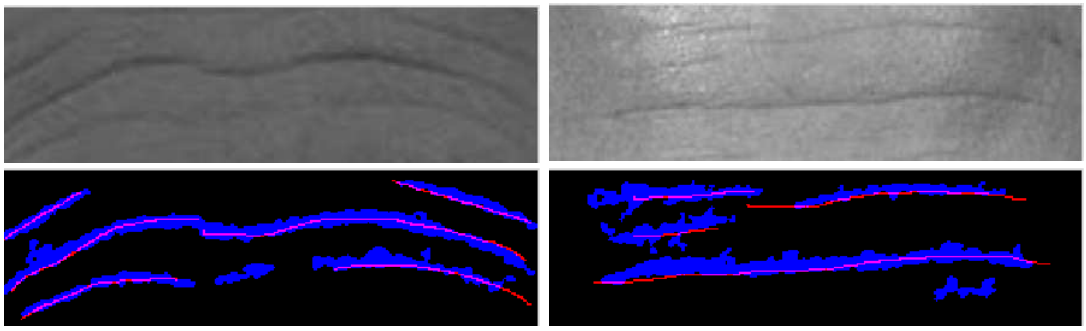


Figure 4.16: Examples of FORERET images. Top row shows the original image cropped from FERET. Bottom row shows the human annotation (red) and automatic wrinkle detection using HLT (blue). Each column represents the same image.

$\mathcal{B}_{\text{mf}}$. After median filtering, some noise may still remain. The binary image $\mathcal{B}_{\text{mf}}$ is transformed using four morphological openings with line structuring elements oriented in four different directions $0°$, $45°$, $90°$ and $135°$. These line structuring elements use a length of 3 pixels to preserve only wrinkle-like structures with a length equal or larger than 3 pixels. $\mathcal{B}_{\text{df}}$ is derived using the logical OR of the four directional filtering responses and is shown as Figure 4.14(d). Next, area thresholding is used to remove spots or roundish marks with a size lower than 250 pixels. In the end, wrinkles, $\mathcal{X}$, are detected as illustrated in Figure 4.14(e) and the center line of each wrinkle is used for validation as presented in Figure 4.14(f). Figure 4.15 depicts a summary of HLT pseudo-code.

4.4 Experimental Results

Two datasets, FORERUS and FORERET, were used for performance assessment of automatic wrinkle detection. Coders were given the FORERUS images and they performed wrinkle annotation by saving the wrinkle mask in the Matlab logical format. Two of them are computer scientists (denoted as A and B) and third coder is an experienced beauty therapist (denoted as C). Meanwhile, another dataset namely FORERET was created using the same procedure as described in Section 4.1.1. It was randomly selected from FERET as the image resolution is about half of Bosphorus. The annotation is based on coder A only. Figure 4.16 shows the examples of FORERET images. Measurements of JSI and ACC were used for performance assessment of automatic wrinkle detection. There are two experiments in this section. First, performance assessment is validated on FORERUS. It is followed by the second experiment on FORERET. The last section discusses the work.

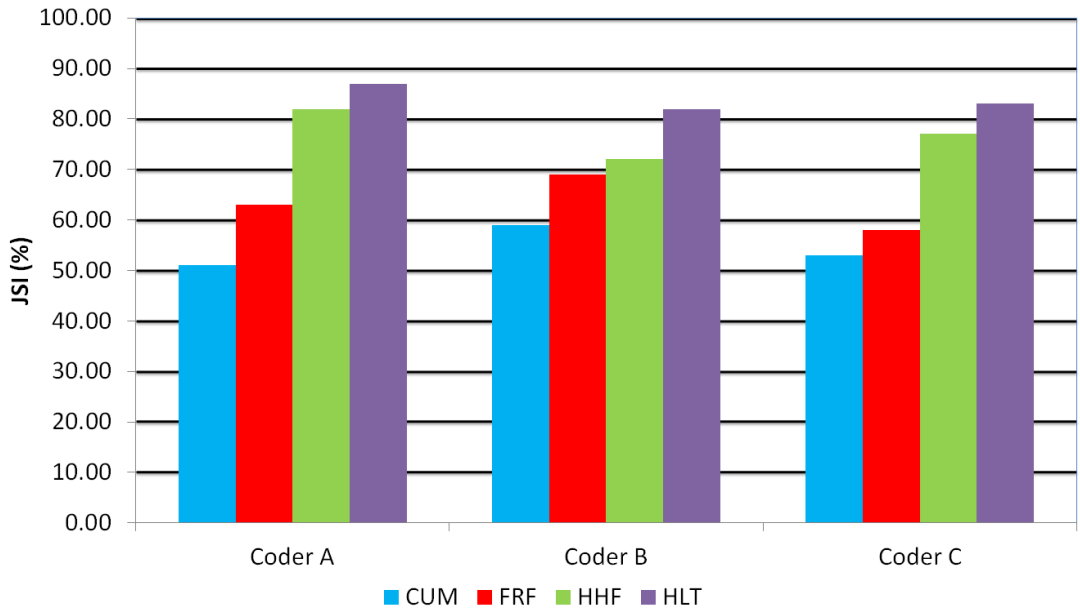


Figure 4.17: JSI of automatic wrinkle detection versus benchmark of coder A, B and C.

4.4.1 Performance Assessment on FORERUS

Figure 4.17 presents the results of wrinkle detection using state-of-the-art methods CLM (Cula et al., 2013), FRF (Frangi, 2001), the proposed methods, HHF and HLT. Overall, it is found that HLT performs better than CLM, FRF and HHF with an accuracy of 84% and STD 2.16%. The accuracies of CLM, FRF and HHF are 54.33%, 63.33% and 77.00% with STD 3.40%, 4.50% and 4.08%, respectively.

Figure 4.18 shows some results of wrinkle detection using HHF and HLT. The first row is the original image, the second row is the output of HHF and the last row is the HLT output. Figure 4.18(a) shows that HHF achieves a JSI of 58.07% while HLT has 87.72%; Figure 4.18(b) shows that HHF has a JSI of 40.07% while HLT has 63.59%; Figure 4.18(c) illustrates that HHF has a JSI of 3.54% while HLT has 50.28%; and Figure 4.18(d) depicts that HHF scores 34.34% while HLT is 87.45%.

On coarse wrinkles (as in Figure 4.18(a) and Figure 4.18(b)) both HHF and HLT

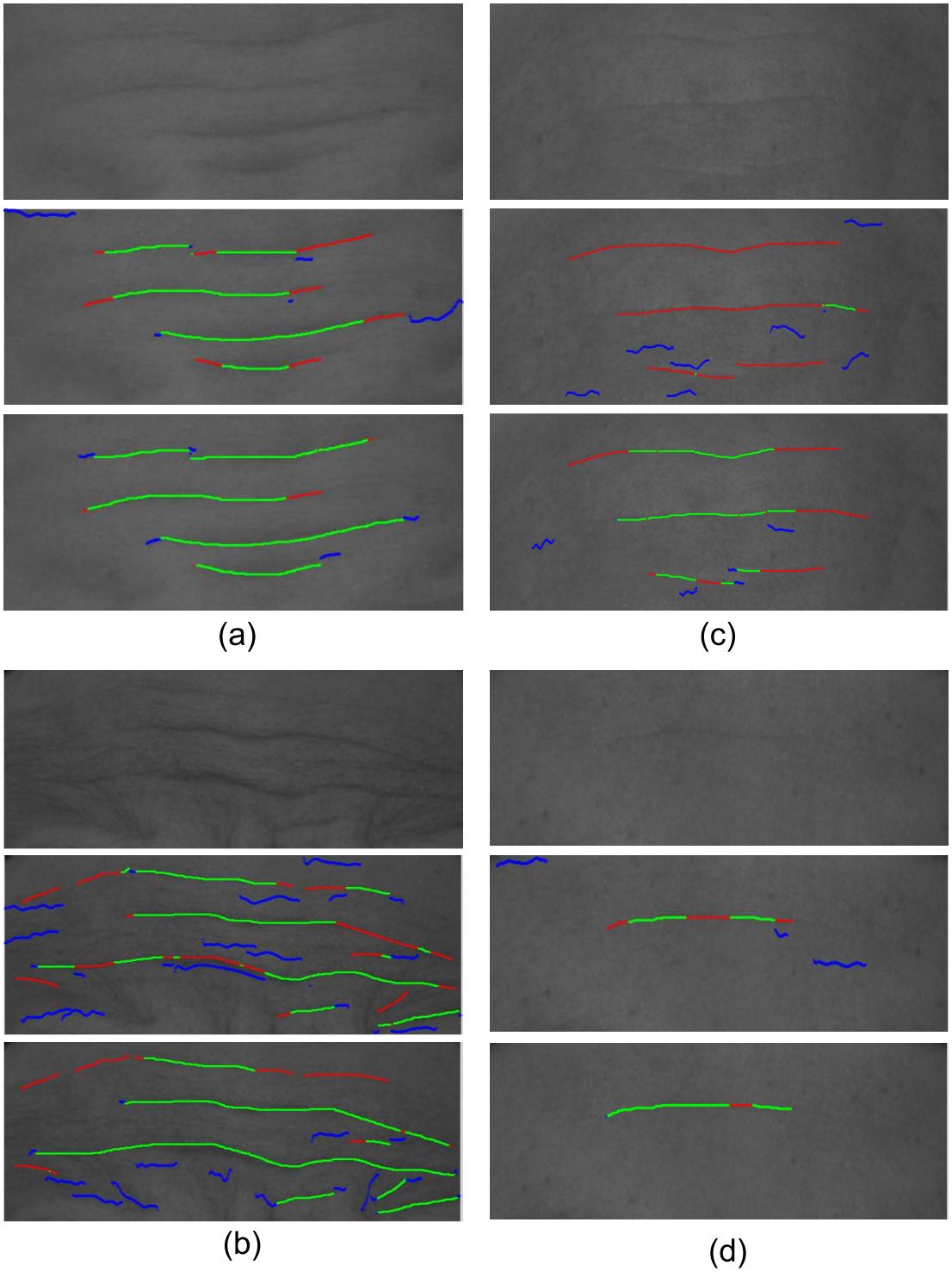


Figure 4.18: HHF vs HLT of different FORERUS images. Green is a hit on ground truth, red is a miss and blue is a false alarm. The first row is the original image, the second row is the output of HHF and the last row is the HLT output. (a) JSI of HHF is 58.07% and HLT is 87.72%. (b) JSI of HHF is 40.07% and HLT is 63.59%. (c) JSI of HHF is 3.54% and HLT is 50.28%. (d) JSI of HHF is 34.34% and HLT is 87.45%.

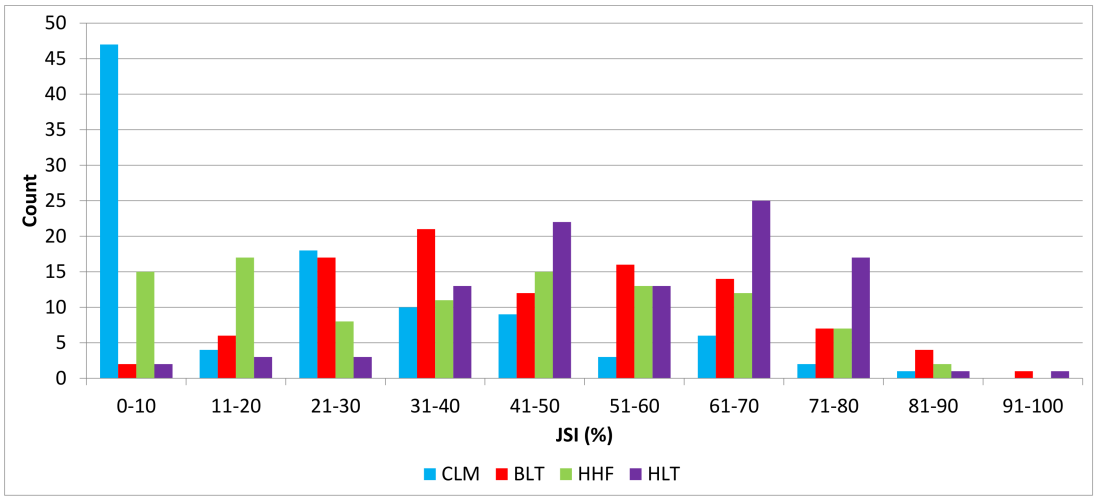


Figure 4.19: Wrinkle detection accuracy between HHF and HLT. In 100 images, HLT achieved accuracy of 79% while HHF only hit 49%.

manage to detect the line with the JSI greater then 40%. However, on a fine wrinkle (as in Figure 4.18(c) and Figure 4.18(d), HHF performs poorly with JSI below 40% but HLT still perform well in both cases with JSI of 50.28% and 87.45%, respectively.

4.4.2 Performance Assessment on FORERET

In this section, the performance of wrinkle detection is validated by comparing CLM, FRF, HHF and HLT on FORERET images. Figure 4.19 shows the accuracy of wrinkle detection. In 100 images, results showed that HLT achieves better than others with accuracy of 79%, where HHF is 49%, FRF is 30% and CLM is 21%, while STD are 17.28%, 23.50%, 27.05% and 23.46%, respectively. Figure 4.20 illustrates the results of wrinkle detection. For coarse wrinkles, both HHF and HLT achieved a good JSI over 75% as shown in Figure 4.20(a). However, HHF is not as reliable as HLT where coarse wrinkles appear as shown in Figure 4.20(b). For fine wrinkles, HLT performs better than HHF in both cases as shown in Figure 4.20(c) and (d). Overall, HLT outperforms than others on FORERET images. Hence, it is used to extract the wrinkles in the following experiments.

A Student's t-test was performed between HHF and HLT, assessing the null hypothesis that there is no difference between HHF and HLT. This found that $t = 2.40$, $d.f. = 198$, $P = 0.02$, mean difference is 6.91, standard error difference is 2.88, upper of 95% confidence is 1.22 and lower is 12.60. As $P < 0.05$, the null hypothesis is rejected. In other words, there is a significant difference between HHF and HLT.

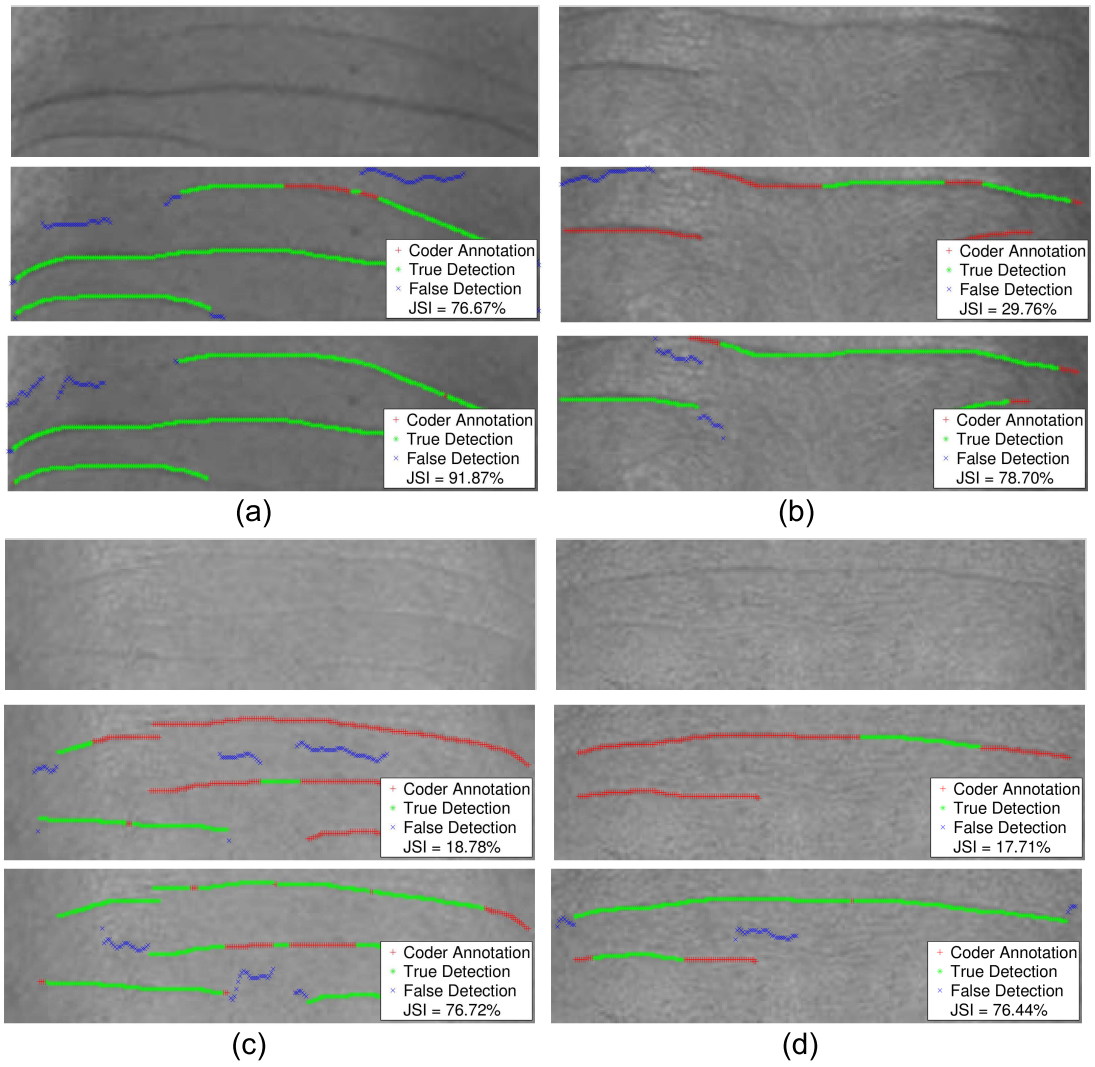


Figure 4.20: Wrinkle detection results of FORERET images. Green is a hit on ground truth, red is a miss and blue is a false alarm. The first row is the original image, the second row is the HHF output and third row is the HLT output. Each column refers to the same image.

4.4.3 Discussion

Experimental results showed that HHF is capable of detecting the wrinkles compared to CLM and FRF. One reason for this might be that in HHF the directional gradient has greatly smoothed the image and preserved the data of interest. Cula et al. (2013) argued that red channel of image and histogram equalisation will strengthen the wrinkle image, but the CLM result was poor. Histogram equalisation is good when the histogram of the image is confined to a particular region, but it is not useful on certain images

which have large intensity variations where the histogram covers a large region.

Wrinkles usually show a large variability in length, width and pattern in either different images and within the same image which challenging the development of automatic wrinkle detection. Therefore, a second novel algorithm, HLT, has been proposed. This method is based on seed extraction by HHF and HLT to overcome the weakness of HHF and capture wrinkle variability in the whole image. The number of seeds affects the robustness of the confidence array. If a small group of seeds is selected from the image, the detection accuracy may decrease because the seeds affect the wrinkle tracking probability. On the other hand, the selection of an excessive number of seeds may increase the false positive rate. Therefore, a pre-processing step must be undertaken in order to ensure that the seeds are representative, sufficient and minimal. HHF highlights potential wrinkle pixels based on a Gaussian distribution between ridge and valley curves. It is found that a wrinkle pattern was similar to the valley sequence. However, the ridge distribution in relation to wrinkles had not the previously been explored. Therefore, in this study, ridge distribution is proposed as the seed for HLT and results showed that it is a robust method for wrinkle detection.

Originally, a line tracking method was proposed by Vlachos and Dermatas (2010), but had some drawbacks if in the context of wrinkle detection. First, the tracking involved all neighbourhood pixels. It computed the confidence array in all directions which is not ideal for directional wrinkles. If the target of interest is a horizontal wrinkle, the tracking should exclude vertical tracking where the noises like hair might appear. Therefore, both 90° and 270° in HLT are excluded from the tracking process with the assumption that the major orientation of forehead wrinkles is horizontal. In addition, the exclusion of the current or reference pixel in calculating the cross-sectional profile parameter $\mathcal{V}$ is not appropriate. It is important to show the relationship between the current and candidate pixels. If both have similar intensity and darker than the background pixels, both pixels are highly similar and possibly belong to a wrinkle. Therefore, the calculation of the cross-sectional profile parameter has been modified by including the current pixel. Thirdly, threshold t is modified for noise adaptation. By default, the value is set to 7 for all cases. However, this value is not appropriate for all kinds of background. If specular light reflection is too strong, it will lead to over segmentation, otherwise, if the image is too smooth or the wrinkle to be detected is too fine, then under segmentation would happen. In order to reduce both problems, in HLT, an adaptive threshold is introduced. If the largest area of connected pixels in the initial wrinkle map is higher than 10% of the image size, then the threshold is set to a higher value of 15 to reduce the search

space. If an empty map is generated, then the threshold is set to a lower value of 3. These parameters were tested in FORERET. For other datasets, it was necessary to adjust the parameters according to the resolution and lighting conditions.

Although the overall performance of HLT is better than HHF, there are a few issues to be addressed. On 18 out of 100 images, HHF performed better than HLT. A few even decreased the JSI of HLT detection lower than 40%. These images have a mixture of coarse and fine wrinkles. Moreover, three images with fine wrinkles resulted zero detection with either method. This issue was caused by the nature of the image where the line is not very clear and even the coders achieved low inter-coder reliability on these images. A topology data analysis could be investigated for different types of wrinkles. In addition, JSI measurement is limited to line-based validation; additional computation concerning features such as wrinkle depth, length and width can be added in the near future. Moreover, this experiment can be extended to detect the wrinkles on the face especially around the eye corner (crows' feet) and mouth corner.

4.5 Summary

In this chapter, two novel methods of wrinkle detection, HHF and HLT, are described. HHF is capable of detecting the ridge and valley patterns, while HLT extends it by exploring the neighbourhood relationship for multi-scale tracking. From the experimental results, HLT outperformed HHF, FRF and CLM for wrinkle detection where the reliability is as close as human annotation. From a preliminary study in Appendix C, the findings showed some evidence backing the use of wrinkles to predict face age. Thus, in the following chapters, HLT is utilised as a wrinkle extractor for face age estimation. A relationship between the detected wrinkles and facial aging is investigated. Wrinkles are extracted and presented in a compact pattern where different regression methods are implemented to predict the face age. A feature pattern, MWP, is proposed and discussed in details in the next chapter.

Chapter 5

Multi-scale Wrinkle Patterns

In order to detect different types of wrinkles, this chapter introduces a novel feature descriptor, MWP, based on the wrinkle statistics, to predict the face age. It involves image resizing, wrinkle extraction using HLT, feature representation as MWP and age estimation using SVR. Benchmark algorithms FACE++, BIF and KLBP are compared to MWP. The performance is accessed by two measurements, MAE and CS.

5.1 Introduction

Wrinkles are interesting features which can be embedded in several image-based applications related to human skin. However, wrinkle-based age estimation research has not been widely addressed. Figure 5.1 illustrates an example of face wrinkle variety. Figure 5.1(a) is the original image and Figure 5.1(b) is the wrinkle detection output of the full face where red shows a high confidence array as wrinkles in HLT. It is noticed that forehead, crows' feet, lower eyelid and nasolabial regions present wrinkles coarser than other regions. Figure 5.1(c) is a cropped forehead image of Figure 5.1(a), Figure 5.1(d) is the result of HLT detection with 50 regions detected as wrinkles, and Figure 5.1(e) is the half size image of Figure 5.1(c) where only 9 regions found as wrinkles. This implies that the different sizes of image demonstrates different amounts of wrinkles. Therefore, it is interesting to find out how the multi-scale wrinkles would contribute to the performance of age estimation.

Conventionally, Cula et al. (2013) proposed the wrinkle features using edge detector. However, the extracted wrinkles were not validated against the ground truths.

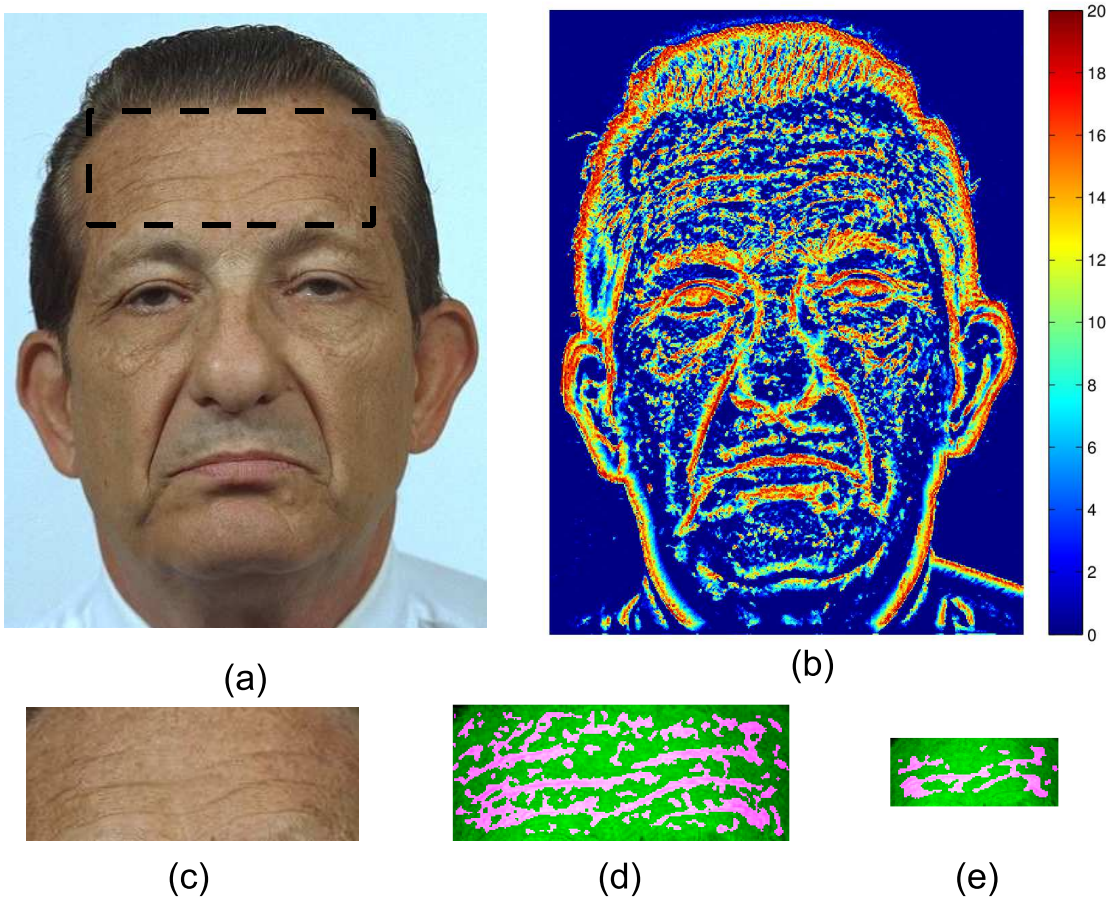


Figure 5.1: Different kind of face wrinkles. (a) original image; (b) wrinkle detection output of the full face where red shows a high confidence array as wrinkles in HLT; (c) cropped forehead image as black dotted line in (a); (d) 50 regions detected as wrinkles; (e) half size image of (c) and 9 regions found as wrinkles. Note that this figure was redrawn from FERET.

Frangi (2001) proposed a vessel enhancement filter for examining the multi-scale second order structure of an image. However, their filter is not reliable for wrinkle detection. Batool and Chellappa (2015) proposed a fast detector of facial wrinkles based on Gabor features using image morphology and geometric constraints. However, the proposed detector cannot deal with different types of wrinkles because a predefined threshold only suitable for either coarse, medium or fine wrinkles.

In this section, a novel feature representation, Multi-scale Wrinkle Patterns (MWP), is proposed for face age estimation. It extracts the features from the local patches without an extensive geometric modelling. First, facial landmarks are located using

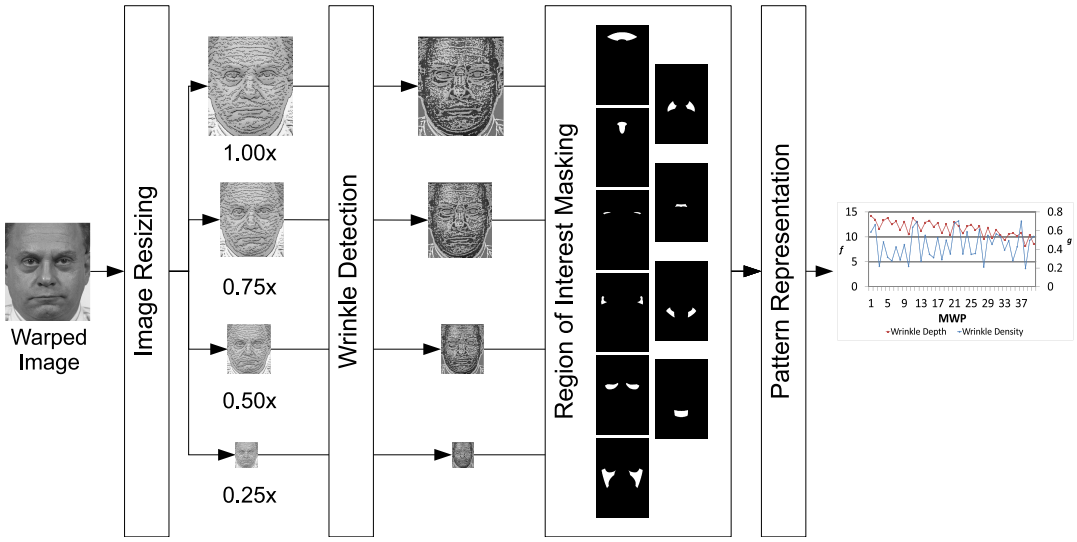


Figure 5.2: Flow chart of MWP. It consists of four steps: image resizing, wrinkle detection, ROI masking and pattern representation. Note that 1.00x means the original size of FERET image and 0.50x presents the half size of 1.00x. Ten regions of wrinkle template are forehead, glabella, upper eyelids, crows feet, lower eyelids, cheeks, nasolabial grooves, upper lids, marionette and lower lips.

the Face++ detector and then the face is normalised using a linear transformation. After that, the face wrinkles are extracted by HLT. A wrinkle template which consists of ten predefined wrinkle regions is manually created and the wrinkles are constructed into a robust feature representation using MWP. Finally, the age is estimated by implementing SVR. The performance of algorithms are assessed using MAE on FERET. It is observed that MWP produces a comparable MAE to KLBP but with lower computational cost than KLBP.

5.2 Multi-scale Wrinkle Patterns Algorithm

Figure 5.2 demonstrates the proposed Multi-scale Wrinkle Patterns (MWP). It consists of four steps: image resizing, wrinkle detection, ROI masking and wrinkle pattern representation.

5.2.1 Image Resizing

Given a warped image $\mathbf{I}$, it is scaled with different ratios $\hat{r} \in \{\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_\tau\}$ where τ indicates the total number of ratios. In this case, $\tau = 4$ and $\hat{r} \in \{1.00, 0.75, 0.50, 0.25\}$ where $\hat{r} = 1.00$ represents the original image and $\hat{r} = 0.50$ represents half size of

the original image. For each scale, the directional gradient (G_x, G_y) of $\mathbf{I}$ is computed from the greyscale image. G_y emphasizes a y-variation and it is used as the input for calculating the Hessian filter, $\mathcal{H}$. After G_y is convoluted by $\mathcal{H}$, a logical mask, $\mathcal{W}$ is generated as Eq. (4.15). Due to the HLT detection focuses on ridge area, $\mathcal{W}_{\text{inv}}$ is used as the seeds (see Section 4.3).

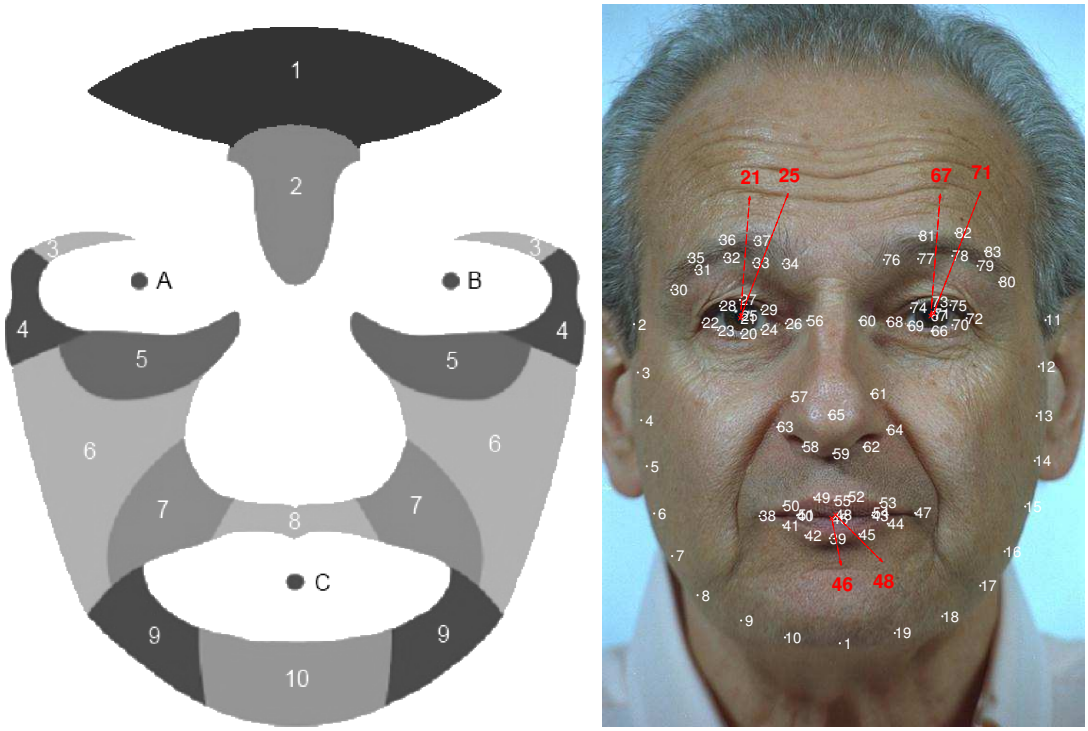
5.2.2 Wrinkle Detection

First, a seed, sds, is drawn from $\mathcal{W}_{\text{inv}}$. Then, candidate pixels $\mathbf{h}$ and background pixels $\mathbf{p}_1, \mathbf{p}_2$ are derived from it with $\hat{s} = 1$ and $\theta = 0^\circ$. Cross-sectional profile $\mathcal{V}\left(x_\theta, y_\theta, \hat{s}\right)$ is computed as Eq. (4.19). After that, the computation is repeated on different degrees in order to determine the winning pixel $\mathbf{z}$. The confidence array $\mathcal{C}$ is incremented by one if a winning pixel is found. Next, the tracking continues with either winning pixel or a seed drawn from $\mathcal{W}_{\text{inv}}$. Once the tracking is completed for the whole image, $\hat{s}$ is increased to the next scale and the calculation continues until the last scale. Finally, the initial wrinkle map $\mathcal{X}$ is generated by validating the confidence array $\mathcal{C}$ as Eq. (4.22). Details of wrinkle detection are given in Section 4.3.

5.2.3 Region of Interest Masking

In order to extract the wrinkle from a specific region, a standard wrinkle template (as shown in Figure 5.3a) is utilised to normalise the face. A mean face is generated from FERET where each face region in the template corresponds to a specific muscle underneath the skin (Ekman et al., 2002). When human ages, a continuous muscle movements on a particular region forms a variety of wrinkles including coarse and fine wrinkles. Therefore, a standard template is used for face alignment and region masking in the following experiments.

Figure 5.3a illustrates the wrinkle template or mask with ten predefined wrinkle regions and fixed coordinates for eyes and mouth. It was cropped and scaled for illustration purposes and the original size of this image is 512 x 768 pixels which is the same size as FERET images. These regions have been identified and created manually through the review (Aznar-Casanova et al., 2010; Choi et al., 2011). The margins of top, bottom, left and right are 70.5, 209.5, 50.5 and 56.5 pixels, respectively. The distance between A and B is 218 pixels, A and C is 236.6 pixels, B and C is 237.49 pixels. Numbers 1 to 10 mean the wrinkle regions of forehead, glabella, upper eyelids, crow's feet (or eye corners), lower eyelids (or eyebag), cheeks, nasolabial grooves (or



(a) Wrinkle template (b) FERET image

Figure 5.3: Wrinkle template and Face++ landmarks. (a) Wrinkle template with ten predefined wrinkle regions and fixed coordinates of eyes and mouth. The wrinkle regions are forehead (1), glabella (2), upper eyelids (3), crow's feet (4), lower eyelids (5), cheeks (6), nasolabial grooves (7), upper lips (8), marionette (9) and lower lips (10). Symbols A, B and C are fixed coordinates used for face normalisation. (b) Result of the Face++ detector (Zhou et al., 2013). White landmarks are detected by the Face++ detector and the red labels represent the center of the eyes and mouth. Note that this image was redrawn from FERET (Phillips et al., 2000).

nasolabial folds), upper lips, marionette and lower lips; the corresponding areas are 15745, 6437, 1475, 4169, 9417, 21759, 9216, 2298, 7376 and 7403 pixels; the corresponding notations are $\mathcal{Z}_1$ to $\mathcal{Z}_\nu$ where $\mathcal{Z}_i$ is a binary image and $\nu = 10$. Due to the area sizes being fixed, a standard feature vector will be produced either in training or testing data. The center coordinates of left eye, right eye and mouth are (145, 247), (363, 247) and (254, 457), respectively.

Prior to perform the face alignment, a set of face landmarks is needed. These landmarks must be obvious and easily identified from a shape. For instance, eyes and mouth could be considered good landmarks for face alignment. Therefore, in this work, the center landmarks of eyes and mouth are used for performing the local transformation. First, a face image is detected by the Face++ detector (Zhou et al., 2013) and a

total of 83 landmarks are obtained as shown in Figure 5.3b. Only six landmarks are used for estimating the center of left eye, right eye and mouth, others are discarded. The center of left eye is averaged from the landmarks of 21 and 25, the center of right eye is averaged from the landmarks of 67 and 71, the center of mouth is averaged from the landmarks of 46 and 48. Symbols A, B and C (as Figure 5.3a) correspond to the center of left eye, center of right eye and the center of mouth, respectively. Due to the limitations of the Face++ detector, it is observed that a small number of detection errors which were identified and corrected manually.

Based on the fixed three points (symbols A, B and C), each face image is normalised to the mask (as shown in Figure 5.3a) using a piecewise affine warping (see Section 3.4.2). With consistent area size for the ten face regions, all regions are used to construct the wrinkle patterns, which produce a standard feature vector subsequently used for training and testing. From the wrinkle map $\mathcal{X}$ and region of interest $\mathcal{Z}_i$, the wrinkle image $\mathcal{Y}_i$ of a particular region, i, is defined as

$$\mathcal{Y}_i(x,y) = \begin{cases} \mathbf{I}(x,y) & \text{if } \mathcal{X}(x,y) \cap \mathcal{Z}_i(x,y) = 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

5.2.4 Pattern Representation

Let MWP $= \{\mathfrak{f}_1, \mathfrak{f}_2, \ldots, \mathfrak{f}_\psi, \mathfrak{g}_1, \mathfrak{g}_2, \ldots, \mathfrak{g}_\psi\}$, where $\psi = \tau \times \nu$. Note that, in this work, total image scale, $\tau = 4$, and total wrinkle regions, $\nu = 10$. The wrinkle intensity $\mathfrak{f}_i$ of one particular region, i, is defined as

$$\mathfrak{f}_i = \log \sum_{x=1}^{wt} \sum_{y=1}^{ht} \mathcal{Y}_i(x,y) \tag{5.2}$$

where wt and ht are the width and height of $\mathcal{Y}$. The wrinkle density $\mathfrak{g}_i$ of one particular region, i, is defined as

$$\mathfrak{g}_i = \frac{area_1(i)}{area_2(i)} \tag{5.3}$$

where $area_1$ is the wrinkle area found in a particular region i and $area_2$ is the area of region i.

Table 5.1: MAE of different age groups.

Age Group	10-19	20-29	30-39	40-49	50-59	60-70
Total images	34	847	483	413	171	52
FACE++ (Zhou et al., 2013)	11.06	7.18	5.57	8.09	6.98	3.56
BIF (Guo et al., 2009a)	7.85	3.99	4.07	4.85	5.36	6.46
KLBP (Ylioinas et al., 2013)	7.59	3.81	3.63	3.93	4.48	7.40
Proposed MWP	8.59	3.35	3.93	4.10	4.92	9.31

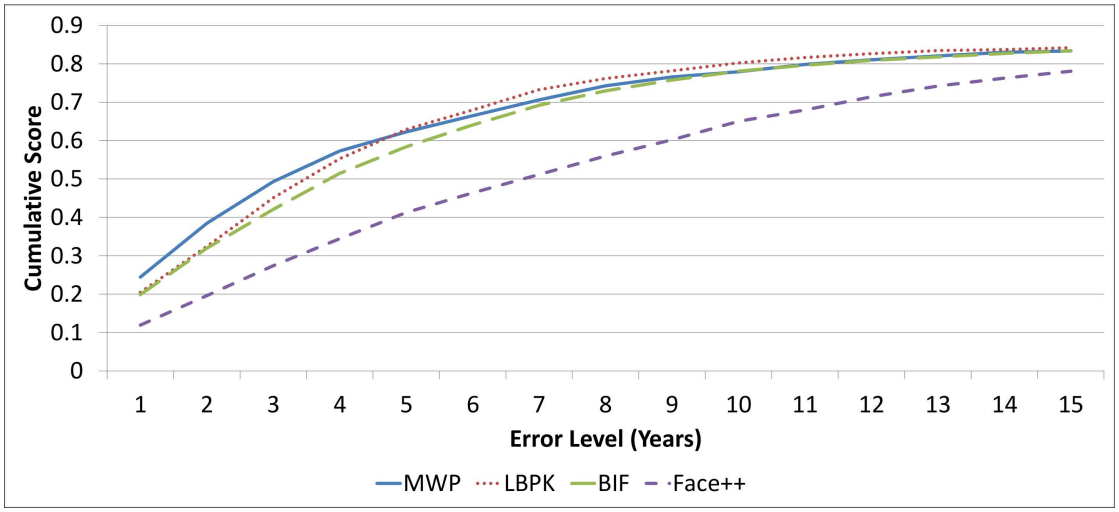


Figure 5.4: Cumulative score of age estimation on FERET. It is a comparison between MWP, KLBP, BIF and FACE++.

5.3 Experimental Results

Batool and Chellappa (2015) claimed that FERET is a low resolution dataset like FGNET. However, due to wrinkle extraction requiring decent resolution, the performance of proposed algorithm is assessed on FERET, the highest resolution amongst the benchmark datasets such as FGNET and MORPH. In addition, FERET consists of a few hundred age-separated face images of subjects with the age difference of 18 months or more and the age range is between 10 and 70. 2000 samples of FERET were randomly selected from the original dataset to compare the performance of three age estimation algorithms - this includes the benchmark algorithms, namely BIF and KLBP, and the proposed MWP. The evaluation protocol was a 10-fold cross validation, where both training and testing sets are disjoint. This section presents three experimental results: automatic age estimation on FERET, the effect of gender on face wrinkle and the effect of facial expression against face wrinkle.

Table 5.2: Comparison of MAE results on FERET. Note that Face++ results were calculated by replacing the fail detection with ground truth and STD means standard deviation.

Method	Face++	BIF	KLBP	MWP
MAE	6.94	4.42	4.01	4.02
STD	5.38	3.77	3.23	3.74

5.3.1 Age Estimation using MWP on FERET

Table 5.1 shows MAE results of different age groups of FERET. It is noticed that KLBP has the lowest MAE on the majority of age groups apart from age group of 20-29 (by MWP) and age group of 60-70 (by FACE++). Table 5.2 presents the comparison of MAE results on the state-of-the-art methods. The MAEs of FACE++, BIF, KLBP and MWP are 6.94 ($\pm$5.38), 4.42 ($\pm$3.77), 4.01 ($\pm$3.23) and 4.02 ($\pm$3.74), respectively. KLBP is a superior method in representing local features with the lowest MAE, at 4.01. However, MWP produced a comparable MAE of 4.02 with lower computational cost where it only requires 80 features compared to KLBP with 32768 features. This is supported by the fact that MWP only takes 11.7 seconds to train the model instead of 891.6 seconds by KLBP. Figure 5.4 illustrates the comparison of CS measurements for FERET using BIF, KLBP, Face++, and MWP. It is observed that MWP achieves the best score for the error level below 5 years but KLBP performs slightly better for 5 years and above. This indicates MWP is suitable for age estimation due to the importance of low error levels (Geng et al., 2007). Based on these results, it is believed that wrinkle features can be used as a key indicator for face age estimation on images with decent resolution.

5.3.2 The Effect of Gender on Wrinkle Patterns

In FERET, there are 1447 images of males and 919 images of females. MWP was applied to each category for validating the effect of gender for age estimation. Figure 5.5 illustrates the MAE results using MWP between gender and the number of images. Overall, the performance of female result is slightly lower than male with a MAE of 3.93 and 4.13, respectively. There are no notable differences between genders in the majority of age groups. However, due to the small sample size particularly in age groups of 10-19 and 60-70, the result produced a high MAE. Further investigation between genders should consider a larger dataset.

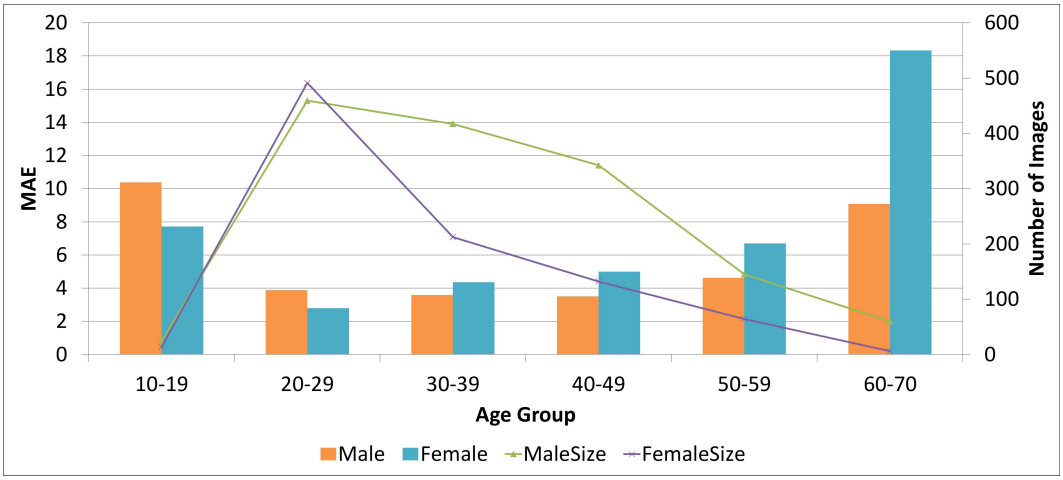


Figure 5.5: MAE results based on gender using MWP.

Table 5.3: Confusion matrix of human identification on FERET.

Coder	A - Expression	A - Neutral
B - Expression	1437	238
B - Neutral	137	554

5.3.3 The Effect of Expression on Wrinkle Patterns

In order to validate the effect of expression on wrinkle patterns, FERET was classified into two categories, neutral and expressive faces, by two FACS coders. Then, age estimation was validated independently on each category using the MWP features. Table 5.3 presents the confusion matrix of human identification by two FACS coders. Based on the result of Coder A, there were 792 images of neutral and 1569 images of expression, while for coder B, there were 691 images of neutral and 1675 images of expression. Human identification correctness of classifying the images into two categories was 84.15%. In the next experiment, results of coder A were used for performance assessment. Figure 5.6 illustrates the MAE results based on neutral and smile categories. Overall, a smile group performed better than a neutral group with a MAE of 4.77 and 5.19, respectively. A separate group of smile or expression produced a higher MAE than a mixed group (MAE = 4.02). This shows that MWP is highly affected by the facial expression.

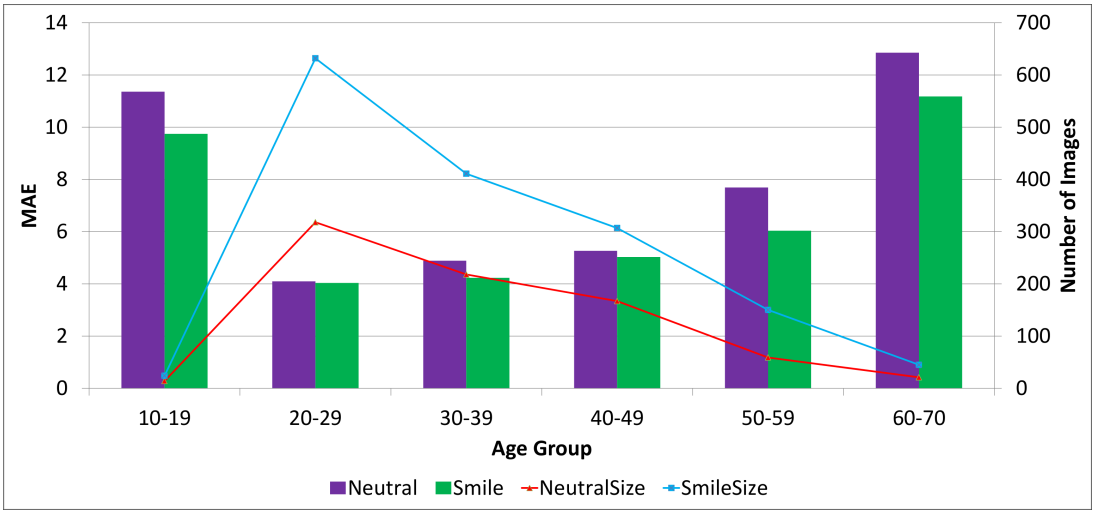


Figure 5.6: MAE results based on expression using MWP.

5.3.4 Discussion

The idea of this work is based on the facial wrinkle distribution on the predefined wrinkle regions. It is useful to identify a simple yet powerful pattern from wrinkles. However, there are few issues that affect the performance of MWP. First, the wrinkle template adaptability is not universal. Although it has been set to a mean face of a dataset, some faces still cannot fit the template very well due to large variations of cranial ratio and occlusions such as hair and moustaches. Second, there are false detections of the Face++ detector in a few images. Although the majority were identified and corrected manually, there are some cases with minor errors in alignment. Finally, the performance of age estimation are highly dependent on a robust wrinkle pattern. Even though wrinkles are highly associated with aging, some individuals have less wrinkles than others.

In general, wrinkle representation is affected by other factors such as facial expressions and cosmetic treatment. In order to assess the factors, FERET is divided into different subsets based on gender and facial expression. Two experiments showed that wrinkle patterns are affected by gender and facial expression. Currently, no dataset contains a comprehensive information with age, gender and facial expression. Manual identification of facial expression was performed by FACS coders. It was time consuming and costly. Thus, these factors remain as a challenging issue for future research.

5.4 Summary

In this chapter, a novel method, MWP, is proposed as a feature descriptor for face age estimation. By deriving wrinkles with the multi-scale filters across ten face regions, wrinkle patterns are generated. They are used to train the classifier and then predict the face age. It is particularly useful when dealing with the images of high resolution and frontal. Experiment results showed the evidence that MWP features able to predict the face age on high resolution images like FERET with the strengths of lower computational cost and better cumulative score at a low error level. With the further improvement of wrinkle detection by HLT, the wrinkle patterns are better represented. Although the proposed MWP is limited to high resolution images, additional fusion concerning face shape and appearance is discussed in the next chapter.

Chapter 6

Wrinkle Pattern Models

This chapter investigates the use of appearance model parameters and different local descriptors for face age estimation. A hybrid pattern, with FAM and MWP as complementary to each other, is demonstrated. Then, a comprehensive hierarchical age estimation method is investigated by a multi-layer age regression approach.

6.1 Introduction

According to Choi et al. (2011), FAM was designed for face modelling, individual characteristics such as the identity variation are reflected stronger than the aging attributes in the FAM parameters. In addition, some crucial aging features such as wrinkles and skin textures are removed from the FAM parameters during dimension reduction by PCA (Cootes et al., 2001). Figure 6.1 shows a comparison between original and reconstructed face images of FAM. It is noticed that the reconstructed image does not sufficiently express the facial aging features such as wrinkles and skin textures. For example, the wrinkles around the eyes, cheeks and mouth of the reconstructed image are reduced or removed. These differences are more obvious on the old subject. This implies that facial aging features cannot be well represented by FAM particularly for older subjects. Therefore, local descriptors are needed to compensate the weakness of FAM.

In Chapter 5, a novel feature representation, MWP, was proposed for face age estimation. It extracted and represented the face wrinkles into a robust aging pattern. Experimental results showed that wrinkle features has the potential for face age estimation. However, such features are limited by two factors. First, people age differently,

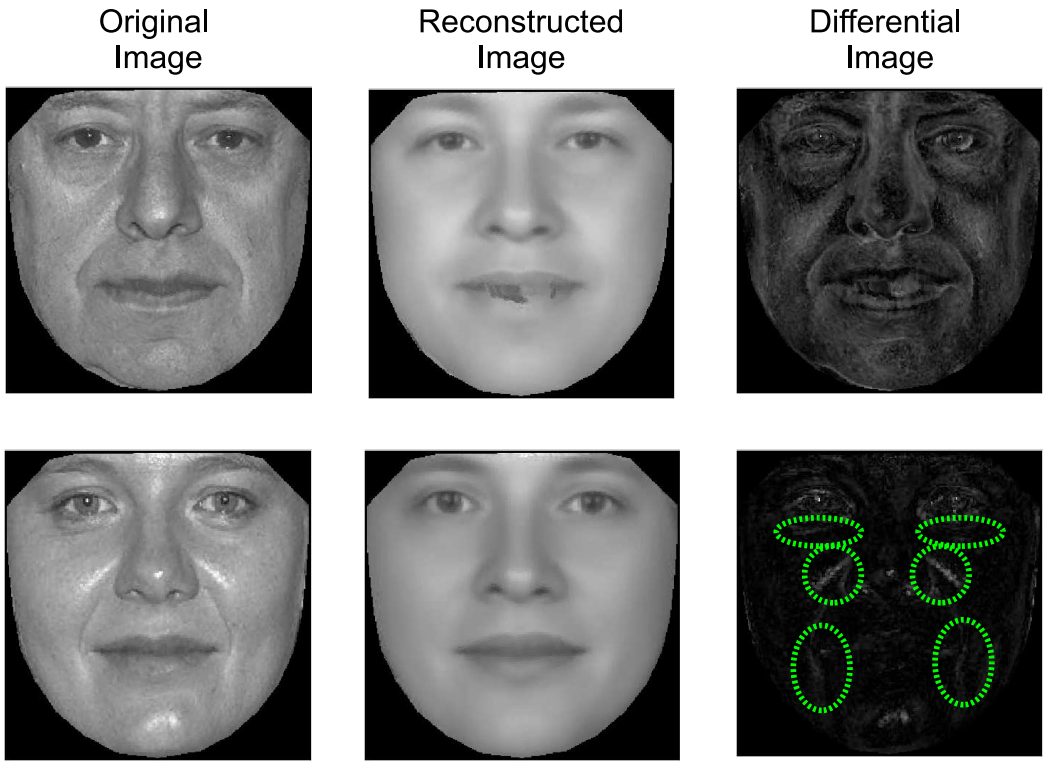


Figure 6.1: A comparison between original and reconstructed face images of FAM. Top row is a male subject of 60 years old and bottom row is a female subject of 30 years old. It appears that wrinkles (dotted green line) are removed on the reconstructed images. For differential image, black means no difference and white presents the differences. Note that this figure was redrawn from FERET.

some with more wrinkles compared to others. Second, wrinkle features only describe the texture difference between adults and consequently these features are not suitable for child age estimation. For these reasons, a combination with global features will improve the performance of wrinkle features for face age estimation. FAM is the complementary of MWP where the face shape and appearance are well defined. Therefore, this chapter presents two novel age estimation methods: a novel hybrid feature representation, HAP, and a novel hierarchical estimation method, MAR, for face age estimation.

In the next section, a hybrid pattern using FAM and MWP is described in Section 6.2. Next, a hierarchical model, MAR, using FAM and different local descriptors is explained in Section 6.3. It is followed by the experimental results for face age estimation in Section 6.4. Finally, a discussion and summary of the work are presented.

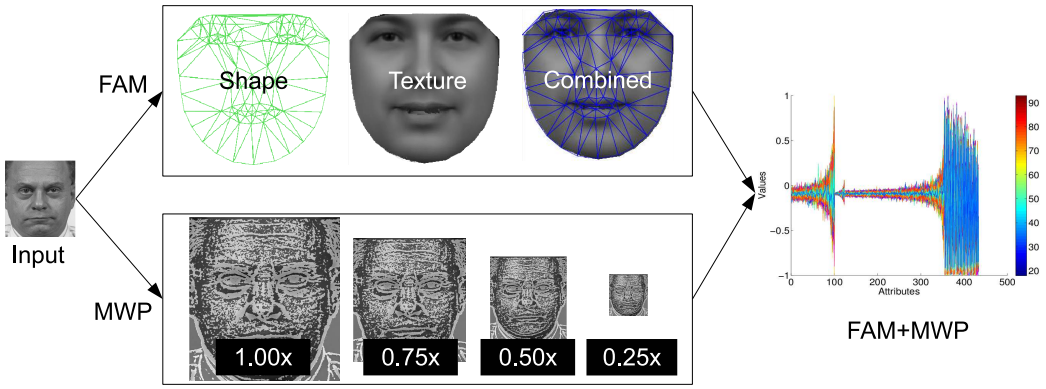


Figure 6.2: An illustration of HAP patterns. First, the features of a warped image (input) are extracted by FAM and MWP. Then, a hybrid pattern is constructed from FAM and MWP as FAM+MWP. The graph of FAM+MWP demonstrates the normalised feature values and the colour bar shows the actual age of each instance. There are approximately 500 instances and 435 features used for this illustration. 1.00x means the original size of input and 0.50x shows the half size of input, and so on. Note that this figure was redrawn from FERET.

6.2 Hybrid Aging Patterns

Chen et al. (2011) explored the facial feature fusion and model selection for age estimation. They found that the feature fusion with model selection can achieve significant improvement for age estimation. However, their results were only tested on FAM and LBP. LBP is a texture descriptor but not designed for facial wrinkle. Therefore, an expansion to the similar idea but a better feature representation is proposed to overcome the limitation. A novel feature representation, Hybrid Aging Patterns (HAP), is proposed for face age estimation. It is a combination of FAM and MWP as shown in Figure 6.2. FAM is capable of representing the face appearance in a set of hidden parameters while MWP is good of describing the facial wrinkles.

As described in Section 3.2.2.3, FAM is a generative parametric model that consists of shape, texture and combined appearance of a human face. It is a model where PCA is used to project high dimension of face shapes and textures into a low dimension of principal component parameters. Manual landmark annotations (FGNET) and automatic detected landmarks (by FACE++ detector) were used to produce a FAM model. The pertinent equations of FAM parameters are repeated here for convenience. Let $\mathbf{s}$ and $\mathbf{t}$ denote a synthesized shape and texture of a face image in the reference frame,

and let $\bar{\mathbf{s}}$ and $\bar{\mathbf{t}}$ denote the corresponding sample means. New instances are now generated by adjusting the principal component scores, $\mathbf{b}_s$ and $\mathbf{b}_t$ in

$$\mathbf{s} = \bar{\mathbf{s}} + \Phi_s \mathbf{b}_s \tag{6.1}$$

$$\mathbf{t} = \bar{\mathbf{t}} + \Phi_t \mathbf{b}_t \tag{6.2}$$

where Φ_s and Φ_t are matrices of column eigenvectors of the shape and texture dispersions estimated from the training set. To obtain a combined shape and texture parameterisation, $\mathbf{c}$, the values of $\mathbf{b}_s$ and $\mathbf{b}_t$ over the training set are combined into

$$\mathbf{b} = \begin{bmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{W}_s \Phi_s^T (\mathbf{s} - \bar{\mathbf{s}}) \\ \Phi_t^T (\mathbf{t} - \bar{\mathbf{t}}) \end{bmatrix} \tag{6.3}$$

A suitable weighting between pixel distances and pixel intensities is carried out through the diagonal matrix $\mathbf{W}_s$. To make the normalised measures of pixel distance and pixel intensities commensurate, the shape model scores are typically weighted by the square root of the ratio between the sums of the texture and shape eigenvalues.

To recover any correlation between shape and texture, the two eigen-spaces are usually coupled through a third principal component transform as

$$\mathbf{b} = \Phi_c \mathbf{c} = \begin{bmatrix} \Phi_{c,s} \\ \Phi_{c,t} \end{bmatrix} \mathbf{c} \tag{6.4}$$

and $\mathbf{b}$ is the FAM features of each image as $\text{FAM} = \{b_1, b_2, \cdots, b_n\}$ where $b_i \in \mathbf{b}$ and n is the total number of FAM features of each image.

Let MWP features of each image as $\text{MWP} = \{c_1, c_2, \cdots, c_m\}$ where m is the total number of MWP features of each image and c_i is defined as Eq. (5.2) and (5.3). As a result, HAP of each image is defined as $\text{HAP} = \{b_1, b_2, \cdots, b_n, c_1, c_2, \cdots, c_m\}$ where $m+n$ is the total number of HAP features of each image.

Let A as the predicted age of each image and it is defined as

$$A = f(\text{HAP}) \tag{6.5}$$

where $f(\cdot)$ is an estimation function by SVR (see Section 3.3.2.2) and the input is HAP. FAM builder configurations[1] of each dataset are given in Appendix D.

[1] Note that the resolution for each dataset is different, therefore the model reduction factor in Appendix D is defined according to the dataset resolution.

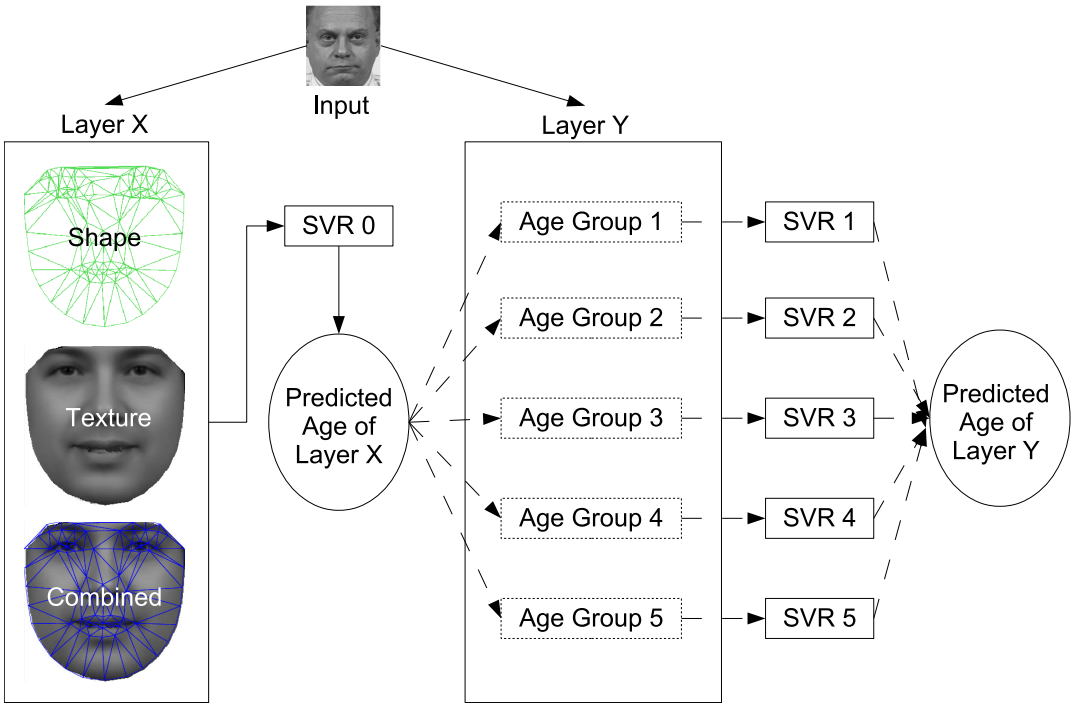


Figure 6.3: Flow chart of MAR. It consists of two layers: layer X and Y. The layer X is the parameters of FAM while the layer Y is the features of local descriptor. First, SVR 0 estimates the age of input using FAM parameters. Then, the predicted age of layer X is used to select an age group of layer Y. Each age group has a fixed range of ages. If age group 3 is selected , then SVR 3 is used to train the local features in age group 3 and finally predict the age of input in the layer Y. The output of both layers is a real number (predicted age) instead of a class. Dotted arrow indicates only one of the age groups will be selected. Note that this figure was redrawn from FERET.

6.3 Multi-layer Age Regression

Traditionally, age estimation is a single-layer process where the face age is predicted in one way by a set of features. This design is simple but it cannot deal effectively with something complex like the human face. Han et al. (2015) claimed that a hierarchical age estimation is a coarse-to-fine estimation. Others (Choi et al., 2011; Thukral et al., 2012) also found that a hierarchical approach is more effective than a single-layer estimation. Therefore, a MAR is proposed to enhance the performance of face age estimation by combining both global and local features into a hierarchical model.

The proposed MAR consists of two layers, layer X and Y as shown in Figure 6.3. In the layer X, features are represented by the FAM parameters, while in the layer Y, features are represented by a local descriptor such as BIF, KLBP and MWP. Moreover,

FAM parameters are repeated in the layer Y as a local descriptor in order to validate the performance of using FAM only. This results four methods of MAR: AAM-AAM, AAM-BIF, AAM-KLBP and AAM-MWP. They are assessed against the single-layer predictions using AAM, BIF, KLBP and MWP.

Let A_{X} as the predicted age of layer X and the prediction is defined as

$$A_{\mathrm{X}} = f(\mathbf{b}) \tag{6.6}$$

where $f(\cdot)$ is a estimation function by SVR. In the layer Y, a between-group classification is implemented based on A_{X}. Let the sorted local features in different age groups as $\{\mathbf{G}_1, \mathbf{G}_2, \cdots, \mathbf{G}_n\}$ where n is the total number of age groups of each dataset and A_{Y} as the predicted age of layer Y. The prediction is defined as

$$A_{\mathrm{Y}} = \begin{cases} f_1(\mathbf{G}_1, \mathbf{f}) & \text{if } A_{\mathrm{X}} < \hat{a}_1 \\ f_2(\mathbf{G}_2, \mathbf{f}) & \text{if } \hat{a}_1 \leq A_{\mathrm{X}} < \hat{a}_2 \\ f_3(\mathbf{G}_3, \mathbf{f}) & \text{if } \hat{a}_2 \leq A_{\mathrm{X}} < \hat{a}_3 \\ f_4(\mathbf{G}_4, \mathbf{f}) & \text{if } \hat{a}_3 \leq A_{\mathrm{X}} < \hat{a}_4 \\ f_5(\mathbf{G}_5, \mathbf{f}) & \text{if } \hat{a}_4 \leq A_{\mathrm{X}} \end{cases} \tag{6.7}$$

where f_i is one particular SVR model, $\mathbf{G}_i$ is the training set, $\mathbf{f}$ is the testing set of an input image, and both $\mathbf{G}_i$ and $\mathbf{f}$ are the same feature selected from either FAM, BIF, KLBP or MWP.

In Figure 6.3, SVR 0 estimates the age of input using FAM parameters. Then, the predicted age of layer X is used to select an age group of layer Y. Each age group has a fixed range. If age group 3 is selected, then SVR 3 is used to train the local features in age group 3 and finally predict the exact age of input in the layer Y. Based on such design, the age of a face image is estimated by a coarse-to-fine approach where the first layer focuses on using regression to predict the age, which will then be used in selecting an age group, while the layer Y concentrates on predicting the exact age within the group.

6.4 Experimental Results

In this work, four descriptors, BIF, KLBP, MWP and FAM are evaluated on four datasets: FGNET (with 1002 images), MORPH (with 2000 images of random selection), FERET (with 2366 images) and PAL (with 576 images). Figure 6.4 shows the age distribution of different datasets. Table 6.1 presents the summary of the experimental datasets which

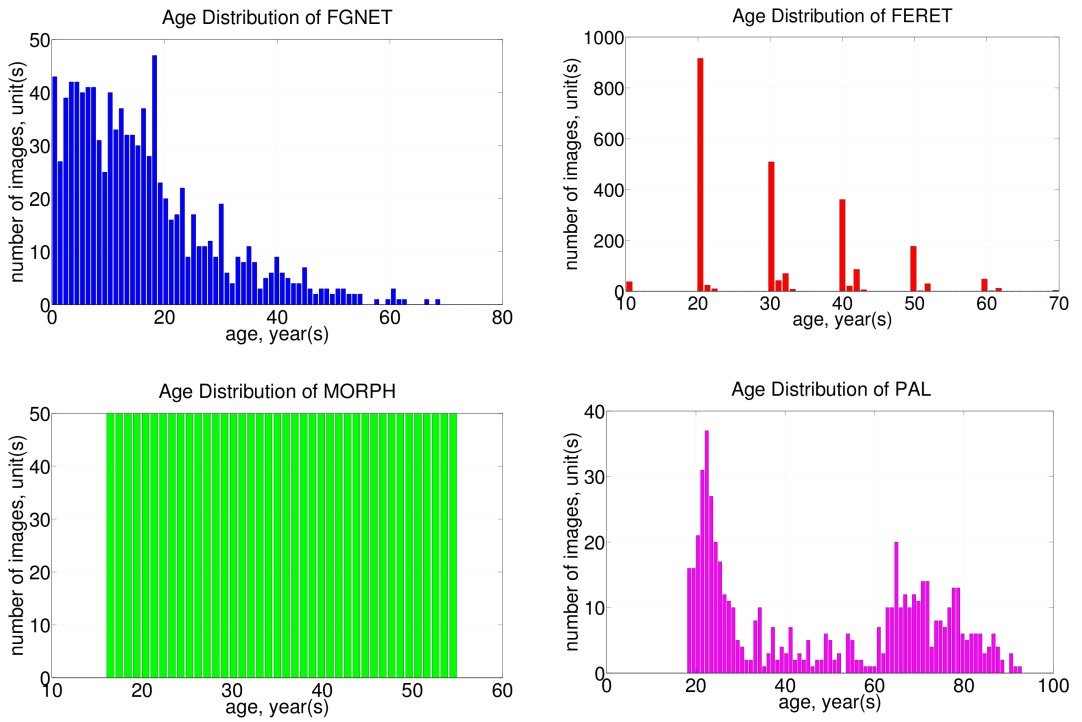


Figure 6.4: Age distribution of different datasets.

includes collection environment, collection era, image type, image size, image compression technique, available annotation, pose, total images used in this work, population trend, evaluation protocol and subset description. For FGNET, a LOPO approach and 68 landmark points are used as the evaluation protocol while the remaining datasets are based on 10-fold cross validation and 83 landmark points of FACE++ detector. Due to the majority of FGNET research using 68 points in the experimental setup, same landmark is used in the following experiments for a fair comparison. Both training and testing datasets are disjoint. In FAM modelling, 95% of the coefficients are preserved during the dimension reduction by PCA as the training set. FAM fitting is not considered in this experiment as the fitting error might increase the estimation error (Choi et al., 2011). Due to the dimension difference in each dataset, the SVR parameters are varied from a grid search approach (Ylioinas et al., 2013).

This section describes two experimental results of face age estimation on four different datasets. First, it is a comparison between HAP with the state of the art. Second, a comprehensive analysis is presented using MAR in terms of the performance assessment, computation time, reliability of age group classification and the performance using soft boundary assessment.

Table 6.1: Summary of the experimental datasets. Note that ** means 2 student unions, a mall, & 2 citizen festivals; JPEG stands for Joint Photographic Experts Group; and PPM indicates Portable Pixmap Format.

Property	FGNET	FERET	MORPH	PAL
Collection Environment	Personal photos	Over the course of 15 sessions	Real-world conditions	**
Collection Era	Unknown	≈1993-1996	≈2003-2007	Unknown
Digital, Paper Scan	Unknown	Digital	Digital & Scan	Digital
Image Size	Mostly $\approx 400 \times 500$	512×768	200×240 & 400×480	640×480
Compression	JPEG	PPM	JPEG	Bitmap, JPEG
Annotation	68 points	None	None	None
Frontal Pose	Uncontrolled	Frontal & Profile	Uncontrolled	Frontal & Profile
Source	Public	Public	Public	Public
Total Used in Experiment	1002	2366	2000	576
Population Trend, [age range]	>50% are ages 0-13, [0-69]	Highly uneven, [10-70]	Even, [16-55]	[18-93]
Evaluation Protocol	LOPO	10-fold cross validation	10-fold cross validation	10-fold cross validation
Subset Disjoint	Yes	Yes	Yes	Yes

6.4.1 Performance Assessment using Hybrid Aging Patterns

In order to compare the performance of the proposed method with other hybrid patterns, FAM+BIF and FAM+KLBP are experimented. Table 6.2 shows the experimental results of age estimation using hybrid features on different datasets compared to single feature representation (FAM, BIF, KLBP and MWP). Overall, a combination between FAM and MWP (FAM+MWP) achieved the best MAE on FERET and this demonstrated that wrinkle patterns are reliable features for face age estimation. Results showed that on FGNET, FAM achieves the lowest MAE of 5.39; on FERET, FAM+MWP scores the lowest MAE of 3.02; on MORPH, FAM+KLBP achieves the lowest MAE of 3.67; and on PAL, BIF hits the lowest MAE of 5.94. It is noticed that the hybrid features of FAM+BIF

Table 6.2: Experimental results of age estimation using hybrid features on different datasets. Bold indicates the lowest MAE of each dataset.

Method	FGNET	FERET	MORPH	PAL
FAM	**5.39 (±5.63)**	3.34 (±3.26)	3.99 (±3.28)	6.96 (±5.92)
BIF	5.59 (±5.59)	3.57 (±3.26)	3.98 (±3.20)	**5.94 (±4.60)**
KLBP	6.09 (±6.09)	3.91 (±3.24)	4.02 (±3.22)	6.10 (±5.18)
MWP	7.34 (±7.54)	4.16 (±3.83)	5.16 (±4.35)	7.65 (±6.61)
FAM+BIF	5.49 (±6.22)	6.54 (±5.68)	7.82 (±6.01)	16.93 (±11.49)
FAM+KLBP	6.06 (±6.16)	4.01 (±3.20)	**3.67 (±2.94)**	6.08 (±5.08)
FAM+MWP	5.66 (±5.88)	**3.02 (±2.92)**	3.68 (±2.98)	6.50 (±5.25)

have the highest MAEs in FERET, MORPH and PAL datasets. It may due to the conflict of feature values between FAM and BIF in a single feature representation. The overall results also demonstrated that the FAM+MWP is more consistent compared to others.

6.4.2 Performance Assessment using Multi-layer Age Regression

Table 6.3 shows a comprehensive analysis of face age estimation with and without MAR. For FGNET and FERET, ages are divided into five groups where $\hat{a}_1 = 15$, $\hat{a}_2 = 35$, $\hat{a}_3 = 45$ and $\hat{a}_4 = 55$; for MORPH, $\hat{a}_1 = 18$, $\hat{a}_2 = 28$, $\hat{a}_3 = 38$ and $\hat{a}_4 = 48$; for PAL, $\hat{a}_1 = 20$, $\hat{a}_2 = 40$, $\hat{a}_3 = 60$ and $\hat{a}_4 = 80$. These parameters were determined based on the age range of dataset.

In a comparison within each dataset, results showed that the lowest MAE of FGNET is 5.39 (±5.63) by FAM; FERET is 3.00 (±4.14) by FAM-FAM; MORPH is 3.98 (±3.20) by BIF; and PAL is 3.43 (±2.71) by FAM-BIF. Overall, FAM, FAM-FAM, BIF, FAM-BIF scored the best MAEs in the datasets of FGNET, FERET, MORPH and PAL, respectively.

In a comparison between age groups, the lowest and the highest MAEs of each dataset are: FGNET - 3.29 (±4.15) and 38.77 (±10.00); FERET - 2.51 (±2.68) and 10.85 (±5.25); MORPH - 3.19 (±2.50) and 10.93 (±6.59); PAL - 0.51 (±0.11) and 10.42 (±6.51), respectively. Two of the lowest MAEs were scored by FAM-FAM, one by FAM-BIF and another by KLBP.

The lowest MAE of each method is obtained on FERET dataset and it is listed as follows: FAM has a MAE of 3.34 (±3.26), BIF is 3.57 (±3.26), KLBP is 3.91 (±3.24), MWP is 4.16 (±3.83), FAM-FAM is 3.00 (±4.14), FAM-BIF is 3.28 (±4.02), FAM-KLBP is 3.29 (±4.07) and FAM-MWP is 3.36 (±4.14). Overall, results showed that MAR

Table 6.3: Experimental results of face age estimation in a single-layer processing. **Bold** - the lowest MAE within one particular dataset, *italic* - the lowest MAE of each method, and the underlined values present the highest and lowest of MAE of age group of each dataset.

Dataset (number of images)	MAE (STD)			
	FAM	BIF	KLBP	MWP
FGNET (1002)	**5.39 (5.63)**	5.59 (5.97)	6.09 (6.43)	7.34 (7.54)
$y_i < \hat{a}_1$ (545)	3.79 (3.42)	3.80 (3.29)	4.17 (3.36)	4.84 (4.09)
$\hat{a}_1 \leq y_i < \hat{a}_2$ (355)	4.94 (3.84)	5.05 (3.93)	5.18 (4.25)	6.62 (5.02)
$\hat{a}_2 \leq y_i < \hat{a}_3$ (61)	10.76 (5.52)	11.92 (6.10)	14.34 (4.90)	17.76 (5.48)
$\hat{a}_3 \leq y_i < \hat{a}_4$ (30)	19.73 (7.39)	21.90 (7.01)	25.27 (6.90)	28.52 (7.40)
$\hat{a}_4 \leq y_i$ (11)	30.15 (7.75)	32.22 (7.93)	32.74 (8.88)	38.77 (10.00)
FERET (2366)	*3.34 (3.26)*	*3.57 (3.26)*	*3.91 (3.24)*	*4.16 (3.83)*
$y_i < \hat{a}_1$ (38)	8.98 (4.09)	9.82 (3.44)	8.64 (4.20)	9.30 (5.18)
$\hat{a}_1 \leq y_i < \hat{a}_2$ (1579)	2.86 (2.67)	3.04 (2.72)	3.44 (2.76)	3.76 (3.40)
$\hat{a}_2 \leq y_i < \hat{a}_3$ (474)	2.99 (2.54)	3.58 (2.98)	3.92 (3.18)	4.01 (3.59)
$\hat{a}_3 \leq y_i < \hat{a}_4$ (209)	4.72 (3.61)	4.82 (3.81)	4.94 (3.65)	5.00 (4.17)
$\hat{a}_4 \leq y_i$ (66)	9.54 (6.74)	8.37 (5.52)	9.14 (4.70)	9.33 (6.34)
MORPH (2000)	3.99 (3.28)	**3.98 (3.20)**	4.02 (3.22)	5.16 (4.35)
$y_i < \hat{a}_1$ (100)	6.74 (5.21)	5.80 (4.58)	6.67 (4.04)	10.93 (6.59)
$\hat{a}_1 \leq y_i < \hat{a}_2$ (500)	3.91 (3.03)	3.99 (2.96)	3.93 (2.92)	4.93 (4.19)
$\hat{a}_2 \leq y_i < \hat{a}_3$ (500)	3.44 (2.87)	3.47 (2.72)	3.19 (2.50)	4.53 (3.42)
$\hat{a}_3 \leq y_i < \hat{a}_4$ (500)	3.65 (2.91)	3.88 (3.07)	3.96 (2.93)	4.94 (3.72)
$\hat{a}_4 \leq y_i$ (400)	4.48 (3.48)	4.25 (3.56)	4.57 (3.97)	5.07 (4.61)
PAL (576)	6.96 (5.92)	5.94 (4.60)	6.09 (5.09)	7.65 (6.61)
$y_i < \hat{a}_1$ (32)	6.75 (4.65)	4.96 (3.88)	5.27 (4.53)	4.84 (4.38)
$\hat{a}_1 \leq y_i < \hat{a}_2$ (234)	7.13 (6.52)	5.01 (4.06)	5.62 (4.91)	6.53 (6.81)
$\hat{a}_2 \leq y_i < \hat{a}_3$ (60)	6.17 (5.58)	7.06 (4.89)	6.66 (5.28)	10.19 (7.22)
$\hat{a}_3 \leq y_i < \hat{a}_4$ (197)	6.32 (4.99)	5.95 (4.41)	5.90 (4.72)	8.08 (6.34)
$\hat{a}_4 \leq y_i$ (53)	9.68 (6.73)	9.29 (5.75)	8.76 (6.42)	9.76 (5.52)

Continued on the next page. . .

Table 6.3: Continued from the previous page. Experimental results of face age estimation with MAR. **Bold** - the lowest MAE within one particular dataset, *italic* - the lowest MAE of each method, and the underlined values present the highest and lowest of MAE of age group of each dataset.

Dataset (number of images)	MAE (STD)			
	FAM-FAM	FAM-BIF	FAM-KLBP	FAM-MWP
FGNET (1002)	5.48 (6.67)	5.49 (6.61)	5.79 (6.67)	6.19 (6.85)
$y_i < \hat{a}_1$ (545)	3.29 (4.15)	3.20 (3.68)	3.54 (3.73)	3.97 (3.73)
$\hat{a}_1 \leq y_i < \hat{a}_2$ (355)	5.06 (4.13)	5.16 (4.15)	5.36 (4.31)	5.53 (4.46)
$\hat{a}_2 \leq y_i < \hat{a}_3$ (61)	13.70 (5.24)	14.07 (5.40)	14.46 (5.56)	15.36 (5.70)
$\hat{a}_3 \leq y_i < \hat{a}_4$ (30)	22.91 (7.99)	23.29 (8.36)	23.48 (8.17)	24.90 (8.80)
$\hat{a}_4 \leq y_i$ (11)	34.38 (7.39)	34.14 (7.65)	34.56 (8.52)	34.93 (8.18)
FERET (2366)	***3.00 (4.14)***	*3.28 (4.02)*	*3.29 (4.07)*	*3.36 (4.14)*
$y_i < \hat{a}_1$ (38)	10.46 (5.19)	10.85 (5.25)	9.86 (4.80)	9.63 (5.03)
$\hat{a}_1 \leq y_i < \hat{a}_2$ (1579)	2.51 (2.68)	2.87 (2.72)	3.01 (2.88)	3.11 (3.02)
$\hat{a}_2 \leq y_i < \hat{a}_3$ (474)	2.75 (5.31)	2.95 (5.07)	2.60 (4.94)	2.65 (5.06)
$\hat{a}_3 \leq y_i < \hat{a}_4$ (209)	3.85 (4.80)	3.86 (4.42)	3.86 (4.74)	3.79 (4.71)
$\hat{a}_4 \leq y_i$ (66)	9.41 (8.45)	9.24 (8.24)	9.39 (8.28)	9.32 (8.15)
MORPH (2000)	4.18 (3.63)	4.06 (3.55)	4.06 (3.47)	4.08 (3.51)
$y_i < \hat{a}_1$ (100)	5.42 (5.45)	5.39 (5.19)	5.35 (5.05)	5.77 (5.46)
$\hat{a}_1 \leq y_i < \hat{a}_2$ (500)	3.76 (3.38)	3.73 (3.32)	3.80 (3.24)	3.77 (3.10)
$\hat{a}_2 \leq y_i < \hat{a}_3$ (500)	3.91 (3.31)	3.74 (3.20)	3.74 (3.11)	3.85 (3.22)
$\hat{a}_3 \leq y_i < \hat{a}_4$ (500)	4.04 (3.32)	4.05 (3.35)	3.96 (3.31)	3.93 (3.32)
$\hat{a}_4 \leq y_i$ (400)	4.92 (3.96)	4.58 (3.87)	4.61 (3.75)	4.53 (3.77)
PAL (576)	6.72 (6.64)	**3.43 (2.71)**	6.33 (6.28)	6.63 (6.65)
$y_i < \hat{a}_1$ (32)	5.31 (2.67)	0.51 (0.11)	4.53 (2.50)	4.34 (2.06)
$\hat{a}_1 \leq y_i < \hat{a}_2$ (234)	5.73 (7.06)	3.18 (2.60)	5.15 (6.67)	5.47 (6.95)
$\hat{a}_2 \leq y_i < \hat{a}_3$ (60)	8.00 (7.19)	4.77 (2.77)	7.81 (6.20)	7.82 (7.00)
$\hat{a}_3 \leq y_i < \hat{a}_4$ (197)	6.83 (6.09)	3.95 (2.85)	6.63 (5.80)	7.00 (6.23)
$\hat{a}_4 \leq y_i$ (53)	10.05 (6.50)	2.83 (1.62)	9.82 (6.32)	10.42 (6.51)

Table 6.4: Computation time (in seconds) for age regression by SVR.

Method	FGNET	FERET	MORPH	PAL	Average
FAM	72.45	168.32	91.47	3.70	83.99
BIF	955.62	4701.11	1869.89	44.15	1892.69
KLBP	4005.18	15660.85	5559.93	184.40	6352.59
MWP	32.11	82.17	30.37	2.32	36.74
FAM-FAM	89.82	205.00	100.25	6.81	100.47
FAM-BIF	520.96	714.61	222.29	23.73	370.40
FAM-KLBP	516.39	2011.61	652.91	87.70	817.15
FAM-MWP	88.25	189.85	95.95	6.06	380.11

improves MAE significantly compared to the existing methods.

Although MAE of FAM-FAM is lower than others, the results are comparable. For example, the results of FERET showed that FAM-FAM hits a MAE of 3.00 compared to FAM-MWP is 3.36, FAM-BIF is 3.28 and FAM-KLBP is 3.29; the results of MORPH showed that FAM-MWP achieves a MAE of 4.08 compared to FAM-FAM is 4.18, FAM-BIF is 4.06 and FAM-KLBP is 4.06. These results showed that wrinkle patterns are as good as state of the art for face age estimation.

6.4.2.1 Computation Time of Age Estimation

Table 6.4 illustrates the computation time (in seconds) for age regression by SVR. These results were based on the settings of Table 6.3. It involved both training and prediction times of SVR. Results showed that MWP demonstrates the fastest performance for age regression with 36.74 seconds, compared to FAM is 83.99 seconds, BIF is 1892.69 seconds, KLBP is 6352.59 seconds, FAM-FAM is 100.47 seconds, FAM-BIF is 370.40 seconds, FAM-KLBP is 817.15 seconds, and FAM-MWP is 380.11 seconds. This correlates with the total number of features of each method where FAM is around 300 units, BIF is 7464 units, KLBP is 32769 units and MWP is the lowest with 160 units.

6.4.2.2 Age Regression versus Age Group Classification

Figure 6.5 shows the age group classification in MAR. In the first layer, features are sorted into a specific age group based on the ground truth. The output of layer X is an age group instead of the predicted age as shown in Figure 6.3. Figure 6.6 and Table 6.5 show a Receiver Operating Characteristic (ROC)[2] curve and confusion matrix

[2]The underlying population curve is theoretically given by varying the cut point used to determine the values of the observed variable to be considered abnormal and then plotting the resulting sensitivities

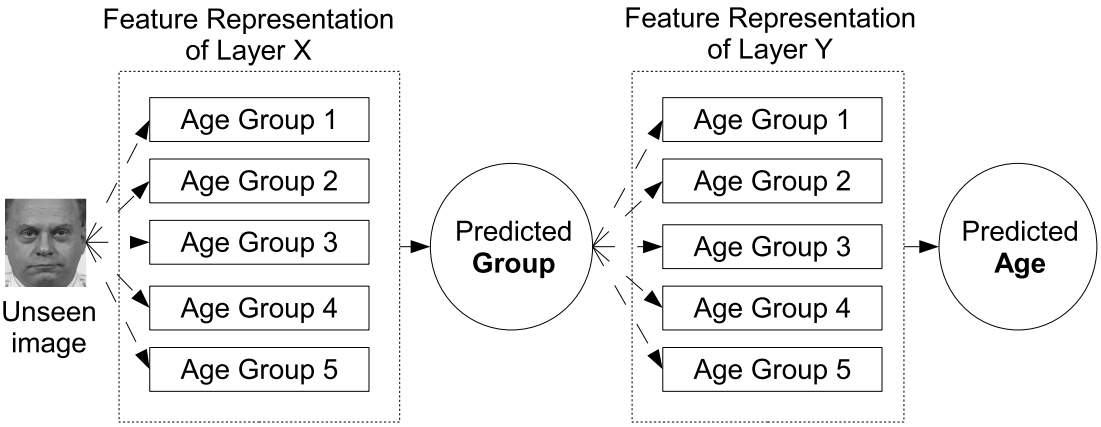


Figure 6.5: An illustration of age group classification in MAR. It consists of two layers: layer X and Y. Layer X is the parameters of FAM. Layer Y is the features of the local descriptor. Dotted arrow indicates only one of the age groups will be selected. Note that this figure was redrawn from FERET.

Table 6.5: Confusion matrix for age regression versus age group classification on MORPH. Left: ROC of SVR regression. Right: ROC of SVM classification. Age group classification was performed by SVM model with a grid search parameters.

Ground Truth	Prediction									
	SVR Regression					SVM Classification				
	$\mathbf{G}_1$	$\mathbf{G}_2$	$\mathbf{G}_3$	$\mathbf{G}_4$	$\mathbf{G}_5$	$\mathbf{G}_1$	$\mathbf{G}_2$	$\mathbf{G}_3$	$\mathbf{G}_4$	$\mathbf{G}_5$
$\mathbf{G}_1$	23	68	8	1	0	0	60	6	34	0
$\mathbf{G}_2$	26	321	152	1	0	0	287	81	132	0
$\mathbf{G}_3$	1	83	351	63	2	0	100	205	195	0
$\mathbf{G}_4$	0	7	153	275	65	0	15	88	396	1
$\mathbf{G}_5$	0	0	7	162	231	0	2	5	383	10

for the comparison between age regression (Figure 6.3) and age group classification (Figure 6.5) on MORPH. SVR regression results were better than the age group classification. For example, if an instance is predicted as 15.4 years old, then it belongs to $\mathbf{G}_1$ where the $\mathbf{G}_1$ range is lower than 18, and so on. For SVM classification, the training and testing sets were sorted according to the predefined age groups. Evaluation was based on a 10-fold cross validation. Overall, SVR regression performed slightly better than SVM classification with a classification accuracy of 60.05% and 44.90%. For the ROC curve, it is noticed that $\mathbf{G}_1$ and $\mathbf{G}_5$ are accurately predicted in SVR regression

against the corresponding false positive rates. The curve would then pass through the point (0,1) on the unit grid. The closer an ROC curve comes to this ideal point, the better its discriminating ability. A test with no discriminating ability will produce a curve that follows the diagonal of the grid (Hanley and McNeil, 1982; DeLong et al., 1988).

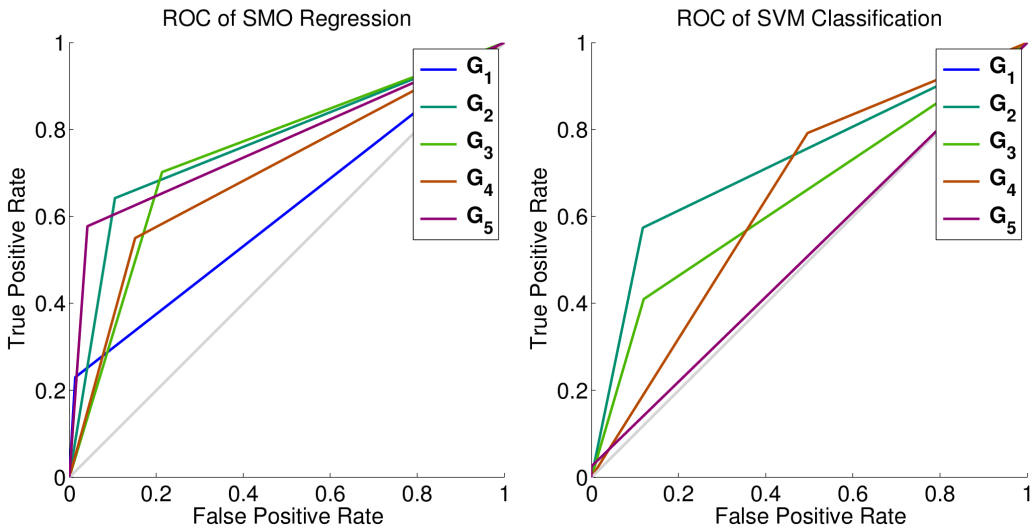


Figure 6.6: ROC curve for age regression versus age group classification. SVR regression was based on MORPH results of Figure 6.3. Age group classification was performed by SVM model with a grid search parameters.

compared to SVM classification. A low classification accuracy of these groups will yield a marginal error of age estimation. This implies that age regression is more reliable than age group classification especially in a multi-layer age regression.

6.4.2.3 Soft Boundary Assessment

A hard boundary is a fixed range where $\hat{a}_1, \cdots, \hat{a}_5$ of Eq. (6.7) are set as the same for both training and testing stages. A soft boundary is a range where the training stage includes a wider range than the testing stage (Choi et al., 2011). In other words, the Eq. (6.7) can be redefined as,

$$A_Y = \begin{cases} f_1(\mathbf{G}_1, \mathbf{f}) & \text{if } A_X < \hat{a}_1 + \Delta \\ f_2(\mathbf{G}_2, \mathbf{f}) & \text{if } \hat{a}_1 - \Delta \leq A_X < \hat{a}_2 + \Delta \\ f_3(\mathbf{G}_3, \mathbf{f}) & \text{if } \hat{a}_2 - \Delta \leq A_X < \hat{a}_3 + \Delta \\ f_4(\mathbf{G}_4, \mathbf{f}) & \text{if } \hat{a}_3 - \Delta \leq A_X < \hat{a}_4 + \Delta \\ f_5(\mathbf{G}_5, \mathbf{f}) & \text{if } \hat{a}_4 - \Delta \leq A_X \end{cases} \tag{6.8}$$

where Δ is a fixed range in year(s). This results an overlapped group where more training samples are learned in a particular group. For example, if an image of a 16 year-old subject is misclassified to the group of 20-30 in the layer Y, the resulting estimation error will be very large (MAE $>$ 5 years). If an overlapped group of 15-35 is used, the result could be improved. Due to an overlapping group only used in the training stage,

Table 6.6: A comparison between hard and soft boundaries on FERET dataset with a fixed range of five years old.

Approach	FAM-FAM	FAM-BIF	FAM-KLBP	FAM-MWP
Hard Boundary	3.00	3.28	3.29	3.36
Soft Boundary	3.00	3.12	3.22	3.36

this does not introduce label ambiguity during group selection in the testing stage. In addition, Han et al. (2015) also claimed that a hierarchical model without using an overlapped group leads to a higher MAE.

Table 6.6 shows a comparison between hard and soft boundary on FERET with a fixed range of 5 years old. Results showed that the soft boundary of FAM-BIF and FAM-KLBP improves the MAEs slightly from 3.28 and 3.29 to 3.12 and 3.22, respectively. Others remain the same. This experiment recommends that a soft boundary is better than a hard boundary in a hierarchical age estimation.

6.5 Discussion

In HAP, a combination between FAM and local descriptors is proposed. Experimental results showed that FAM+MWP performs better than others on FERET with a MAE of 3.02. This implies that FAM+MWP is a robust feature representation where it complements each other for a reliable face age estimation. It is noticed that HAP achieves a good performance on FERET and MORPH where both datasets have a better resolution and the dataset sizes are larger than others. This implies that a high resolution dataset and a large dataset are important factors of face age estimation.

In MAR, the first layer is trained by FAM parameters instead of a local descriptor. There are two reasons of this design. First, FAM is a comprehensive model which comprises both shape and texture, while a local descriptor is based on texture analysis only. Thus, FAM is considered more representative than local descriptor. Second, in a single-layer processing, MAE of FAM is lower than local descriptors in most cases (refer to Table 6.3), with the lowest MAE of 5.39 on FGNET, 3.34 on FERET. This indicates that FAM is more suitable than the local descriptors.

This chapter demonstrates that age regression is more feasible than age group classification. Imagine a face image with an age of 10 years old and a model is trained by FGNET images. This image should be predicted as the first age group ($age < 15$) or the second age group. However, an age group classification model might predict this

image as the fourth group or last group. Consequently, this prediction results a large error between the predicted age and the ground truth. Such an error can be reduced by a regression model where the prediction is a real number instead of a group and normally it is close to the ground truth.

It is observed that the number of training images affect the performance of age estimation. For FERET, there are 38 images in the first age group ($age < 15$) and 66 images in the last age group ($55 \leq age$). MAE results of these groups are higher than other groups. Same goes for MORPH where the age group ($age < 18$) suffered a high MAE compared to other groups.

Age range of a particular age group also influences the performance of age estimation. There are two factors when considering an age range: the nature of features used and the total number of images available in a group. Tsukahara et al. (2007) found that wrinkles appear in various facial sites with the age as low as 10 years and above. However, Mark et al. (1980) claimed that, under normal circumstances, wrinkling does not begin prior to the 30th years. These reviews show the conflict in deciding an age range. Although the results seem promising, but still this issue remains an open question for future research.

6.6 Summary

This chapter outlines a hybrid pattern, HAP, and a novel hierarchical approach, MAR, for age estimation. HAP is a combination of FAM and MWP, while MAR is achieved by a multi-layer age regression where the estimation is divided into two different stages. Experiments were performed on four different datasets: FGNET, FERET, MORPH and PAL. Overall, estimation results showed that FAM-FAM (MAR approach) outperforms with a MAE of 3.00 and FAM+MWP (HAP approach) achieves a comparable MAE of 3.02 on FERET. This demonstrates that wrinkle patterns are robust features for age estimation.

Chapter 7

Conclusion

This chapter summarises the original contributions made by the thesis to the research area of face age estimation. It also highlights possible future improvements and directions of research related to the work presented. The contributions of this thesis are described, their relative merits are discussed and future work is proposed.

7.1 Introduction

Age estimation is a core technology in various applications for facial image analysis, especially in face recognition. This is a challenging task due to the complexity in human face modelling. There are a tremendous number of face images emerging everyday. The problem is further aggravated with the lack of a complete chronological aging dataset. In this work, different methods for face age estimation are investigated to improve the performance. This thesis proposed two novel wrinkle detection methods (HHF and HLT), two novel feature representations for age estimation (MWP and HAP) and one novel age estimation method (MAR).

In general, wrinkles appear when a person ages. The amount of wrinkles become larger and the wrinkle depth appears deeper. In this work, different methods are proposed for wrinkle detection and face age estimation. First, HHF identifies the ridge and valley patterns of wrinkle images using a second order derivative. It is a combination between a directional gradient and a Hessian filter where the eigenvalues of Hessian are explored for wrinkle segmentation. Second, HLT extends the HHF method by looking into the relationship between potential wrinkle pixels and its neighbourhood. It is a multi-scale tracking on different orientations and scales. Third, MWP is proposed as a

Table 7.1: Research objective versus outcome.

No.	Objective	Outcome
1	To explore a novel method for wrinkle detection where the pattern is an arbitrary curve and discontinuous.	Two novel methods, HHF and HLT, are proposed.
2	To investigate the use of discriminative features such as wrinkles and face manifolds, for facial aging estimation.	Three novel methods, MWP, HAP and MAR, are proposed.
3	To evaluate the proposed methods using the benchmarks and human judgement.	The experiments are validated on five different datasets named as Bosphorus, FGNET, FERET, MORPH and PAL. The performance are measured by JSI, MAE and CS.

feature representation of face age where the wrinkle location, intensity and density are take into consideration. However, MWP is a local feature and it has the limitations in outlining the face shape and appearance as in FAM model. Therefore, a hybrid pattern, HAP and a hierarchical model, MAR, are proposed for face age estimation in order to compensate the weakness in both global and local features. Experimental results showed that HAP and MAR performed better than using state-of-the-art algorithms. The following is a summary of research findings.

7.2 Research Findings

Table 7.1 shows the objectives of this work and the corresponding research outcomes. The first objective is to explore a novel method for wrinkle detection where the pattern is an arbitrary curve and discontinuous. This objective is achieved by two proposed methods, HHF and HLT.

In Section 4.2, an automatic wrinkle detection method, HHF, is proposed. Traditionally, the objective quantification was missing to validate the wrinkle detection accuracy. Such implementation requires significant user interactions where results are subjective, depending on user's expertise. This work contributes to the groundwork of automated wrinkle detection using a new approach with objective quantification, which will bring

advancements to the research in facial aging, cosmetics and soft biometrics. The proposed HHF is based on the directional gradient and Hessian matrix. This matrix is computed for all pixels of the image. Eigenvalues of the Hessian matrix indicate that a point belongs to a ridge or valley pattern. HHF is compared with CLM and FRF. Experimental results demonstrated that HHF outperforms state-of-the-art methods with an average JSI of 75.67%. Although there is a gap between human observation and automatic wrinkle detection, the proposed HHF significantly increases the number of true detection and wrinkles can be located more accurately.

In Section 4.3, a novel method, HLT, is proposed to improve HHF. It begins with a group of seeds, extracted from the ridge area of the Hessian matrix. Then, a multi-scale tracking is applied recursively to all seeds. Once completed, each pixel confidence is validated over the scales to produce an initial wrinkle map. Finally, post-processing includes median filtering, directional filtering and area thresholding are applied for noise reduction. In the experimental setup, three coders were instructed to annotate the wrinkle manually on 100 cropped forehead images from Bosphorus. As an assessment of manual annotation, both intra- and inter-coder reliability achieved accuracies of 94% and above, demonstrating that the manual annotation is reliable between and within coders. In the performance assessment against benchmark algorithms, HLT performs better than CLM, FRF and HHF with accuracies of 84.00%, 54.33%, 63.33% and 77.00%, respectively. Such performance is close to that of manual annotation. HLT is a superior method that explores wrinkle connectivity in addition to the curve and valley pattern.

The second objective is to investigate the use of discriminative features such as wrinkles and face manifolds, for facial aging estimation. This objective is achieved by three proposed methods, MWP, HAP and MAR. Wrinkles are important features which can be embedded in several image-based applications related to human skin. However, wrinkle-based age estimation research has not been widely addressed. In Chapter 5, a novel method based on MWP, is proposed for face age estimation. This directly extract the features from local patches without extensive geometric modelling. First, facial landmarks are located by a Face++ detector and then the face is normalised using a linear transformation. A wrinkle template which consists of ten predefined wrinkle regions is created. Then, for each region, wrinkles are detected using HLT and the input patterns are constructed by MWP. Finally, the age is estimated by implementing SVR. The performance of the algorithms is assessed using MAE on FERET. It is observed that MWP produces a comparable MAE value to KLBP but with lower computational

cost than KLBP. Therefore, it is concluded that wrinkles can be used as a aging pattern for face age estimation.

Third objective is to evaluate the proposed methods using the benchmarks and human judgements. The experiments of this work are tested on four different datasets: FGNET, FERET, MORPH and PAL. Results are validated using standard measurements of JSI, MAE and CS. In Chapter 6, a novel feature representation, HAP, and a novel age estimation method, MAR, are proposed for face age estimation. It makes use of the strength of global feature (FAM) in representing the shape and appearance of face and benefits local features (BIF, KLBP and MWP), which is advanced in local representation. Results showed that the proposed MAR achieves the lowest MAE of 3.00 while HAP scores a comparable MAE of 3.02 on FERET. For full results, refer to Appendix A.

7.3 Future Work

This section suggests further improvement on the proposed methods and highlights the potential application areas.

7.3.1 Wrinkle Analysis

Although two methods, HHF and HLT are proposed for wrinkle detection, there is still room for improvement. The experimental datasets have a mixture of coarse and fine wrinkles. It is noticed that some fine wrinkles were filtered out. This issue could be caused by the nature of the image where the line is not very clear and even coders achieved low inter-coder reliability on these images. If wrinkle amount or density is one of the features, then this might lead to a poor age estimation. Due to majority of the wrinkles orientation are horizontal, this thesis is only focusing on the horizontal lines. Future work involves investigation on the impact of vertical lines on face age estimation. In addition, the issues such as over-segmentation, sensitivity to the fine wrinkles, second order structure of Hessian matrix and the distribution of ridge likeliness could be explored.

The potential use of wrinkles in the soft biometric applications, e.g., face verification in the presence of age progression. Aging variation poses a serious problem to automatic face recognition system. As the demand for automatic recognition and

surveillance systems is increasing in the last few decades, it would be interesting to explore how wrinkle features could contribute to this field. In cosmetology, dermatologists are interested in locating and removing face wrinkles in order to achieve skin rejuvenation. Currently, most systems require expert intervention to manually locate and identify wrinkles. Such process is time-consuming and prone to human error. Therefore, an automatic wrinkle quantification system will aid to human decision in cosmetology.

7.3.2 Feature Representation of Face Age

The appearance of wrinkle is affected by other factors such as facial expressions and cosmetic treatment. Wrinkles are highly associated with facial expression. It is worth investigating the expression effect on wrinkle and how it would affect the performance of age estimation. Even though wrinkles are highly associated with aging, it is observed that some individuals have less wrinkles than others. Wrinkle growing pattern and its rates are still not well understood. A relationship between wrinkle and aging should be studied in order to comprehend the difference across age. This pattern is really important in designing a typical aging representation especially for face age estimation. Moreover, different ethnicities and genders might present different types of skin texture. Review showed that a human estimation of facial images of different nationalities, the variance generally tended to be larger. Therefore, the impact of skin texture difference for age estimation is a potential research area. In addition, there is a new trend for age estimation under unconstrained imaging conditions. It would be interesting to see how wrinkle-based pattern can be used under such a challenging condition.

In this work, wrinkle intensity and density were used for feature representation, it can be extended to other measurements such as wrinkle depth, length and width. A curve fitting algorithm and shape modelling in relation to muscle movement and a patch-based representation of found wrinkle in grey-scale values and its orientation could be studied as well. It is observed that face alignment with a global transformation might produce a better result than a local transformation. However, a better way is needed to cope with the distortion after a global transformation. Moreover, it is worth to explore other aging signatures such as spots, freckle and pigments. Although these signs are less common than wrinkles, but it would definitely contribute to a prominent aging representation.

7.3.3 Face Synthesis

Face synthesis renders a face image aesthetically with natural aging and rejuvenating effects on the individual face. It is often implemented by first building a generic face model. Typically, there are two common categories of face models: empirical knowledge-based (with subclasses of geometry based and image based) and statistical learning-based (appearance-based). Geometry-based model generates automatic facial animations with active contours and anthropometric growth. They are mainly designed for non photo-realistic rendering. It digitises facial mesh through geometric units representing face muscles, tissues, and skin in either 2D or 3D. Image-based models focus on generating photo-realistic face images from other images rather than from geometric primitives. A heuristic technique generates texture details on the given face images to simulate human traits, e.g., face skin rendering with creases and aging wrinkles. Appearance-based models consider both shape and texture rendering to achieve highly realistic results. The shape and texture are both vectorised for image representation. Instead of heavily using empirical knowledge like previous two models, this kind of model usually uses a statistical learning to build a model. In addition, it can be extended to face wrinkle modelling for producing a realistic face.

7.4 Concluding Remarks

This thesis has presented a collection of work aimed at bridging a gap between state-of-the-art methods for face age estimation. This has included: (i) exploring novel methods for wrinkle detection of a 2D face image; (ii) presenting novel wrinkle features as a aging pattern for face age estimation; (iii) enhancing the face age estimation using a combination pattern between global and local features. Each of these contributions has been tested with different methods, measurements and datasets. Experimental results showed that face age estimation using wrinkle patterns is feasible on a good resolution dataset such as FERET.

Bibliography

J. Aitchison and C. G. Aitken, "Multivariate binary discrimination by the kernel method," *Biometrika*, vol. 63, no. 3, pp. 413–420, 1976.

S. Akazaki, H. Nakagawa, H. Kazama, O. Osanai, M. Kawai, Y. Takema, and G. Imokawa, "Age-related changes in skin wrinkles assessed by a novel three-dimensional morphometric analysis," *British Journal of Dermatology*, vol. 147, no. 4, pp. 689–695, 2002.

A. Albert and K. Ricanek Jr, "The morph database: investigating the effects of adult craniofacial aging on automated face-recognition technology," *Forensic Science Communications*, vol. 10, no. 2, 2008.

A. Albert, K. Ricanek, and E. Patterson, "A review of the literature on the aging adult skull and face: implications for forensic science research and applications," *Forensic Science International*, vol. 172, no. 1, pp. 1–9, 2007.

M. Albert, A. Sethuram, and K. Ricanek, "Implications of adult facial aging on biometrics," *Biometrics–Unique and Diverse Applications in Nature, Science, and Technology*, pp. 89–106, 2011.

T. Alley, "Head shape and the perception of cuteness." *Developmental Psychology*, vol. 17, no. 5, p. 650, 1981.

T. Alley, *Social and applied aspects of perceiving faces*. Lawrence Erlbaum, 1988.

F. Alnajar, C. Shan, T. Gevers, and J.-M. Geusebroek, "Learning-based encoding with soft assignment for age estimation under unconstrained imaging conditions," *Image and Vision Computing*, vol. 30, no. 12, pp. 946–953, 2012.

E. Alpaydin, *Introduction to machine learning*. MIT press, 2014.

S. J. Axler, *Linear algebra done right*. Springer, 1997, vol. 2.

J. Aznar-Casanova, N. Torro-Alves, and S. Fukusima, "How much older do you get when a wrinkle appears on your face? Modifying age estimates by number of wrinkles," *Aging, Neuropsychology, and Cognition*, vol. 17, no. 4, pp. 406–421, 2010.

Y. Bae, J. S. Nelson, and B. Jung, "Multimodal facial color imaging modality for objective analysis of skin lesions," *Journal of biomedical optics*, vol. 13, no. 6, pp. 064 007–064 007, 2008.

N. Batool and R. Chellappa, "Modeling and detection of wrinkles in aging human faces using marked point processes," in *European Conference on Computer Vision Workshops*. Springer, 2012, pp. 178–188.

N. Batool and R. Chellappa, "Detection and inpainting of facial wrinkles using texture orientation fields and Markov random field modeling," *IEEE Trans. On Image Processing*, vol. 23, no. 9, pp. 3773–3788, 2014.

N. Batool and R. Chellappa, "Fast detection of facial wrinkles based on gabor features using image morphology and geometric constraints," *Pattern Recognition*, vol. 48, no. 3, pp. 642–658, 2015.

A. Ben-Hur and J. Weston, "A users guide to support vector machines," in *Data Mining Techniques for the Life Sciences*. Springer, 2010, pp. 223–239.

D. Berry and L. McArthur, "Perceiving character in faces: the impact of age-related craniofacial changes on social perception," *Psychological Bulletin; Psychological Bulletin*, vol. 100, no. 1, p. 3, 1986.

L. Boissieux, G. Kiss, N. Magnenat-Thalmann, and P. Kalra, "Simulation of skin aging and wrinkles with cosmetics insight," in *Proceedings of the Eurographics Workshop Animation and Simulation*, 2000, pp. 15–27.

C. Braun, M. Gruendl, C. Marberger, and C. Scherber, "Beauty check - causes and consequences of human facial attractiveness (summary)," 2001, the German Students Award. [Online]. Available: http://www.beautycheck.de/index_eng.php

V. Bruce, M. Burton, T. Doyle, and N. Dench, "Further experiments on the perception of growth in three dimensions," *Perception & psychophysics*, vol. 46, no. 6, pp. 528–536, 1989.

S. Buchala, N. Davey, T. M. Gale, and R. J. Frank, "Principal component analysis of gender, ethnicity, age, and identity of face images," in *Proceedings of the IEEE ICMI*. Citeseer, 2005.

L. Cai, L. Huang, and C. Liu, "Age estimation based on improved discriminative Gaussian process latent variable model," *Multimedia Tools and Applications*, pp. 1–18, 2015.

J. Canny, "A computational approach to edge detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, no. 6, pp. 679–698, 1986.

C.-Y. Chang, H.-Y. Liao *et al.*, "Automatic facial spots and acnes detection system," *Journal of Cosmetics, Dermatological Sciences and Applications*, vol. 3, no. 01, p. 28, 2013.

K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2011, pp. 585–592.

W.-L. Chao, J.-Z. Liu, and J.-J. Ding, "Facial age estimation based on label-sensitive learning and age-oriented regression," *Pattern Recognition*, vol. 46, no. 3, pp. 628 – 641, 2013.

S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Transactions on medical imaging*, vol. 8, no. 3, pp. 263–269, 1989.

C. Chen, Y. Chang, K. Ricanek, and Y. Wang, "Face age estimation using model selection," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2010, pp. 93–99.

C. Chen, W. Yang, Y. Wang, K. Ricanek, and K. Luu, "Facial feature fusion and model selection for age estimation," in *IEEE Int. Conf. on Automatic Face & Gesture Recognition and Workshops*, 2011, pp. 200–205.

K. Chen, S. Gong, T. Xiang, and C. C. Loy, "Cumulative attribute space for age and crowd density estimation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2013, pp. 2467–2474.

X. Chen, J. Tian, Y. Zhang, and X. Yang, "A robust orientation estimation algorithm for low quality fingerprints," in *Advances in Biometric Person Authentication*. Springer, 2005, pp. 95–102.

V. Cherkassky and Y. Ma, "Comparison of model selection for regression," *Neural computation*, vol. 15, no. 7, pp. 1691–1714, 2003.

S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, "A comparative study of local feature extraction for age estimation," in *11th International Conference on Control Automation Robotics & Vision (ICARCV)*. IEEE, 2010, pp. 1280–1284.

S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, "Age estimation using a hierarchical classifier based on global and local facial features," *Pattern Recognition*, vol. 44, no. 6, pp. 1262–1281, 2011.

J. F. Cohn and F. De la Torre, "Automated face analysis for affective," *The Oxford Handbook of Affective Computing*, pp. 131–150, 2014.

T. Cootes and C. Taylor, "Statistical models of appearance for computer vision," University of Manchester, Tech. Rep., 2001. [Online]. Available: http://www.itu.dk/stud/projects_f2004/handtracking/referencer/Cootes%20den%20lange%20-%20app_model.pdf

T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

Corel Corp, "Aging of the face," 2012. [Online]. Available: http://face-and-emotion.com/dataface/facets/aging.jsp

C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

G. O. Cula, P. R. Bargo, A. Nkengne, and N. Kollias, "Assessing facial wrinkles: automatic detection and quantification," *Skin Research and Technology*, vol. 19, no. 1, pp. e243–e251, 2013.

N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2005, pp. 886–893.

M. Das and A. Loui, “Automatic face-based image grouping for albuming,” in *IEEE Int. Conf. on Systems, Man and Cybernetics*, vol. 4. IEEE, 2003, pp. 3726–3731.

K. Deffenbacher, T. Vetter, J. Johanson, A. O’Toole *et al.*, “Facial aging, attractiveness, and distinctiveness,” *Perception*, vol. 27, pp. 1233–1244, 1998.

M. M. Dehshibi and A. Bastanfard, “A new algorithm for age recognition from facial images,” *Signal Processing*, vol. 90, no. 8, pp. 2431–2444, 2010.

E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, “Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach,” *Biometrics*, pp. 837–845, 1988.

O. E. Dictionary, “Oxford english dictionary online,” *http://www.oed.com/*, 2015.

K. Drukker, M. L. Giger, K. Horsch, M. A. Kupinski, C. J. Vyborny, and E. B. Mendelson, “Computerized lesion detection on breast ultrasound,” *Medical Physics*, vol. 29, no. 7, pp. 1438–1446, 2002.

C. N. Duong, K. G. Quach, K. Luu, H. B. Le, and K. Ricanek Jr, “Fine tuning age-estimation with global and local facial features,” in *IEEE ICASSP*, 2011, pp. 2032–2035.

P. Ekman and W. V. Friesen, “Facial action coding system,” 1977.

P. Ekman, W. V. Friesen, and J. C. Hager, *Facial Action Coding System, The Manual*, 2nd ed. Research Nexus division of Network Information Research Corporation, 2002.

M. Y. El Dib and M. El-Saban, “Human age estimation using enhanced bio-inspired features (EBIF),” in *17th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2010, pp. 1589–1592.

T. Ezure, J. Hosoi, M. Ooba, H. Miyake, T. Hiroi, E. Hara, C. Matsumoto, and S. Amano, “New approach to anti-aging skincare by mechanism-based improvement of facial sagging: role of subcutaneous tissues,” in *IFSCC*, 2012.

H. Fan, M. Yang, Z. Cao, Y. Jiang, and Q. Yin, "Learning compact face representation: packing a face into an int32," in *Proceedings of the ACM International Conference on Multimedia*. ACM, 2014, pp. 933–936.

J.-J. Fang and S.-Y. Fang, "An automatic method for computerized head and facial anthropometry," in *Digital Human Modeling*. Springer, 2011, pp. 12–21.

J. P. Farkas, J. E. Pessa, B. Hubbard, and R. J. Rohrich, "The science and theory behind facial aging," *Plastic and Reconstructive Surgery Global Open*, vol. 1, no. 1, 2013.

L. G. Farkas, *Anthropometry of the Head and Face*. Raven Pr, 1994.

L. G. Farkas, M. J. Katic, and C. R. Forrest, "International anthropometric study of facial morphology in various ethnic groups/races," *Journal of Craniofacial Surgery*, vol. 16, no. 4, pp. 615–646, 2005.

L. Farkas and S. Schendel, "Anthropometry of the head and face," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 107, no. 1, pp. 112–112, 1995.

C. Fernández, I. Huerta, and A. Prati, "A comparative evaluation of regression learning algorithms for facial age estimation," in *Face and Facial Expression Recognition from Real World Videos*. Springer, 2015, pp. 133–144.

FGNET aging dataset, http://www-prima.inrialpes.fr/FGnet/, accessed on September 2012.

A. F. Frangi, "Three-dimensional model-based analysis of vascular and cardiac images," PhD Thesis (Chapter 2), University Medical Center Utrecht, 2001.

A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Interventation—MICCAI98*. Springer, 1998, pp. 130–137.

Y. Fu and T. Huang, "Human age estimation with regression on discriminative aging manifold," *IEEE Trans. on Multimedia*, vol. 10, no. 4, pp. 578–584, 2008.

Y. Fu, Y. Xu, and T. Huang, "Estimating human age by manifold analysis of face pictures and regression on aging features," in *IEEE Int. Conf. on Multimedia and Expo*. IEEE, 2007, pp. 1383–1386.

Y. Fu, G. Guo, and T. Huang, "Age synthesis and estimation via faces: a survey," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010.

A. Gallagher and T. Chen, "Understanding images of groups of people," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 256–263.

A. Gallagher, M. Das, and A. Loui, "Determining a particular person from a collection," Oct. 31 2005, uS Patent App. 11/263,156.

F. Gao and H. Ai, "Face age classification on consumer images with gabor feature and fuzzy LDA method," in *Advances in biometrics*. Springer, 2009, pp. 132–141.

X. Gao, Y. Su, X. Li, and D. Tao, "A review of active appearance models," *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 40, no. 2, pp. 145–158, 2010.

V. Gartstein and S. A. Shaya, "Assessment of visual signs of skin aging," *Bioengineering of the Skin: Cutaneous Blood Flow and Erythema*, pp. 331–344, 1996.

S. Geisser, *Author predictive inference: an introduction*. New York: Chapman & Hall, 1993.

X. Geng, Z. Zhou, Y. Zhang, G. Li, and H. Dai, "Learning from facial aging patterns for automatic age estimation," in *Proceedings of the 14th annual ACM international conference on Multimedia*. ACM, 2006, pp. 307–316.

X. Geng, Z. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234–2240, 2007.

X. Geng, K. Smith-Miles, and Z. Zhou, "Facial age estimation by nonlinear aging pattern subspace," in *Proceedings of the 16th ACM international conference on Multimedia*. ACM, 2008, pp. 721–724.

X. Geng, K. Smith-Miles, Z.-H. Zhou, and L. Wang, "Face image modeling by multilinear subspace analysis with missing values," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 3, pp. 881–892, 2011.

B. Gilchrest, "A review of skin ageing and its medical therapy," *British Journal of Dermatology*, vol. 135, no. 6, pp. 867–875, 1996.

R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital image processing using MATLAB*, ser. Chapter 10, Image Segmentation. Pearson Education India, 2004.

M. Gonzalez-Ulloa and E. FLORES, "Senility of the face-basic study to understand its causes and effects," *Plastic and reconstructive surgery*, vol. 36, no. 2, pp. 239–246, 1965.

C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 285–339, 1991.

C. Griffiths, A. N. Russman, G. Majmudar, R. S. Singer, T. A. Hamilton, and J. J. Voorhees, "Restoration of collagen formation in photodamaged human skin by tretinoin (retinoic acid)," *New England Journal of Medicine*, vol. 329, no. 8, pp. 530–535, 1993.

A. Günay and V. V. Nabiyev, "Automatic age classification with LBP," in *International Symposium on Computer and Information Sciences*, 2008, pp. 1–4.

A. Günay and V. V. Nabiyev, "Age estimation based on local radon features of facial images," in *Computer and Information Sciences III*. Springer, 2013, pp. 183–190.

A. Günay and V. V. Nabiyev, "Age estimation based on AAM and 2D-DCT features of facial images," *International Journal of Computer Science and Applications*, vol. 6, no. 2, pp. 113–119, 2015.

G. Guo, Y. Fu, C. Dyer, and T. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Trans. on Image Processing*, vol. 17, no. 7, pp. 1178–1188, 2008.

G. Guo, Y. Fu, C. Dyer, and T. Huang, "A probabilistic fusion approach to human age prediction," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2008, pp. 1–6.

G. Guo, Y. Fu, T. Huang, and C. Dyer, "Locally adjusted robust regression for human age estimation," in *IEEE Workshop on Applications of Computer Vision*. IEEE, 2008, pp. 1–6.

G. Guo and G. Mu, "A framework for joint estimation of age, gender and ethnicity on a large database," *Image and Vision Computing*, vol. 32, no. 10, pp. 761–770, 2014.

G. Guo, G. Mu, Y. Fu, C. Dyer, and T. Huang, "A study on automatic age estimation using a large database," in *IEEE 12th Int. Conf. on Computer Vision*, 2009, pp. 1986–1991.

G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 112–119.

T. Hakozaki, C. Swanson, and D. Bissett, "Hyperpigmentation in aging skin," in *Textbook of Aging Skin*, M. Farage, K. Miller, and H. Maibach, Eds. Springer Berlin Heidelberg, 2010, pp. 495–501.

M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

H. Han, C. Otto, X. Liu, and A. Jain, "Demographic estimation from face images: human vs. machine performance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 37, no. 6, pp. 1148–1161, June 2015.

J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve." *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

J.-I. Hayashi, H. Koshimizu, and S. Hata, "Age and gender estimation based on facial image analysis," in *Knowledge-Based Intelligent Information and Engineering Systems*. Springer, 2003, pp. 863–869.

X. He and P. Niyogi, "Locality preserving projections," in *Neural information processing systems*, vol. 16. MIT, 2004, pp. 153–160.

A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE Trans. on Medical Imaging*, vol. 22, no. 8, pp. 951–958, 2003.

W.-B. Horng, C.-P. Lee, and C.-W. Chen, "Classification of age groups based on facial features," *Tamkang Journal of Science and Engineering*, vol. 4, no. 3, pp. 183–192, 2001.

C.-W. Hsu, C.-C. Chang, C.-J. Lin *et al.*, "A practical guide to support vector classification," pp. 1–16, 2003.

I. Huerta, C. Fernández, C. Segura, J. Hernando, and A. Prati, "A deep analysis on age estimation," *Pattern Recognition Letters*, 2015.

R. Iga, K. Izumi, H. Hayashi, G. Fukano, and T. Ohtani, "A gender and age estimation system from face images," in *SICE 2003 Annual Conference*, vol. 1. IEEE, 2003, pp. 756–761.

M. Inc., "Face++ research toolkit," www.faceplusplus.com, Dec. 2013.

P. Jaccard, "Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines (in French)," *Bulletin de la Socit Vaudoise des Sciences Naturelles*, vol. 37, no. 140, pp. 241–272, 1901.

A. Jain, S. Dass, and K. Nandakumar, "Soft biometric traits for personal recognition systems," in *Proceedings of Int. Conf. on Biometric Authentication*, 2004, pp. 731–738.

R. Jana, D. Datta, and R. Saha, "Age estimation from face image using wrinkle features," *Procedia Computer Science*, vol. 46, pp. 1754–1761, 2015.

N. Jayaraman and G. Kulanthaivel, "Neural network based age classification using linear wavelet transforms," *International Journal of Internet Computing*, vol. 1, no. 3, pp. 62–67, 2012.

R. Jdid, B. Summers, G. Cazorla, E. Mauger, C. Guinot, L. Raaff, P. Masson, E. Tschachler, and F. Morizot, "Does skin aging differ in women with dark skin living in different geographic areas?" in *IFSCC*, 2012.

S.-G. Jeong, Y. Tarabalka, and J. Zerubia, "Marked point process model for facial wrinkle detection," in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 1391–1394.

I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.

C.-H. Ju and Y.-H. Wang, "Automatic age estimation based on local feature of face image and regresion," in *International Conference on Machine Learning and Cybernetics*, vol. 2. IEEE, 2009, pp. 885–888.

M. Kilinc and Y. S. Akgul, "Human age estimation via geometric and textural features." in *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2012, pp. 531–538.

J. Kloeppel, “Step right up, let the computer look at your face and tell you your age,” 2010. [Online]. Available: http://news.illinois.edu/news/08/0923age.html

R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proceedings of the International Joint Conference on Artificial Intelligence*, vol. 14, 1995, pp. 1137–1145.

A. Kong, D. Zhang, and M. Kamel, “A survey of palmprint recognition,” *Pattern Recognition*, vol. 42, no. 7, pp. 1408–1418, 2009.

J. Kou, J.-X. Du, and C.-M. Zhai, “Integration of global and local feature for age estimation of facial images,” in *Intelligent Computing Theories and Applications*. Springer, 2012, pp. 455–462.

Y. Kwon and N. da Vitoria Lobo, “Age classification from facial images,” in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1994, pp. 762–767.

Y. Kwon and N. da Vitoria Lobo, “Age classification from facial images,” *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1–21, 1999.

J. Langlois and L. Roggman, “Attractive faces are only average,” *Psychological science*, vol. 1, no. 2, pp. 115–121, 1990.

A. Lanitis, C. Taylor, and T. Cootes, “Toward automatic simulation of aging effects on face images,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442–455, 2002.

A. Lanitis, C. Draganova, and C. Christodoulou, “Comparing different classifiers for automatic age estimation,” *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 34, no. 1, pp. 621–628, 2004.

A. Lanitis, “Age estimation based on head movements: a feasibility study,” in *4th International Symposium on Communications, Control and Signal Processing (ISCCSP)*. IEEE, 2010, pp. 1–6.

T. T. Le, L. G. Farkas, R. C. Ngim, L. S. Levin, and C. R. Forrest, “Proportionality in Asian and North American Caucasian faces using neoclassical facial canons as criteria,” *Aesthetic plastic surgery*, vol. 26, no. 1, pp. 64–69, 2002.

L. Li, S. Mac-Mary, D. Marsaut, J. M. Sainthillier, S. Nouveau, T. Gharbi, O. de Lacharriere, and P. Humbert, "Age-related changes in skin topography and microcirculation," *Archives of dermatological research*, vol. 297, no. 9, pp. 412–416, 2006.

W. Li, H. Su, F. Pan, Q. Gao, and S. Guo, "Facial age classification based on weighted decision fusion," in *33rd Chinese Control Conference (CCC)*. IEEE, 2014, pp. 4870–4874.

Y. Li, Z. Peng, D. Liang, H. Chang, and Z. Cai, "Facial age estimation by using stacked feature composition and selection," *The Visual Computer*, pp. 1–12, 2015.

Z. Li, U. Park, and A. K. Jain, "A discriminative model for age invariant face recognition," *IEEE Trans. on Information Forensics and Security*, vol. 6, no. 3, pp. 1028–1037, 2011.

T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 117–156, 1998.

S. Liu and M. Silverman, "A practical guide to biometric security technology," *IT Professional*, vol. 3, no. 1, pp. 27–32, 2001.

K. Luu, K. Ricanek, T. D. Bui, and C. Y. Suen, "Age estimation using active appearance models and support vector machine regression," in *IEEE 3rd Int. Conf. on Biometrics: Theory, Applications, and Systems*, 2009, pp. 1–5.

K. Luu, T. D. Bui, C. Y. Suen, and K. Ricanek, "Spectral regression based age determination," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2010, pp. 103–107.

K. Luu, K. Seshadri, M. Savvides, T. D. Bui, and C. Y. Suen, "Contourlet appearance model for facial age estimation," in *International Joint Conference on Biometrics (IJCB)*. IEEE, 2011, pp. 1–8.

L. S. Mark and J. T. Todd, "The perception of growth in three dimensions," *Attention, Perception, & Psychophysics*, vol. 33, no. 2, pp. 193–196, 1983.

L. Mark, J. Pittenger, H. Hines, C. Carello, R. Shaw, and J. Todd, "Wrinkling and head shape as coordinated sources of age-level information," *Attention, Perception, & Psychophysics*, vol. 27, no. 2, pp. 117–124, 1980.

M. E. Martinez-Perez, A. D. Hughes, S. A. Thom, A. A. Bharath, and K. H. Parker, "Segmentation of blood vessels from red-free and fluorescein retinal images," *Medical image analysis*, vol. 11, no. 1, pp. 47–61, 2007.

I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.

E. Merinville, G. Grennan, J. Gillbro, J. Mathieu, and A. Mavon, "Influence of facial skin ageing characteristics on the perceived age in a Russian female population," *International journal of cosmetic science*, 2015.

P. Meyer-Marcotty, H. Böhm, C. Linz, J. Kochel, A. Stellzig-Eisenhauer, and T. Schweitzer, "Three-dimensional analysis of cranial growth from 6 to 12 months of age," *The European Journal of Orthodontics*, pp. 1–8, 2013.

E. Meyers and L. Wolf, "Using biologically inspired features for face processing," *International Journal of Computer Vision*, vol. 76, no. 1, pp. 93–104, 2008.

T. Michaud, V. Gassia, and L. Belhaouari, "Facial dynamics and emotional expressions in facial aging treatments," *Journal of cosmetic dermatology*, vol. 14, no. 1, pp. 9–21, 2015.

M. Minear and D. C. Park, "A lifespan database of adult facial stimuli," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 4, pp. 630–633, 2004.

N. Miyamoto, Y. Jinnouchi, N. Nagata, and S. Inokuchi, "Subjective age estimation system using facial images - how old we feel compared to others," in *IEEE Int. Conf. on Systems, Man and Cybernetics*, vol. 4. IEEE, 2005, pp. 3449–3453.

J. Mutch and D. G. Lowe, "Object class recognition and localization using sparse features with limited receptive fields," *International Journal of Computer Vision*, vol. 80, no. 1, pp. 45–57, 2008.

K. Negishi, S. Wakamatsu, N. Kushikata, Y. Tezuka, Y. Kotani, and K. Shiba, "Full-face photorejuvenation of photodamaged skin by intense pulsed light with integrated contact cooling: initial experiences in Asian patients," *Lasers in surgery and medicine*, vol. 30, no. 4, pp. 298–305, 2002.

M. L. Ngan and P. J. Grother, "Face recognition vendor test (FRVT) - Performance of automated age estimation algorithms," *NIST Interagency/Internal Report (NISTIR)*, vol. 7995, pp. 1–31, 2014.

J. Nip, S. B. Potterf, S. Rocha, S. Vora, and C. Bosko, "The new face of pigmentation and aging," in *Textbook of Aging Skin*. Springer, 2010, pp. 509–521.

T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," in *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, no. 1, 1994, pp. 582–585.

E. Osuna, R. Freund, and F. Girosi, "An improved training algorithm for support vector machines," in *Neural Networks for Signal Processing [1997] VII. Proceedings of the 1997 IEEE Workshop*. IEEE, 1997, pp. 276–285.

A. O'Toole, T. Vetter, H. Volz, E. Salter *et al.*, "Three-dimensional caricatures of human heads: distinctiveness and the perception of facial age," *Perception*, vol. 26, pp. 719–732, 1997.

A. O'Toole, T. Price, T. Vetter, J. Bartlett, and V. Blanz, "3D shape and 2D surface textures of human faces: the role of "averages" in attractiveness and age," *Image and Vision Computing*, vol. 18, no. 1, pp. 9–19, 1999.

N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

G. Panis, A. Lanitis, N. Tsapatsoulis, and T. F. Cootes, "Overview of research on facial ageing using the FGNET ageing database," *IET Biometrics*, May 2015.

E. Patterson, A. Sethuram, M. Albert, K. Ricanek, and M. King, "Aspects of age variation in facial morphology affecting biometrics," in *First IEEE International Conference on Biometrics: Theory, Applications, and Systems*. IEEE, 2007, pp. 1–6.

A. Perez-Bernal, M. A. Muñoz-Pérez, and F. Camacho, "Management of facial hyperpigmentation," *American journal of clinical dermatology*, vol. 1, no. 5, pp. 261–268, 2000.

P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *IEEE computer society conference on Computer vision and pattern recognition*, vol. 1. IEEE, 2005, pp. 947–954.

P. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The feret evaluation methodology for face-recognition algorithms," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

G. E. Piérard, I. Uhoda, and C. Piérard-Franchimont, "From skin microrelief to wrinkles. An area ripe for investigation," *Journal of cosmetic dermatology*, vol. 2, no. 1, pp. 21–28, 2003.

J. B. Pittenger, R. E. Shaw, and L. S. Mark, "Perceptual information for the age level of faces as a higher order invariant of growth." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 5, no. 3, pp. 478–493, 1979.

J. Platt *et al.*, "Fast training of support vector machines using sequential minimal optimization," *Advances in kernel methodssupport vector learning*, vol. 3, 1999.

N. Ramanathan and R. Chellappa, "Face verification across age progression," *IEEE Trans. on Image Processing*, vol. 15, no. 11, pp. 3349–3361, 2006.

N. Ramanathan, R. Chellappa, and S. Biswas, "Computational methods for modeling facial aging: a survey," *Journal of Visual Languages & Computing*, vol. 20, no. 3, pp. 131–144, 2009.

N. Ramanathan and R. Chellappa, "Modeling shape and textural variations in aging faces," in *IEEE International Conference on Automatic Face & Gesture Recognition*, 2008, pp. 1–8.

R. Real, "Tables of significant values of Jaccard's index of similarity," *Misc. Zool.*, vol. 22, no. 1, pp. 29–40, 1999.

H. Rexbye and J. Povlsen, "Visual signs of ageing: what are we looking at," *Int J Ageing Later Life*, vol. 2, no. 1, pp. 61–83, 2007.

H. Rexbye, I. Petersen, M. Johansens, L. Klitkou, B. Jeune, and K. Christensen, "Influence of environmental factors on facial ageing," *Age and Ageing*, vol. 35, no. 2, pp. 110–115, 2006.

G. Rhodes and T. Tremewan, "Averageness, exaggeration, and facial attractiveness," *Psychological science*, pp. 105–110, 1996.

M. G. Rhodes, "Age estimation of faces: a review," *Applied Cognitive Psychology*, vol. 23, no. 1, pp. 1–12, 2009.

K. Ricanek and T. Tesafaye, "Morph: a longitudinal image database of normal adult age-progression," in *7th Int. Conf. on Automatic Face and Gesture Recognition*. IEEE, 2006, pp. 341–345.

K. Ricanek Jr, Y. Wang, C. Chen, and S. J. Simmons, "Generalized multi-ethnic face age-estimation," in *3rd International Conference on Biometrics: Theory, Applications, and Systems*. IEEE, 2009, pp. 1–6.

M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature neuroscience*, vol. 2, no. 11, pp. 1019–1025, 1999.

J. A. Ruiz-Hernandez, J. L. Crowley, A. Lux, and L. dInformatique de Grenoble, ""How old are you?": age estimation with tensors of binary Gaussian receptive maps." in *BMVC*, 2010, pp. 1–11.

D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Cognitive modeling*, vol. 1, p. 213, 2002.

A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in *Biometrics and Identity Management*. Springer, 2008, pp. 47–56.

A. Savran, B. Sankur, and M. Taha Bilge, "Comparative evaluation of 3D vs. 2D modality for automatic detection of facial action units," *Pattern recognition*, vol. 45, no. 2, pp. 767–782, 2012.

A. Savran, B. Sankur, and M. Taha Bilge, "Regression-based intensity estimation of facial action units," *Image and Vision Computing*, vol. 30, no. 10, pp. 774–784, 2012.

E. Scali and A. Iorio, *Minimally invasive procedures for facial rejuvenation*. OMICS Group eBooks, 2014, ch. Anatomy and the Aging Process of the Face, pp. 1–7.

A. Selamat, I. M. I. Subroto, and C.-C. Ng, "Arabic script web page language identification using hybrid-KNN method," *International Journal of Computational Intelligence and Applications*, vol. 8, no. 03, pp. 315–343, 2009.

T. Serre, L. Wolf, and T. Poggio, "Object recognition with features inspired by visual cortex," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2005, pp. 994–1000.

C. Shan, "Learning local features for age estimation on real-life faces," in *Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis*. ACM, 2010, pp. 23–28.

J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis*. Cambridge university press, 2004.

S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to the SMO algorithm for SVM regression," *IEEE Trans. on Neural Networks*, vol. 11, no. 5, pp. 1188–1193, 2000.

G. Sinha and B. Patel, *Medical image processing*. PHI Learning, 2014. [Online]. Available: https://books.google.co.uk/books?id=4a3xBQAAQBAJ

A. J. Smola and B. Schölkopf, *Learning with kernels*. Citeseer, 1998.

A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

I. Sobel and G. Feldman, "A 3x3 isotropic gradient operator for image processing," presented at the Stanford Artificial Intelligence Project (SAIL) in 1968.

M. Sofka and C. V. Stewart, "Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures," *IEEE Trans. on Medical Imaging*, vol. 25, no. 12, pp. 1531–1546, 2006.

M. B. Stegmann, "The AAM-API: an open source active appearance model implementation," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003*. Springer, 2003, pp. 951–952.

M. B. Stegmann, B. K. Ersboll, and R. Larsen, "Fame-a flexible appearance modeling environment," *IEEE Trans. on Medical Imaging*, vol. 22, no. 10, pp. 1319–1331, 2003.

A. Stone, "The aging process of the face & techniques of rejuvenation," 2012. [Online]. Available: http://www.aaronstonemd.com/Facial_Aging_Rejuvenation.shtm

J. Suo, F. Min, S. Zhu, S. Shan, and X. Chen, "A multi-resolution dynamic model for face aging simulation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

J. Suo, T. Wu, S. Zhu, S. Shan, X. Chen, and W. Gao, "Design sparse features for age estimation using hierarchical face model," in *IEEE Int. Conf. on Automatic Face & Gesture Recognition*, 2008, pp. 1–6.

J. Suo, X. Chen, S. Shan, and W. Gao, "Learning long term face aging patterns from partially dense aging databases," in *IEEE 12th Int. Conf. on Computer Vision*. IEEE, 2009, pp. 622–629.

J. Suo, S.-C. Zhu, S. Shan, and X. Chen, "A compositional and dynamic model for face aging," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 3, pp. 385–401, 2010.

J. Suo, X. Chen, S. Shan, W. Gao, and Q. Dai, "A concatenational graph evolution aging model," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2083–2096, 2012.

M. Suppa, F. Elliott, J. Mikeljevic, Y. Mukasa, M. Chan, S. Leake, B. Karpavicius, S. Haynes, E. Bakker, K. Peris *et al.*, "The determinants of periorbital skin ageing in participants of a melanoma case-control study in the UK," *British Journal of Dermatology*, vol. 165, no. 5, pp. 1011–1021, 2011.

M. A. Taister, S. D. Holliday, and H. Borrman, "Comments on facial aging in law enforcement investigation," *Forensic science communications*, vol. 2, no. 2, 2000.

H. Takimoto, Y. Mitsukura, M. Fukumi, and N. Akamatsu, "A design of gender and age estimation system based on facial knowledge," in *International Joint Conference SICE-ICASE*. IEEE, 2006, pp. 3883–3886.

H. Takimoto, T. Kuwano, Y. Mitsukura, H. Fukai, and M. Fukumi, "Appearance-age feature extraction from facial image based on age perception," in *Annual Conference of SICE*. IEEE, 2007, pp. 2813–2818.

Y.-M. Tang and B.-L. Lu, "Age classification combining contour and texture feature," in *Neural Information Processing. Models and Applications*. Springer, 2010, pp. 493–500.

K. T. Taylor, *Forensic art and illustration*. CRC Press, 2000.

P. Thukral, K. Mitra, and R. Chellappa, "A hierarchical approach for human age estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 1529–1532.

Q. Tian and S. Chen, "Cumulative attribute relation regularization learning for human age estimation," *Neurocomputing*, vol. 165, pp. 456–467, 2015.

I. W. Tsang, J. T. Kwok, and P.-M. Cheung, "Core vector machines: fast SVM training on very large data sets," in *Journal of Machine Learning Research*, 2005, pp. 363–392.

K. Tsukahara, Y. Takema, H. Kazama, Y. Yorimoto, T. Fujimura, S. Moriwaki, T. Kitahara, M. Kawai, and G. Imokawa, "A photographic scale for the assessment of human facial wrinkles," *Journal of Cosmetic Science*, vol. 51, no. 2, pp. 127–139, 2000.

K. Tsukahara, S. Moriwaki, M. Hotta, T. Fujimura, and T. Kitahara, "A study of diurnal variation in wrinkles on the human face," *Archives of dermatological research*, vol. 296, no. 4, pp. 169–174, 2004.

K. Tsukahara, K. Sugata, O. Osanai, A. Ohuchi, Y. Miyauchi, M. Takizawa, M. Hotta, and T. Kitahara, "Comparison of age-related changes in facial wrinkles and sagging in the skin of Japanese, Chinese and Thai women," *Journal of dermatological science*, vol. 47, no. 1, pp. 19–28, 2007.

P. Turaga, S. Biswas, and R. Chellappa, "The role of geometry in age estimation," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2010, pp. 946–949.

K. Ueki, T. Hayashida, and T. Kobayashi, "Subspace-based age-group classification using facial images under various lighting conditions," in *7th Int. Conf. on Automatic Face and Gesture Recognition*. IEEE, 2006, pp. 6–11.

T. Valentine, "A unified account of the effects of distinctiveness, inversion, and race in face recognition," *The Quarterly Journal of Experimental Psychology*, vol. 43, no. 2, pp. 161–204, 1991.

V. N. Vapnik, *Statistical learning theory*. Wiley New York, 1998, vol. 1.

M. Vlachos and E. Dermatas, "Multi-scale retinal vessel segmentation using line tracking," *Computerized Medical Imaging and Graphics*, vol. 34, no. 3, pp. 213–227, 2010.

X. Wang, R. Guo, and C. Kambhamettu, "Deeply-learned feature for age estimation," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2015, pp. 534–541.

R. Weng, J. Lu, G. Yang, and Y.-P. Tan, "Multi-feature ordinal ranking for facial age estimation," in *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE, 2013, pp. 1–6.

T. Wu, P. Turaga, and R. Chellappa, "Age estimation and face verification across aging using landmarks," *IEEE Trans. on Information Forensics and Security*, vol. 7, no. 6, pp. 1780–1788, 2012.

Z. Xu, H. Chen, and S.-C. Zhu, "A high resolution grammatical model for face representation and sketching," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2005, pp. 470–477.

M. Yacenda *et al.*, "Location and age verification for mobile lottery play," 2013, wO Patent 2,013,040,474.

S. Yan, H. Wang, Y. Fu, J. Yan, X. Tang, and T. Huang, "Synchronized submanifold embedding for person-independent pose estimation and beyond," *IEEE Trans. on Image Processing*, vol. 18, no. 1, pp. 202–210, 2009.

S. Yan, M. Liu, and T. S. Huang, "Extracting age information from local spatially flexible patches," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2008, pp. 737–740.

S. Yan, X. Zhou, M. Liu, M. Hasegawa-Johnson, and T. S. Huang, "Regression from patch-kernel," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

W. Yang, C. Chen, K. Ricanek, and C. Sun, "Ensemble of global and local features for face age estimation," in *Advances in Neural Networks*. Springer, 2011, pp. 251–259.

X. Yang, J. Liu, Y. Ma, and J. Xue, "Facial age estimation from web photos using multiple-instance learning," in *IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2014, pp. 1–6.

Z. Yang and H. Ai, "Demographic classification with local binary patterns," *Advances in Biometrics*, pp. 464–473, 2007.

M. H. Yap, "Enhanced algorithms for lesion detection & recognition in ultrasound breast images," PhD Thesis, Loughborough University, 2008.

J. Ylioinas, A. Hadid, X. Hong, and M. Pietikäinen, "Age estimation using local binary pattern kernel density estimate," in *Image Analysis and Processing–ICIAP*. Springer, 2013, pp. 141–150.

L. A. Zebrowitz and J. M. Montepare, "Impressions of babyfaced individuals across the life span." *Developmental psychology*, vol. 28, no. 6, pp. 1143–1152, 1992.

B. Zhang, S. Shan, X. Chen, and W. Gao, "Histogram of Gabor phase patterns (HGPP): a novel object representation approach for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 57–68, 2007.

G. Zhang and H. Ge, "Support vector machine with a Pearson VII function kernel for discriminating halophilic and non-halophilic proteins," *Computational biology and chemistry*, vol. 46, pp. 16–22, 2013.

Y. Zhang and D.-Y. Yeung, "Multi-task warped Gaussian process for personalized age estimation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 2622–2629.

E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Extensive facial landmark localization with coarse-to-fine convolutional network cascade," in *IEEE Int. Conf. on Computer Vision Workshops (ICCVW)*. IEEE, 2013, pp. 386–391.

S. Zhou, B. Georgescu, X. Zhou, and D. Comaniciu, "Image based regression using boosting method," in *Tenth IEEE Int. Conf. on Computer Vision*, vol. 1. IEEE, 2005, pp. 541–548.

M. Zimbler, M. Kokoska, and J. Thomas, "Anatomy and pathophysiology of facial aging," *Facial plastic surgery clinics of North America*, vol. 9, no. 2, pp. 17987, vii, May 2001.

Appendix A

Review of Detailed Age Estimation

Descriptor	Authors	Method Description	MAE
FGNET	Lanitis et al. (2004)	AAS	14.83
	Lanitis et al. (2002)	WAS	8.06
	Ju and Wang (2009)	LBP	6.85
	Geng et al. (2006)	AGES	6.77
	Geng et al. (2007)	AGES-LDA	6.22
	Geng et al. (2008)	KAGES	6.18
	Günay and Nabiyev (2013)	Radon Features	6.18
	Alnajar et al. (2012)	Soft Encoding	6.14
	Suo et al. (2008)	TGPC	5.97
	Zhou et al. (2005)	IBR, Haar-like features	5.81
	Kou et al. (2012)	PCA+DFT	5.72
	Proposed HAP	AAM+BIF	5.49
	Proposed MAR	AAM-AAM	5.48
	Günay and Nabiyev (2015)	AAM+2D-DCT	5.39
	Geng et al. (2011)	M^2SA	5.36
	Yan et al. (2009)	SSE	5.21
	Ylioinas et al. (2013)	KLBP	5.09
	Kilinc and Akgul (2012)	Geometric+LGBP	5.05
	Guo et al. (2008)	PFA	4.97
	Ruiz-Hernandez et al. (2010)	BGRM	4.96
	Yan et al. (2008b)	RPK	4.95

	Suo et al. (2009)	mkNN	4.93
	Zhang and Yeung (2010)	MTWGP	4.83
	Guo et al. (2009a)	BIF	4.77
	Duong et al. (2011)	AAM-LBP	4.74
	Choi et al. (2011)	GLA	4.66
	Cai et al. (2015)	DGPLVM	4.64
	Chang et al. (2011)	OHR	4.48
	Chao et al. (2013)	LPP	4.38
	Tian and Chen (2015)	CA-AAM	4.37
	Wang et al. (2015)	CNN	4.26
	Weng et al. (2013)	MFOR	4.25
	Luu et al. (2010)	Spectral Regression-AAM	4.25
	Luu et al. (2011)	CAM	4.12
	Chen et al. (2010)	PCA+LPP+sequential selection	4.04
	Li et al. (2015)	SFCS-BIF	3.76
	Thukral et al. (2012)	A Hierarchical Approach	3.40
	El Dib and El-Saban (2010)	**Enhanced BIF**	**3.17**
FERET	Günay and Nabiyev (2013)	Radon Features	6.98
	Zhou et al. (2013)	Face++	6.94
	Proposed HAP	AAM+MWP	3.02
	Proposed MAR	**AAM-AAM**	**3.00**
MORPH	Günay and Nabiyev (2013)	Radon Features	6.65
	Kilinc and Akgul (2012)	Geometric+LGBP	6.28
	Ruiz-Hernandez et al. (2010)	BGRM	6.19
	Yang et al. (2014)	WMIR	5.44
	Li et al. (2015)	SFCS-BIF	5.31
	Fernández et al. (2015)	HOG	4.83
	Wang et al. (2015)	CNN	4.77
	Tian and Chen (2015)	CA-AAM	4.69
	Cai et al. (2015)	DGPLVM	4.66
	Weng et al. (2013)	MFOR	4.20
	Proposed MAR	AAM-BIF	4.06
	Huerta et al. (2015)	CNN	3.88

	Proposed HAP	**AAM+KLBP**	**3.67**
PAL	Choi et al. (2010)	GHPF	8.44
	Chen et al. (2011)	LBP	7.70
	Yang et al. (2011)	Gabor	7.10
	Yang et al. (2011)	AAM	6.29
	Chen et al. (2010)	PCA+LPP+sequential selection	6.25
	Proposed HAP	AAM+KLBP	6.08
	Luu et al. (2011)	CAM	6.00
	Yang et al. (2011)	AAM-Gabor	5.88
	Chen et al. (2011)	AAM-LBP-2	5.65
	Proposed MAR	**AAM-BIF**	**3.43**

Note: IBR - Image-Based Regression; OHR - Ordinal Hyperplanes Ranker; WAS - Weighted Appearance Specific; AAS - Appearance and Age Specific; LDA - Linear Discriminant Analysis; RUN - Ranking with Uncertain Labels; RPK - Regression with Patch-Kernel; LARR - Locally Adjusted Robust Regression; PFA - Probabilistic Fusion Approach; KAGES - Kernel AGES; MSA-Multilinear Subspace Analysis; SSE - Synchronized Submanifold Embedding; mkNN - learned Metric with K-Nearest Neighbour; LLD - Learning from Label Distributions; M^2SA - Multilinear Subspace Analysis With Missing Values; MTWGP - Multi-TaskWarped Gaussian Process; LPP - Locality Preserving Projection; TGPC - Topological, Geometric, Photometric and Configural features; GLA - Gabor, LBP and AAM-based features; DGPLVM - Discriminative Gaussian Process Latent Variable Model; DCT - Discrete Cosine Transform; SFCS - Stacked Features Composition and Selection; CAM - Contourlet Appearance Model; BGRM - Binary Gaussian Receptive Maps; CA - Cumulative Attribute; MFOR - Multi-Feature Ordinal Ranking; WMIR - Witness-based Multiple-Instance Regression; GHPF - Gaussian High Pass Filter; LGBP - Local Gabor Binary Patterns; DFT - Discrete Fourier Transform; CAM - Contourlet Appearance Model; HOG - Histogram of Gradient.

Appendix B

Review of Age Groups Classification

Authors	Dataset	Method	Group	Accuracy
Kwon and da Vitoria Lobo (1999)	unknown	AM, Snakes	Baby, Adult, Senior	100%
Horng et al. (2001)	unknown	AM	4 classes	85.96%
Iga et al. (2003)	unknown	AM	5 classes	58.4%
Ueki et al. (2006)	WIT-DB	APM	11 classes	50% (M), 43% (F)
Takimoto et al. (2006)	HOIP	APM	6 classes	57.3% (M), 54.7% (F)
Geng et al. (2006)	FGNET	AGES	14 classes	40.92% (hit rate)
Guo et al. (2009b)	unknown	APM (BIF)	3 classes	89.7%
Gao and Ai (2009)	unknown	Gabor features + LDA	4 classes	86.64%
Shan (2010)	unknown	boosted LBP + SVM	7 classes	50.3%
Tang and Lu (2010)	unknown	LBP + SVM + fuzzy integral	4 classes	80.23%
Lanitis (2010)	FGNET	Head Movements	3 classes	65.00%
Dehshibi and Bastanfard (2010)	IFDB	seven facial ratios and three wrinkle areas	4 classes	91.00%
Li et al. (2014)	FGNET, MORPH	LBP + ASM	4 classes	82.00%

Note: Age groups classification on various datasets. AM - Anthropometric Model; APM - Appearance Model; ASM - Active Shape Model; M - Male; F - Female; LDA - Linear Discriminant Analysis; IFDB - Iranian Face DataBase.

Appendix C

A Preliminary Study of Face Age Estimation

This appendix presents a preliminary study for face age estimation using wrinkle patterns as shown in [PUB4]. It involves a preprocessing step for finding the mean shape, threshold detection for determining the Canny parameters, wrinkle extraction using Canny edge detector, and age group classification using a MLP.

C.1 Introduction

Research related to age estimation using face images has become increasingly important due to its potential use in various applications such as age group estimation in advertising and age estimation in access control. In contrast to other facial properties, age variation has several unique characteristics which make it a challenging task. As we age, the most pronounced facial change is the appearance of wrinkles (skin creases), which is the focus of aging research in cosmetic and nutrition studies. In this preliminary study, a method named as Local Wrinkle-based Extractor (LOWEX) is proposed for feature representation. This method detects and classifies facial age groups based on wrinkles. First, each face image is divided into several convex regions where wrinkles are distributed. Second, these regions are analysed using a Canny filter and then concatenated into a feature vector. Finally, the face is classified into an age group using a supervised learning algorithm, MLP. Results showed that the proposed LOWEX has a classification accuracy of 80% on FGNET dataset. This investigation showed that the wrinkle patterns have great potential for age estimation. At the end, the advantages

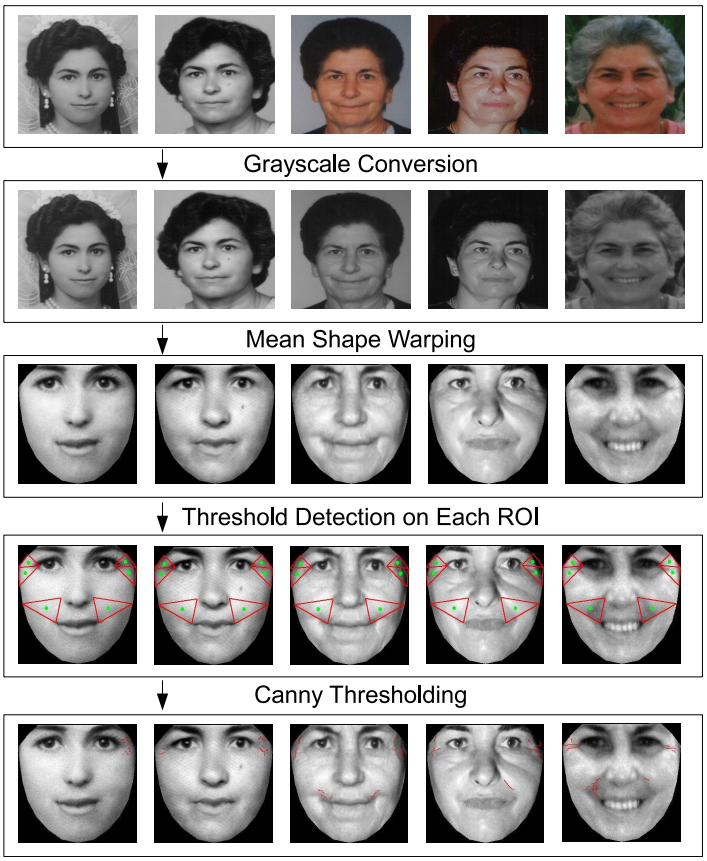


Figure C.1: Information flow of LOWEX for a single individual. The first row shows the original image, 2nd row presents the converted image in grey-scale, 3rd row is the shape-free patches, 4th row is the illustration of threshold detection and the last row is the result of canny thresholding. Note that this figure was redrawn from FGNET.

and challenges of using wrinkles are discussed.

C.2 Local Wrinkle-based Extractor

This section presents the LOWEX algorithm which extracts wrinkles of interest from the facial images. Using the wrinkles patterns, a MLP classifier is implemented to predict the age groups. Figure C.1 illustrates the flow of the proposed LOWEX. It consists of four steps: grey-scale conversion, mean shape warping, threshold detection and Canny thresholding. First, all images are converted to a grey-scale level ($\mathbf{I}_g$) for reducing dimensionality and avoiding issues concerning colour balance. The shape is then modelled based on manually labelled landmarks. For 2D image, the shape, $\mathbf{s}$, is represented by n landmarks, (x_i, y_i), as $\mathbf{s} = (x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n)^T$, $\{\mathbf{s}\} \in$

$\mathbf{D}$ where n is the total number of landmarks and $\mathbf{D}$ is the dataset. The shapes are normalised by the GPA where all the similarity is removed. Subsequently, the images in the training set are warped to the mean shape $\bar{\mathbf{s}}$ to produce a shape-free patch (see Section 3.4.3). From the mean shape, a set of triangular regions, denoted as $\mathcal{Z}_k$ (k is the index of each region) which reflects to the wrinkle growing region is identified, as suggested by Kwon and da Vitoria Lobo (1999). In each triangle, a square image is cropped, denoted as Q_k with the size $l \times l$ and its centre point is equivalent to the centroid of Q_k. There are two reasons a square image is applied for threshold detection: avoid noise around the triangle corners and allow derivation of stable thresholds for wrinkles. If the l value is either too large or too small, it will exceed the ROI or lost the focus of detection. Hence, the l value is set as 20 pixels. In the next step, the Otsu algorithm is applied to each Q_k in order to discover an appropriate threshold, $t_{1,k}$ and $t_{2,k}$, for that region. The Otsu algorithm is a global thresholding technique where a threshold can make the value of interclass variance the highest among all the possible values (see Section 3.1.1.4). Finally, Canny edge detection, f, with the detected thresholds (see Section 3.1.1.2) is utilised as

$$\mathbb{W}_k = f\left(\mathcal{Z}_k, t_{1,k}, t_{2,k}\right) \tag{C.1}$$

where $\mathbb{W}$ is the detected edges (wrinkles) of each region. Length, $\mathbf{g}$, and amount, $\mathbf{u}$, of $\mathbb{W}$ are combined into a feature pattern as

$$\mathbf{f} = [g_1, g_k, \cdots, g_m,\ u_1, u_k, \cdots, u_m], \quad \{\mathbf{g}_k, \mathbf{u}_k\} \in \mathbb{W}_k \tag{C.2}$$

where m is the total number of regions and $k \leq m$.

C.3 Age Group Classification

In this experiment, a similar architecture of MLP as described in Section 3.3.1.2 is adopted. Total nodes of input layer is according to the feature size, s, used for capturing input patterns. If the feature size, s, is 24 then the total number of input layer will be set correspondingly. In this work, $s = m * 2$. Usually, the total number of hidden layer is $1/2$ or $1/3$ of input layer. In this experiment, it is set as $s/2$. The output layer consists of 2 units which is based on the number of classes needed. The input values to the MLP are represented by $i_1, i_2, ..., i_s$, where input is normalised between -1 and 1. The output values to the backpropagation neural network are represented by binary codes, $[0, 1]$

Table C.1: MLP parameters in WEKA.

Description	Value
Function	Backpropagation
Input Node	s
Hidden Node	s / 2
Output Node	Group A / Group B
Learning Rate	0.3
Momentum Rate	0.2
Epochs	500
Features Normalised	-1 to 1
Output Normalised	-1 to 1

and $[1,0]$, which correspond to the age groups of young and old.

Table C.1 shows the MLP parameters used in this experiment. The input node, hidden node and output node are s, $s/2$, and 2, respectively; learning rate is 0.3; momentum rate is 0.2; epochs is 500; input and output are normalised between -1 and 1. These parameters were determined through a grid search approach (Hsu et al., 2003) and it is implemented in open source software Waikato Environment for Knowledge Analysis (WEKA) with classifier "MultilayerPerceptron" (Hall et al., 2009).

C.4 Experimental Settings

The FGNET dataset was used to evaluate the performance of the proposed method (see Section 3.6.2 for dataset description). All individuals in the dataset have more than one image included with different ages. Each image has 68 annotated facial feature points; these were used for the shape normalisation. According to Albert et al. (2007), normal aging of the facial soft tissues begins in the 20's with the fine facial lines appearing horizontally across forehead, vertical lines emerging between eyebrows, and faint lines developing around the outer corners of the eyes. With this in mind, the experiment set was divided into two groups: group A was between age 0 and 20 while group B was above 20 years old. In the experiment 1 and 2, 20 images were selected from FGNET for group A and another 20 images for group B, all of which had frontal pose and clear texture. In the experiment 3, these images were used as the training set and the remaining images of FGNET were used for testing. A 10-fold cross validation was performed to evaluate the performance of the age group classification. For ROI extraction, all 68 facial feature points were used to produce the shape-free patches.

In order to evaluate the performance of age group classification, measurements such as precision, recall and F1-measure are defined as,

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{C.3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{C.4}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Recall} + \text{Precision}} \tag{C.5a}$$

$$= \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{C.5b}$$

where TP, FP, FN are true positive, false positive and false negative. If both ground truth and predicted answer are yes, then it is a TP. If the ground truth is yes and predicted answer is no, then it is a FP. If the ground truth is no and predicted as yes, then it is a FN. Otherwise, it is a true negative (TN).

The precision describes the probability that an estimated object (randomly selected) is relevant to a desired group. The recall describes the probability of a relevant object being classified correctly. F1 describes the average between precision and recall which is often used in characterizing performance (Yap, 2008).

C.5 Experimental Results

This experiment used a Canny edge detector to find the wrinkles in the region of interest. The output of Canny edge detection is a binary image in which the true values represent the edges. In Matlab, either the thresholds of the Canny function be defined or the default threshold is used. Figure C.2 shows the effects of using different thresholds of the Canny method. The first row shows a 17-year-old participant with no wrinkles on face, while the second row shows a 22 year-old participant with some noticeable wrinkles. It is observed that the Canny function with threshold detection performs better than Canny with default threshold. A default threshold might not suitable for different sizes of wrinkles where it may extract excessive amounts of detail. The experiment showed that an adaptive method like the Otsu algorithm decides a better threshold for wrinkle detection. (Otsu, 1979).

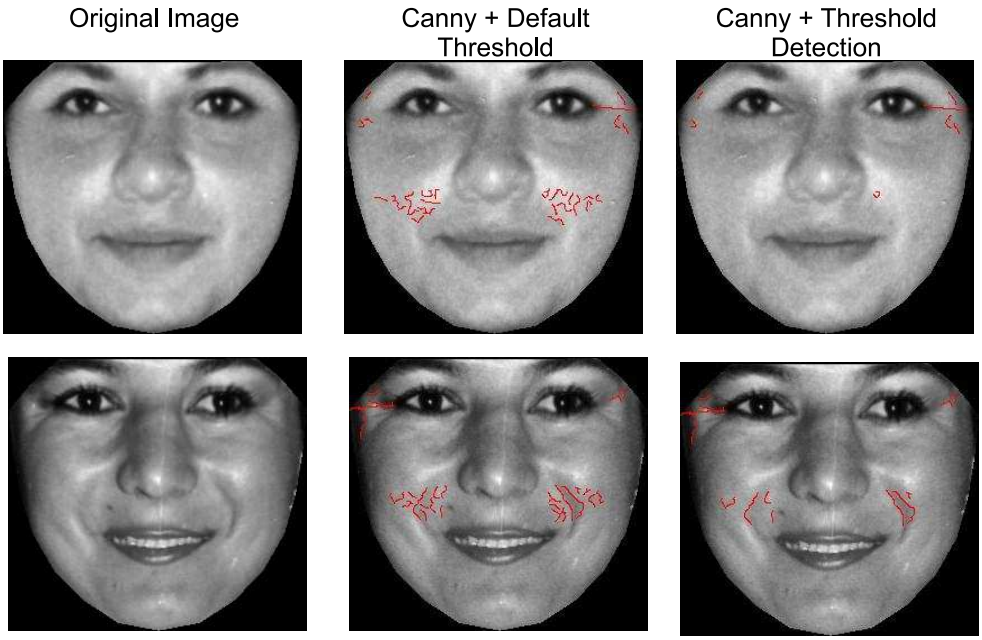


Figure C.2: Comparison between manual threshold and automatic threshold detection for Canny edge detection. Note that this figure was redrawn from FGNET.

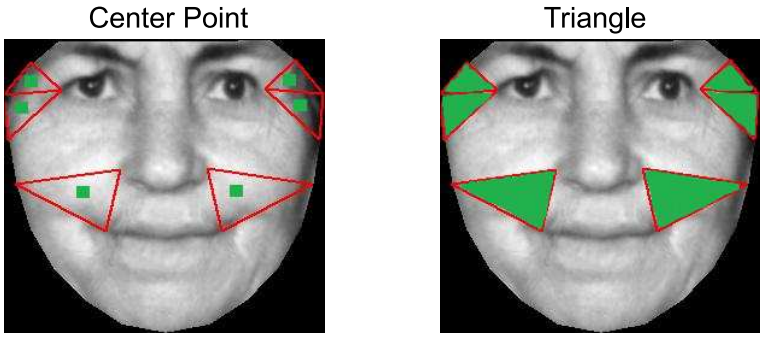


Figure C.3: Center point versus triangle region. Green is the area of interest for threshold detection. Note that this figure was redrawn from FGNET.

In Experiment 1, threshold detection by centre point region ($\tilde{c}_k$) was compared with that from the whole region ($\mathcal{Z}_k$). Figure C.3 shows the different areas (green area) used for threshold detection. Table C.2 shows the comparative results for centre point and triangle region threshold determination. The classification accuracy for triangle threshold is 67.50% and it is improved by centre point threshold to an accuracy of 80%. It seems that triangle threshold is less consistent due to noise around the triangle corners which diverts the Otsu algorithm from finding appropriate threshold. Results showed that a centre point threshold which focuses on the wrinkle area gives a better result than using the whole region.

Table C.2: Experiment 1, a comparison of threshold type.

Threshold Type	Group A	Group B	Accuracy
Centre Point	16	16	80.00%
Triangle	13	14	67.50%

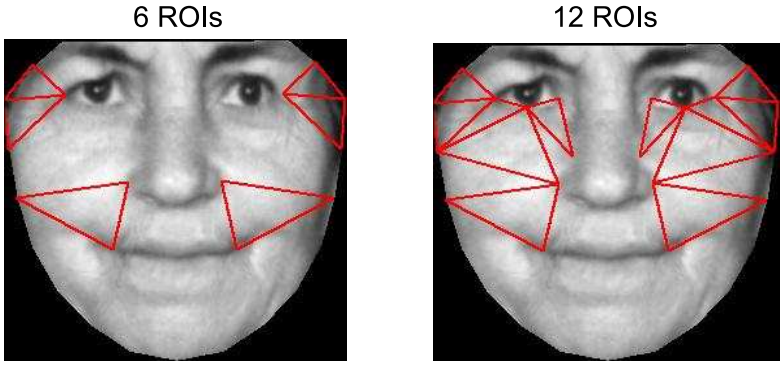


Figure C.4: Different wrinkle region amount. Note that this figure was redrawn from FGNET.

In Experiment 2, effects of ROI size were considered. In general, facial wrinkles are created by repeated facial muscular movements and expressions, therefore it formed increasingly as the person gets older. It was assumed that the amount of wrinkles is a cue that strengthens the feature for discriminating the age groups. The experiment settings are the same as Experiment 1 but the number of ROI was increased from 6 to 12, which included the eye bag area. Figure C.4 presents the different ROI used in this test.

Table C.3 shows the experimental results of using different amount of ROIs. Results showed that the region amount affects the age group classification where 12 ROIs hits an accuracy of 82.50% compared to 6 ROIs with an accuracy of 80.00%.

In Experiment 3, age group classification was performed on the whole FGNET dataset. The selected 40 images from Experiment 1 were used as a training set and the remaining images of FGNET as the test set. Other settings were same as Experiment 1. Table C.4 illustrates the detailed accuracy by class. Overall, the F1-measure of group A was 0.731 and group B was 0.466. This implies that 618 of 962 samples were correctly classified. It was expected that the classification performance might drop significantly due to the very poor quality and imbalanced nature of the FGNET images.

Table C.3: Experiment 2, variation in numbers of ROIs.

Threshold Type	Group A	Group B	Accuracy
6 ROIs	16	16	80.00%
12 ROIs	16	17	82.50%

Table C.4: Experiment 3, detailed accuracy by class.

Description	Group A	Group B
TP Rate	0.678	0.551
FP Rate	0.449	0.322
Precision	0.793	0.403
Recall	0.678	0.551
F1-measure	0.731	0.466

C.6 Discussion

Overall, although the accuracy is reasonable, there are a number of issues to be investigated. For the experimental set-up, the age groups were divided into two groups, one is between age 0 and 20 and others are above 20. This arbitrary choice of cut-off point between age groups makes for a highly unbalanced dataset. It would be interesting if the set-up could be narrowed into a smaller age interval, say 10 years old. Second, it has been noticed that the manual selection of higher quality images for the training set makes it unrepresentative of the data as a whole. Statistical test would be a better way to show that this dataset is uniform. Third, a validation on the correctness of wrinkle detection is needed. It is well known that the Canny edge detector extracts the boundary of an object. It would be good to see a validation between human annotation and prediction by the edge detector. Fourth, FGNET is a noisy dataset where most of the face wrinkles are distorted due to the poor image quality. A better resolution is important for local feature analysis. Therefore, different datasets should be used for validating the wrinkle analysis. Fifth, face age estimation is a complex modelling process which involves many parameters such as ethnicity, gender and expression. Thus, the use of wrinkles as key features for predicting the face age is an interesting area to be explored.

C.7 Summary

In this appendix, a local feature representation namely LOWEX is explored for age estimation. Features produced by LOWEX are wrinkle length and amount. Then, a MLP

is designed to classify the age groups based on the feature descriptor. Results showed that the proposed method produces an effective feature descriptor for age group classification with high accuracy. This experiment demonstrated that wrinkle pattern is a key feature in discriminating between young and old people. However, it is observed that the Canny edge detector only segments the wrinkle boundary instead of wrinkle itself. It might affect the performance of age group classification which depends on wrinkle statistics. Therefore, it would be interesting to see how wrinkles can be extracted accurately and further used for detailed age estimation.

Appendix D

AAM Builder Configuration

The AAM-API is a free, open-source, C++ implementation of the AAM framework targeted for education, further AAM research and for pure analysis and segmentation purposes. AAM-API details can be found at Stegmann (2003). The builder configuration of each datasets is depicted as follows.

FGNET

```
#############################################################################
#
#   Active Appearance Model Builder Configuration File
#
#############################################################################

3   # Model reduction               [1-n]   (reduction factor = 1/x)

3   # Model expansion               [0-n]   (pixels along the point normal)

1   # Use convex hull               [0|1]   (off/on)

1   # Verbose mode                  [0|1]   (off/on)

1   # Write registration movie      [0|1]   (off/on)

1   # Write variance image          [0|1]   (off/on)

1   # Produce model documentation   [0|1]   (off/on)
```

```
1   # Use tangent space projection  [0|1]   (off/on)

1   # Training method [ 0=PC Regression, 1=Jacobian (recommended) ]

95  # Shape model truncation (percentage [0-100], -1=parallel analysis)

95  # Texture model truncation (percentage [0-100], -1=parallel analysis)

95  # Combined model truncation (percentage [0-100], -2=no combined model)

1   # Subsampling of the training set (during training) [1-n]

1   # Warping method [ 0=benchmark, 1=software, 2=hardware (requires OpenGL) ]
```

FERET

```
##############################################################################
#
#   Active Appearance Model Builder Configuration File
#
##############################################################################

5   # Model reduction               [1-n]   (reduction factor = 1/x)

3   # Model expansion               [0-n]   (pixels along the point normal)

1   # Use convex hull               [0|1]   (off/on)

1   # Verbose mode                  [0|1]   (off/on)

1   # Write registration movie      [0|1]   (off/on)

1   # Write variance image          [0|1]   (off/on)

1   # Produce model documentation   [0|1]   (off/on)

1   # Use tangent space projection  [0|1]   (off/on)
```

```
1   # Training method [ 0=PC Regression, 1=Jacobian (recommended) ]

95  # Shape model truncation (percentage [0-100], -1=parallel analysis)

95  # Texture model truncation (percentage [0-100], -1=parallel analysis)

95  # Combined model truncation (percentage [0-100], -2=no combined model)

1   # Subsampling of the training set (during training) [1-n]

1   # Warping method [ 0=benchmark, 1=software, 2=hardware (requires OpenGL) ]
```

MORPH

```
##############################################################################
#
#   Active Appearance Model Builder Configuration File
#
##############################################################################

3   # Model reduction               [1-n]   (reduction factor = 1/x)

3   # Model expansion               [0-n]   (pixels along the point normal)

1   # Use convex hull               [0|1]   (off/on)

1   # Verbose mode                  [0|1]   (off/on)

1   # Write registration movie      [0|1]   (off/on)

1   # Write variance image          [0|1]   (off/on)

1   # Produce model documentation   [0|1]   (off/on)

1   # Use tangent space projection  [0|1]   (off/on)

1   # Training method [ 0=PC Regression, 1=Jacobian (recommended) ]
```

```
95  # Shape model truncation (percentage [0-100], -1=parallel analysis)

95  # Texture model truncation (percentage [0-100], -1=parallel analysis)

95  # Combined model truncation (percentage [0-100], -2=no combined model)

1   # Subsampling of the training set (during training) [1-n]

1   # Warping method [ 0=benchmark, 1=software, 2=hardware (requires OpenGL) ]
```

PAL

```
##############################################################################
#
#   Active Appearance Model Builder Configuration File
#
##############################################################################

2   # Model reduction               [1-n]   (reduction factor = 1/x)

3   # Model expansion               [0-n]   (pixels along the point normal)

1   # Use convex hull               [0|1]   (off/on)

1   # Verbose mode                  [0|1]   (off/on)

1   # Write registration movie      [0|1]   (off/on)

1   # Write variance image          [0|1]   (off/on)

1   # Produce model documentation   [0|1]   (off/on)

1   # Use tangent space projection  [0|1]   (off/on)

1   # Training method [ 0=PC Regression, 1=Jacobian (recommended) ]

95  # Shape model truncation (percentage [0-100], -1=parallel analysis)
```

```
95  # Texture model truncation (percentage [0-100], -1=parallel analysis)

95  # Combined model truncation (percentage [0-100], -2=no combined model)

1   # Subsampling of the training set (during training) [1-n]

1   # Warping method [ 0=benchmark, 1=software, 2=hardware (requires OpenGL) ]
```