UNIVERSITY OF
KWAZULU-NATAL

INYUVESI
YAKWAZULU-NATALI

Hierarchical Age Estimation Using Enhanced Facial Features

Author:
Raphael ANGULU
215082164

Supervisor:
Prof. Jules-Raymond TAPAMO
Co-Supervisor:
Prof. Aderemi O. ADEWUMI

A thesis submitted to the University of KwaZulu-Natal, College of Agriculture, Engineering and Science, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Science

School of Mathematics, Statistics and Computer Science
University of KwaZulu-Natal

March, 2018

Declaration of Authorship

I, Raphael ANGULU, declare that this thesis titled, "Hierarchical Age Estimation Using Enhanced Facial Features" and the work presented in it are my own. I declare that:

1. The research reported in this thesis, except where otherwise indicated or acknowledged, is my original work;
2. This thesis has not been submitted in full or in part for any degree or examination to any other university;
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
   (a) their words have been re-written but the general information attributed to them has been referenced;
   (b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;
5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Candidate:

Signed :_____ *Date* :_____
Raphael ANGULU

As the candidates' supervisor I approve the submission of this thesis for examination

Supervisor:

Signed :_____ *Date* :_____
Prof. Jules-Raymond TAPAMO

Co-supervisor:

Signed :_____ *Date* :_____
Prof. Aderemi O. ADEWUMI

List of Publications

I, Raphael ANGULU, declare that the following are publications from this thesis

1. Angulu, R, Tapamo, J-R, Adewumi, O. A, "Landmark localization approach for facial computing", *in Proceedings of IEEE Conference on Information and Communication Technology and Society (ICTAS)*, Durban, South Africa, pp. 1-6, March 2017, Link http://ieeexplore.ieee.org/document/7920664/

2. Angulu, R, Tapamo, J-R, Adewumi, O. A, "Human age estimation using multi-frequency biologically inspired features (MF-BIF)", *in Proceedings of 2017 IEEE AFRICON Conference*, Cape Town, South Africa, pp. 32-37, September 2017, Link http://ieeexplore.ieee.org/document/8095450/

3. Angulu, R, Tapamo, J-R, Adewumi, O. A, "Frontal Face Landmark Displacement across Age", *IEEE Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech)*, CUT, Bloemfontein, South Africa, November 2017, Link http://ieeexplore.ieee.org/document/8261119/

4. Angulu, R, Tapamo, J-R, Adewumi, O. A, "Age estimation with local statistical biologically inspired features", *in Proceedings of 2017 IEEE International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, USA, December 2017

5. Angulu R., Tapamo J.R., Adewumi A.O. (2018) Age Estimation with Local Ternary Directional Patterns. In: Paul M., Hitoshi C., Huang Q. (eds) Image and Video Technology. PSIVT 2018. Lecture Notes in Computer Science, vol 10749. Springer, Cham Link: https://link.springer.com/chapter/10.1007%2F978-3-319-75786-5_34

6. Angulu, R, Tapamo, J-R, Adewumi, O. A, "Age-group Estimation Using Feature and Decision Level Fusion", *Computer Journal*, 2018
[Under Revision]

7. Angulu, R, Tapamo, J-R, Adewumi, O. A, "Human age estimation via face images: A survey", *EURASIP Journal of Image and Video Processing*, 2018
[Under Revision]

8. Angulu, R, Tapamo, J-R, Adewumi, O. A, "Age estimation using significant orientation response local directional patterns", *IET Biometrics*, 2018 **[Under Review]**

"All Computer Science problems can be solved by another layer of abstraction, except for the problem of too many layers of abstraction."

David J. Wheeler

"We cannot seek achievement for ourselves and forget about progress and prosperity for our community... Our ambitions must be broad enough to include the aspirations and needs of others, for their sakes and for our own."

Cesar Chavez

Abstract

Ageing is a stochastic, inevitable and uncontrollable process that constantly affect shape, texture and general appearance of the human face. Humans can easily determine ones' gender, identity and ethnicity with highest accuracy as compared to age. This makes development of automatic age estimation techniques that surpass human performance an attractive yet challenging task. Automatic age estimation requires extraction of robust and reliable age discriminative features. Local binary patterns (LBP) sensitivity to noise makes it insufficiently reliable in capturing age discriminative features. Although local ternary patterns (LTP) is insensitive to noise, it uses a single static threshold for all images regardless of varied image conditions. Local directional patterns (LDP) uses k directional responses to encode image gradient and disregards not only central pixel in the local neighborhood but also $8-k$ directional responses. Every pixel in an image carry subtle information. Discarding $8-k$ directional responses lead to lose of discriminative texture features. This study proposes two variations of LDP operator for texture extraction. Significant-orientation response LDP (SOR-LDP) encodes image gradient by grouping eight directional responses into four pairs. Each pair represents orientation of an edge with respect to central reference pixel. Values in each pair are compared and the bit corresponding to the maximum value in the pair is set to 1 while the other is set to 0. The resultant binary code is converted to decimal and assigned to the central pixel as its' SOR-LDP code. Texture features are contained in the histogram of SOR-LDP encoded image. Local ternary directional patterns (LTDP) first gets the difference between neighboring pixels and central pixel in 3×3 image region. These differential values are convolved with Kirsch edge detectors to obtain directional responses. These responses are normalized and used as probability of an edge occurring towards a respective direction. An adaptive threshold is applied to derive LTDP code. The LTDP code is split into its positive and negative LTDP codes. Histograms of negative and positive LTDP encoded images are concatenated to obtain texture feature. Regardless of there being evidence of spatial frequency processing in primary visual cortex, biologically inspired features (BIF) that model visual cortex uses only scale and orientation selectivity in feature extraction. Furthermore, these BIF are extracted using holistic (global) pooling across scale and orientations leading to lose of substantive information. This study proposes multi-frequency BIF (MF-BIF) where frequency selectivity is introduced in BIF modelling. Local statistical BIF (LS-BIF) uses local pooling within scale, orientation and frequency in $n \times n$ region for BIF extraction. Using Leave-one-person-out (LOPO) validation protocol, this study investigated performance of proposed feature extractors in age estimation in a hierarchical way by performing age-group classification using Multi-layer Perceptron (MLP) followed by within age-group exact age regression using support vector regression (SVR). Mean absolute error (MAE) and cumulative score (CS) were used to evaluate performance of proposed face descriptors. Experimental results on FG-NET ageing dataset show that SOR-LDP, LTDP, MF-BIF and LS-BIF outperform state-of-the-art feature descriptors in age estimation. Experimental results show that performing gender discrimination before age-group and age estimation further improves age estimation accuracies. Shape, appearance, wrinkle and texture features are simultaneously extracted by visual system in primates for the brain to process and understand an image or a scene. However, age estimation systems in the literature use a single feature for age estimation. A single feature is not sufficient enough to capture subtle age discriminative traits due to stochastic and personalized nature of ageing. This study propose fusion of different facial features to enhance their discriminative power. Experimental results show that fusing shape, texture, wrinkle and appearance result into robust age discriminative features that achieve lower MAE compared to single feature performance.

Acknowledgements

I thank the Almighty God of all creations for the gift of life and good health, guidance and protection He has unconditionally granted me.

I am obliged to Masinde Muliro University of Science and Technology for offering me a chance to pursue this research under proficient guidance, support and advice of Prof. J.-R. Tapamo and Prof. A. O. Adewumi. I am indebted to Prof. J. R. Tapamo for literally teaching me Image Processing, skills that helped me settle in this hitherto strange field. This research would never have been done smoothly without superb technical support of the crazy tech man Soren Greenwoods, thank you Soren.

I am grateful to my dear mother Jane Wakhu Angulu, a courages Mum who taught me that whenever you step into the unknown, you either find a solid base or you learn to fly. And I must admit in most cases, I learnt how to fly. Thank you mum for your support, love and guidance without which I would never have found a definition of myself. I am indebted to my wife Volian Nekesa and my daughter Finswin Shimuli Angulu who allowed me to travel miles to pursue this study. Your love and emotional support is very much acknowledged! Among the best, you two, come second after Christ.

Finally but not least, I thank my fellow researchers in the Image Processing, Data Mining, Machine Learning and Optimization research groups for their support, critic and comments that helped shape not only this thesis but also my mind.

Angulu, R.

Contents

List of Figures

List of Tables

List of Algorithms

List of Abbreviations

1D	**1-D**imensional
2D	**2-D**imensional
3D	**3-D**imensional
AAM	**A**ctive **A**ppearance **M**odel
ACM	**A**ctive **C**ontour **M**odel
AE	**A**bsolute **E**rror
AGES	**A**ging **p**attern **S**ubspace
ANN	**A**rtificial **N**eural **N**etwork
ASM	**A**ctive **S**hape **M**odel
BIF	**B**iologically **I**nspired **F**eatures
BoW	**B**ag-**o**f-**W**ords
CCA	**C**anonical **C**orrelation **A**nalysis
CNN	**C**onvolutional **N**eural **N**etwork
CS	**C**umulative **S**core
CUDA	**C**ompute **U**nified **D**evice **A**rchitecture
DPS	**D**ensity-**P**reserving-**S**ampling
EM	**E**xpectation **M**aximization
EMF-BIF	**E**nhanced **M**ulti **F**requency **BIF**
FG-NET	**F**ace and **G**esture **NET**work
GLCM	**G**ray **L**evel **C**ooccurence **M**atrix
GNN	**G**rassmann **N**earest **N**eighbor
GOP	**G**radient **O**riented **P**yramids
GPU	**G**raphical **P**rocessing **U**nit
HOG	**H**istogram of **O**riented **G**radients
HPC	**H**igh **P**erformance **C**omputing
KNN	**K**-**N**earest **N**eighbor
LARR	**L**ocally **A**djusted **R**obbust **R**egression
LBP	**L**ocal **B**inary **P**atterns
LDA	**L**inear **D**iscriminant **A**nalysis
LDP	**L**ocal **D**irectional **P**atterns
LLE	**L**ocally **L**inear **E**mbedding
LOO	**L**eave-**O**ne-**O**ut
LOPO	**L**eave-**O**ne-**P**erson-**O**ut
LPO	**L**eave-**P**-**O**ut
LPP	**L**ocallity **P**reserving **P**rojection
LS-BIF	**L**ocal **S**tatistical **BIF**
LTDP	**L**ocal **T**ernary **D**irectional **P**atterns
LTP	**L**ocal **T**ernary **P**atterns
MAE	**M**ean **A**bsolute **E**rror
MF-BIF	**M**ulti **F**requency **BIF**
MMML	**M**ulti **M**anifold **M**etric **L**earning
MLP	**M**ulti-**L**ayer **P**erceptron
OLPP	**O**rthogonal **L**ocallity **P**reserving **P**rojection

PCA	**P**rincipal **C**omponent **A**nalysis
PLS	**P**artial **L**east **S**quares
RBF	**R**adial **B**asis **F**unction
ROI	**R**egion **O**f **I**nterest
SDP	**S**emi**d**efinite **P**rogramming
SFP	**S**patially **F**lexible **P**atches
SOM	**S**elf-**O**rganizing **M**aps
SOR-LDP	**S**ignificant **O**rientation **R**esponse LDP
SURF	**S**peeded-**U**p **R**obbust **F**eatures
SVM	**S**upport **V**ector **M**achines
SVR	**S**upport **V**ector **R**egression
UTP	**U**niform **T**ernary **P**atterns
YGA	**Y**amaha **G**ender and **A**ge

This thesis is dedicated to my lovely daughter Finswin Shimuli Angulu

Chapter 1

Introduction

Someone must have asked you: *how old are you?* This could have happened in a doctors' consultation room, school or college admission, at a job recruitment interview, at workplace or in a normal chat with a friend. These are just but a few scenarios where automatic age estimation can be used. When date of birth is prompted, in most of the cases, it is a subject's age being derived. Most of the time people try to estimate your age from your facial appearance even before they ask; how old are you? However, facial appearance varies more dynamically as it is affected by several intrinsic and extrinsic factors including genetic make-up, bony structure, pose, facial expression, head profile, illumination, ageing, occlusion, mustache, beards, makeup (cosmetics), diet, weather and hair style. Regardless of these variations, humans can still estimate one's age from facial appearance, but not with ideal accuracy.

Humans develop early in life the ability to estimate age in the range of 20-60 years [1] but with lower accuracy as compared to age estimation in the range of 0-20 years [2]. Humans cannot accurately estimate age of individuals aged above 60 years. Although humans can estimate age from facial visual cues, they do not achieve better performance as they do with gender, ethnicity and identity estimation. The commonly used measure for age estimation is mean absolute error (MAE) and recent study [2] found that humans can estimate age with MAE of about 7.4 years. This makes developing age estimation techniques that surpass or equivalent to human performance an attractive yet challenging task [3]. Developments in engineering and computer science[1] provide positive answer to the question; Can a computer estimate age with same accuracy as a human? With these advances in technology, and vast application areas, there is need to develop automatic age estimation techniques that can match or surpass human accuracy.

Age related research can be categorized into two; 1) age estimation and 2) age simulation. Age estimation involves training a model that can estimate age in an image while age simulation involve modelling of ageing effects on facial appearance. Age estimation is a technique of automatically labelling face image with exact age (integer showing years elapsed since birth) or age-group (range within which actual age falls). There are four types of ages [3]

1. *Actual age*- This is the chronological age signified by the total number of years elapsed since an individual was born. This is the real age of an individual

[1] Advances is software and hardware. Techniques like Convolutional Neural Networks (CNN), High Performance Computing (HPCs), Graphics Processing Units (GPUs), Open Graphics Library (OpenGL), Compute Unified Device Architecture (CUDA) and Open Computing Language (OpenCL) platforms

2. *Appearance age*- Individual's age as it can be depicted from facial appearance. It is the age as shown by visual appearance of an individual

3. *Perceived age*- This is individual's age as estimated by a human expert with cues from appearance age

4. *Estimated age*-Individual's age as estimated by a computer with cues from visual appearance

Automated age estimation techniques use visual cues (appearance age features) to estimate actual age. It could be thought that appearance age is consistent to actual age but variations are visible due to personalized ageing patterns as well as artificial/environmental factors. Estimated age and perceived age are both based on visual artifacts of appearance age. Actual age represent the ground truth of an individual's age. Facial skin appearance determines appearance and perceived age [4] [5]. Twigg and Majima [6] found that females seem to be very concerned about their perceived age and facial appearance as indicated by high-cost cosmetic and aesthetic procedures they incur.

Ageing is a stochastic inevitable and irreversible process that causes variation in facial shape and texture. Ageing introduces significant change in facial shape in formative years and relatively large texture variations with minor change in shape in older age groups [7] [3]. Shape variations in younger age groups is caused by craniofacial growth. As one age, facial blemishes like wrinkles, freckles and age-spots appear. Underneath the skin, melanin producing cells are damaged due to exposure to sun's Ultra Violet (UV) rays leading to over production of melanin. Consequently, light reflecting collagen not only decreases but also become non-uniformly distributed making facial skin tone non-uniform [8]. Parts adversely affected by sunlight are upper cheek, nose, nose-bridge and forehead. Consequently, these regions are selected for skin wrinkle and texture analysis for age-group and age estimation. The most visible variations in adulthood to old age are skin variations exhibited in texture variations. Biologically, as the skin grow old, collagen underneath the skin is lost [3]. Loss of collagen and effect of gravity makes the skin to become darker, thinner, leathery and less elastic. Facial spots and wrinkles appear gradually. The framework of bones beneath the skin may also start deteriorating leading to accelerated development of wrinkles and variations in skin texture. More details about facial ageing can be found in [9].

Bowen and Atwood [10] asserts that, ageing is accumulation of physical, psychological and social changes in humans. For instance, reflex may slow as one age while general knowledge may increase with age. It has been found that ageing is one of the risk factors in contacting most human illness [11] as it causes a number of effects in one's life. For instance teenagers lose ability to hear high-frequency sounds exhibited in infants [12] while skin wrinkles start appearing in sun-exposed regions as one age [13]. Ageing also affects face recognition and verification system due to large variations between age separated images of the same subject [14] [15]. Age estimation is a soft biometric that can be used to improve face recognition and verification accuracies. This could be done by estimating one's age prior to face recognition and verification.

There are a number of age associated traits in humans, but facial appearance is the most common characteristic people rely on for estimation of one's age. Human

face provides prior perceptible information about one's age, gender, identity, ethnicity, expression and mood. Alley [16] asserts that attributes derived from human facial appearance like mood and perceived age significantly impact interpersonal behaviour as is considered as essential contextual cue in social networks [17], [18] and social perception during craniofacial development [19]. Facial appearance may influence impressions if they reveal psychological characteristics. These are indications of enormous prospective applications of automatic age estimation.

Ageing is affected by both intrinsic factors like genetic makeup and extrinsic factors like environment (dry and windy weather), lifestyle (diet, smoking, drug abuse), artificial (use of cosmetics), exposure to ultraviolet rays and air pollution (cigarette smoke, oxides, particulate matter) [20] [21] [22]. These extrinsic factors lead to development of deeper wrinkle lines and change in skin pigmentation [23] [24]. For instance, frequent facial expressions intensify formation of wrinkle lines [25]. Facial ageing has three unique attributes [5]

1. *Ageing is inevitable and uncontrollable*. No one can avoid ageing, advance or delay it. Ageing process is slow but irreversible.

2. *Ageing patterns are personalized*. People age differently. Individuals ageing pattern is dependent on her/his genetic makeup as well as various extrinsic factors such as health, environmental conditions, lifestyle.

3. *Achieved ageing patterns are temporal*. Facial variations caused by ageing are not permanent. Furthermore, facial variation at a particular point in time affects future appearance and does not affect previous appearance of these faces.

These facial ageing attributes, among other factors, make automatic age estimation a difficult and challenging task. Since individuals cannot voluntarily control ageing, automatic age estimation data collection becomes a hard task to do. This problem was slightly alleviated by dissemination of Face and Gesture Network (FG-NET) Ageing Dataset [26] in 2002. Although this dataset has images of subjects at different ages, there are several missing images hence making the ageing patterns incomplete. Fortunately, a complete ageing face dataset is not needed since humans, who computers try to mimic, also learn how to process face image patterns from incomplete patterns. Age estimation technique should be capable of considering various ageing patterns since each individual has his/her own ageing pattern.

Information rendered by human face has attracted significant attention in face image processing research community. Face image based age and age-group estimation have vast applications areas in age invariant face recognition, face verification across age, commercial and law enforcement areas [27] [28] [29] [30] [31], security control and surveillance [3] [32], age based image retrieval [33], biometrics [3] [34] [35] human computer interaction [36] [37] and electronic customer relationship management (ECRM) [3]. The main aim of studying age estimation is to find out ageing patterns and variations in facial appearance and how to best characterize an ageing face for accurate age estimation. Although this problem has attracted significant research, still automatic age estimation accuracies are far below the ideal.

Age estimation from face images require a great deal of information. Extraction of facial ageing features is vital since age estimator performance rely on the quality of these features [3]. A number of ageing feature extraction techniques have been proposed in the literature such as active appearance model (AAM) [38], age manifold

[39] [40], anthropometric model [41], Ageing pattern Subspace (AGES) [36] [5], biologically inspired features (BIF) [42], patch-based appearance model [43] [43], local binary patterns (LBP) [44][45][46], local directional patterns (LDP) [47], local ternary patterns (LTP) [48] among others. After ageing feature extraction, the next important aspect is building a reliable and robust age estimator. A machine learning approach could be used to learn a model and estimate age of a probe image from the learnt model.

Age estimation is not a classical classification problem. This problem can be approached as a multi-class classification problem [32] [5] [49] [50] or a regression problem [39] [40] [42] [51] [52] [53] [54] [55] [56] or a combination of classification and regression [49] [57] [58]. It could also be viewed as an ordinal process thereby warranting age estimation to be treated as a ranking problem [59] [60] [61]. Some of the factors that affect age estimation research is difficulty in collecting large database with chronological images of a subject. Prolific and diverse information conveyed by faces also make special attributes of ageing variations not accurately captured [3]. Uncontrollable, personalized age progression and diverse information displayed on faces further complicates age estimation problem [39] [5] [62]. In this thesis, age estimation is approached as a combination of both classification and regression analysis.

1.2 Motivation

Even though there are several image feature descriptors, extracting robust and reliable features from an image for pattern recognition remains an open problem. LBP is sensitive to noise an monotonic illumination. LTP uses a static user-chosen threshold for all images regardless of image condition. LDP only considers top k responses and discard remaining $8 - k$ responses. Furthermore, LDP does not consider central pixel while calculating image gradient. The questions that motivates improvement of LDP operator are 1) Does considering all directional responses improve discriminative power of LDP? 2) How does incorporation of central pixel influence and adaptive threshold improve discriminative power of LTP and LDP operators? Can LDP operator be improved by extending it to a ternary pattern with an adaptive threshold? Although previous studies [63], [64] [65] [66] show that there is spatial frequency analysis in primary visual cortex, previous biologically inspired image processing studies do not incorporate frequency in visual cortex modelling. This study sort to find out how does incorporation of frequency selectivity improve robustness of BIF features? Furthermore, previous BIF studies mainly use holistic pooling across feature selectors. This study is motivated to find out effect of local pooling within feature selectors on robustness of BIF?. Existing facial component detection techniques only detect eyes, nose and mouth. Other facial components like cheeks, nose-bridge and forehead provide rich texture information. Therefore, how can these secondary facial components be accurately and swiftly located for subsequent processing? Can spatial and geometric information of fiducial component help is detection of these secondary components? Craniofacial development cause facial components to drift as they expand to occupy space created on the cranium. What is the trend of facial component displacement across age? By what magnitude does facial components drift towards or away from each other? Ageing is a stochastic, inevitable and personalized process that slowly cause irreversible effects of facial shape and appearance. This study was motivated to find out what factors affects ageing, what algorithms are used for age estimation, how are face images

represented for age estimation, what are the challenges facing age estimation, and how does age estimation state-of-the-art techniques perform? Although ageing affects both facial shape and appearance, previous studies use a single feature for age estimation. Single feature may not be enough to capture subtle age variations. This study sort to find out how does feature enhancement through fusion improve age estimation accuracies? Facial appearance varies across gender and age. Therefore, could performing gender and age-group discrimination improve age estimation accuracies? Looking at the preceding questions, the motivation of this study could be summarized in the question, how does use of enhanced facial features and hierarchical age estimation improve age estimation accuracies?

1.3 Problem statement

There has been relatively few age estimation techniques proposed in the literature [67]. This could be attributed to age estimation not being a classical classification problem, lack of large database with chronological images of subjects and difficulty in accurately capturing ageing features from prolific and diverse information conveyed by the face. Uncontrollability and personalized ageing patterns further complicates age estimation problem [39] [5] [62].

Age estimation techniques have not yet achieved the ideal zero MAE. This makes age estimation a viable research problem. Commonly used ageing feature extraction techniques are Local Binary Pattern (LBP) [44] [45], Active Appearance Model (AAM) [38], Active Shape Model (ASM) [68], Gabor filters [69], Biologically Inspired Features (BIF) [70] [42], Local Ternary Patterns (LTP) [48], Local Directional Patterns [47]. Although these local texture descriptors are proven to be effective in various pattern recognition tasks, their performance in age estimation is not comprehensively investigated. LBP sensitivity to noise makes it not sufficiently reliable in capturing age discriminative features. Although LTP is insensitive to noise, it uses a single static threshold for all images and datasets regardless of image and dataset conditions. Limitation of LDP is that it uses k directional responses to encode image gradient and disregards not only central pixel in the local neighbourhood but also $8 - k$ directional responses. Every pixel in an image carry subtle information about the image. Discarding $8 - k$ directional responses result into lose of discriminative texture features. Furthermore, these top k responses could be aligned towards a particular orientation hence encoding texture features at that orientation only. To capture subtle ageing features, every pixel in the image must be considered in all orientations.

Although there is evidence of spatial frequency analysis in primary visual cortex [63], [64] [65] [66], BIF used in previous studies is extracted using scale and orientation selectivity only. Gabor filters are used to model simple cells [71] [72] which have basic attributes of multi-orientation, multi-frequency and multi-scale selection [73]. Although it has been found that using Gabor filters with different orientation and frequency selectivity results into useful discriminating features [64], BIF extraction techniques in literature do not include frequency selectivity. The question this study seeks to answer is, Does inclusion of frequency selectivity in BIF extraction improve its' discriminative power? Furthermore, BIF extraction techniques in literature use holistic (global) pooling across scales and orientations. Holistic pooling across feature selectors discard a lot of information that could improve robustness of BIF. This

study seek to find out how local pooling within feature selectors improve robustness and discriminative power of BIF extracted.

During craniofacial development, the forehead slopes back releasing space on the cranium. Eyes, ears, mouth and nose expand to cover interstitial space created [9] thereby causing fiducial landmarks to drift away or towards each other [74]. What is the trend in fiducial landmark displacement across age? This study track positions of these landmarks across age in an effort to answer the preceding question.

Exposure to UV rays cause ageing effects like freckles and age-spots on some parts of the face like upper cheek, nose, nose-bridge and forehead [75] due to over production of melanin. Although these parts could provide age discriminative features, there is no robust technique for their localization. Landmark localization techniques in literature only locate eyes, nose and mouth. This study seeks to answer the question, could spatial and geometric information of fiducial landmarks be used to locate forehead, nose-bridge and cheeks for subsequent face image processing?

Although it has been shown that facial geometry (and shape) has significant influence on facial appearance and perceived age [76] [77], age estimation using facial anthropometric measurements has not been significantly investigated in the literature. Age estimation techniques in literature use a single feature and/or classifier. Although facial shape provide discriminative features about ones' age [14], it may not be sufficiently reliable in adulthood and old age because in these stages, shape remains relatively constant compared to facial texture and appearance. This study seeks to find out how fusing shape, wrinkle, appearance and texture features improve age estimation accuracies. Furthermore, ageing is different across gender [40], [6] and age-groups [9]. This study seeks to answer the question, could performing gender and age-group discrimination improve age estimation accuracies?

1.4 Objectives of the study

This study sought to achieve the following objectives

1. To develop an automatic landmark localization technique that can locate both fiducial and secondary facial landmarks for facial computing
2. To analyze previous studies in age and age-group estimation: factors affecting ageing, algorithms used for age estimation, results achieved, age estimation challenges, image representation techniques for age modelling, age estimation evaluation protocols
3. To find out the trend of facial landmarks displacement across age
4. To develop a model for age-group estimation based on fused facial and facial component features and ensemble of classifiers
5. To define LDP variant that consider significant orientation responses in encoding image gradient rather than top k responses
6. To extend LDP to a ternary directional pattern based on pixel differential values and an adaptive threshold for encoding image gradient

7. To find out effect of frequency selectivity on the performance of BIF in age estimation
8. To investigate effect of local pooling, within feature selectors, on performance of BIF in age estimation
9. To develop a model for hierarchical age estimation using enhanced facial features with gender and age-group discrimination

1.5 Research questions

The following are the questions posed and investigated in this research

1. How can fiducial and secondary facial landmarks be localized for subsequent texture and shape face image processing?
2. What are the factors affecting ageing in persons, what algorithms are used for age estimation, what are the results achieved, what are the challenges facing age estimation and what are the limitations of existing age estimation approaches?
3. How are facial landmarks displaced from each other across age?
4. How does the use of fused facial and facial component features and ensemble of classifiers improve age-group estimation?
5. How does consideration of significant orientation response rather than top k responses improve performance of LDP operator in age estimation?
6. Does extending LDP to a ternary directional pattern descriptor with an adaptive threshold improve its' performance in age estimation?
7. To what extend does frequency selectivity improve performance of BIF in pattern recognition?
8. To what extend does local pooling within feature selectors improve performance of BIF in age estimation? How does local pooling within feature selectors compare to holistic pooling across feature selectors in BIF extraction?
9. Does gender and age-group discrimination improve age estimation accuracies?

1.6 Contribution of the thesis

Research work in this thesis makes the following contributions to the pattern recognition knowledge base

1. Localization of fiducial and secondary facial components for subsequent shape and texture extraction for pattern recognition
2. Comprehensive and analytical review of human facial ageing, factors affecting facial ageing, ageing feature extraction techniques, age estimation approaches, evaluation and performance of age estimators

3. Facial landmarks displacement across age is modelled as a time series problem. This can be used in facial shape modelling and simulation.

4. The study demonstrates that fusing different features (shape, texture, appearance e.t.c) and fusing of decisions from ensemble of classifiers lead to better accuracies in pattern recognition tasks

5. Extension of two local texture descriptor operators, LTP and LDP, for effective and robust texture features encoding for pattern recognition

6. Relevance of frequency selectivity in Biologically Inspired Feature (BIF) extraction is investigated and demonstrated

7. Importance of local pooling within feature selectors as compared to holistic (global) pooling across selectors is demonstrated

8. The study demonstrates that performing gender and age-group profiling prior to exact age estimation improves achieved accuracies

1.7 Scope of the study

This study focuses on extraction of ageing features from the face for age estimation. Age estimation systems often consist of various modules that include user interface, facial image acquisition, image preprocessing, storage servers, ageing feature extraction and age estimator. This study focuses on face and facial component detection and ageing feature extraction process with light considerations accorded to other modules. Age-group and age estimation are approached as hierarchical multi-class classification and regression problems respectively. Both shape and appearance ageing facial features are extracted from Face and Gesture Network (FG-NET) ageing dataset. This dataset has 1002 images of 82 subjects aged between 0 and 69 years. Support Vector Machines (SVM), Support Vector Regression (SVR) and Artificial Neural Network (ANN) machine learning techniques are used to learn ageing patterns from the extracted facial ageing features.

1.8 Thesis overview

The rest of this thesis is organized as follows: **`Chapter 2`** gives background in facial ageing, factors and challenges affecting age estimation via faces. This chapter also presents a review of related work in age estimation and ageing feature extraction techniques. **`Chapter 3`** describes the proposed hierarchical age estimation and presents materials and techniques used for facial component localization, local ternary directional patterns (LTDP), significant-orientation response local directional patters (SOR-LDP), multi-frequency biologically inspired features (MS-BIF) local statistical biologically inspired features (LS-BIF) and enhanced multi-frequency biologically inspired features (EMF-BIF). **`Chapter 4`** presents a comprehensive analysis and discussion of results including comparison with related work. The thesis is concluded in **`Chapter 5`** and recommendations for further research presented.

Chapter 2

Background and Literature Review

Facial ageing has received substantial research attention with increasing attention in age invariant face recognition, age estimation and face verification across age among other areas [27] [28] [29] [30] [31]. Ageing involves both variations in facial soft tissues and bony structure. Facial appearance at one age is different from facial appearance of the same individual at a different age. Age estimation has been extensively studied with the aim of finding out ageing patterns, pattern variations among individuals, factors that cause or accelerate facial ageing and how to best characterize human face for automatic age estimation. In this chapter present an analysis of facial ageing process, factors affecting facial ageing, age estimation techniques developed, and analysis of algorithms used for face representation and age estimation. Comparative analysis of previous research in age estimation is also presented.

2.1.1 Facial ageing

Ageing is a stochastic, uncontrollable, inevitable and irreversible process that cause variations in facial appearance, shape and texture. Although ageing is stochastic with different people having different ageing patterns, there are some general variations and similarities that can be modelled [78] [9]. There are two stages in human life that are distinct with regard to facial growth: formative or childhood stage and adulthood or old age stage [79].

Ageing introduce significant change in facial shape in formative years and relatively large texture variations with still minor shape variation in older age groups [3] [7]. Shape variations in younger age groups is caused by craniofacial growth. Craniofacial studies have shown that human face change from circular to oval as one age [80]. These change lead to variations in the positions of fiducial landmarks [74]. During craniofacial development, the forehead slopes back releasing space on the cranium. Eyes, ears, mouth and nose expand to cover interstitial space created. The chin become protrusive as cheeks extend. Facial skin remain moderately unchanged compared to shape. More literature on craniofacial development is found in [9].

As one age, facial blemishes like wrinkles, freckles and age-spots appear. Underneath the skin, melanin producing cells are damaged due to exposure to suns' Ultra Violet (UV) rays. Freckles and age-spots appear due to over production of melanin. Consequently, light reflecting collagen not only decreases but also become non-uniformly distributed making facial skin tone non-uniform [81]. Parts adversely affected by sunlight are upper cheek, nose, nose-bridge and forehead.

The most visible variations in adulthood to old age are skin variations exhibited

in texture change, although there is still minimal change in facial shape. Biologically, as the skin grow old, collagen underneath the skin is lost [3]. Loss of collagen and effect of gravity causes the skin become darker, thinner, leathery and less elastic. Facial spots and wrinkles appear gradually. Bone framework beneath the skin may also start deteriorating leading to accelerated development of wrinkles and variations in skin texture. More details about facial ageing in adulthood is found in [9]. These variations in shape and texture across age could be modelled and used to automatically estimate someone's age.

2.1.2 Factors affecting facial ageing

Facial ageing is affected by several factors ranging from life style, natural, occupation, psychological and environmental. Factors affecting facial ageing can be categorized as intrinsic and extrinsic. Extrinsic factors are those that are external to human body like environmental and occupation factors while intrinsic are internal factors like bone structure and genetic influence which occur naturally over time [81], [82]. In childhood, facial variation is mainly caused by craniofacial development leading to change in facial shape [9] due to growth, modelling and deposition of bony tissues in the face. This leads to change in height and shape of the face [83]. The forehead slopes back releasing space on the cranium. Drifting and expansion of facial landmarks to occupy this space cause variations in facial shape in childhood. In adulthood, facial ageing is mainly manifested in texture variation which is caused by wide variety of factors.

Taister *et al.* [83] found that general exposure to wind and arid air influence facial ageing. Arid environment and wind dehydrates the skin leading to wrinkle formation. Air pollution has also been found to affect ageing by accelerating wrinkle development [20] [21] [22]. Research on air pollution and ageing has shown that city dwellers who are exposed to air pollution from industries develop deep wrinkles than individuals who are not exposed to pollution. Smoking influence on ageing has also been cited in [83], [84], [85], [86] although [75] asserts that smoking has negligible effect to facial wrinkling as compared to effect of UV rays. However, smoking interrupts skin micro-vasculature which affects elastin and collagen production and functioning leading to wrinkles around the mouth but photoaging effects leads to more facial wrinkling compared to smoking [83], [75]. It is therefore evident that facial skin ageing does not provide objective analysis of cumulative exposure to UV rays. Taister *et al.* [83] also asserts that exposure to drug and psychological stress affects skin texture and color making skin complexion spotted and blemished.

Exposure to UV rays influences production of collagen making the skin darker. UV rays dries and destroys cells and underlying skin structure, giving the skin a furrowed and thickened appearance hastening development of wrinkles especially around the eyes due to squinting effects [87]. Long exposure to UV rays lead to variations in photoaging like skin wrinkling, elastosis, actinic keratosis and irregular pigmentation [88]. With long exposure to UV rays, skin texture and color change becoming blotchy, yellowish, leathery, loose, inelastic and hyper-pigmented. Blood veins close to skin surface become protrusive forming "spider veins" network in addition to overall speckled skin appearance [89]. Naturally, with lower production of collagen and elastin, the skin becomes leathery and less elastic. Loss of skin elasticity makes the skin leathery leading to formation of wrinkles [90]. Fat cells begin

to disappear leading to skin sagging. Fat deposits in some areas like eye lobe region also affect skin texture. Force of gravity makes the skin leathery and less elastic hence accelerating skin wrinkling. Internally, changes in bone structure and subsequent variations in musculature causes skin wrinkling [9]. Ageing was also found to be different between males and females with female faces tending to age faster compared to male faces [9].

Ageing in males and females share many common characteristics but there are some differences. Although it is generally acknowledged that females age faster compared to men, it is not yet clear whether these gender differences are caused by rate of ageing or sexual dimorphism [9]. Investigation into differences in ageing between males and females is necessary [91]. Difference in male facial ageing include manifestation of facial hair like beards, increased thickness, facial vascularity, sebaceous content, and potential differences in fat and bone absorption rates [92]. Development of deeper wrinkles around perioral region is high in women compared to men [92] since women skin has few appendages compared to men [93]. Some women look younger than their actual age, and have large lips and are genetically protected from wrinkle and gray hair development [94].

Other factors affecting perceived facial ageing include diet, genetic makeup, ethnicity (race), skin infections, and cosmetics. Cosmetics are generally used to hide perceived age of an individual by hiding wrinkles and age spots and brightening wrinkle shadows around eyes, mouth and nose regions [95]. Chen *et al.* [95] found that facial makeup significantly impacts age estimation. Guo and Wang [96], Nguyen *et al.* [97] investigated effect of facial expression in age estimation. By quantitative evaluations on Lifespan [98] and FACES [99] datasets, Guo and Wang [96] found that facial expression influences age estimation. Same findings were reported by Nguyen *et al.* [97]. Voelkle and Ebner [100] investigated the effect of age, gender and facial expression on perceived age. They found that facial expression influences age estimation with faces with happy facial expressions most underestimated. Some facial expressions like smiling, frowning, surprise, laughing may introduce wrinkle-like lines on some regions of the face like forehead, cheek-bone area, mouth region and nose-bridge regions. These wrinkle-like lines may be registered as wrinkles during age estimation hence having an impact on age estimation performance.

2.1.3 Application areas of age estimation

Characterizing variations in facial appearance across age has many significant real world applications. Computer based age estimation is useful in situations where one's age is to be determined. There are several application areas for age estimation including

2.1.3.1 Face recognition and verification

Age estimation can be used to improve accuracies of face recognition and verification. Face recognition and verification systems could be trained for various age-groups. During recognition phase, the age-group of the face is first estimated to determine the age-group based model to be used for face recognition or verification. This is because age separated face images of the same individual are different in both shape and appearance.

2.1.3.2 Age simulation

Characterization of facial appearance at different ages could be effectively used in simulating or modelling one's age at a particular point in time. By observing ageing patterns at different ages, unseen appearance could be simulated.

2.1.3.3 Electronic customer relationship management (ECRM)

ECRM [3] is the use of Internet-based technologies such as websites, emails, forums and chat rooms, for effective managing of distinguished interactions with clients and individually communicating to them. Customers in different ages may have diverse preferences and expectations of a product [101]. Therefore, companies may use automatic age estimation to monitor market trends and customize their products and services to meet needs and preferences of customers in different age groups. The problem here is how to acquire and analyze substantive personal data from all client groups without infringing on their privacy rights. With automatic age estimation, a camera can snap pictures of clients and automatically estimate their age in addition to collection of other demographic data.

2.1.3.4 Security and surveillance

Age estimation can be used in surveillance and monitoring of alcohol and cigarette vending machines and bars for preventing underage from accessing alcoholic drinks and cigarettes or restricting children access to adult websites and movies [49] [33]. Age estimation can also be significant in controlling ATM money transfer fraud by monitoring particular age group that is apt to the vice [3]. Age estimation can also be used to improve accuracy and robustness of face recognition hence improving homeland security. Age estimation can also be used in health-care systems like robotic nurse and doctors expert system for customized medical services. For instance, a customized Avatar can be automatically selected from a database for interacting with patients from various age groups depending on preferences.

2.1.3.5 Biometrics

Age estimation via faces is a soft biometric [102] that can be used to compliment biometric techniques like face recognition, fingerprints or iris in order to improve recognition, verification or authentication accuracies. Age estimation can be applied in age-invariant face recognition [67], iris, hand geometry and fingerprint in order to improve accuracy of hard (primary) biometric system [3].

2.1.3.6 Employment

Some government employment like military and police consider one's age as a requirement. Age estimation systems could be used to determine age of the recruits during recruitment process. It is also a policy of several governments that employee should retire after reaching a particular age. Age estimation systems could also play a significant role in finding if one has reached retirement age.

2.1.3.7 Content access

With the proliferation of diverse content in televisions (TV) and the Internet (Internet of Things (IoT)), age estimation can be used to control access to unwanted content to

children. A camera could be mounted on a TV to monitor people looking at it such that it switches off the TV if at a particular time unwanted content is streamed and people watching are children.

2.1.3.8 Missing persons

Age estimation role in age simulation go a step further in aiding identification of missing persons. Age simulation can be used to identify old people from their previous images for purposes of identification.

2.2 Image representation for age modelling

This section present different image representation approaches for age estimation. Age estimation can be modelled using anthropometric data, Active Appearance Model (AAM) parameters, Ageing Pattern Subspace (AGES), manifold learning, appearance features or a hybrid of two or more modelling technique. An overview of these modelling techniques is presented in the subsequent sections.

2.2.1 Anthropometric models

Anthropometric modelling of facial ageing focus on distance measurements between facial points. Face anthropometry is the study of measuring sizes and proportions on human faces [103]. Farkas [103] defined face anthropometry based on measurements taken from 57 landmark points on human faces. Figure 2.1 shows some of the points used to describe a face. Landmark points are identified by abbreviation of their respective anatomical names. For instance, eye inner corner is *en* for *endocanthion* while front of the ear is *t* for *tragion*.

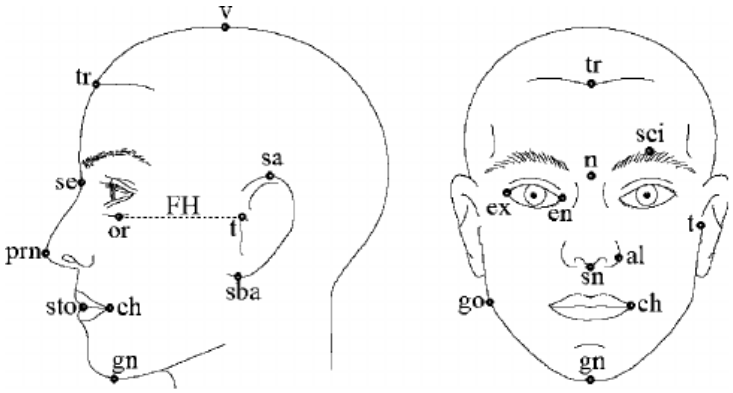


FIGURE 2.1: Anthropometric points on the face.
Source: [103]

Farkas defined five measurements between landmarks; *shortest distance, axial distance, tangential distance, angle of inclination* and *angle between locations*. Figure 2.2 show sample measurements of these distances.

A total of 132 facial measurements were defined by Farkas[103], where by some corresponding measurements on left and right of the face were paired. The measurements can be taken by hand by experienced anthropometrists or 3D scanners [104] [103] [105].

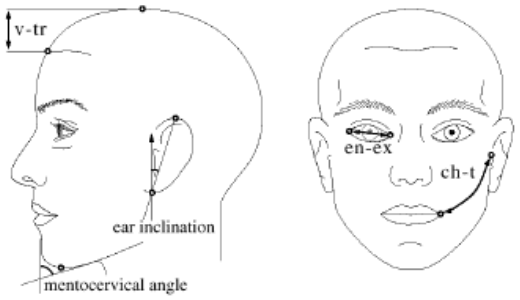


FIGURE 2.2: Sample of anthropometric measurements.
Source: [103]

These measurements could be taken at different ages for instance from childhood to old age. Ratios of distances between facial landmark like eyes, nose, mouth, ear, chin, and forehead are measured across age. These measurements are used to determine ageing pattern of an individuals at particular age and hence used to discriminant between ages and age groups. This approach embraces studies in craniofacial development theory [106].

Craniofacial development theory uses cardioid strain transformation mathematical model to describe a person's facial growth from infancy to adult age. This model defines a circle to track facial growth. By tracking variations in radius of the circle as

$$R' = R\,(1 + k\,(1 - \cos\theta)) \tag{2.1}$$

where R is the initial radius of the circle, θ is the initial angle formed with the vertical axis, k is a parameter that increases with time and R' is the successive growth of the circle over time. Figure 2.3 shows simulated face profiles using cardioidal strain transformations.

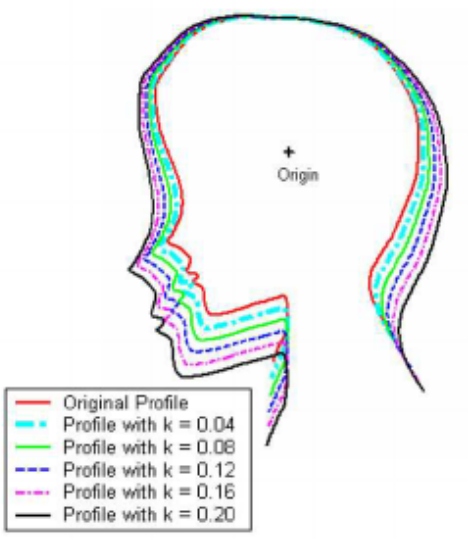


FIGURE 2.3: Simulation of facial growth using cardioidal strain transformations. Sequence proceeds from infancy (inner most profile) to adulthood (outermost profile).
Source: [107]

The mathematical formulation in equation 2.1 is not commonly used for age estimation because it does not encode head profile, especially in adults [108] and head profiles are hard to estimate from 2D facial images [3]. Furthermore, anthropometric models cannot be used for age modelling in adult and old age face images since there is no significant changes in facial shape at this stages. This approach is also only appropriate for frontal face images since distance between landmarks are sensitive to head poses. This modelling technique has not been experimented on a large publicly available database, with few studies reported in the literature working on small private datasets. Another limitation of this approach is that it only considers distance between facial landmarks with no consideration for facial appearance. Measurements and landmark points defined by Farkas in [103], which often guide anthropometric modelling are from people in one ethnic group (European) and may not be representative of all other races.

2.2.2 Active shape models

Active Shape Models (ASM) [68] is a statistical model that characterize shape of an object. ASM builds a model by learning patterns of variability from a training set of correctly annotated images. ASMs are able to capture natural variability of images of the same class unlike Active Contour Models (ACM) [109]. ASM are specific to images of the class of objects they represent. Face image shape is denoted by a collection of landmark points. Good choice for landmark points are points at clear corners of the face and facial landmark boundaries. These points can be determined by use of appropriate 2D landmarking algorithm like one proposed in [110]. The sets of points are automatically aligned to reduce the variance in distance between equivalent points. The number of landmark points must be adequate enough to show overall shape of the face images. Each face is then represented by a predefined number of landmark points depending on complexity of the facial shape and the desired level of descriptive information. A Point Distribution Model (PDM) is derived by examining spatial statistics of labeled points. PDM gives mean locations of points and a set of parameters that control main variability modes found in the training set.

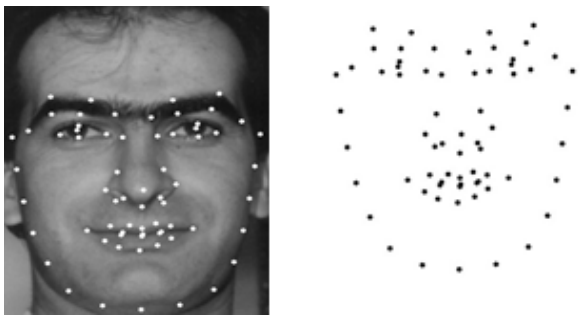
Given such a model and test image, image interpretation involves choosing values for each of the parameters such that the best fit of the model to the image is found. ASM allows initial rough guess of best shape, orientation, scale and position which is refined by comparing hypothesized model instance to image data and using difference between model and image to deform to shape. ASM is more similar to AAM but differ in the sense that instances in ASM can only deform in according to variations found in the training set. ASM is not commonly used in age estimation hence more investigation adopting this modelling strategy are necessary.

2.2.3 Active appearance model

Active Appearance Models (AAM) [38] are statistical facial image coding models. Using Principal Component Analysis (PCA), AAM learns shape model and intensity model from a set of training images. AAMs have been used extensively in modelling facial shape for face recognition, face verification, age estimation, and gender estimation among other tasks. AAM considers both facial shape and texture unlike anthropometric models that consider shape parameters only. This makes AAMs appropriate for age estimation modelling at all stages from infancy to old age. Labeling

each test image with a definite age label from continuous age range makes AAMs approaches give precise age estimations [3].

Annotated set of training images are marked with points defining facial main features are needed to build AAM. Figure 2.4 shows sample of annotated face and points used for annotation. These points can be determined by use of appropriate



(a) Annotated Face (b) Facial Points

FIGURE 2.4: Facial shape and appearance annotation

2D landmarking algorithm like one proposed in [110]. These set of points are represented as a vector and aligned before a statistical shape model built. Each training image is then warped so that the annotated points match points of mean shape and obtain a shape-free image patch. The shape-free raster is pushed into a texture vector, g, which is normalized by applying a linear transformation, $g \leftarrow \frac{(g-\mu_g 1)}{\sigma_g}$, where 1 is a vector of ones, μ_g and σ_g^2 are the mean and variance of elements of g. After normalization, $g^T 1 = 0$ and $|g| = 1$. Principal Component Analysis (PCA) is then used to build a texture model. Finally, connections between shape and texture are learned to produce a combined appearance model as detailed in [111].

The generated appearance model has parameters, c, controlling the shape and texture according to

$$\begin{aligned} x &= \bar{x} + Q_s c \\ g &= \bar{g} + Q_g c \end{aligned} \tag{2.2}$$

where $\bar{x}$ is the mean shape, $\bar{g}$ is the mean texture in a mean shaped patch, and Q_s, Q_g are matrices describing modes of variation derived from training set. AAM are slower compared to Active Shape Models (ASM) [68]. Details of AAM implementation could be found in [38].

Lanitis *et al.* [112] extended AAM by proposing and ageing function $age = f(b)$. In this function, age is the real subject's age, b is AAM-learned vector of 50 raw model parameters, and f is ageing function. The function f describes the association between an individual's age and vector of parameters.

AAM face encoding consider both shape and texture unlike anthropometric techniques that only represent shape. This makes AAM approaches appropriate for age estimation since both texture and shape features necessitate precise age estimation. However, evidence is needed to show that ageing patterns can be modelled as a

quadratic function and highlight effect of outliers in age estimation.

2.2.4 Ageing pattern subspace

Geng *et al.* [5] [36] proposed ageing pattern subspace (AGES) for automatic age estimation using appearance of face images. A series of individuals images arranged in temporal order make up ageing pattern. Ageing pattern is defined in [5] as

"...a sequence of personal face images sorted in time order."

All images in a pattern must come from the same individual and must be ordered by time. This ageing pattern is called a complete pattern if images at all ages for an individual are available or else it is referred to as incomplete pattern. AGES compensate missing ages by learning a subspace representation of one's images when modelling a series of a subject's ageing face. To estimate age, test image is positioned at each possible location in the ageing pattern to find a point that can best reconstruct it. Ageing subspace that minimizes reconstruction error determines age of the test image. Figure 2.5 show vectorization of ageing pattern with missing images in the ageing pattern vector marked with *m*. Available face images in the pattern (age 2, 5 and 8) are placed at their respective positions and ages at which images are not available their positions are left blank. After vectorization of the ageing pattern, face

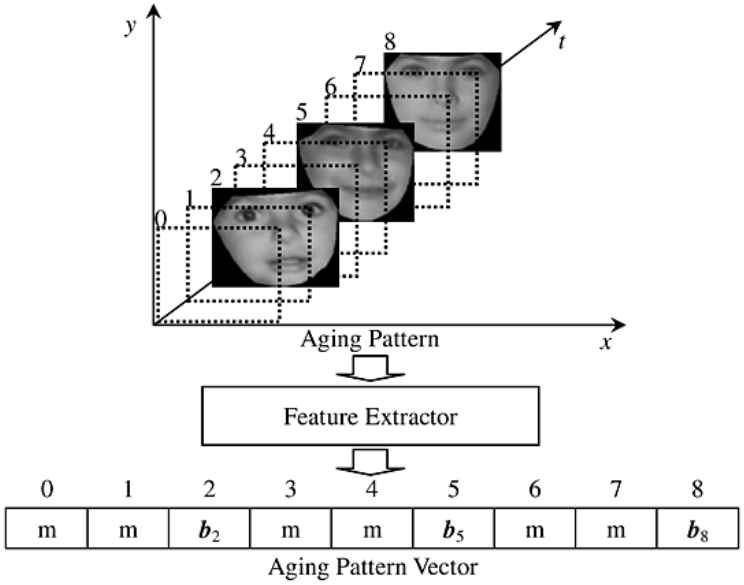


FIGURE 2.5: Ageing pattern vectorization. Age is marked at top-left corner of the corresponding feature.
Source: [5]

images at ages 2, 5 and 8 are represented by feature vectors b_2, b_5 and b_8 respectively.Representing ageing pattern using AGES ensures that label $age\,(I)$ and $id\,(I)$ are integrated into the data where by each pattern implies an ID, each age is fixed at a particular time-ordered position in the ageing pattern.

The first step of AGES is learning, where ageing pattern is learnt then followed by age estimation. Subspace representation is obtained in the learning stage using PCA. Due to possibility of missing age images, reconstruction error between available age and reconstructed face image is minimized by Expectation maximization (EM) iterative learning technique. Average of the available face images is used to initialize

values for missing faces. Thereafter, mean, covariance matrix and eigenvectors of all face images are computed. Faces are then reconstructed using mean face and eigenvectors. This process is repeated until the reconstruction error is significantly small. During age estimation, the test image finds ageing pattern subspace and position in that pattern that can minimize its reconstruction error. The position that gives minimal reconstruction error is returned as the estimated age of the probe image. Ghost-like twisted faces are reconstructed when test image is positioned at wrong location in the ageing pattern subspace [5] [36].

AGES was evaluated on FG-NET [26] and a MAE of 6.77 years was reported [36], [5]. This performance was superior to previously used approaches reported in literature. In AGES face images are first encoded with AAM. AGES undertakes existence of multiple images of the same person at various ages or ageing pattern of the face is similar in given training dataset. This assumption may not be satisfied in ageing datasets like Yamaha Gender and Age (YGA) [39]. Collecting face dataset with individuals' face images at several ages with some image quality may not be possible. AAM cannot encode wrinkles on the face since AAM only encode image gray values without spatial neighbourhood information for texture pattern calculation. Intensities of individual pixels cannot describe local texture. This affects applicability of AGES for age and age-group estimation since single pixel values cannot represent local texture. Techniques like Gabor Filter [69] may be appropriate to encode wrinkle features on elderly faces.

2.2.5 Age manifold

In age manifold, a common ageing pattern is learned from images of many individuals and different ages. Several face images are adopted to represent an age. Each subject may be represented by one image or several images at different ages. These images make a set referred to as a manifold which make up points in a high dimensional vector space. Age manifold learning face representation offers flexible means of face representation as compared to AGES [5]. Age manifold [40] can be used to learn ageing pattern by learning low-dimensional ageing pattern from several faces at every age. Individuals may have as low as one image at each age in the dataset which makes it simpler to collect enormous facial ageing dataset. Scherbaum *et al.* [113] proposed statistical age estimation using manifold learning on 3D morphable model. Isosurfaces of non-linear Support Vector Regression (SVR) function formed the manifold and ageing pattern was found by identifying a trajectory orthogonal to the isosurfaces. Discriminative subspace learning based on manifold criterion for low-dimensional representation of ageing manifold was proposed by Guo *et al.* in [49]. Coded face representation and age is learnt by applying regression on ageing manifold patterns. This approach consisted of two Support Vector Regression (SVR) with one used for rough age-group estimation followed by refined age estimation within the initially obtained age-group.

Given age-ordered image space $X = \{x_i : x_i \in \mathbb{R}^D\}_{i=1}^n$ with image dimension D and a vector $L = \{l_i : l_i \in \mathbb{N}^D\}_{i=1}^n$ of labels associated with the images in the image space, the objective is to learn a low-dimensional manifold in the embedded subspace, data distribution and its representation $Y = \{x_i : x_i \in \mathbb{R}^D\}_{i=1}^n$ with $d \leq D$, which is a direct mapping to X. Therefore, image space to manifold space projection can be modelled as $Y = P(X, L)$, where $P(\cdot)$ denotes the projection function which can be linear or nonlinear. Figure 2.6 show a simple nonlinear projection function

that model an image space into a 2D age manifold. Respective ages are shown on top-left corner of each image. Objective of manifold embedding is to find $n \times d$

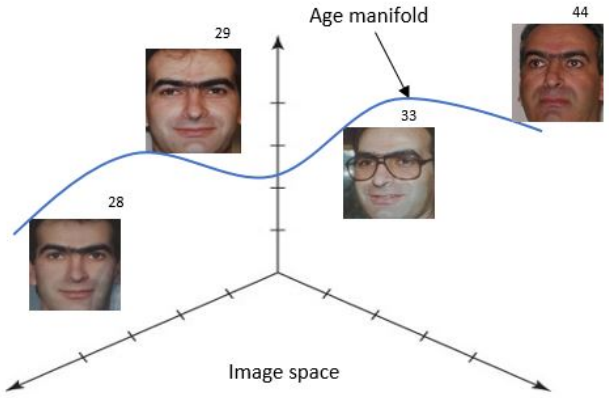


FIGURE 2.6: Simple nonlinear age manifold

matrix P that satisfies $Y = P^T XX$ or directly finds Y where $Y = \{y_1, y_2 \ldots, y_n\}$, $X = \{x_1, x_2 \ldots, x_n\}$, $P = \{p_1, p_2 \ldots, p_n\}$ and $d \leq n$. PCA, Locally Linear Embedding (LLE) and Orthogonal Locality Preserving Projections (OLPP) are examples of techniques used for dimensionality reduction and embedding manifold. PCA finds the embedding that maximizes the projected variance $P = \arg\max_{|p|=1} P^T p$ where $S = \sum_{i=1}^{n} (x_i - \bar{x}) (x_i - \bar{x})^T$ is the scatter matrix and $\bar{x}$ is the mean of vector $\{x_i\}_{i=1}^{n}$. LLE technique seeks a nonlinear embedding in a neighbourhood-preserving way by using local linear image class reconstruction symmetries while seeking local reconstruction optimal weights. Based on LPP, OLPP technique produces orthogonal basis functions [84], [85] to find additional discerning information for embedding. LPP looks for the embedding that will preserves essential manifold structure by measuring distance information in local neighbourhood. Affinity weights are defined as $s_{ij} = \exp\left(\frac{|x_i - x_j|^2}{t}\right)$ where x_i and x_j are k nearest neighbours of each other, otherwise $s_{ij} = 0$, and $(i, j) = s_{ij}$ is a symmetric matrix. LPP similarly defines diagonal matrix $D\,(i, j)$ and a Laplacian matrix $L = D - S$. LPP represents age manifold well and performs better in age estimation compared to traditional PCA.

There is a connection between age manifold and subspace analysis for ageing patterns. This technique finds embedded low-dimensional when each age is represented by many faces in the database. By using LPP for manifold embedding, age labels can be incorporated to the embedding process in a supervised manner which improves results compared to PCA embedding. Age manifold, Unlike AGES [5], does not learn subject-specific ageing pattern rather it uses all available ages from different individuals. However, age manifold requires large dataset in order to satisfactorily learn the embedded manifold.

Huang *et al.* [114] proposed a multi-manifold metric learning (MMML) for for face recognition based on image sets. In MMML, several person-specific distance metrics in different manifolds are learned by modelling each image set as a manifold minimizing intra-class variations and maximizing inter-class manifold variations. Figure 2.7 show the multi-manifold metric learning. MMML could be applied to age estimation by grouping images at the same age into one set and learn distance metrics between these sets. Each class (as shown in Figure 2.7) could consist of images at a

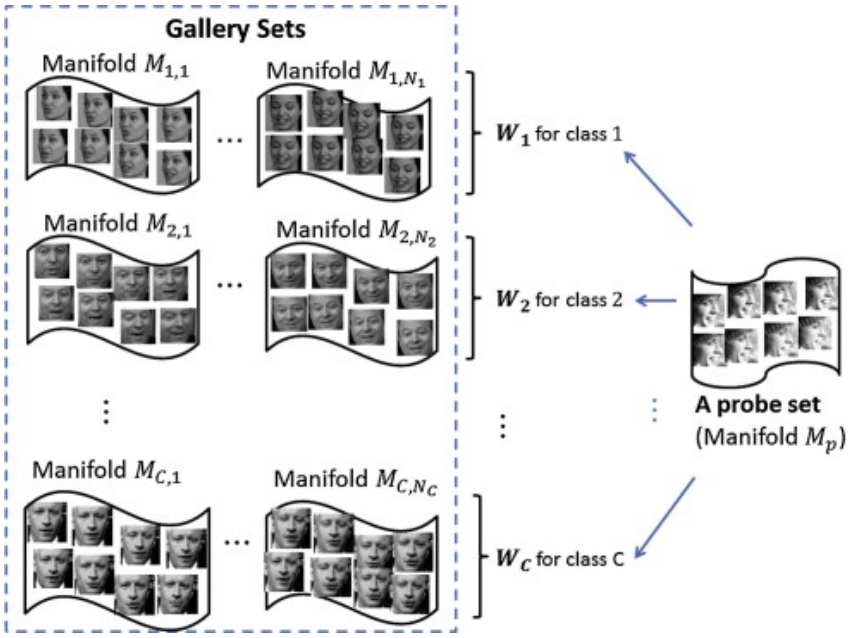


FIGURE 2.7: Multi-Manifold Metric Learning.
Source: [114]

particular age.

2.2.6 Appearance models

Appearance models mainly model facial appearance using texture, shape and wrinkle features for age estimation, face recognition, face verification and gender estimation among other tasks. Image is represented by vectoring both shape and texture [115]. Appearance models are more like AAM [38] that builds a statistical model using shape and texture of the face. Both global and local texture, shape and wrinkle features are extracted and modelled for age estimation. Texture and shape has been used for age and gender estimation [116] [117]. Age estimation using appearance features can be improved by performing gender estimation prior since males and females exhibit varied ageing patterns.

Given a set of facial images $X = \{x_i : x_i \in \mathbb{R}\}_{i=1}^{n}$ and a vector of age labels $X = \{l_i : l_i \in \mathbb{N}\}_{i=1}^{n}$ facial features are extracted from vector $\{x_i\}_{i=1}^{n}$ of images at a particular age. Every feature $\mathbf{F}_i$ has a one-to-one mapping with one of the age label l_i. After features are extracted and associated with age label, they are used for age estimation either using a regression model or classification. Effectiveness of LBP [46] in texture characterization has made it popular in extraction of appearance features for age estimation. LBP has been used in [118] and achieved 80% accuracy in age estimation with nearest neighbour classifier, and 80%-90% accuracy with AdaBoost classifier [119]. Gao and Ai [120] used Gabor filter [69] appearance feature extraction technique for age estimation and reported better results compared to LBP technique. BIF [70], [71] is also used in appearance-based models as used in [42]. Using age manifold, BIF and SVM classifier, MAE of 2.61 and 2.58 years for females and males respectively can be achieved on YGA database [3]. This shows BIFs superior performance in age estimation. Spatially Flexible Patch (SFP) proposed in [51] and [43] is another feature descriptor that can be used for characterizing appearance for age estimation. Other techniques that can be used to build appearance models for

age estimation are Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA). Detailed description of these techniques is presented in feature extraction techniques section.

2.2.7 Hybrid models

What is the best modelling approach for age estimation? It is hard to certainly answer this question since each of the modelling approaches discussed have their inherent strengths and limitations. To get the answer to the question, one may try different modelling approaches on the representative images and compare their performance. By comparing different modelling approaches, strengths and limitations of each of the models can be found. Modelling approaches that are complementary of each other can be combined to form a hybrid modelling approach. Hybrid age estimation modelling combines several modelling techniques to take advantage of strengths of each technique used. By combining different modelling techniques, age estimation accuracies are expected to not only improve but also be robust. These models could be combined in a hierarchical manner or parallel and results from different models combined for final age estimation.

2.3 Ageing feature extraction techniques

2.3.1 Gabor filters

Originally introduce by Denis Gabor in 1946 [69], Gabor filters has been extensively used for wrinkle, edge and texture feature extraction due to its capability of determining orientation and magnitude of wrinkles [8]. Gabor filter has been regarded the best texture descriptor in object recognition, segmentation, tracking of motion, and image registration [121]. Gabor features has been used in age estimation [37] and demonstrated to be effective texture descriptor compared to LBP. Since wrinkles appear as edge-like components with high frequency, Gabor edge analysis technique has been commonly used for wrinkle features extraction. Sobel filter [122], [123] Hough transform [116] and active contours [41] are among most commonly used texture edge descriptors. Though, edges in a face image also consist of noise such as beards, mustache, hairs and shadows. To reduce the effect of this noise, [8] proposes use of predominant orientation of wrinkles to be considered in wrinkle feature extraction. 2D spatial domain Gabor is defined as

$$g\left(x,y\right)=\left(\frac{1}{2\pi\sigma_x\sigma_y}\right)\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jWx\right] \tag{2.3}$$

where σ_x and σ_y are the standard deviations of the distribution along x and y axes respectively and W is the sinusoidal radial frequency. Fourier transform of the function in equation 2.3 is expressed as

$$G\left(u,v\right)=\exp\left[-\frac{1}{2}\left(\frac{\left(u-W\right)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}\right)\right] \tag{2.4}$$

where $\sigma_u=\frac{1}{2}\pi\sigma_x$ and $\sigma_v=\frac{1}{2}\pi\sigma_y$. Gabor filter bank is obtained by rotations and dilations of $g\left(x,y\right)$. The general equation for creating Gabor filter bank could be expressed as

$$g_b\left(x,y\right)=a^{-m}g\left(\bar{x},\bar{y}\right) \tag{2.5}$$

where $\bar{x} = x\cos\theta + y\sin\theta$ and $\bar{y} = -x\sin\theta + y\cos\theta$ where $\theta_k = \pi\frac{(k-1)}{n}, k = 1, 2, 3 \ldots n$ where n is the number of orientations used and a^{-m} is filter scale for $m = 0, 1, 2 \ldots S$ for S scales. Redundancy in the frequency domain is prevented by designing Gabor wavelets as

$$\sigma_u = \frac{\left(\left(\frac{U_h}{U_l}\right)^{\frac{1}{(s-1)}} - 1\right) U_h}{\left(\left(\frac{U_h}{U_l}\right)^{\frac{1}{(s-1)}} + 1\right)\sqrt{2\ln 2}}$$
$$\sigma_v = \tan\left(\frac{\pi}{2k}\right)\left[U_h - 2\ln\left(\frac{\sigma_u^2}{U_h}\right)\right]\left[2\ln 2 - \frac{(2\ln 2)^2\,\sigma_u^2}{U_h^2}\right]^{0.5} \tag{2.6}$$

where U_l and U_h denote lower and higher average frequencies respectively, and $W = U_h$. We refer readers to [121] and [124] for more details on Gabor wavelets.

2.3.2 Linear discriminant analysis

Linear Discriminant Analysis (LDA) [125] [126] is a feature extraction technique that searches for features that best discriminant between classes. Given a set of independent features, LDA creates a linear combination of these features such that largest mean differences between classes is achieved. LDA defines two measures; with-in class scatter matrix, given by

$$S_w = \sum_{j=1}^{c}\sum_{i=1}^{N_j}\left(x_i^j - \mu\right)\left(x_i^j - \mu_j\right)^T \tag{2.7}$$

where x_i^j is i^{th} sample of class j,μ_j is the mean of class j, c is number of classes and N_j is the number of samples in class j; and between-class scatter matrix, given by

$$S_b = \sum_{j=1}^{c}\left(\mu_j - \mu\right)\left(\mu_j - \mu\right)^T \tag{2.8}$$

where μ is the mean of all classes. LDA main objective is to maximize between-class scatter matrix while minimizing within-class scatter matrix.

One way of doing this is maximizing the ratio $\frac{det|S_b|}{det|S_w|}$. Given that S_w is non-singular, it has been proven [125] that this ratio is maximized when column vectors of projection matrix are the eigenvectors of $S_w^{-1}S_b$. S_w maximum rank is $N - c$ with N samples and c classes. This therefore requires $N = t + c$ samples to guarantee that S_w does not become singular, where t is the dimensionality of input data. The number of samples N is almost always smaller than t, making the scatter matrix S_w singular. To solve this problem [127] and [128] propose projecting input data to PCA subspace, to reduce dimensionality to $N - c$, or less, before applying LDA. PCA and LDA are widely used appearance feature extraction methods in pattern recognition [129]. Consequently, we adopt LDA for extraction of global face appearance features for age-group estimation.

2.3.3 Histogram of Oriented Gradients

Originally proposed by Dalal and Triggs [130], Histogram of Oriented Gradients (HOG) computes image gradient by counting occurrences of gradient orientations in local image regions. HOG uses 1D centred derivative mask $[-1, 0, +1]$ and $[-1, 0, +1]^{\mathbf{T}}$ for vertical (g_x) and horizontal (g_y) directions respectively to calculate image gradient of 64×128[1] image divided into blocks of 16×16 with 28×8 cells making up a block. The magnitude (g) and direction of the gradient (θ) are calculated as

$$g = \sqrt{g_x^2 + g_y^2}$$
$$\theta = \arctan \frac{g_y}{g_y} \tag{2.9}$$

where g_x and g_y are vertical and horizontal gradients of the image respectively, g is the gradient magnitude and θ is gradient direction. Figure 2.8 shows the x and y gradients, magnitude (g) and direction (θ) visualization of an input image.

(a) Input image (b) x-gradient (g_x) (c) y-gradient (g_y) (d) Magnitude (g) (e) Direction (θ)

FIGURE 2.8: Histogram of Gradients image responses

At every position (x, y), the image gradient has magnitude and direction. The gradients of colour images are evaluated for every colour channel. The gradient at point (x, y) is the maximum among magnitude of colour channels and the direction (angle) is one that correspond to the maximum gradient. Histogram of gradients is created in 8×8 cells. The resultant histogram has 9 bins corresponding to angles $0^o, 20^o, 40^o, 60^o \ldots 160^o$. Figure 2.9 shows a sample histogram for single 8×8 cell where **5** shows number of edges at orientation **120°**.

8.75	2			13		5		26.25
0°	**20°**	**40°**	**60°**	**80°**	**100°**	**120°**	**140°**	**160°**

FIGURE 2.9: Histogram of Gradients

If direction θ^o is between any two angles θ_1^o and θ_2^o shown in Figure 2.9, a voting scheme is employed to determine in which bin does θ^o fall. Generally, if there are n entries in direction matrix with θ^o where $\theta \mod 20 \neq 0$, the entries are shared

[1] Any dimensions can be used as long as the aspect ratio is maintained.

between θ_1^o and θ_2^o. The entry at θ_1 and θ_2 will be incremented by

$$\begin{aligned} \theta_1 &= 1 - \left(\frac{\|\theta_1^o - \theta^o\|}{20} \right) \times n \\ \theta_2 &= 1 - \left(\frac{\|\theta_2^o - \theta^o\|}{20} \right) \times n \end{aligned} \tag{2.10}$$

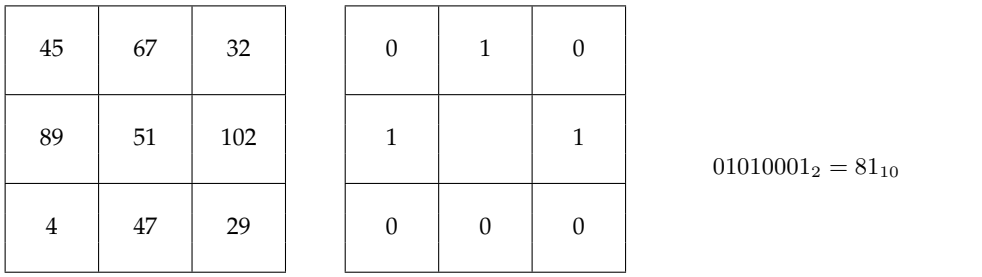
respectively. Sample of this voting scheme is shown in Figure 2.9 where it shows how 35 entries at orientation 165^o are share between 0^o and 160^o. These histograms are calculated for every 8×8 cell and resultant histograms are concatenated to make up the HOG feature vector. This vector is normalized and used for pattern recognition.

HOG is mostly used for object detection since it differentiates situations where a bright object is against a dark background or a dark object being against a bright background [131], making it mostly suitable for shape description rather than texture feature descriptor. Texture descriptors should capture both image edges and subtle textural information like spots for better surface characterization. HOG is a very high dimensional feature and it only count occurrences of gradient in localized portions of an image without considering the direction of the gradient. A gradient at orientation 0^o could be firing towards east of west with regard to reference pixel. It is important to encode all orientation, magnitude and direction of a gradient to make resultant feature vector more discriminative. Furthermore, HOG only calculates horizontal and vertical gradients of an image with the same weights for the edges. More discriminative image gradient is calculated using filters that consider more directions with different weights for each direction like Kirsch masks [132] which considers eight directions or four orientations.

2.3.4 Local binary patterns

Texture features have been extensively used in age estimation techniques [67]. Local Binary patterns (LBP) is a texture description technique that can detect microstructure patterns like spots, edges, lines and flat areas on the skin [46]. LBP is used to describe texture for face recognition, gender classification, age estimation, face detection, face and facial component tracking. Gunay and Nabiyev [107] used LBP to characterize texture features for age estimation. They reported accuracy of 80% on FERET [118] dataset using nearest neighbour classifier and 80 – 90% accuracy on FERET and PIE datasets using AdaBoost classifier [119]. Figure 2.10 show sample 3 x 3 LBP operation. Concatenating all 8 bits gives a binary number. Resulting binary number is converted to a decimal and assigned to center pixel as its LBP code.

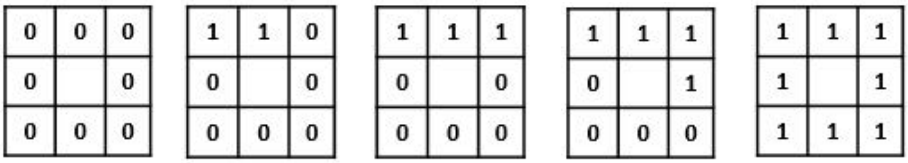
Ojala *et al.* [45] found that when using 8 neighbours and radius 1, 90% of all patterns are made up uniform patterns. The original LBP operator had limitation in capturing dominant features with large scale structures. The operator was latter extended to capture texture features with neighbourhood of different radii [45]. A set of sampling pixels distributed evenly along circle circumference centred at the pixel to be labelled define the neighbourhood. Bilinear interpolation of points that do not fall within the pixels is done to allow any radii and any number of sampling pixels.



(a) Sample image region (b) Thresholding (c) Resultant LBP code

FIGURE 2.10: LBP operation with P = 8 R = 1

Uniform patterns may represent micro-structures as line, spot, edge or flat area. Figure 2.11 show micro-structure pattern representation.



(a) Spot (b) Line end (c) Corner (d) Edge (e) Flat/Spot

FIGURE 2.11: Microstructure Pattern LBP code with P = 8 R = 1

Ojala *et al.* [46] further categorized LBP codes as uniform and non-uniform patterns. LBP pattern with at most two bitwise transition from 0 to 1 or 1 to 0 is categorized as a uniform pattern. For instance, $00000000, 00010000$ and 11011111 patterns are uniform while $01010000, 11100101,$ and 10101001 are non-uniform patterns. Figure 2.12 show LBP codes for all uniform patterns in $LBP_{(8,1)}$ neighbourhood.

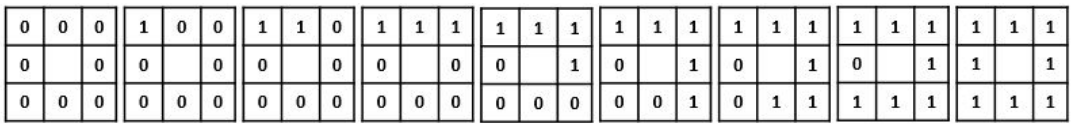


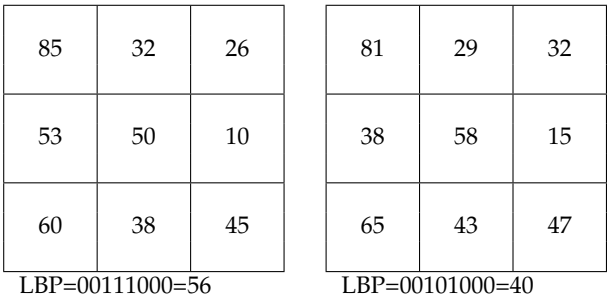
FIGURE 2.12: Example of LBP codes that represent uniform patter with P = 8, R = 1

In order to extract rotational invariant features using LBP, the generated LBP code is circularly rotated until its minimum value is obtained [133].

Extended LBP operator could capture more texture features on an image but still it could not preserve spatial information about these features. Ahonen *et al.* [134] proposed a technique of dividing a face image into n cells. Histograms are generated for each cell then concatenated to a single spatial histogram. Spatial histogram preserves both spatial and texture description of an image. Image texture features are finally represented by histogram of LBP codes. LBP histogram contains detailed texture descriptor for all structures on the face image like spots, lines, edges and flat areas.

2.3.5 Local directional pattern

LBP [44] was found to be unstable to image noise and variations in illumination. Jabid *et al.* [47] proposed Local Directional Pattern (LDP) which is robust to image noise and non-monotonic variations in illumination. Figure 2.13 show robustness of LDP operator to noise compared to LBP



(a) Original image (b) Noisy image

FIGURE 2.13: Robustness of LDP compared to LBP

LDP computes 8-bit binary code for each pixel in the image by comparing edge response of each pixel in different orientations instead of comparing raw pixel intensities as LBP. Kirsch edge detector [132], Prewitt edge detector [135] and Sobel edge detector [136] are some of the edge detectors that can be used [137]. Among them, the Kirsch edge detector has been known to detect different directional edge responses more accurately than others because the Kirsch edge detector considers all eight neighbours [138].

2.3.5.1 Kirsch edge detector

Kirsch is first-order derivative edge detector that get image gradients by convolving 3×3 image regions with a set of masks. Kirsch defined a nonlinear edge detector technique as [137]

$$P(x,y) = \max\left\{1, \max_{k=0}^{7}\left[|5S_k - 3T_k|\right]\right\} \tag{2.11}$$

where

$$S_k = P_k + P_{k+1} + P_{k+2}$$

and

$$T_k = P_{k+3} + P_{k+4} + P_{k+5} + P_{k+6} + P_{k+7}$$

where $P(x,y)$ is the Kirsch gradient, a in k_a is evaluated as $a = a\%8$ and P_k $[k = 0, 1, 2 \ldots, 7]$ are eight neighbouring pixels of $P(x,y)$ as shown in Figure 2.14.
The Kirsch gradient in a particular direction is found by convolving 3×3 image region with the respective mask M_k.

Figure 2.15 shows Kirsch edge detector response masks (kernels) for 8 orientations. As in [138], 4 directional edge-features could be computed by taking the maximum response of each of the two responses for each direction. The orientations of these edges are defined as horizontal (H), vertical (V), left-diagonal (LD) and right-diagonal

P_3	P_2	P_1
P_4	$p(x,y)$	P_0
P_5	P_6	P_7

(a) Eight neighbours of pixel p(x, y)

M_3	M_2	M_1
M_4		M_0
M_5	M_6	M_7

(b) Eight directional Kirsch mask positions

FIGURE 2.14: (a) Eight neighbours of pixel p(x, y) and (b)corresponding Kirsch mask positions

$$\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} \quad \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$$

(a) East M_0 (b) North East M_1 (c) North M_2 (d) North West M_3

$$\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}$$

(e) West M_4 (f) South West M_5 (g) South M_6 (h) South East M_7

FIGURE 2.15: Kirsch edge response masks in eight directions

(RD) and could be defined as

$$\begin{aligned} H &= \max\{|5S_0 - 3T_0|, |5S_4 - 3T_4|\} \\ V &= \max\{|5S_6 - 3T_6|, |5S_2 - 3T_2|\} \\ LD &= \max\{|5S_3 - 3T_3|, |5S_7 - 3T_7|\} \\ RD &= \max\{|5S_1 - 3T_1|, |5S_5 - 3T_5|\} \end{aligned} \tag{2.12}$$

This way, each direction is encoded in the final image representation. These responses could be combined and use for ageing pattern recognition.

2.3.5.2 Encoding an image with LDP

Given a center pixel in an image $P(i, j)$, 8-directional responses are computed by convolving the neighbouring pixels, 3×3 image region, with each of the Kirsch masks. For each center pixel, there will be 8 directional response values. Presence of an edge or a corner will show high (absolute) response values in that particular direction. The interest of LDP is to determine k significant directional responses and set there corresponding bit-value to 1 and set the rest of $8 - k$ bits to 0. These binary bits are converted to decimal and assigned to the center pixel. This process is

repeated for all pixels in an image to obtain LDP representation of the image. Figure 2.16 show process of encoding an image using LDP operator.

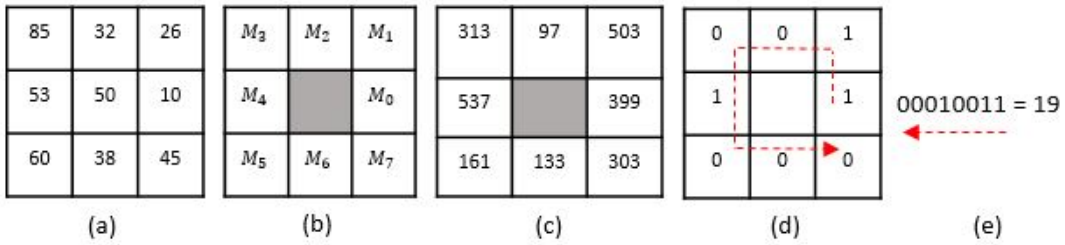


FIGURE 2.16: Process of encoding an image with LDP operator (a) Image region. (b) Kirsch masks as presented in Figure 2.15. (c) Result of convolving each pixel in (a) with 8 Kirsch masks. (d) Pick top $k = 3$ significant responses, set there corresponding bit to 1 and the rest to 0 (e) Resultant LDP code

Given an image region as shown in Figure 2.16(a), LDP response in east direction is obtained by convolving the 3×3 image region with the *East* M_0 mask shown in Figure 2.15 top-left corner as

$$\begin{aligned} M_0 = &(85 \times -3) + (32 \times -3) + (26 \times 5) + (10 \times 5) + \\ &(45 \times 5) + (38 \times -3) + (60 \times -3) + (53 \times -3) \\ = &-399 \end{aligned} \tag{2.13}$$

The absolute values of the directional responses are arranged in descending order. For $k = 3$ significant responses, the binary response bit for each of the 8 neighbouring pixels shown in Figure 2.16 (d) is calculated as

$$\begin{aligned} LDP_k &= \sum_{i=0}^{i=7} b_i \left((m_i - m_k) \times 2^i \right) \\ b_i(a) &= \begin{cases} 1, & \text{if } a \geq 0 \\ 0, & \text{if } a < 0 \end{cases} \end{aligned} \tag{2.14}$$

where m_k is the k^{th} significant directional response, for instance, in the example in Figure 2.16 $m_k = 399$ and m_i is response of Kirsch mask M_i.

For $k = 3$, LDP operator generates $C_3^8 = \frac{8!}{3! \times (8-3)!} = 56$ distinct patterns in the LDP encoded image. A histogram $H(i)$ with C_k^8 bins can be used to represent the input image of size $M \times N$ as

$$\begin{aligned} H(i) &= \sum_{m=0}^{M} \sum_{n=0}^{N} f\left(LDP_k(m,n), i\right) \\ f(p,i) &= \begin{cases} 1 & \text{if } p = i \\ 0 & \text{if } p \neq i \end{cases} \end{aligned} \tag{2.15}$$

where $f(p,i)$ is a logical function that compare if the LDP code at location $p(m,n)$ of the LDP-encoded image is equal to current LDP pattern i for all i in the range $0 \leq i \leq C_k^8$. The resultant histogram has dimensions $1 \times C_k^8$ and is used to represent the image. The resultant feature has spots, corners, edges and texture information about the image [139]. The limitation of LDP with $k = 3$ is that it uses responses of at most 3 directions out of the possible 8 directions. These directional responses could possibly be one sided as South-East, East and North-East. The eight directional responses could be paired as in [138] (as shown in Equation 3.32) and guarantee that each directional response is likely to be used to determine the image gradient.

2.3.6 Local ternary patterns

LBP is sensitive to noise and illumination especially in nearly uniform image blocks. Local Ternary Patterns (LTP) [48] seeks to improve robustness of image features in a fairly uniform region. LTP extends LBP to a 3-value code by comparing pixel values of the neighbouring pixels with a preset threshold value τ. Values that lie within $\pm\tau$ are set to 0, values above τ are set to $+1$ while values below τ are set to -1. The thresholding function is defined as

$$f(x_i, x_c, \tau) = \begin{cases} 1 & \text{if} \quad x_i \geq x_c + \tau \\ 0 & \text{if} \quad |x_c - x_i| < \tau \\ -1 & \text{if} \quad x_i \leq x_c - \tau \end{cases} \tag{2.16}$$

where τ is a preset threshold, x_c is the value of the central pixel and x_i for $i = 0, 1, 2 \ldots 7$ are the neighbouring pixels of x_c. Although this extension makes LBP robust to noise and encode more patterns, it is not easy to practically select an optimum τ for all images in a dataset or for all datasets and the resultant code is not invariant to pixel value transformations. LTP can encode 3^8 patterns. The LTP codes are split into its positive and negative parts and two histograms are generated, one for the negative part and the other for the positive part. These histograms are concatenated and used as feature descriptor for pattern recognition. Figure 2.17 shows LTP codes for a 3×3 sample image region. LTP could be combined with LDP to

85	32	71
53	50	47
64	38	15

(a) Original image

1	-1	1
0	x	0
1	-1	-1

(b) LTP code $\tau \pm 5$

0	1	0
0	x	0
0	1	1

(c) Negative LTP code $\tau \pm 5$

1	0	1
0	x	0
1	0	0

(d) Positive LTP code $\tau \pm 5$

FIGURE 2.17: LTP code with $\tau = \pm 5$ and corresponding positive and negative LBP codes

make it robust to noise and illumination. This could be done by comparing edge responses rather than comparing raw pixel values. The threshold τ could also be made dynamic by deriving it from the weights of the neighbouring pixels rather

than setting it to a static value that could not be optimum to all images in a dataset or to all datasets under experiment.

2.3.7 Gray Level Co-occurrence Matrix

Statistical moments of histogram intensities of an image are commonly used to describe texture of an image [140]. Use of histograms to describe texture results to texture descriptors that convey information about gray level intensity distribution with no spatial relative information of pixel with each other. Haralick *et al.* [141] introduced Gray-Level Co-occurrence Matrix (GLCM) back in 1973.

GLCM describes image texture by comparing each pixel with its neighbouring pixel at a specified distance and orientation. This technique extracts second order statistical texture features from gray scale images. GLCM is a square matrix whose rows and columns is equal to the number of quantized gray levels, N_g. The entry $p\left(i,j\right)$ is the second order statistical probability for changes between gray level values i and j at a particular distance d and orientation θ.

Suppose there is an $N \times N$ image $I\left(i,j\right)$, with N_x columns and N_y rows. N_g is quantization of gray level appearing at each pixel in the image. Let the rows of the image be $N_y = (1,2,\ldots N_y)$, the columns be $N_x = (1,2,\ldots N_x)$, and set of N_g quantized gray levels be $G_x = (1,2,3\ldots N_{g-1})$. The image can be represented as a function that assigns some gray level in G to each pixel or pair of coordinates in $L_y \times L_x$;$G \leftarrow L_y \times L_x$. Texture information is specified by GLCM matrix of relative frequencies $C\left(i,j\right)$. The value at $GLCM\left(i,j\right)$ represent the number of occurrences of gray level value i at reference pixel and gray level value j at a neighbour pixel, a certain distance d and orientation θ^o. The probability measure can be defined as

$$P_{d,\theta} = p\left(i,j\right) \tag{2.17}$$

where $p\left(i,j\right)$ is defined as

$$p\left(i,j\right) = \frac{GLCM\left(i,j\right)}{\sum_{i=0}^{N}\sum_{j=0}^{N} GLCM\left(i,j\right)} \tag{2.18}$$

The sum in the denominator represent total number of gray level pairs (i,j) within the image and is bounded by $N_g \times N_g$. Dividing every pixel in the GLCM matrix with the denominator results into a normalized GLCM matrix. Figure 2.18 show an example of calculating GLCM from an image region at distance 1 and angle $\theta = 0^o$ and Figure 2.19 show an example of calculating GLCM from an image region at distance 1 and angle $\theta = 45^o$ The orientation of neighbour pixel from reference pixel can be $\theta = (0^o, 45^o, 90^o, 135^o)$ and distance can vary from $d = (1,2,3\ldots n)$ where n is any reasonable distance bounded by M_x and M_y.

Haralick *et al.* [141] defined 14 statistical features that can be used to describe texture. Table 2.1 show some of the Haralick features used for texture description [142].

where $\mu_x = \sum_i \sum_j ip\left(i,j\right), \mu_y = \sum_i \sum_j jp\left(i,j\right), \sigma_x = \sqrt{\sum_i \sum_j \left(i-\mu_x\right)^2 p\left(i,j\right)}$ and $\sigma_y = \sqrt{\sum_i \sum_j \left(j-\mu_x\right)^2 p\left(i,j\right)}$

Harlick features have been successfully used in brain tumor classification [143],texture description [144], remote sensing [145] among other fields. GLCM has not been

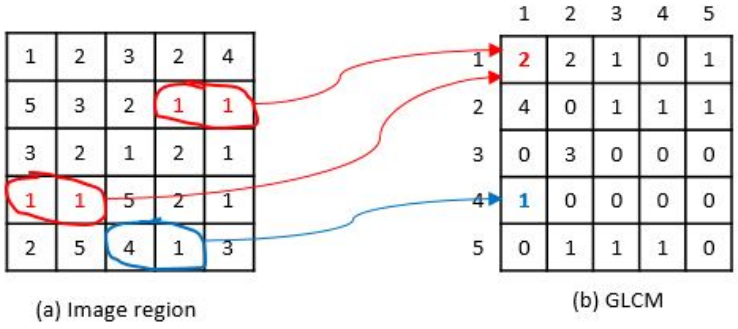


FIGURE 2.18: GLCM calculation with d = 1, $\theta = 0^o$

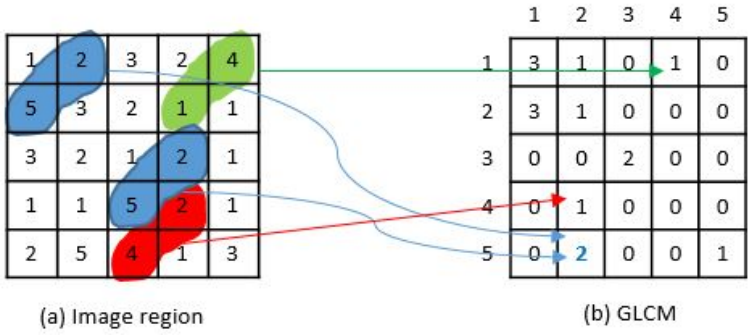


FIGURE 2.19: GLCM calculation with d = 1, $\theta = 45^o$

investigated in ageing feature extraction. Haralick features like homogeneity, variance, correlation could be extracted from age separated faces and used for age estimation.

2.3.8 Spatially flexible patch

Spatially Flexible Patch (SFP) proposed in [51] and [43] is another feature descriptor that can be used for feature extraction for age estimation. SFP is effective for capturing local variations in facial appearance as one ages. SFP encodes local appearance and its spatial information. SFP solves the problem of local variations in appearance during ageing since SFPs similar in appearance and slightly different in position can provide similar confidence for age estimation. By considering local patches and their spatial information, SFP can effectively characterize facial images with slight disorientation, occlusion and head pose disparities. Another advantage of SFP is that it alleviates the problem of insufficient samples by enriching the discriminating characteristics of the feature vector.

2.3.9 Grassmann manifold

Grassmann manifold is the space $G(k,n)$ of all k-planes through the origin in $\mathbb{R}^n$,$k \leq n$, generalize real projective spaces [146]. Consist of set of all k-dimensional subspaces of $\mathbb{R}^n$. To each k-plane v in $\mathbb{R}^n$, a matrix $n \times k$ can be associated with orthogonal matrix Y such that columns of matrix Y form an orthonormal basis vectors that span the same subspace. Therefore, each k-plane v in $G(k,n)$ is connected with a correspondence class of $n \times k$ matrices YR in $\mathbb{R}^{n\times k}$, for $\mathbb{R} \in SO(k)$, where Y is an orthonormal basis for the k-plane. $G(k,n)$ is not a vector space, but points on $G(k,n)$ can be projected onto the tangent space at mean-point and standard vector-space methods can be used on tangent space. Geodesic distance between points on

TABLE 2.1: Summary Haralick features

Feature Name	Feature Description	Feature Formula
Angular second moment (ASM)	Shows how uniform a texture is by measuring local homogeneity	$ASM = \sum_i \sum_j p(i,j)^2$
Energy (E)	Measure homogeneity	$E = \sqrt{\sum_i \sum_j p(i,j)^2}$
Contrast (C)	Shows variation in texture	$C = \sum_i \sum_j \lvert i-j \rvert^2 p(i,j)$
Dissimilarity (D)	Variation in texture	$D = \sum_i \sum_j \lvert i-j \rvert p(i,j)$
Homogeneity (H)	Uniformity of non-zero entries	$H = \sum_i \sum_j \frac{1}{1+(1-j)^2} p(i,j)$
Entropy (En)	Spatial disorder of texture	$En = \sum_i \sum_j p(i,j) \log(p(i,j))$
Correlation (Cr)	Linear relationship of texture	$Cr = \sum_i \sum_j p(i,j) \frac{(i-\mu_x)(j-\mu_y)}{\sigma_x \sigma_y}$
Auto correlation (ACr)	Measure repeating patterns	$ACr = \sum_i \sum_j (i \cdot j) p(i,j)$
Variance (V)	Measure of texture heterogeneity	$V = \sum_i \sum_j (i-\mu_x)^2 \cdot p(i,j) + \sum_i \sum_j (i-\mu_y)^2 \cdot p(i,j)$
Cluster shade (Cs)	Measure perceptual uniformity	$Cs = \sum_i \sum_j (i+j-\mu_x-\mu_y)^3 p(i,j)$
Cluster prominence (Cp)	Measure image symmetry	$Cp = \sum_i \sum_j (i+j-\mu_x-\mu_y)^4 p(i,j)$
Maximum probability (Mp)	Maximum co-occurrence	$Mp = \max p(i,j)$

the manifold are used for classification or regression problems. Wu [14] used Grassmann manifold tangent-space regression approach for age estimation.

Grassmann manifold can be used in age estimation by representing each face by a deformation that warps an average face to a given face. This requires defining what an average face is and how to quantify the deformation between the average face and the given face. Average face can be represented by computing a mean point from all the (landmark) points on $G(k,n)$. This can be done by calculating Karcher mean [147]. Age estimation can be performed using Grassmann Nearest Neighbor (GNN) classification approach. In GNN, Karcher mean is computed for every age. During testing, compare the Karcher mean of the probe image with mean of every age using one defined distance on Grassmann manifold. The closest mean to the probe gives the target age.

2.3.10 Biologically inspired features

BIF were first proposed in 1999 by Riesenhuber and Poggio (R & P Model) [70]. These BIF features are derivative of primates feed-forward model of visual object recognition pipeline, referred to as, HMAX model [148]. Primates are known to be able in recognizing visual patterns with high accuracy. Recent studies in computer vision and brain cognition show that Biologically Inspired Models (BIM) improve face identification performance [149], object recognition [150] and scene classification [151]. Visual cortex application in age estimation tasks saw some improvement in age estimation accuracies.

BIF were first proposed in 1999 by Riesenhuber and Poggio (R & P Model) [70]. These BIF features are derivative of primates feed-forward model of visual object recognition pipeline, referred to as, HMAX model [148]. The visual model of primates contains alternating layers of simple (S) and complex (C) cell units. Complexity of these cells increase as layers advance from primary visual cortex (V1) to inferior temporal cortex (IT). In primary visual cortex, S units use a bell-shaped tuning function to combine input intensities to increase scale and orientation selectivity. Using MAX, STD, AVG or any other pooling operation, C units pool inputs from S units, thereby introducing gradual invariance to scale, rotation and translation.

Gabor functions [72] [152] are used to model simple cells (S) in the visual cortex of mammalian brains. Frequencies and orientation illustration in Gabor filters are alike to frequencies and orientations in human visual system. It is therefore thought that Gabor filter image analysis is similar to perception in visual system of humans. BIF have demonstrated success in age estimation tasks [52] [42] [153]. BIF feature extraction encompass two layers of computational units with simple cell units (S1) in layer one followed by complex cells units (C1) in the subsequent layer.

S1 Unites-Simple Cells: They represent the receptive field in primary visual cortex (V1) [72] which has basic attributes of multi-orientation, multi-frequency and multi-scale selection [73]. SI units are commonly described by a bank of Gabor filters [71]. Gabor filters are appropriate for modelling of cortical simple cell receptive fields. 2D spatial domain Gabor is defined as

$$G(x,y) = \exp\left(-\frac{X^2+\gamma^2Y^2}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda}X\right) \tag{2.19}$$

where $X = x\cos\theta + y\sin\theta$ and $Y = -x\sin\theta + y\cos\theta$ are angle of rotations of Gabor filters, θ varies from 0 to π, γ and σ are aspect ratio and standard deviation of the Gaussian envelop respectively and λ is the wavelength and determines spatial frequency $1/\lambda$.

Useful discriminating features are extracted using Gabor filters with different orientation and frequencies [64]. Consequently, previous studies [63] [64] suggest that spatial frequency processing is done in primary visual cortex. Spatial frequency analysis extracts discriminative features that are more robust to distortions [154]. Daugman [65] found that visual system in primates extracts information both in 2D spatial and frequency domains, and Shapley [38] proved that spatial frequency analysis help the brain understand an image.

C1 Units-Cortical Complex Cells: These units receive responses from S1 units and perform linear feature integration. C1 units represent complex cells that are shift invariant. Lampl *et al.* [155] proposed that spatial integration of complex cell in visual cortex that can be described by a series of pooling operations. Riesenhuber and Poggio [70] demonstrated merits of using MAX pooling operator compared to SUM while Guo *et al.* [42] showed that standard deviation (STD) pooling operator outperforms MAX operator. Cai *et al.* [73] improved on STD by using a cell grid of 4×4 in normalization. The MAX operator returns maximum values at each index i of the two consecutive scale features. Given a feature at scale S_x and scale S_{x+1}, the maximum value F_i at index i is given by

$$F_i = \begin{cases} S_x^i, & \text{if } S_x^i \geq S_{x+1}^i \\ \\ S_{x+1}^i, & \text{if } S_{x+1}^i > S_x \end{cases} \tag{2.20}$$

where S_x^i and S_{x+1}^i are the filtered values at the position i of features from scale x and $x+1$ respectively.

Guo *et al.* [42] defined the STD operator to incorporate mean of values in a particular neighbourhood. The STD operator was defined as:

$$STD = \sqrt{\frac{1}{n_s \times n_s}\sum_{i=1}^{n_s \times n_s}\left(F_i - \bar{F}\right)} \tag{2.21}$$

where maximum value at index i between two consecutive S_1 scales is represented by F_i, $\bar{F}$ is the mean mean of filtered values within $n_s \times n_s$ neighbourhood. Given two $N \times N$ features at scales S_x and S_{x+1}, STD operator with $n_s \times n_s$ grid returns $\lfloor N/n_s \rfloor \times \lfloor N/n_s \rfloor$ features. STD operator captures local texture and wrinkle variations which are significant for subtle age estimation.

Serre *et al.* [71] [156] extended the HMAX model [70] to include two layers; S_2 and C_2 for object recognition. In S_2, template matching is done to match patches from C_1 layer with some pre-learned patches extracted from images. S_2 layer get more selective intermediate features capable of discriminating between object classes. The S_2 units are convolved over entire image and maximum response value of S_2 are assigned to C_2 units. Mutch and Lowe [157] extended model in [71] by reducing number of output units in S_1 and C_1 and picking features that are highly weighted by support vector machines (SVM) [158].

2.4 Age estimation algorithms

Once ageing features are extracted and represented, the subsequent phase is age estimation. Age estimation is a special patter recognition task where age labels can be viewed as a class or a set of sequential value. When age labels are viewed as classes, age estimation is approached as a classification problem whereas when age labels are viewed as sequential chronological series, regression approach is used for age estimation. Hybrid approach can also be employed for age estimation where both classification and regression techniques integrated, mostly hierarchically, to find the relationship between extracted feature vectors and age labels. An analysis of existing approaches is presented and suggestions of effective age estimation.

2.4.1 Classification

Lanitis *et al.* [33] explored performance of Nearest Neighbor, Artificial Neural Network and Quadratic function in age estimation tasks. Although the quadratic function used to relate face representations to face labels is a regression function, the authors referred to it as a quadratic function classifier [33]. The quadratic function reported MAE of 5.04, which was superior to MAEs reported by Nearest Neighbor. ANN and Self-Organizing Maps (SOM) reported better performance compared to quadratic function. The authors proposed clustering and hierarchical age estimation for improving performance. The error rates in the extended techniques reduced although evaluations were done on small datasets. Comparison between humans and computers in age estimation was also done and found that computers can estimate age almost as reliable as humans.

Ueki *et al.* [50] built eleven Gaussian models in low-dimensional 2DLDA and LDA feature space using Expectation Maximization (EM). Age group estimation was determined by fitting probe image to each cluster and comparing the probabilities. They reported a higher accuracy, 82% male, 74% female, with wide age-groups of 15 years as compared to 50% male, 43% female in age-groups of five-year range. This demonstrates that this approach can only post better accuracies where age groups have wide ranges and hence not applicable in a narrow-range age group estimation.

Fusing texture and local appearance, Huerta *et al.* [159] used a deep learning classification for age estimation. Using LBP [45], Speeded-Up Robust Features (SURF) [160], Histogram of Oriented Gradients (HOG) [161], he evaluated performance of deep learning on two large datasets and achieved MAE of 3.31. Hu *et al.* [162] used Kullback-Leibler / raw intensities for face representation before using Convolutional Neural Network (CNN) for age estimation. Their approach achieved

MAE of 2.8 on FG-NET and 2.78 on MORPH II. This demonstrates that deep learning (Deep Neural Networks or CNN) achieves better MAE compared to traditional classification methods.

2.4.2 Regression

Using 50 raw model parameters, Lanitis *et al.* [112] investigated linear, quadratic and cubic formulation of ageing function. Genetic algorithm is used to learn optimal model parameters from training face images of different ages. Quadratic and cubic ageing function achieved better MAE 0.86 and 0.75 respectively compared to 1.39 of linear function. This suggest that quadratic function offers best alternative since its Mae was not significantly different from that of cubic function and it is not computationally intensive as cubic function. Guo *et al.* [49], [57] used linear Support Vector Regression (SVR) on age manifold for age estimation. They reported MAE of 7.47 years and 7.00 years for males and females respectively on YGA dataset and MAE of 5.16 on FG-NET dataset. Yan *et al.* [54] formulated regression problem for age estimation using Semidefinite Programming (SDP). The regressor was learnt from uncertain nonnegative labels. They reported MAE of 10.36 and 9.79 years for males and females on YGA. They further demonstrated that age estimation by SDP formulation achieves better results compared to ANN. The limitation of SDP is that it is computationally expensive especially when training set is large.

Nguyen *et al.* [163] used a regression model for age estimation. The face image was represented by a Multi-Level Local Binary Pattern (MLBP). Their approach achieved a MAE of 6.6. Guo and Mu 2013 [153] achieved MAE of 4.0 by use BIF to model a regression model for age estimation. Using manifold of raw pixel intensities to represent face image, Lu and Tan 2013 [164] evaluated their regression model on MORPH II dataset and obtained MAE of 5.2 White ethnic group and 4.2 for Black ethnic group. Onifade *et al.* [165] applied a boosted regressor on Age-rank Local Binary Patterns (arLBP). They reported a MAE of 2.34 on FG-NET using LOPO validation protocol. Their approach demonstrated that age ranking with correlation of ageing patterns across age groups improves performance of age estimation. Using raw pixel features, Akinyemi and Onifade [166] ethnic-specific age group ranking for age estimation. This approach learns ethnic parameter in addition to the parameters learnt in [165]. They evaluated this technique on FG-NET and FAGE datasets and reported MAE of 3.19 years. Their findings show that incorporating ethnic parameters improves performance of age estimation approaches. This could be attributed to the fact that people in different ethnic groups age differently.

2.4.3 Hybrid approach

As discussed in preceding sections, age estimation task can be approached as either classification or a regression problem. To choose between the two, one may perform an experiment by selecting representative classifiers and regressors to compare their performance on the same dataset using same features. Guo *et al.* [49] [57] compared SVM classifier to SVR regressor. This experiment showed that SVM performs better compared to SVR on YGA dataset with SVM achieving MAE of 5.55 for females and 7.00 for males while SVR achieving 5.52 for females and 7.47 for males. It was also reported that SVM performed poorly on FG-NET compared to SVR (MAE 7.16 against 5.16 years). This experiment shows that classification approach to age estimation may perform better or worse than regression approach depending on other aspects like quality of images in the dataset used, feature selection and feature extraction techniques used, distribution of images across age among other factors.

Combining classification and regression may result into robust and more accurate age estimation systems. Guo *et al.* [49] [57] therefore proposed age estimation using Locally Adjusted Robust Regression (LARR). LARR first performs regression using

all existing ageing images. Regression results are then used to limit a classifier with small search range. They demonstrated that better age estimation performance can be achieved by combining classification and regression schemes. By combining regression and classification, the MAE improved to 5.30 and 5.25 years for females and males respectively on YGA dataset and 5.07 on FG-NET dataset. The limitation of LARR method [57] is that it cannot automatically determine local search range for a classifier. The range is determined by heuristically trying different ranges and requires that the user to experimentally choose best solution. To automatically determine limited search range, Guo *et al.* [58] proposed a likelihood-based approach for combining classification and regression outcomes. Using a uniform distribution, regression results are transformed into likelihoods, then likelihoods from classification outcome are cut off by the uniform distribution. This further improved accuracies by achieving MAE 5.12 and 5.11 for males and females respectively on YGA and 4.97 on FG-NET.

Gunay *et al.* [167] represented ageing face by fusing AAM, LBP and Gabor features. They used an ensemble of three SVMs arranged in a hierarchical manner to build an age estimation model. The first step of their model was to perform age-group estimation by SVM classification. A linear regression was is then performed to estimate age within the age-group. Their approach achieved MAE of 4.13 on FG-NET. These results show that feature and decision fusion used in a hybrid hierarchical age estimation can improve estimation errors compared to classification approaches.

Han et al [2] performed hierarchical demographic estimation and compared machine and human performance. They extracted BIF features and demographic informative features using a boosting algorithm. They then perform a hierarchical age estimation using between-group classification followed by with-in group regression. Evaluation this technique on MORPH II and FG-NET, they achieved MAE of 3.6 and 3.8 on MORPH II and FG-NET datasets respectively. Choi *et al.* 2011 [8] used AAM, Gabor and LBP to represent face image. Their hybrid age estimation model achieved MAE of 4.7 on FG-NET, 4.3 on PAL and 4.7 on BERC datasets.

Hybrid approach to age estimation demonstrate better performance compared to regression and classification when used alone. To combine classification and regression, one may test extracted features on both techniques separately before combining them. Arrange regression and classification in arbitrary hierarchical order and compare performance when regression is done before classification and when done after classification.

2.5 Facial ageing databases

Precise age and age group estimation require a database with good quality facial images of several individuals from different ethnic groups at different ages and across gender. It is hard to collect large ageing database with a series of chronometric images from an individual. Age and age group estimation often use databases collected and published earlier. A brief description of ageing databases that have been used in the literature or can be used in future investigations is presented.

2.5.1 FG-NET ageing database

FG-NET [26] contains 1,002 both colour and grayscale images of 82 individuals from age 0 to 69 years. Each individual has on average 12 images. Images are collected from multi-race subjects and have great inconsistencies in head pose, facial expression and illumination. Some images have adverse condition because they were scanned. There are 68 landmark points provided which can be used to model facial shape. Age features can be modelled as AAM or as appearance model using texture and wrinkle features.

2.5.2 MORPH database

MORPH [168] is a publicly available ageing database created by Face Ageing Group at the University of North Carolina. This dataset is split into two sets. Album 1 has 1724 images collected between 1962 and 1998 from 515 individuals. Images in this dataset range from 27 to 68 years. There are 1,430 images for males and 294 images for females with age gap ranging from 46 days to 29 years. Set 2 contains 55, 134 images of 13,000 individuals collected over 4 years. Both albums contain metadata for race, gender, date of birth and date of acquisition. The eye coordinates of the dataset can be requested. A commercial version of album 2 contains larger set of images collected over a longer time span and include information like height and weight of individual.

2.5.3 Yamaha Gender and Age (YGA) database

YGA [39] [40] database has 8000 high resolution colored images of 1,600 individuals consisting of 800 males and 800 females of Asian race, aged between 0 and 93 years. Each subject has approximately five nearly frontal face images at the same age and a label of his or her approximated age. The images have high variations in expression, illumination, and facial expression. Haar cascade face detector [169] is used to crop and resize images to 60 x 60 gray scale patches.

2.5.4 WIT-DB database

Waseda human-computer Interaction Technology [170] dataset consist of 12,008 face images of 2,500 females and about 14,214 images of 3,000 males from Japanese race, with age ranging between 3 and 85 years. The ages are arranged in 11 non-overlapping age groups. The dataset has wide variations in illumination on un-occluded frontal view faces with neutral facial expression. Face images are cropped and resized to 32 x 32 gray-scale patches.

2.5.5 AI & R Asian Face database

AI & R Asian [171] dataset contains images of different expression, age, pose and illumination. There are 34 frontal-view images collected from 17 individuals with ages ranging from 22 to 61 years. There is on average 2 images per individual making this database is not suitable for age or age-group estimation.

2.5.6 Burt's Caucasian Face database

Collected and used in [172] by Burt and Perrett to investigate visual cues to age by blending colour and shape of facial components. The database contains 147 images of European males aged between 20 and 62 years. Faces had neutral expression with beards shaved with no glasses and makeups. There are 208 landmark points placed manually in standardized positions. These points can be used to encode facial shape.

2.5.7 LHI face database

Lotus Hill Research Institute (LHI) database contains 50,000 images of Asian adults at different ages. The images have slight dissimilarities in pose and lighting. Part of this database was used in [173] by Suo *et al.* to model a hierarchical face model for age estimation. The part used consists of 8,000 colour images of individuals aged between 9 and 89 years with one image per person. This database could not be appropriate for subject-based age estimation since it does not provide multiple face images of the same individual at different ages.

2.5.8 HOIP face database

Human and Object Interaction Processing (HOIP) database consist of 306,600 images of 300 individuals aged between 15 and 64 years. The database in divided in 10 age groups. Each age group has got 30 subjects, 15 females and 15 males [3].

2.5.9 Iranian face database

Iranian Face Database [174] has 3,600 colour images from 616 individuals aged between 2 and 85 years of which 487 are males and 129 females. The images have variations in pose and facial expression. At least one image with glasses was also taken. Majority of the images are of subjects in the age group of 1 – 40 years. This database can therefore be appropriate in modelling ageing and age estimation in formative and middle-age years.

2.5.10 Gallagher's Web-Collected database

This database was collected by Gallagher and Chen [175] from Flickr.com image search engine. The database has 28,231 faces in 5,080 images. It divided into seven age groups as 0 - 2, 3 – 7, 8 – 12, 13 – 19, 20 – 36, 37 – 65 and 66+. This dataset is suitable for age-group estimation although the age-groups are wider in older ages.

2.5.11 Ni's Web-Collected database

This database was collected from the web by Ni *et al.* [176] [177] using Google.com and Flickr.com image search engines. The database has 219, 892 faces in 77, 021 images with age range between 1 and 80 years. This is the largest database ageing database ever reported. The wide age range in this database makes it suitable for age estimation both in child, adult and old age-groups.

2.5.12 Kyaw's Web-Collected Database

This database was collected from the web by Kyaw *et al.* [178] using API services provided by Microsoft Search Engine Bing. The images in the collected database are aligned with eye corner points captured manually and cropped to 65 by 75 patches. The database contains 963 images divided in four age groups of 3 – 13, 23 – 33, 43 – 53 and 63 – 73. The database is not appropriate for age-group estimation since there are missing images between age groups.

2.5.13 BERC database

BERC database [8] was collected by the Biometric Engineering Research Center (BERC). The database contains images of 390 subjects with age ranging from 3 – 83 years. Images are of high resolution 3648 x 2736 pixels. There is no variations in light and facial expression on all the images and subjects are uniformly distributed with respect to age and gender. These makes the database suitable for age estimation, although it is comparatively small.

2.5.14 3D Morphable Database

The database contain 3D scans of 100 male adults and 100 female adults' faces and 238 teenage faces aged between 8 and 16 years consisting of 113 females and 125 males [179] [113]. All faces were without makeup, accessories and facial hair. In 3D Morphable Face Models, individual faces are represented as face vector in a 3D. By caricaturing texture a shape feature vectors, the model can transform one's face. As one ages, each face will transform along a curved trajectory in a high dimensional

space. Faces are represented by shape and texture vectors such that each linear combination of different faces is a new realistic face.

2.5.15 Summary

FG-NET, MORPH and Web-collected Gallagher's databases are publicly available. Other databases can be found by contacting owners. MORPH, Ni's, YGA, LHI and Gallagher's Web-collected databases are large databases and well suited for regression based age estimation using statistical algorithms like AAM and age manifold. FG-NET is suitable database for evaluations with several age estimation methods like AGES. AI & R, LHI and Iranian datasets comprise comparatively high resolution 2D face images. Other datasets stated here were not extensively used but may be appropriate for some application areas.

2.6 Biometric information fusion

A biometric system is made up of *sensor, feature extraction* and *matching* modules. Performance of a biometric system is affected by reliability of the sensors used and robustness of the features extracted from acquired samples [180]. Biometric information fusion is a technique of combining multiple biometric inputs or biometric data processing techniques in order to improve system efficiency, accuracy [181], robustness and applicability [182]. Fusion can be in the form of

- *multiple representation of a single biometric trait*
- *multiple matching algorithms applied on a single biometric trait*
- *multiple biometric traits fused/combined*

Biometric systems that use fusion techniques are called multi-biometric systems. Multi-biometric systems help to overcome limitations of unimodal biometric systems.

Fusion can be categorized into [180] [181]

- ***Multi-sample:*** this is fusion of multiple images (samples) acquired from the same biometric trait, for instance multiple images of a single face, fingerprint or audio recording. A single sensor is used to acquire these samples
- ***Multi-instance:*** this involves fusing multiple instance of the same biometric trait e.g. multiple fingerprints from different fingers or images of both irises (eyes)
- ***Multi-sensor:*** this is fusion of multiple instances of a single biometric trait acquired using multiple sensors. Multiple sensor can be used to alleviate problem of noisy sample.
- ***Multi-algorithm:*** fusing multiple processing methods for each sample. Could mean use of multiple matchers (classifiers) or multiple feature extraction methods
- ***Multi-modal:*** comprise using data collected from multiple biometric traits like fusing data from face and fingerprints

There are three levels in a biometric system where fusion can be performed

- ***Feature level fusion:*** features (samples) from multiple sample (instances) or sensors can be combined (concatenated) to form a single feature vector for classification. The resultant feature vector will have a higher dimensionality and may have more discriminating attributes.

- ***Matching score fusion:*** Each algorithm provide a matching score (probability) that indicate closeness of the feature vector to the template feature. These scores are combined.
- ***Decision level fusion:*** Multiple algorithms are used to learn pattern from features. Decisions from these algorithms are combined using some rule (AND rule, OR rule or majority voting rule) to generate a final decision.

Features are supposed to be normalized to ensure their values lie within the same range before fusing them. The most commonly used normalization scheme is *z-score* normalization. Z-score normalization is performed as

$$f_{norm} = \frac{\left(f - \bar{f}\right)}{\sigma_f} \tag{2.22}$$

where f_{norm} is the normalized feature, f is the un-normalized feature, $\bar{f}$ is the mean of the feature and σ_f is the standard deviation of the feature. Normalization is only done if dimensionality d of the features is greater than 1.

Features can also be normalized using min-max normalization as

$$f_i^{norm} = \frac{f_i - min}{max - min} \tag{2.23}$$

where f_i is a feature value at index i in the feature vector f, min and max are the minimum and maximum values in the feature vector respectively, and f_i^{norm} is the normalized value of feature f_i. It can be seen that the normalized value will be 1 when $f_i = max$ and it will be zero when $f_i = min$.

2.7 Age estimation evaluation protocols

Evaluation protocol determines system test, criteria for test data selection and system performance measure. A good validation strategy should be independent of training data and representative of the population from which it has been drawn [183]. Age estimation technique need to be validated using previously unseen data to avoid over-fitting age estimation technique and improve its generalization capability. Cross validation is a popular strategy for age estimation evaluation. In cross-validation, data is split into two subsets; one segment is used to train or learn age estimation model and the other segment is used to validate or evaluate the model. In classic cross-validation, training and validation datasets must cross-over in consecutive rounds such that every data point has equal chance of being validated or evaluated against the other. The basic form of validation is holdout.

Holdout strategy is the simplest and computational efficient strategy [184] used for validating age estimation techniques. The dataset is randomly split into two sets; training subset and validation subset. Commonly training subset consist of $\frac{2}{3}$ of the original data and remaining $\frac{1}{3}$ samples constitute validation subset. Age estimation model is then fitted using the training subset and validated on the test subset. In this strategy the model is trained and validated only once. Although this method is preferred and takes shorter time to compute, its evaluation depends on the data in respective subsets and results into high variance hence making this strategy give difference evaluation results depending on how the dataset is divided [185]. Another validation strategy commonly used is repeated random sub-sampling (RSS) [186] [187]. In RSS validation technique, the holdout strategy is iterated a number of times and results averaged. The dataset is randomly split into two subsets (train and validation) with fixed number of samples for each phase of validation. For each data split, age estimation model is retrained on train subset and validated using

test subset. The advantage of this strategy over k-fold validation is that the size of training and validation is independent to number of validation iterations. However, this strategy has a limitation such that some samples may never be selected for validation while other samples may be selected repetitively leading to overlapping of validation subsets [188]. But with significantly large number of iterations done, RSS is likely to achieve better results as k-fold validation [189].

Cross validation [188] is a standard statistical technique used for model generalization ability with wide application in classification and regression problems [190]. It involves dividing dataset into two subsets, one subset is used to train an estimator while the other subset used to test an estimator [191]. Cross validation is used to assess how a model generalizes to initially unseen data [188] [192]. Cross-validation strategies can be categorized into two: i) exhaustive (compute all possible ways of data splitting) and ii) non-exhaustive (does not compute all possible ways on data splitting). Exhaustive cross-validation algorithms include Leave-One-Out (LOO) and Leave-P-Out (LPO) while non-exhaustive include k-fold and repeated random subsampling (RSS) [185] [193]. Cross-validation [194] consist of averaging multiple holdout validation results from different subsets of data.

K-fold cross-validation is the basic form of cross-validation. Other forms of cross-validation are just but special cases of k-fold cross-validation or involve repeated rounds of k-fold validation. In k-fold cross-validation [194], original data is randomly split into k equal subsets. Then, k iterations of training and validation are performed such that in every iteration, a different fold of data is reserved for validation while the remaining k-1 are used to learn a model. The estimated error is the mean of all validation errors. Standard deviation of these errors can be used to approximate the confidence range of the estimate. The main advantage of k-fold cross-validation is that eventually all samples will used for both learning and validating a model. Common value of k used in various techniques is 10 as a compromise between efficiency and accuracy. A stratified cross-validation is commonly used in order to improve accuracy of the estimation [188].

Leave-One-Out (LOO) [195] [194] [191] is a special type of cross-validation that given a dataset with C classes, $C-1$ validation experiments are performed. For each experiment, data from $C-1$ classes is used for training and data from one class that was left out is used for validation. Therefore, given a dataset of S subjects from age $0 \rightarrow A_n$, LOO cross-validation will perform $S-1$ validation experiments. In each experiment i, facial images of subject S_i are used for validation while images of the rest $S-1$ subjects are used for learning a model. In this approach, images of each subject will be used for both training and validation. This way, the technique is validated in the same way as its application scenario where the subject whose age is to be estimated is previously unseen in the system. Although LOO is almost unbiased, it may give unreliable estimates due to its high variance [196]. Leave-p-Out (LPO) [197] with $p \in \{1, 2, 3 \ldots, n-1\}$ successively leaves out every possible subset of p data samples to be used for validation. In age estimation, given a set of images of N subjects, LPO can be used by leaving out images of p where $p \leq (N-1)$ subjects to be used for validation and use images of $N-p$ subjects for training. Elisseef and Pontil [198] showed that LPO cross-validation is less biased compared to LOO. LPO will have $\binom{n}{k}$ iterations where n is the number of images. These iteration are almost always much higher compared to $n-1$ iterations in LOO, leading to high computation time. LPO with $p = 1$ is same as LOO. LOO and LPO are exhaustive cross-validation strategies compared to other methods. Further information on LPO can be found in [199]. Detailed information on cross-validation can be found in [197] and [200].

Bootstrap is a strategy introduced by Efron and Tibshirani [201] [202]. Bootstrap is commonly used when working on a small dataset [184]. In this strategy, a bootstrap set is created by uniformly sampling, with replacement, n instances from the

original data to make a training set. The remaining samples not selected are used as testing set. The value n of selected samples is likely to change from fold to fold. Since data is sampled with replacement, the probability of any data sample not being selected given by $\left(1-\frac{1}{n}\right)^n \approx e^{-1} \approx 0.368$. Chances of a data sample being selected into a train set is $(1-0.368) = 0.632$. Therefore, expected number of distinct samples appearing in the train set is $0.632 \times n$. Since error estimate obtained by using test data will be too pessimistic (since only 62.3% of instances are used for training), error is calculated as $error = 0.632 \times e_0 + 0.368 \times e_{bs}$ where e_0 is rate of error obtained from bootstrap sets not having the instance being predicted (test set error) and e_{bs} is the error obtained on bootstrap sets themselves, both averaged over all data samples and bootstrap samples. Estimate accuracy is directly proportional to number of times the process is repeated. More details on bootstrap validation technique can be found in [202]. Bootstrapping increases the variance that can occur in each fold which makes this strategy more realistic of the real application situation [202]. This validation strategy is rarely used in age estimation.

In most cases a dataset is split into 3 subsets: validation subset, training subset and testing subset [192]. In this approach, the validation subset is used to tune the system to determine termination point of the training phase when overfitting start occurring on the training subset. The testing subset is used to validate the trained model using data samples not initially in validation and training subsets. Kiline and Uysal [189] proposed a technique of splitting the dataset with samples from specific subjects rotationally left out of training and validation sets. Budka and Gabrys [183] proposed a density-preserving sampling (DPS) technique that eliminates need for repeating error estimation procedures by diving dataset into subsets that are guaranteed to be representative of the population the dataset is drawn from. These new proposed approaches of model validation could be experimented in age estimation problem and results compared with other common methods. Cross-validation and bootstrap strategies are commonly used when one has limited data such that hold-out strategy cannot be sufficient for data representativeness in both training and test sets. With abundant data with stable distribution over time, single stratified random split is able to provide required representativeness [183].

For purposes of comparing performance metric of two or more learning algorithms, Salzberg [203] proposed use of k-fold cross-validation followed by appropriate hypothesis testing instead of comparing their average accuracies. This strategy can be used to compare two age estimation techniques.

In each iteration of validation, Absolute Error (AE) for each estimated age is defined as

$$AE = |a_i - \bar{a}_i| \tag{2.24}$$

where is a_i is the ground truth age and $\bar{a}_i$ is the estimated age. After all validation iterations, Mean Absolute Error (MAE) is defined as the average of all absolute errors between estimated and ground truth age as

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|a_i - \bar{a}_i| \tag{2.25}$$

where N is total number of test images, a_i is the ground truth age of image i, and $\bar{a}_i$ the estimated age of image i. Although this performance evaluation is commonly used, it does not give age estimation performance for specific age but rather gives general performance of the technique for all ages. This approach could be slightly modified such that it gives MAE for every age and general MAE of the technique.

Given a set of testing images $a_1^{n_1}, a_2^{n_2} \ldots a_k^{n_k}$ belonging to k ages to be estimated with n_i representing number of test images known to belong to age a_i, MAE for every

age can be defined as

$$MAE_k = \frac{1}{n}\sum_{i=1}^{n}|a_k - \bar{a}_i| \tag{2.26}$$

where $\bar{a}_i$ is the estimated age for image i of age a_k and n is the number of test images belonging to age a_k. This will give age specific performance of age estimation technique. Overall MAE can be found by summing all the MAE for all ages tested and divide by sum of number of test images in each age as

$$MAE_{TOTAL} = \sum_{i=1}^{k}\frac{(MAE_i \times n_i)}{N} \tag{2.27}$$

where $N = n_1 + n_2 + \cdots + n_k$

Age estimation techniques performance is evaluated based MAE. The smaller the MAE the better the age estimation performance. MAE only shows average performance of the age estimation technique. MAE is appropriate measure of age estimation when the training data has missing images [10]. The overall accuracy of the estimator is given by cumulative score (CS) [49] [39] which is defined as

$$CS\,(x) = \frac{N_{e\leq x}}{N} \times 100\% \tag{2.28}$$

where $N_{e\leq x}$ is the number of images on which the age estimation technique makes an absolute error no higher than x years error tolerance and N is the total number of test images.

In age group estimation, the age-group label represent a range of ages, hence the cumulative scores are compared at error level 0, i.e. the percentage of exactly correct age-group estimation. Therefore, the CS equation becomes

$$CS\,(x) = \frac{n_x}{N_x} \times 100\% \tag{2.29}$$

where n_x is the number of test images correctly recognized as belonging to age-group x and N_x is the total number of test images in age-group x. Therefore, CS is used as an indicator of accuracy of age-group estimator [5]. CS is useful measure of performance in age estimation when the training dataset has samples at almost every age [3]. MAE is best evaluation technique when the training set has a lot of missing ages. However, in age estimation, both MAE and CS are used since different techniques, datasets, systems may be extremely imbalanced or skewed for evaluation.

2.8 Literature review

In this section we present a review of the previous studies in landmark localization, age-group and age estimation. Comparative analysis of the techniques used and results achieved is also presented.

2.8.1 Facial landmark localization

Facial landmarks are those features on the face that can easily be recognized and they stand out from its immediate environment and hence can provide discriminative information. These features vary from person to person but there are those that are common to all persons like eyes, nose, mouth, ears, beards, facial spots, eye corners, chin, mouth corners, nostril corners, eye-brow arcs, cheeks and cheek-bone

areas, forehead, nose-bridge among others. Facial landmarks play a crucial role in face recognition because they can provide discriminative information about individuals who may seem to appear identical from a global face image.

It is important to accurately detect facial landmarks for subsequent facial feature computing like feature extraction for component-based face recognition, landmark and/or face recognition and tracking, facial expression modelling [204]. Fiducial landmarks are regions on the face that can easily be recognized like eye, nose, mouths, eye-corners, and mouth-corners. These regions are hardly affected by facial expressions hence regarded as more reliable regions for feature extraction for subsequent face computing.

There are several landmark localization techniques proposed. Most of these techniques learn geometric and appearance features using a machine learning tool like Support Vector Machines (SVM) [205]. These learning techniques improve detection accuracy at the expense of detection speed. Accurate and fast localization of facial landmark is important in application areas like facial expression recognition [206] [207], face simulation [208], 3D reconstruction of face [204], face recording [209], component-based face recognition, face verification [210] [211] [212] and face tracking [213] [214]. Facial landmarks provide vital spatial information for extracting region of interest (ROI) for shape and texture based face recognition. Shape and Texture based image processing often require an image region. However, the landmarks detected on human faces are often point descriptors like nose tip, eyebrows, eye-corners which hardly provide shape and texture information for discriminant recognition of faces. There is need for a robust technique to automatically detect and crop the needed ROI for subsequent processing. The choice of these regions of interest must be very inclusive such that the chosen regions (landmarks) have highest likelihood to be found in all human faces.

2.8.1.1 Types of facial landmarks

There are two general types of facial landmarks. Primary/fiducial landmarks and secondary/ancillary landmarks. Primary landmarks are those that are directly detected and significantly guide facial detection and detection of secondary landmarks. Primary landmarks like eyes corners, nose center, mouth corners and eyebrows can be located comparatively easy using Scale Invariant Feature Transform (SIFT) [215] extracted features or features extracted by Histogram of Gradients (HOG) [130]. Secondary landmarks like nose-bridge, chin, cheek bone area, eye midpoints, and eye lids often provide scanty information and their localization directed by fiducial landmarks' information.

2.8.1.2 Facial landmark localization techniques

There has been significant development in research on face detection in recent past. One of the most popular techniques being the Haar-cascade face detection proposed by Viola and Jones [216]. Viola and Jones used Adaboost to train a cascade of classifiers that search wavelet features in a rectangular region. In every stage, the classifier discards sizeable portion of negative features that survive to that stage. This ensures that majority of rectangular regions that do not contain faces are dropped in early stages of the cascade. This classifier is popularly used in face detection due to its accuracy and high speed that makes it ideal for real-time applications [217]. The results of the detection gives the location of the face, eye centres, nose, and mouth among other landmarks. Zhu *et al.* [218] used local regions around facial landmarks to detect a face, localize landmarks and estimate facial pose. Hatem *et al.* [217] found that the technique presented in [218] never captured global structural information of the face as well as information on local variations of local parts. The technique was also found not to be robust to occlusion. Recently, Yan *et al.* [219] proposed

cascade classifiers to improve speed of facial part detection.

Adaboost algorithm creates an ensemble of cascade classifiers by combining selected Haar-like features [218]. The cascade structure makes the detector faster. The accuracy of the detector is improved by the use of Adaboost [217]. Currently, majority of face detection techniques adopt Viola and Jones Haar cascade classifier. This classifier uses a sliding window to generates, in every stage, many weak classifiers from Haar-like features extracted from different positions and scales. Adaboost is used to boost these weak classifiers into a strong classifier that predict presence or absence of a face in a window [220]. Mita *et al.* [221] used joint Haar-like features to build a weak classifier for capturing co-occurrence of multiple Haar-like features. This technique had minimal improvement on performance [217]. Hatem *et al.* [217] proposed a combined technique for face detection, face tracking and head-pose estimation. They used Haar-like features and cascade object detector to detect face, eye, mouth and nose.

However, all these techniques do not explicitly provide facial landmarks for further processing for face recognition. Efraty *et al.* [222] using cascade classifiers and Bag-of-Words (BoW) representation developed facial component-landmark detection method that is robust to pose and illumination variations. Hakan *et al.* [223] proposed SVM based landmark detection technique. These proposed techniques although posting higher accuracy, they take relatively longer time to detect eyes, nose and mouth. These techniques also do not detect forehead, nose-bridge and cheek bone area despite these regions having more texture discriminant information on frontal faces.

Despite there being numerous work in the area of face landmark detection, there is still need for improved detection techniques [204]. Most application areas like component-based face recognition for surveillance require real-time landmark detection and localization. This makes the speed at which landmarks are detected or localized an important factor. The required landmark detection techniques also need to be robust to factors like illumination, occlusion, and facial expressions among others.

Knowledge of spatial relationship between landmarks and general knowledge on shape and appearance of these landmarks help to guide their localization. This knowledge could be converted to a rule-base by incorporating point-to-point statistics like distances and angles between landmarks to improve speed and accuracy of detection [204]. Fiducial facial landmarks like eyes, nose and mouth can be detected using algorithms like Viola-Jones [216], Gabor filters [69], projection histograms [224] or specifically trained machine learning tools like SVMs [225]. Spatial and geometric information of these fiducial landmarks can be used to locate secondary landmarks like nose-bridge, cheek bone area, forehead in a minimised search area.

2.8.1.3 Facial landmark localization challenges

There are a number of challenges posed by facial landmark detection and localization. Differences in landmark appearance is attributed to face variability between subjects, illumination, facial expressions, camera resolution and face angle among other factors. These factors affect landmark detection and localization performance making a detector trained on one dataset performing poorly on another dataset. For instance, Sankur [226] shows performance drop of about 20%-30% in the Face Recognition Grand Challenge 1 and 2 training and testing when the datasets are used interchangeably for training and testing. Landmark detection performance has been extensively reported by Dibeklioglu *et al.* [227]. Apart from facial expression, image quality, partial occlusion, choice of model and dataset, it was also shown that accuracy needed and landmarks required also affect landmark detection and localization [204]. This is because facial landmarks needed and their accuracy are

application dependent. For instance, detecting a face or recognizing a face may need eyes, nose and mouth corners or their bounding rectangles and another application may need more landmarks with higher spatial accuracy.

2.8.2 Landmark displacement across age

Facial anthropometry investigations provide quantifiable explanation of craniofacial complexity using distance measurements between facial fiducial landmarks across age. Age-based face measurements and proportions have been widely used to create anthropometric face models for face recognition. Six ratios were computed by Kwon and Lobo [228] [41] and used to differentiate children from grown-ups. The experiment was performed on a small dataset of 47 and the authors did not report overall performance of this small dataset.

Craniofacial studies have shown that human faces change from circular to oval as one age [80]. These changes lead to variations in the position of fiducial landmarks [74]. Ageing introduce significant change in facial shape in formative years and relatively large texture variations with still minor change in shape in older age groups [3] [7]. Shape variations in younger age groups is caused by craniofacial growth. These changes lead to variations in the position of fiducial landmarks [74]. During craniofacial development, forehead slopes back releasing space on the cranium. Eyes, ears, mouth and nose expand to cover interstitial space created. The chin become protrusive as cheeks extend. Facial skin remain moderately unchanged than shape. More literature on craniofacial development is found in [9].

In adulthood, facial sides vary a little and displacements between main facial landmarks remain constant. For example, in children, displacement between eyes is nearly equal to displacement between nose and eye as displacement between eye and nose is near displacement between mouth and nose [74]. These information could be tracked across age to determine displacement of landmarks from each other as one grow. This affects the shape and size of face as one ages. In old age, human face appear to be rectangular or trapezoid shaped unlike V-shaped in childhood [229]. This change in facial shape and size as one grow could be as a result of drifting of landmarks from each other. More literature on craniofacial development is presented in [9].

Ageing affects facial size and shape as anthropometry of the face varies with age. This change in shape and size could make fiducial points like eyes, nose and mouth drift from each other. This leads to variations in distances between these landmarks. These distance variations could be measured and tracked for purposes of face recognition, age group classification, gender classification or ethnic grouping among other application areas. Difficulty in measuring face profiles on 2D face images lead to use of distances and ratios, instead of mathematical model, for modelling facial shape [230].

Dehshibi and Bastanfard [74] used distance ratios between landmarks to classify human faces in various age groups using back propagation neural network with distance ratios as inputs. For this to happen, it means these ratios are distinct in different age groups. Therefore, these ratios could be tracked across age to find the trend in landmark drift from each other as one grow. Farkas *et al.* [231] used 10 anthropometric measurements of the face to classify individuals in various ethnic groups. They analyzed these measurements and identified ones that contribute significantly to diversity in facial shape in different ethnic groups. They also found that horizontal measurements differed between ethnic groups than vertical measurements.

Most commonly used distances in face morphology are Euclidean and Geodesic distances since they are capable of characterizing facial morphological traits [103].

Investigations have shown that important information is conveyed by measuring Euclidean and geodesic distances, and angles between fiducial points on the face.

Morphological based face recognition technique was developed by Tiwari *et al.* [232] using Euclidean distance measurements between fiducial facial landmarks as inputs to back propagation neural network. This technique recognized faces but it was independent of ageing factor due to variations in distances between landmarks as one age.

2.8.3 Age group estimation

Global, local and hybrid features have been previously used in age and age group estimation. Ramanathan *et al.* [233] present a recent survey in automated age estimation techniques.

Age group is a range of ages. Persons whose real age are within the defined ranges are said to be in the same age group. Significant amount of research has been done to automatically extract visual artifacts from faces and group persons in respective age groups. Kwon and Lobo [41] estimated age-group based on anthropometry and density of wrinkles. They separated adults from babies using distance ratios between frontal facial landmarks on a small dataset of 47 images. They also extracted wrinkle features from specific regions using snakes. Young adults were differentiated from senior adults using these wrinkle indices. Baby group classification accuracy was lower than 68% but overall performance of their experiments was not reported. Furthermore, ratios used were mainly from baby faces. Horng *et al.* [122] used geometric features and Sobel filter [136] for texture analysis to classify face images into four groups. They used Sobel edge magnitude to extract and analyze wrinkles and skin variance. They achieved an accuracy of 81.6% on subjectively labelled age-groups.

Ramanathan and Chellappa [108] computed eight distance ratios for modelling age progression in young faces like 0 to 18 years. Their objective was to predict one's appearance and face recognition across age progression. Using 233 images of which 109 were from FG-NET ageing dataset, and the rest from their private dataset, they reported improvement in face recognition from 8% to 15%. Dehshibi and Bastanfard [74] used distance ratios between landmarks to classify human faces in various age groups. Using back propagation neural network with distance ratios as inputs, they classified face images into four age-groups of below 15, 16-30, 31-50 and above 50 years on a private dataset. They reported 86% accuracy. Thukral *et al.* [234] used geometric features and decision fusion for age-group estimation. They achieved 70% overall performance for 0-15, 15-30 and above 30 years age-groups. Farkas *et al.* [231] used 10 anthropometric measurements of the face to classify individuals in various ethnic groups. They analyzed these measurements and identified ones that contribute significantly to diversity in facial shape in different ethnic groups. They also found that horizontal measurements differed between ethnic groups than vertical measurements.

Tiwari *et al.* [232] developed a morphological based face recognition technique using Euclidean distance measurements between fiducial facial landmarks. Using morphological features with back propagation neural network, they reported superior recognition rate than performance of Principal Component Analysis (PCA) [126] with back propagation neural network. This technique recognized faces but it was independent of ageing factor due to variations in these distances as one age. This signifies that distances between facial landmarks differ at different ages, especially in young age-groups. Therefore, these distances could be used in age estimation. Gunay and Nabiyev [235] used spatial LBP [46] histograms to classify faces into six age-groups. Using nearest neighbour classifiers, they achieved accuracy of 80% on age-groups $10\pm5, 20\pm5, 30\pm5, 40\pm5, 50\pm5$ and 60 ± 5. In [167], Gunay and Nabiyev

trained three Support Vector Machines (SVM) model for age-group estimation using AAM [38], LBP and Gabor filter [69] features. They fuse decisions from these classifiers to obtain final decision. Although they reported 90% accuracy of subsequent age estimation, overall performance of age-group estimation was not reported.

Hajizadeh and Ebrahimnezhad [236] represented facial features using Histogram of Oriented Gradients (HOG) [130]. Using probabilistic neural network (PNN) to classify HOG features extracted from several regions, they achieved 87% accuracy in classifying face images into four groups. Liu *et al.* [237] build a region of certainty (ROC) to link uncertainty-driven shape features with particular surface features. Two shape features are first designed to determine face certainty and classify it. Thereafter, SVM is trained on gradient orient pyramid (GOP) [238] features for age-group classification. Testing this method on three age-groups, 95% accuracy was reported. They further used GOP in [239] with Analysis of Variance (ANOVA) for feature selection to classify faces into age-groups using linear SVM [240] to model features from eyes, nose and mouth regions. They achieved 91% on four age groups on FG-NET dataset and 82% on MORPH dataset. It was also found that the overall performance of age-estimation decreases as the number of age-groups increase. This is because the number of images in each age-group reduce drastically as the number of groups increase.

Lanitis *et al.* [112] adopted AAM to represent face image as a vector of combined shape and texture parameters. They defined ageing as a linear, cubic or quadratic function. For automatic age estimation, they further evaluated quadratic function, nearest neighbour and Artificial Neural Network (ANN) in [33]. They found that hierarchical age estimation achieve better results with quadratic function and ANN classifiers. Although AAM have been extensively used, it does not extract texture information. This problem is avoided by using hybrid features extraction techniques to combine both shape and texture features for age and age group estimation.

Sai *et al.* [241] used LBP, Gabor and Biologically Inspired Features for face representation. They used Extreme Learning Machines (ELM) [242] for age group estimation. Their approach achieved accuracy of about 70%. Using LBP and a bank of Gabor filters, Wang *et al.* [243] classified images into 4 age-groups. They used SVM, Error Correcting Output Codes (ECOC) and AdaBoost for age-group estimation. Table 2.2 show summary of age and age-group estimation studies.

2.8.4 Age estimation

Age is a real number that signify number of years elapsed since birth to a point in life. Age estimation is the process of estimation one's actual age using visual artifacts on face. These visual artifacts are extracted and used to estimate one's age.

Lanitis *et al.* [112] adapted Active Appearance Model (AAM) for ageing face by proposing ageing function. They defined age as a function $age = f(b)$ to cater for age introduced variations. In this function, age is the real estimated age of a subject, b consist of 50 AAM-learned-parameters feature vector, f is the ageing function. They performed experiments on 500 images of 60 individuals of which 45 subjects had images at different ages. Focusing on small age variations, they demonstrated that simulation of age improves performance of face recognition from 63% up to 71% and from 51% to 66% when training and testing datasets are used interchangeably.

Adopting Ageing Pattern Subspace (AGES), Geng *et al.* [5] [36] proposed automatic age estimation using appearance of face images. Evaluating AGES on FG-NET ageing database, they used 200 AAM parameters to characterize each image for age estimation. They reported 6.77 years Mean Absolute Error (MAE). Fu and Huang [39]

used age separated face images to model a low dimensional manifold. Age was estimated by linear and quadratic regression analysis of feature vectors derived from respective low dimensional manifold. Same approach of manifold learning was used by Guo *et al.* in [49]. They extracted face ageing features using age learning manifold scheme and performed learning and age prediction using locally adjusted regressor. Their approach reported better performance than Support Vector Regression (SVR) and SVM.

Guo *et al.* [49] used Locally Adjusted Robust Regression (LARR) to estimate age. Evaluating their approach on a large dataset, they reported MAE of 5.30 years and 5.07 years on FG-NET. Guo *et al.* [42] further proposed age estimation using Biologically Inspired Features (BIF) [70] [71]. BIF features with Support Vector Machine (SVM) achieved MAE of 4.77 years on FG-NET ageing dataset and 3.91 years and 3.47 years on females and males respectively on YGA dataset. Combining gender and age estimation, Guo *et al.* [244] used BIF and age manifold feature extraction with SVM classifier. They reported superior MAE of 2.61 for females and 2.58 for males on YGA database. Yan *et al.* [245] performed person-independent age image encoding using Synchronized Submanifold Embedding (SME). SME considers both individuals identities and age labels to improve generalization ability on age estimation. Evaluating this technique on FG-NET, they reported MAE of 5.21 years. Yan *et al.* [51] [43] used Spatially Flexible Patch (SFP) for feature description. SFP does not only consider local patches only but also their spatial information. With SFP, slight misalignment, pose variations and occlusion can be effectively handled. Furthermore, this technique can improve discriminating characteristics of the feature vector when limited samples are available. Adopting Gaussian Mixture Model (GMM), they achieved MAE of 4.95 years on FG-NET ageing dataset, 4.94 years and 4.38 years on females and males respectively on YGA dataset. Combining BIF and age manifold features and SVM for age estimation achieves MAE of 2.61 years and 2.58 years for males and females respectively on YGA dataset [3].

Suo *et al.* [173] designed graphical facial features topology based on hierarchical face model [246]. They used particular filters to diverse features at various stages of their hierarchical feature extraction design. Using Multilayer Perceptron (MLP), they reported MAE of 5.97 years on FG-NET and 4.68 years on their private dataset.

Craniofacial ageing model that combines psychophysical and anthropometric evidences was proposed by Ramanathan *et al.* [108]. The model was used to simulate perceived age of a subject across age for improving accuracy of face recognition. Choi *et al.* [8] proposed age estimation approach using hierarchical classifiers with local and global facial features. Using Gabor filters for wrinkle extraction and LBP for skin feature extraction, they classified face images into age groups with SVM. This approach is error prone because only depend on a single classifier. Wrong age group classification leads to wrong age estimation. For accurate age estimation, age group classification must be robust and this can be achieved by use of ensemble of classifiers. Chao *et al.* [247] determined the relationship between age labels and facial features by merging distance metric, learning and dimensionality reduction. They used label-sensitive and Nearest Neighbor (KNN) and SVR for age estimation. Chang *et al.* [59] proposed ordinal hyperplanes ranker for age estimation. Using AAM and SVM, their approach achieved 4.48 MAE on FG-NET and MORPH II datasets. Guo *et al.* [52] build a regression model using BIF and Partial Least Squares (PLS) for age estimation. Their approach achieved 4.43 MAE on MORPH II dataset and showed that learning label distribution improves age estimation. Lu and Tan [164] investigated age estimation using ordinary preserving manifold analysis approach. They found that gait can be used as an effective cue for age estimation at a distance for purposes of enhancing understanding capabilities of existing visual surveillance systems. They further found that discriminating age information can be better exploited in the low-dimensional manifold for achieving better age estimation performance.

Using Uniform Ternary Patterns (UTP), Tan *et al.* [48] and AAM, Luu *et al.* [248] proposed a spectral regressor for age estimation. Evaluating their technique they achieved a MAE of 6.17. Further work by Luu *et al.* [249] using Contourlet Transform achieved a MAE 6.0 on FG-NET and PAL datasets which was better compared to using UTP. Using Gabor wavelets and Orthogonal Locality Preserving Projections (OLPP), Lin *et al.* [250] developed an automatic age estimation system. They evaluated their technique on FG-NET dataset and SVM as a classifier and achieved a MAE of 5.71 years. Wu *et al.* [14] used 2D points to model facial shape for age estimation. Choober *et al.* [251] proposed use of an ensemble of classifiers for improving automatic age estimation. The limitation of this work is that only neural network was used to make the ensemble. An ensemble can be made robust if different classifiers are used so as each act as a complimentary to the other. Guo and Mu [153] compared canonical correlation analysis (CCA) and partial least squares (PLS) performance in age, gender and ethnicity estimation. Using BIF as a feature extractor, they found that CCA performs better compared to PLS. Hadid and Pietikainen [252] experimented manifold learning on age and gender estimation. They reported 83.1% accuracy age estimation on images extracted from video. Geng *et al.* [253] learnt label distribution and used them for age estimation. Their technique was evaluated on both FG-NET and MORPH datasets.

Guo *et al.* [42] first introduced BIF in image-based age estimation domain. They reported that using Gabor bank starting from smaller sizes like 5×5 can characterize ageing. Later Guo and Mu [52] used KPLS for simultaneous dimensionality reduction of BIF features for age estimation using a regressor. They also showed that Partial Least Squares (PLS) performs better in dimensionality reduction compared to traditional dimensionality reduction techniques like Principal Component Analysis (PCA). They later [153] used Canonical Correlation Analysis (CCA) for modelling age estimation as multiple-label regression problem. They reported that CCA based methods works better compared to KPLS based methods. Spizhevoi and Bovyrin [254] used RBF SVM to learn BIF features for age estimation. Han *et al.* [255] proposed a hierarchical age estimation and analyzed how ageing affect distinct facial components. They used SVM for both classification and regression to classify each face component. Their component localization was not accurate, thereby affecting subsequent features extracted from these components. They later [2] compared human and machine performance on demographic (age, gender and ethnicity) estimation. They modelled age estimation in particular as a hierarchical problem that consists of between-class classification and with-in class regression of boosted BIF and demographic informative features extracted from a face image.

Deep learning schemes, especially Convolutional Neural Network (CNN) have been successfully used in face analysis tasks including face detection, face alignment [256], face verification [257] and demographic estimation [258]. Wang *et al.* [259] extracted feature maps obtained in different layers as age features based on deep learning model. Huerta *et al.* [159] provide a thorough evaluation on deep learning for age estimation using fused features and compare it with hand-crafted fusion features. CNN have been used in different recent studies on age estimation and have demonstrated superior performance compared to other methods. Niu *et al.* [260] used ordinal regression and multiple output CNN for age estimation and reported a MAE 3.27 on MORPH II and a private Asian Face Age Dataset (AFAD). Chen *et al.* [261] presented a cascaded CNN that had 0.297 Gaussian error on age estimation. As further demonstrated in [262] [263] [264] CNN have posted better results in age estimation tasks. Although CNN performs better that other traditional methods, their applicability is limited by high processing demand required for their implementation. Table 2.2 show a summary of studies in age-group and age estimation.

TABLE 2.2: Summary of age and age-group estimation studies

Publication	**Feature Representation**	**Face Database/ Database Size**	**Algorithm**	**Evaluation Protocol**	**Performance / Accuracy** (MAE, CS)
Kwon and Lobo 1999 [41]	Anthropometric Model	Private/47	Classification	15 images for testing	CS:100%
Kanno *et al.* 2001 [265]	Appearance Model	Private/440	Classification	N/A	CS:80%
Lanitis *et al.* 2002 [112]	2D shape, raw pixel values	Private/500	Regression	500 train, 65 test	MAE: 4.3
Iga *et al.* 2003 [266]	GWT, 2D shape	Private/101	Classification	N/A	CS:58.4%
Lanitis *et al.* 2004 [33]	Active Appearance Model (AAM)	Private/400	Classification and Regression	50% train, 50% test	MAE: 3.82-5.58
Zhou *et al.* 2005 [56]	Anthropometric Model	FG-NET/1002	Regression	5-fold cross-validation	MAE:5.81
Ueki *et al.* 2006 [170]	Anthropometric Model	WIT-DB	Classification	2-fold cross-validation	CS:50%(M) 43%(F)
Takimoto *et al.* 2006 [267]	Anthropometric Model	HOIP	Classification	Leave-One-Out	CS:57.3%(M) 54.7%(F)
Geng *et al.* 2006 [36]	AGES	FG-NET/1002	Regression	Leave-One-Person-Out	MAE:6.77 CS:81%
Takimoto *et al.* 2007 [268]	Anthropometric Model	HOIP	Regression	Leave-One-Out	MAE: 3.0(M) 4.4(F)
Gen *et al.* 2007 [5]	AGES	FG-NET/1002 MORPH	Regression	Train on FG-NET, Test on MORPH	MAE:8.83 CS:70%
Yan *et al.* 2007 [53]	Active Appearance Model	FG-NET/1002	Regression	Leave-One-Person-Out	MAE:5.33
Yan *et al.* 2007 [53]	Appearance Model (Raw image)	YGA/8000	Regression	4-fold cross-validation	MAE:6.95(M) 6.95(F) CS:79%(M) 78%(F)
Yan *et al.* 2007 [54]	Active Appearance Model	FG-NET/1002	Regression	Leave-One-Person-Out	MAE:5.78
Yan *et al.* 2007 [54]	Appearance Model (Raw image)	YGA/8000	Regression	1000 train, 3000 test	MAE: 10.36 (M), 9.79 (F) CS: 61%(M), 63%(F)
Fu *et al.* 2007 [39]	Age Manifold	YGA	Regression	50% Test, 50% Train	MAE: 8.0 (M), 7.8 (F) CS: 70% (M), 70% (F)
Yan *et al.* 2008 [51]	Appearance Model (Patches)	FG-NET/1002	Regression	Leave-One-Person-Out	MAE: 4.95
Yan *et al.* 2008 [51]	Appearance Model (Patches)	YGA/8000	Regression	1000 train, 3000 test	MAE: 4.38 (M), 4.94 (F) CS: 88% (M), 85% (F)
Guo *et al.* 2008 [49]	Active Appearance Model	FG-NET/1002	Regression	Leave-One-Person-Out	MAE: 5.08
Guo *et al.* 2008 [58]	Active Appearance Model	FG-NET/1002	Hybrid	Leave-One-Person-Out	MAE: 4.97, CS: 88%

Table 2.2 Summary of. . . **Continued**

Publication	**Feature Representation**	**Face Database/ Database Size**	**Algorithm**	**Evaluation Protocol**	**Performance / Accuracy** (MAE, CS)
Guo *et al.* 2008 [58]	Age Manifold	YGA/8000	Hybrid	4-fold cross-validation	MAE: 5.12 (M), 5.11 (F) CS: 83% (M), 82% (F)
Yan *et al.* 2008 [43]	Appearance Model (Patches)	YGA/8000	Regression	1000 train, 3000 test	MAE: 7.82 (M), 8.53 (F) CS: 75% (M), 70% (F)
Fu and Huang 2008 [39]	Age Manifold	YGA/8000	Regression	50% Test, 50% Train	MAE: 6.0 (M), 5.5 (F) CS: 82% (M), 83% (F)
Suo *et al.* 2008 [173]	Active Appearance Model, Photometric	FG-NET/1002, Private/8000	Regression	N/A	MAE: 6.45 (Private), 6.97 (FG-NET)
Zhuang *et al.* 2008 [269]	Appearance Model (Patches)	YGA/8000	Regression	50% Test, 50% Train	MAE: 5.40 (M), 6.33 (F) CS: 82% (M), 76% (F)
Ni *et al.* 2009 [270]	Appearance Model (Patches)	MORPH, Web Images	Regression	Train on Web Images, Test on MORPH	MAE: 8.60
Yan *et al.* 2009 [245]	Active Appearance Model	FG-NET/1002	Classification	Leave-One-Person-Out	MAE: 5.21
Xiao *et al.* 2009 [271]	Active Appearance Model	FG-NET/1002	Classification	Train:300 Test:702	MAE: 5.04
Guo *et al.* 2009 [42]	Age Manifold, BIF	FG-NET/1002	Regression	Leave-One-Person-Out	MAE: 4.77
Guo *et al.* 2009 [42]	Appearance Model, BIF	YGA/8000	Classification	4-fold cross validation	MAE: 3.47 (M), 3.91 (F) CS: 88% (M), 85% (F)
Guo *et al.* 2009 [244]	Appearance Model(BIF) + Age Manifold	YGA/8000	Classification	4-fold cross-validation	MAE: 2.58 (M), 2.61 (F)
Guo and Mu 2011 [52]	BIF	MORPH II/55,132	Regression	50% Train, 50% Test	MAE:4.2
Choi *et al.* 2011 [8]	Active Appearance Model, Gabor, LBP	FG-NET/1002, PAL/430, BERC/390	Hybrid	Leave-One-Person-Out	MAE: 4.7(FG-NET), 4.3(PAL), 4.7(BERC) CS: 73%, 70%, 65% respectively
Luu *et al.* 2011 [249]	Contourlet Appearance Model	FG-NET/1002 PAL/443	Hybrid	Leave-One-Person-Out	MAE:4.1 (FG-NET), 6.0(PAL) CS: 74% (FG-NET)
Chang *et al.* 2011 [59]	Active Appearance Model	FG-NET/1002 MORPH II/5492	Classification	20% Test, 80% Train	MAE:4.5(FG-NET), 6.1(MORPH) CS:74.4%, 56.3% respectively

Table 2.2 Summary of. . . **Continued**

Publication	**Feature Representation**	**Face Database/ Database Size**	**Algorithm**	**Evaluation Protocol**	**Performance / Accuracy** (MAE, CS)
Wu *et al.* 2012 [14]	Grassmann Manifold of 2D shape	FG-NET/1002 Passport/223	Hybrid	Leave-One-Person-Out	MAE: 8.84 (Passport), 5.9 (FG-NET) CS: 40% (passport), 62% (FG-NET)
Thukral *et al.* 2012 [234]	Grassmann Manifold of 2D shape	FG-NET/1002	Hybrid	Leave-One-Person-Out	MAE:6.2
Chao *et al.* 2013 [247]	Active Appearance Model with distance metric adjustment	FG-NET/1002	Regression	Leave-One-Person-Out	MAE: 4.4
Lu and Tan 2013 [164]	Manifold of raw pixel intensities	MORPH II/20000	Regression	50% Test, 50% Train	MAE: 5.2(White), 4.2(Black) CS: 67%(White), 59%(Black)
Hadid	Pietikainen 2013 [252]	Raw intensities, Volume LBP	Internet videos/2000	Classification	10% Test, 90% Train CS: 83.1%
Guo and Mu 2013 [153]	BIF	MORPH II/55132	Regression	N/A	MAE: 4.0
Geng *et al.* 2013 [253]	Active Appearance Model, Label Distribution	FG-NET/1002 MORPH II/55132	Classification	Leave-One-Person-One (FG-NET), 10-fold cross-validation (MORPH II)	MAE: 4.8(FG-NET), 4.8(MORPH II)
Han and Jain 2014 [272]	BIF	LFW/143 FG-NET/1002	Classification	98.8% Train, 1.2% Test	MAE: 4.5(FG-NET) CS: 66.7% (LFW), 68.1% (FG-NET)
Han *et al.* 2015 [2]	Demographic informative features	FG-NET/1002 MORPH II/78, 207, PCSO/100,012, LFW/4,211	Hybrid	N/A	MAE:3.8(FG-NET), 3.6(MORPH II), 4.1(PCSO), 7.8(LFW) CS: 78%(FG-NET), 77.4%(MORPH II), 72.6%(PCSO), 42.5%(LFW)
Nguyen *et al.* 2015 [163]	MLBP, Gabor	PAL/1160	Regression	50% Test, 50% Train	MAE: 6.6
Li *et al.* 2015 [273]	2D Age Manifold, BIF	FACES/1,026 FG-NET/1002	Classification	N/A	MAE: 1.3(FG-NET), 8.2(FACES)
Onifade *et al.* 2015 [165]	Age-Rank LBP	FG-NET/1002	Regression	Leave-One-Person-Out	MAE: 2.34
Yang *et al.* 2015 [274]	Raw intensities	Private/3,615	Classification	2479 Train, 1,136 Test	MAE: 0.31
Huerta *et al.* 2015 [159]	HOG, LBP, SURF	FG-NET/1002	Classification	Leave-One-Person-Out	MAE: 3.31

Table 2.2 Summary of... **Continued**

Publication	**Feature Representation**	**Face Database/ Database Size**	**Algorithm**	**Evaluation Protocol**	**Performance / Accuracy** (MAE, CS)
Hu *et al.* 2016 [162]	Kullback-Leibler / raw intensities	LFW/150,000 MORPH II/55,132 FG-NET/1002	Classification (CNN)	N/A	MAE: 2.8(FG-NET), 2.78(MORPH)
Akinyemi and Onifade 2016 [166]	Raw pixel intensities	FG-NET/1002 FAGE/238	Regression	Leave-One-Person-Out	MAE: 3.19

2.9 Summary

Comprehensive survey of various techniques and approaches used for age estimation has been presented. There has been enormous effort from both academia and industry dedicated towards modelling age estimation, designing of algorithms, ageing face dataset collection and protocols for evaluating system performance. Table 2.2 summarizes the findings of recent studies in age estimation, evaluation protocol used, dataset used, age estimation approach used (regression, classification or hybrid) and feature extraction or age face representation used.

The main issues to consider in age estimation via faces is image representation and estimation techniques. AAM provide a parametric modelling for face representation. A face is represented as set of shape and texture parameters learnt from a face image. AAM can represent both young and old faces since model parameters encodes both facial shape and texture. AAM is often used in line with regression-based age estimation approaches. Anthropometric face representation encode change in facial shape. Anthropometric approaches to facial representation can be very significant in capturing change in facial shape in young faces. AGES can be used to extract subjects ageing patterns when a dataset has sequential ageing face images while age manifold is convenient when a dataset has missing ageing face images in a large age dataset with wide age ranges. Age manifold learning entwines ageing feature extraction and dimensionality reduction. Age manifold can be used both in classification and regression based approaches. Appearance models often extract facial features that can be used in regression or classification based age estimation approach. These features represent facial appearance. These features could be texture, shape or wrinkle. Feature extraction techniques like LBP, Gabor, BIF, LDA, PCA, and LDP have been often used for appearance face modelling.

Age estimation can be either approached as age-group estimation or exact age estimation. Age-group estimation approaches approximate age range in which a face image can fall. Exact age estimation approaches estimate a single label (value) that represent the age of a face image. Both exact age and age-group estimation can be either classification-based, regression-based or hybrid of both classification and regression. Choice between regression and classification may be guided by face image representation and size and age distribution of the dataset. For big datasets with sequential age-labels, both classification and regression can be used while for datasets with only age-group labels or significantly missing images at some ages, classification-based approach may be more appropriate. Both classification and regression can be combined in a hierarchical manner. In this hybrid approach, often classification is used for age-group estimation followed by exact age estimation within the age-group using regression techniques.

Age estimation techniques can be evaluated using Mean Absolute Error (MAE) or Cumulative Score (CS). MAE is appropriate when training set has a lot of missing

ages while CS is used when training dataset has samples at almost every age. Overall performance of the system is represented by CS. In practice, both MAE and CS are used because different techniques, datasets may be biased for evaluation. The most often used evaluation protocols are LOPO and Cross-Validation. This chapter has been published as a survey paper in [275].

Chapter 3

Materials and Methods

This chapter presents an overview on hierarchical age estimation model adopted for this research. Techniques used for ageing feature extraction are also described and dataset used presented.

3.1 Hierarchical age estimation model

This section presents Hierarchical Age Estimation Model (HAEM). HAEM is a conceptual model that characterizes functions of age estimation system modules without regard of the technologies used in each of the module. HAEM has 6 layers with each layer using output of the previous layer. The modules that make up HAEM are image preprocessing, feature extraction, feature enhancement, gender estimation, age-group estimation and age estimation. Figure 3.1 shows the structure of proposed hierarchical age estimation model.

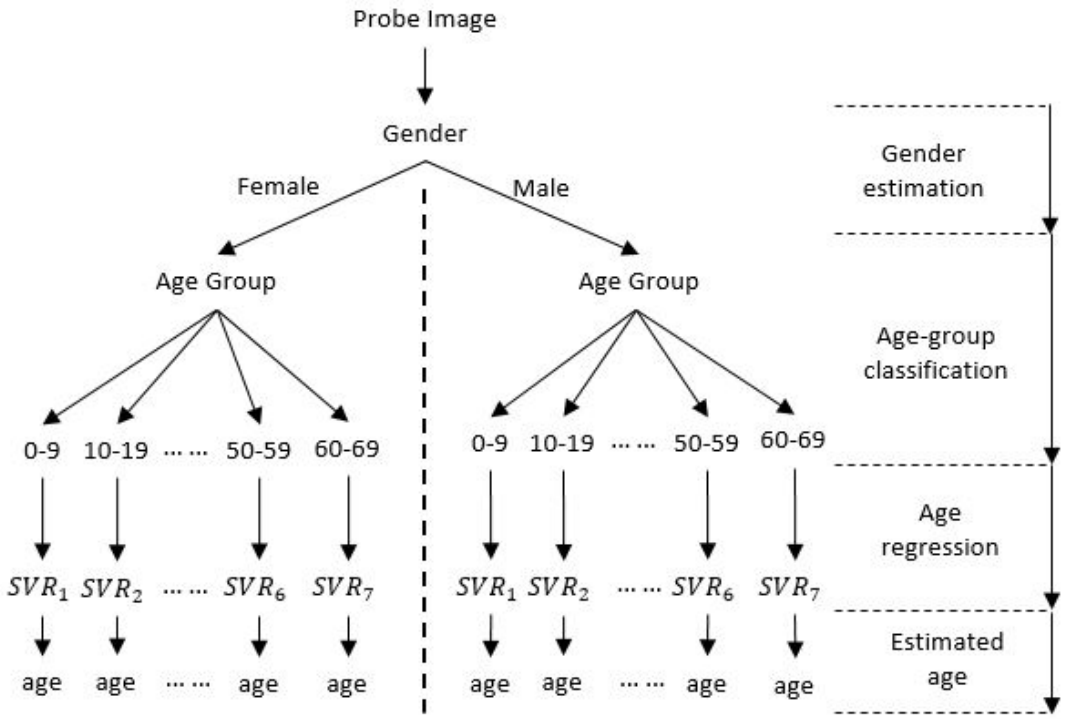


FIGURE 3.1: Hierarchical age prediction

In this model, the first step is gender estimation followed by within-gender age-group estimation. Age-group estimation is followed by exact age estimation.

3.1.1 Image preprocessing

The objective of image preprocessing is to enhance the image such that the resultant image is more suitable than the original image in specific application [140]. Image preprocessing can be done is spatial or frequency domain or in combination of both. Image preprocessing module enhances image quality before face and landmark localization. Preprocessing may include image alignment, rotation, illumination equalization and smoothing. It is important to smooth an image before feature extraction in order to remove noise and improve features like edges and line.

3.1.2 Face detection and landmark localization

After an image has been preprocessed, facial part of the image is automatically detected an cropped. There are a number of face detection algorithms that can be used such as Viola and Jones [169] Haar-cascade classifiers. After detection of face, facial features can be extracted using holistic or local feature extraction techniques. Facial landmarks like eyes, nose, mouth, cheeks, nose-bridge, forehead, eye-brows e.t.c could be located in the detected face. Holistic and local feature extraction techniques are applied on the whole face or the detected landmarks for ageing features extraction.

3.1.3 Feature extraction

The feature extraction module implements holistic and local feature extraction techniques. For both face and facial components, appearance and shape features could be extracted. Shape features could be extracted using techniques like ASM, SIFT, SURF, Zernike moments among others. Holistic appearance features could be extracted using PCA, LDA, AAM among other holistic feature extraction techniques. Local appearance features could be extracted using techniques like LBP, LTP, LDP, Gabor wavelets, LTDP, SOR-LDP, LTDP, LS-BIF, MF-BIF, LFA among others. Both appearance and shape features could be enhanced before being used in classification.

3.1.4 Feature enhancement

Feature enhancement involves improving discriminative power of the extracted features. This enhancement could be done using weak classifiers like Haar-cascade or random trees for feature selection before using the features for pattern recognition. Feature fusion is also another technique that could be used to improve discriminative power of features. A boosting algorithm like AdaBoost could also be used to enhance features before pattern recognition. Dimensionality reduction is treated as a feature enhancement process because it is aimed at reducing the number of features while still maintaining or even improving its discriminative capabilities.

3.1.5 Gender estimation

It has been found that males and females age differently [40]. Ageing patterns observed in males are not similar to those observed in females with females appearing younger than their actual age. Due to these variations in ageing between males and females, it is important to perform gender profiling prior to age estimation. Intuitively, several factors are considered by humans when estimating one's age and one of these factors is gender. Learning gender specific ageing patterns results into improvement of age estimation accuracies since the learnt patterns are not scattered in a wide diverse space. In automatic age estimation, gender could be estimated using classification approaches based on holistic or local features or a combination of the two. Higher accuracies in gender estimation results into improved age estimation results.

3.1.6 Age-group estimation

After gender estimation, estimate the age-group in which the face image belongs. With higher accuracies in age-group estimation, age estimation errors are expected to reduce. Narrow age-groups preferable of below 5 years range could post better results compared to wider age-groups of above 10 years range. However, narrow age-groups require dataset with many individuals with images at almost every age forming a complete ageing pattern. Missing images in a dataset results to an incomplete ageing pattern and is one of the reasons why wider age-groups may be preferred to narrow age-groups. However, narrow age-groups can still be used in a dataset that has images at almost every age, regardless whether these images belong to the same subject or not.

3.1.7 Age estimation

After gender and age-group profiling, exact age is estimated within the age-groups of males and females respectively. Classification, regression or hybrid approach that consists of both can be used to build age estimation model. The age estimation model could be made using an ensemble of classifiers with features from different landmarks.

3.2 Colour and gray scale image

An image may be defined as a two-dimensional function, $f(x, y)$, where x and y are spatial coordinates, and the value of f at any (x, y) is a positive scalar quantity that represent the intensity (colour value) of the image at that point. When x, y, and the amplitude values of f are all finite discrete quantities, the image is refereed to as a digital image. In digital images, entries in coordinates (x, y) are refereed to as *pixels* and they represent a colour value at that coordinate. In gray-scale images, the pixels represent different shades of gray colour. Therefore, a digital gray-scale image can be defined as $0 \leq f(x, y) \leq 255$ with 0 representing black and 255 representing white. Colour images are represented as a tensor of order 3. The third order of this tensor is an array whose dimensional is equal to number of colours used to represent the image. For instance, RGB color image can be represented as order 3 tensor as there are three entries at each coordinate (x, y); for red, green and blue, while CMYK images have four entries at each coordinate; for cyan, magenta, yellow and black. Although color images may give more information compared to gray-scale images, it is computationally expensive to process these images; hence they are often converted to gray-scale before processing. Figure 3.2 shows how color and gray-scale images are represented.

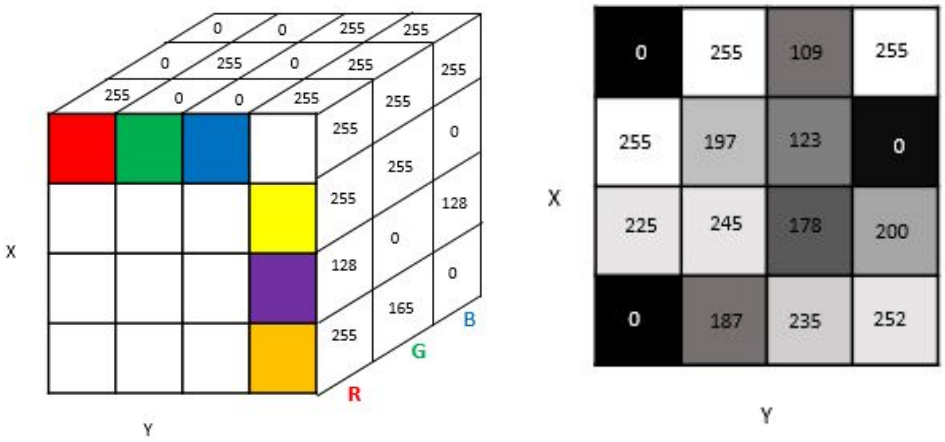
There are several methods used to convert color images to gray-scale among them lightness, average, and luminosity. Lightness methods represents the image as an average of most prominent and least prominent colors as

$$f(x, y) = \frac{(\max(R, G, B) + \min(R, G, B))}{2} \tag{3.1}$$

while average method gets average of the channels as

$$f(x, y) = \frac{(R + G + B)}{3} \tag{3.2}$$

Luminosity gets a weighted average of the color values to account for human perception. Human visual system is more sensitive to green compared to red and blue,



(a) RGB color image represented as tensor (b) Gray scale image represented as a matrix

FIGURE 3.2: Representation of digital color and gray-scale images

therefore, green channel is always weighted higher than red and blue channels[1]. Luminosity can be defined as

$$f(x, y) = 0.299R + 0.587G + 0.144B \tag{3.3}$$

Information contained in the image can be used to determine some visual artifacts like texture and shape of a scene or object. However, this information could be affected by noise, poor illumination, and contrast. To extract meaningful information from digital images, image processing techniques are applied to enhance image information. Image processing techniques are used for noise removal, contrast adjustments, edge detection, region detection and segmentation. An enhanced image is then analyzed. Image analysis is the processing of extracting meaningful information from digital images [276]. The information extracted from these images are used to train computer models for computer vision and pattern recognition among other tasks. A typical computer vision or pattern recognition system consist on image acquisition module, region of interest recognition or segmentation, image enhancement, feature extraction and analysis, model learning and data storage. This study focused on image processing module, feature extraction and model learning modules for pattern recognition.

3.3 Image preprocessing

The image is enhanced through preprocessing to make it more suitable for ageing feature extraction. In this study, images are processed in spatial domain where all operations are done directly on the image pixels.

The image is first converted to single channel gray scale using luminosity method and then smoothened using 2D Gaussian spatial filter defined as

$$F(a, b) = \frac{1}{2\pi\sigma^2} e^{\left(-\frac{a^2+b^2}{2\sigma^2}\right)} \tag{3.4}$$

[1]The weights vary from weighted method to another. But common thing in all the weighted methods is that green is always weighted higher than the other two colors, followed by red and blue receives lowest weight. This is the commonly used method for converting color images to gray scale images

where a and b are the displacements from origin in the horizontal and vertical axes respectively and σ is the standard deviation of the Gaussian envelope. Given an image $I(x, y)$, an enhanced image $I_G(x, y)$ of size $m \times n$ is found by performing a convolution of $F(a, b)$ with $I(x, y)$ as

$$I_G(x, y) = \sum_{a=0}^{m-1} \sum_{b=0}^{n-1} F(a, b)\, I(x - a, y - b) \tag{3.5}$$

The smoothed image of $I(x, y)$ is $I_G(x, y)$.

Face is detected using AdaBoost face detector proposed in [277]. In some situations the face was detected using Haar-cascade [216] [169] [278]. Figure 3.3 shows input image, smoothened image and detected face. Ageing facial features are extracted

(a) Input image (b) Detected Face (c) Gray Face (d) Smoothened

FIGURE 3.3: Image pre-processing

from each of the smoothened gray-scale face image and its components and used for age estimation.

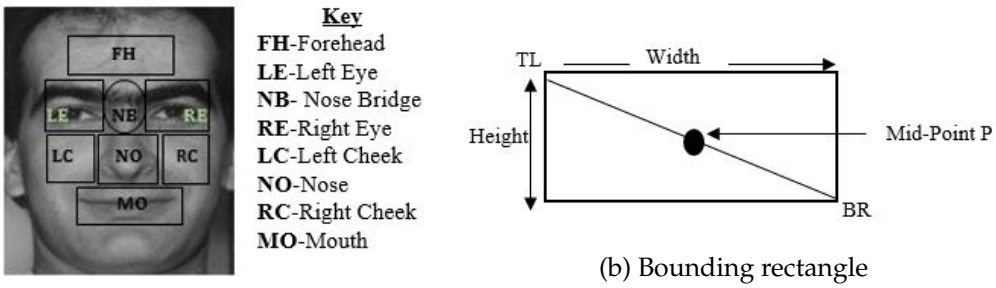
3.4 Landmark localization approach

This section describes the technique used to achieve the first objective of developing an automatic landmark localization technique that can locate both fiducial and secondary facial landmarks for facial computing.

Fast ROI detection technique that uses Haar-Cascade classifiers to detect eyes, nose and mouth and geometric approximations to locate cheek bone area and forehead is proposed. Haar-cascade [277] and LBP-cascade [279] are used for face detection. Once the face is detected, the technique start by detecting eyes using Viola and Jones [216] Haar-Cascade eye detection classifier. High speed and high detection rate of Viola and Jones classifier guided its choice for this study. The rectangular region around, and including the eyes is detected and cropped. Haar-Cascade classifier [280] is used to detect nose and mouth separately. The bounding rectangles around the nose and the mouth are found and ROIs related to them are detected. Due to high chances of false detections for nose and mouth, threshold is used to ensure that the mouth is always below the nose. The center point between the two detected eyes p is calculated. The Euclidean distance d between the centre of gravity of the two bounding rectangles around the eyes is calculated. Nose-bridge is located as a circular region centred at point p with radius $0.25 \times d$. Forehead is defined as rectangular region above eyes with top-left corner at point (x_1, y_1) bottom-right corner at point (x_2, y_2). Coordinate values x_1 and x_2 correspond to x coordinates of left eye and right eye respectively. Coordinate values and y_1 and y_2 are experimentally derived from geometric information of eyes.

The top-left corner of the cheek bone area is found as point $(x, y + h)$ where x and y are the x and y coordinate of the top-left corner of respective eye, h is the height of the eye. Bottom-right corner of cheek bone area is found as point $(x + w, y + 1.5 \times h)$

where w is the width of the respective bounding rectangle around the eye. Shifting factor of y-coordinate is determined experimentally and is found to be closer to 1 for top-left corner and varies between 1.5 to 2.0 for bottom right corner. As shown in Algorithm 1, the contribution of this work is the localization of left and right cheek bone areas, forehead and nose-bridge using geometric and spatial information of fiducial landmarks. Figure 3.4 the facial landmarks localized and how they are annotated in this study. For each landmark, a rectangular bounding box shown in Figure 3.4b is returned.



(a) Landmarks localized

(b) Bounding rectangle

FIGURE 3.4: Annotation of landmarks to be located

Top-left corner (TL) and bottom-right corner (BR) of fiducial landmark rectangles are tracked because they are used to geometrically approximate TL and BR of secondary landmarks (LC, RC and FH). Top-left corner of RE is considered as RE.TL and its coordinates are $RE.TL_x$ and $RE.TL_y$ respectively.

Left and right eyes are detected using Haar-cascade classifier [216] and geometric and spatial information of LE is used to detect LC and that of RE to detect RC. Distance d between LE and RE is calculated. The mid-point p between LE and RE is gotten and used to locate nose bridge. Nose bridge is located as rectangular region that inscribe a circle of radius $0.25 \times d$ centred at p. The centre of LE denoted as $LE.C$ and centre of RE denoted as $RE.C$ are derived. Forehead is located as a rectangular region above eye-brows and tightly bounded between eye centres. The y-coordinates of the forehead top left corner $FH.TL$ and bottom right $FH.BR$ are empirically derived from the y-coordinates of left eye centre and right eye centre respectively. The y coordinate of top left corner of forehead is found to vary between $LE.C_y - (1.40 \times h)$ and $LE.C_y - (1.50 \times h)$ where h is height of LE. $FH.BR_y$ is found to vary between $RE.C_y - (0.50 \times h)$ and $RE.C_y - (0.60 \times h)$ where h is height of RE.

Top-left corner of right cheek RC.TL is set to have same x-coordinate as $RE.TL_x$ and y-coordinate equal to $RE.BR_y$. Therefore, RC.TL coordinates are found to be $(RE.TL_x, RE.BR_y)$. RC.BR is found by shifting $RC.TL_x$ by width of RE width and $RC.TL_y$ by height of RE respectively. Therefore, $RC.BR$ is found to be $(RC.TL_x + w, RC.TL_y + h)$ where h and w are height and width of RE respectively. The same approach was followed to locate LC using spatial and geometric information of LE. NO and MO are detected using Haar-cascade classifier in [280] [281]. The summary of the proposed method is presented in Algorithm 1.

In Algorithm 1, the image is loaded and converted into gray scale, single channel image. Before any detection is done the image is first smoothened to remove noise. The image is then passed to a face detection cascade classifier that returns a bounding rectangle around a detected face. The face region of the image is cropped and landmarks detected. Spatial and geometric information of eyes are found and used to locate cheeks, forehead and nose-bridge. Experimental results of this technique are presented in section 4.1. The proposed landmark localization approach is published in [282].

Algorithm 1: Face landmark localization and extraction

input : Face image I
output: LM(FH, LE, NB, RE, LC, RC, NO, MO)

Convert image I to grayscale;
Smooth image I: Reduce noise by applying a filter;
$F \leftarrow extractFace(I)$: Haar-Cascade(I) | LBP-Cascade(I);
$LE|RE \leftarrow extractEyes(F)$: $Haar - Cascade(F)$;
for *Every eye i locate cheek i* **do**
    $ETL_x \leftarrow (LE.TL_x|RE.TL_x)$: Current eye TL_x;
    $ETL_y \leftarrow (LE.TL_y|RE.TL_y)$: Current eye TL_y;
    $EBR_x \leftarrow (LE.BR_x|RE.BR_x)$: Current eye BR_x;
    $EBR_y \leftarrow (LE.BR_y|RE.BR_y)$: Current eye BR_y;
    $\mathbf{LM} \leftarrow \mathbf{RectangleCE}(\mathbf{ETL_x}, \mathbf{ETL_y}, \mathbf{EBR_x}, \mathbf{EBR_y})$: eye^i;
    $Eye - Length(EW) \leftarrow EBR_x - ETL_x$: eye^i width;
    $Eye - Width(EH) \leftarrow ETL_y - EBR_y$: eye^i height;
    $EC_i \leftarrow (ETL_x + 0.5 \times EW, ETL_y + 0.5 \times EH)$;
    $CTL_x \leftarrow (ETL_x)$: Current cheek TL_x;
    $CTL_y \leftarrow (ETL_y + EW)$:Current cheek TL_y;
    $CBR_x \leftarrow (EBR_x)$: Current cheek BR_x;
    $CBR_y \leftarrow (EBR_y + EW)$:Current cheek BR_y;
    $\mathbf{LM} \leftarrow \mathbf{RectangleCC}(\mathbf{CTL_x}, \mathbf{CTL_y}, \mathbf{CBR_x}, \mathbf{CBR_y})$: $cheek^i$;
$d \leftarrow EuclideanDistance(EC_i, EC_{i+1})$: distance between eyes;
$Point\ P \leftarrow ((LE.EMP + RE.EMP)/2)$: Midpoint of 2 eyes;
$NB.TL \leftarrow ((P_x - (0.25 \times d)), (P_y - (0.25 \times d)))$;
$NB.BR \leftarrow ((P_x + (0.25 \times d)), (P_y + (0.25 \times d)))$;
$\mathbf{LM} \leftarrow \mathbf{RectangleNB}(\mathbf{NB.TL}, \mathbf{NB.BR})$;
$FH.TL \leftarrow (LE.C_x, (LC.C_y - 1.50 \times EH))$;
$FH.BR \leftarrow (RE.C_x, (RE.C_y - 0.50 \times EH))$;
$\mathbf{LM} \leftarrow \mathbf{RectangleFH}(\mathbf{FH.TL}, \mathbf{FH.BR})$;
$NO.TL \leftarrow Haar\_cascade(I)$; $NO.BR \leftarrow Haar\_cascade(I)$;
$\mathbf{LM} \leftarrow \mathbf{RectangleNO}(\mathbf{NO.TL}, \mathbf{NO.BR})$;
$MO.TL \leftarrow Haar\_cascade(I)$; $MO.BR \leftarrow Haar\_cascade(I)$;
$\mathbf{LM} \leftarrow \mathbf{RectangleMO}(\mathbf{MO.TL}, \mathbf{MO.BR})$;

3.5 Landmark displacement approach

This section presents the description of the technique used to find out the trend of facial landmark displacement across age. The technique described here was used to achieve objective 3 of the study.

After face detection and preprocessing presented in section 3.3, landmark localization is performed as detailed in section 3.4. Landmark centroids[2] are then determined using spatial and geometric information of the bounding rectangle (circle) returned by Algorithm 1. For instance, the centroids of the Left-Eye (LE) $LE_c\,(x, y)$ is determined as

$$LE_c\,(x, y) = LE.TL_x + \frac{W}{2}, LE.TL_y + \frac{H}{2} \tag{3.6}$$

[2]The centre of mass of a geometric object of uniform density.

where $LE.TL_x$ and $LE.TL_y$ are x and y coordinates of top-left corner of the LE bounding rectangle, W and H are the width and height of the rectangle respectively. As denoted in Figure 3.5, nose (NO), mouth (MO), left-eye (LE, right-eye (RE), forehead (FH) centroids and midpoint of the two eyes (ME) are used to investigate landmark displacement across age. Ten Euclidean distances between landmark

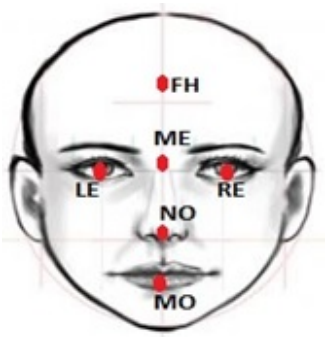


FIGURE 3.5: Annotation of landmark centroids

centroids are calculated. Table 3.1 shows landmark centroid pairs used for each distance and it description.

TABLE 3.1: Definition of distances between landmark centroids

Distance	**Description**
D(LE, RE)	Distance from left-eye to right-eye
D(LE,NO)	Distance from left-eye to nose
D(LE,MO)	Distance from left-eye to mouth
D(RE,NO)	Distance from right-eye to nose
D(RE,MO)	Distance from right-eye to mouth
D(MO,FH)	Distance from mouth to forehead
D(NO,FH)	Distance from nose to forehead
D(ME,MO)	Distance from middle of eyes to mouth
D(ME,NO)	Distance from middle of eyes to nose
D(MO,NO)	Distance from mouth to nose

From the distances in Table 3.1, 7 ratios were calculated, 4 for horizontal distances and 3 for vertical distances. These ratios were calculated as shown in equation 3.7 through 3.13. Using D (X,Y) to denote Euclidean distance between X and Y, where X and Y are arbitrary facial landmark centroids, the ratios defined are

$$Ratio_1 = \frac{D\,(LE, RE)}{D\,(LE, NO)} \tag{3.7}$$

$$Ratio_2 = \frac{D\,(LE, RE)}{D\,(RE, NO)} \tag{3.8}$$

$$Ratio_3 = \frac{D\,(LE, RE)}{D\,(LE, MO)} \tag{3.9}$$

$$Ratio_4 = \frac{D\,(LE, RE)}{D\,(RE, MO)} \tag{3.10}$$

$$Ratio_5 = \frac{D\,(ME, NO)}{D\,(NO, FH)} \tag{3.11}$$

$$Ratio_6 = \frac{D\left(ME, MO\right)}{D\left(MO, FH\right)} \tag{3.12}$$

$$Ratio_7 = \frac{D\left(MO, NO\right)}{D\left(MO, ME\right)} \tag{3.13}$$

These ratios were tracked across age to establish landmark displacement across age. Distances ratios were used to make the measurements to be invariant to changes in image scale and aspect ratio. This is because the two aspects would alter distance between landmarks.

Distance ratios were used to establish how landmarks drift from each other as one grows. Landmark displacement investigated are the drifting of the eyes from the nose, eyes from the mouth, forehead from the nose, forehead from the mouth and nose from mouth. The images used were grouped into age groups of 5 years. The distances and ratios were calculated for each individual for all the age groups. Each ratio in every age group was summed and averaged as per the number of images in that age group to get mean ratio for each of the 7 ratios in a particular age group. The mean ratio was then analysed against each age group. Means of each ratio were tracked across all age groups to establish the trend in landmark displacement across age.

3.5.1 Facial landmark relativity

Facial landmarks are located at some point in a particular direction from another landmark. Generally nose and mouth are located below eyes with nose being above the mouth. Dehshibi and Bastanfard [74] showed that nose is located $0.6 \times D_{eyes}$ from eyes where D_{eyes} the distance between eyes is. They further showed that the mouth is at position $0.85 \times D_{eyes}$ below the eyes. This findings do not specify exact position of a landmark as a landmark is not only at particular distance from another landmark, but also at a particular angle relative to other landmark. Therefore, an accurate location of a landmark from another landmark should be a vector consisting of a magnitude (distance), and direction (angle). In this study, relative positions of a nose and mouth are represented from the position of eyes. Euclidean distance between landmarks were calculated. Ratios of distance between eye centroids and distance between nose, mouth centroid and eye centroids were calculated as

$$Ratio_{nose\_eye} = \frac{D_{nose\_eye}}{D_{eye\_eye}} \tag{3.14}$$

$$Ratio_{mouth\_eye} = \frac{D_{mouth\_eye}}{D_{eye\_eye}} \tag{3.15}$$

where D_{nose_eye} is the distance between eyes centroids and nose centroid, D_{mouth_eye} is distance between mouth centroid and eyes centroids and D_{eye_eye} is distance between left and right eye.

In order to give comprehensive relative position of nose and mouth from the eyes, the angle between nose and eyes, mouth and eyes were also calculated. Figure 3.6 annotates eye, nose and mouth relative positions.
Angle θ, between nose and right eye and angle β, between mouth and right eye were calculated as

$$\beta|\theta = \left(\arctan\left(\frac{|\Delta y|}{|\Delta x|}\right)\right) \times \frac{180}{\pi} \tag{3.16}$$

Where for θ, $\Delta y = RE.y - NO.y$ and $\Delta x = RE.x - NO.x$ and for β, $\Delta y = RE.y - MO.y$ and $\Delta x = RE.x - MO.x$, and $\pi = 3.1415926536$.

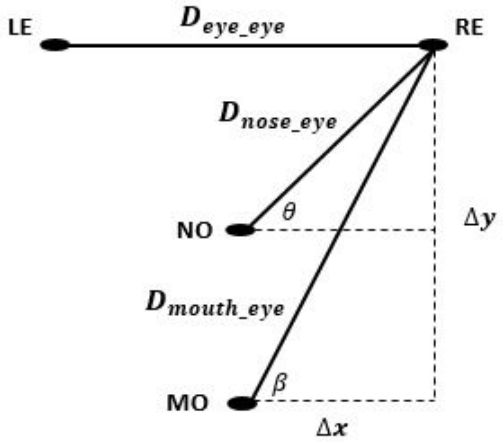


FIGURE 3.6: Landmark relativity; Distance and direction

3.5.2 Achieving rotation and translation invariance

An image is defined as a function $f(x, y)$ where x and y represent spatial coordinates of each pixel in the image in Euclidean space. Rotation occurs when every point in a plane turns about a fixed point and with the same angle. Image rotation causes change in x and y pixel positions and angle between facial landmarks. Nose centroid N, midpoint of two eyes P and mouth centroid M should be on straight vertical line such that there coordinates are $N(x, y_n)$,$P(x, y_p)$ and $M(x, y_m)$. The angle δ between these points should be zero before calculating angles between nose and eyes, and mouth and eyes. Translation and rotation is performed before calculating angles between landmarks if $\delta \neq 0$.

Translation $T(-x, -y_n)$ is applied to the image such that the nose centroid $N(x, y_n)$ is at the origin of Euclidean plane and the angle δ between point $N(x, y_n)$ and point $P(x, y_p)$ is calculated. If $\delta \neq 0$, then the image is rotated by angle δ. Image transformation M is represented as

$$M = T(-x, -y_n) R \begin{bmatrix} \cos\delta & -\sin\delta \\ \sin\delta & \cos\delta \end{bmatrix} \tag{3.17}$$

Translating the nose centroid to the origin makes the process of calculating angles between landmarks translation invariant. If $\delta \neq 0$, the image is rotated making angles between landmarks obtained rotation invariant.

3.5.3 Modelling age related data as a time series

Age is the number of years elapsed since birth to a particular time in one's life. This value is expressed as an integer. Time series is a sequence of data points taken at successive equally spaced time intervals. When information extracted from age separated images is arranged in the order of the subjects' age, it can be modelled as a time series. Age of an individual at time t_1 is greater than age at time t_0 given that $t_1 > t_0$. Therefore, information extracted from the face image of a person at time $t_0, t_1, t_3, \ldots t_n$ can be organized in a time series with ages of these individuals at each respective time representing a time point.

It is not easy to get a data set with sequential age of each individual. It is because of this reason that in this study, subjects are organized in age-groups of 5 years defined

as

$$Age_{group} = \lfloor \frac{age}{5} \rfloor + 1 \tag{3.18}$$

where $\lfloor \cdot \rfloor$ is a floor operator. Applying equation 3.18 on FG-NET dataset, the subjects were grouped into 14 age groups with first age group being $0-4$ years and last age group $65-69$. The label for the first and last group is 1 and 14 respectively. Data obtained from the images could now be organized as a time series in the form of $t_1, t_2, t_3 \ldots t_{14}$.

Each ratio was calculated and put in a feature vector $\vec{R_i} = [x, y]$ where x represents the ratio, y represent age-group label and i is the ratio index for $i = 1, 2, \ldots, 7$. This results into 7 feature vectors. Each feature vector was supplied into an SVR for training a regression model. Seven models were trained and evaluated. Experimental results of this technique are presented in section 4.2. Proposed landmark displacement across age technique is published in [283].

3.6 Age-group estimation approach

This section describe the approach used to achieve objective four. A model for age-group estimation based on fused facial and facial component features and ensemble of classifiers is described.

Given a face image, global shape features S_i^j and global appearance features A_i^j for subject i at age-group j are extracted. A_i^j and S_i^j are fused to get a single global feature vector f_G. f_G is used to train a classifier for age-group estimation based on global feature vector. Face image is divided into R where $R > 1$, regions of interest (ROI).

For each ROI R_r for $r = 1, 2, \ldots R$, where R is number of ROIs, local wrinkle feature $W_{i,r}^j$ and local texture feature $T_{i,r}^j$ of subject i at age-group j are extracted. $W_{i,r}^j$ and $T_{i,r}^j$ are fused into a single local feature vector f_L^r for ROI r. A classifier is trained on each of f_L^r A total of $T = R+1$ classifiers are trained. Each of these classifiers returns age-group label for respective feature vector. Figure 3.7 shows a block diagram of the proposed age-group estimation method. Feature level fusion is adopted since

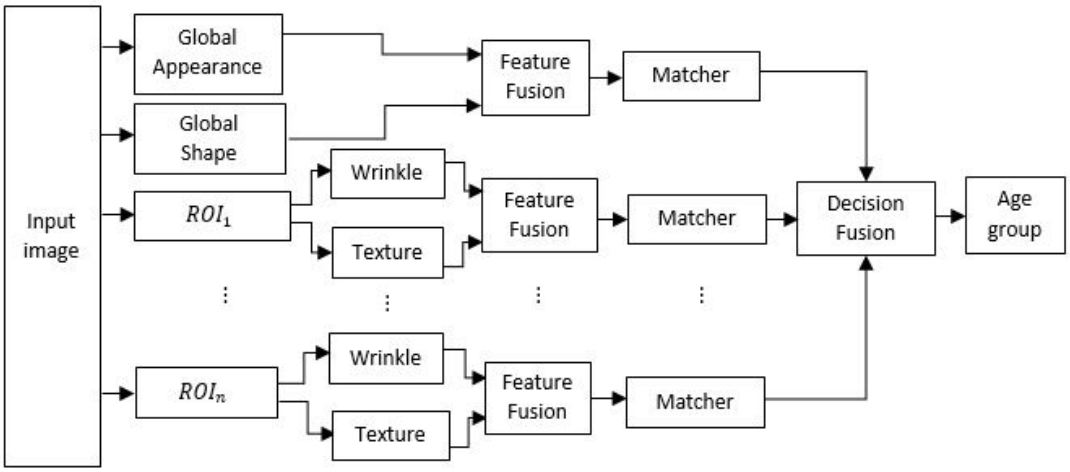


FIGURE 3.7: Age-group estimation model

it results into a richer feature representation for better classification [284]. Decision fusion is done using simple majority vote rule.

Global shape features $f_1 = (s_1, s_2, s_3 \ldots s_n)$, global appearance feature vector $f_2 =$

$(a_1, a_2, a_3 \dots a_x)$, local wrinkle features $f_3 = (w_1, w_2, w_3 \dots w_p)$ and local texture features $f_4 = (t_1, t_2, t_3 \dots t_m)$ are extracted. The values $n > 0, x > 0, p > 0$ and $m > 0$ represent dimensionality of each respective feature vector. Feature dimensionality is normalized using LDA. Z-score normalization is done on these features before fusion. A feature f_i is normalized as

$$\bar{f_i} = \frac{f_i - \mu_i}{\sigma_i}, i = 1, 2, 3, 4 \tag{3.19}$$

where f_i is i^{th} feature vector, μ_i is mean and σ_i is the standard deviation of f_i, $\bar{f_i}$ is the i^{th} normalized feature. We thereafter fused feature f_1 with feature f_2 and feature f_3 with feature f_4. Fusion is done by concatenating features as

$$\begin{aligned} f_g &= [f_1 \oplus f_2] \\ f_l &= [f_3 \oplus f_4] \end{aligned} \tag{3.20}$$

where f_g and f_l represent fused global feature vector and fused local feature vector respectively and $\oplus$ is a concatenation operator. Feature fusion often lead to high dimensional feature vectors leading to 'curse of dimensionality' problem. Thus, each of the fused feature vectors dimensionality is reduced using PCA.

Decision fusion is performed using majority voting scheme. Given number of classifiers as T with each classifier having a single vote, predicted class is one with T votes, or one with $\lfloor \frac{T}{2} \rfloor + 1$ votes or the class with simple majority vote [285]. The final decision is; select class ω_x if

$$x = \arg \max_{j=1}^{c} \sum_{t=1}^{T} d_{t,j} \tag{3.21}$$

Assumption in this technique are that T is odd and that for any instance x, probability p that a classifier will give correct class label is independent for each classifier [181]. Simple majority voting rule is used to get final output of ensemble. Experimental results of the proposed technique are presented in 4.3 and the approach is published in [286].

3.7 Age estimation approach

Three approaches of hierarchical age estimation are investigated

- Hierarchical age estimation using a single feature vector for age-group classification followed by within-group age regression
- Hierarchical age estimation using fused features for age-group classification followed by within-group age regression
- Hierarchical age estimation using fused features for gender classification, followed by gender-based age-group classification followed by within gender and age-group age regression

Figure 3.8 shows hierarchical age estimation using MF-BIF features.
Using MF-BIF for hierarchical age estimation, MF-BIF features are extracted from the face and used for age-group classification with Artificial Neural Network (ANN) Multilayer Perceptron (MLP) followed by age regression with SVR. Images are classified into groups of 10 years as shown in Table 3.2. Age-group classification is followed by age regression where ageing patterns are learnt for each group in a narrow range of 10 years. MF-BIF was enhanced using fusion techniques and used as

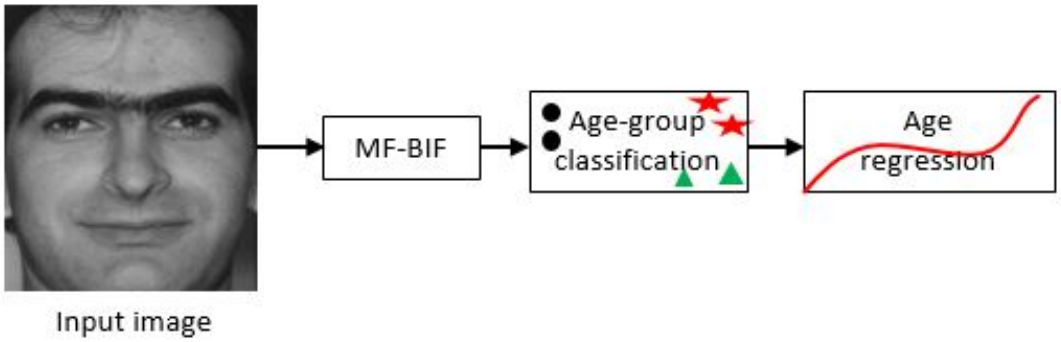


FIGURE 3.8: Hierarchical age estimation using MF-BIF

shown in Figure 3.9 for hierarchical age estimation. MF-BIF is enhanced by fusing

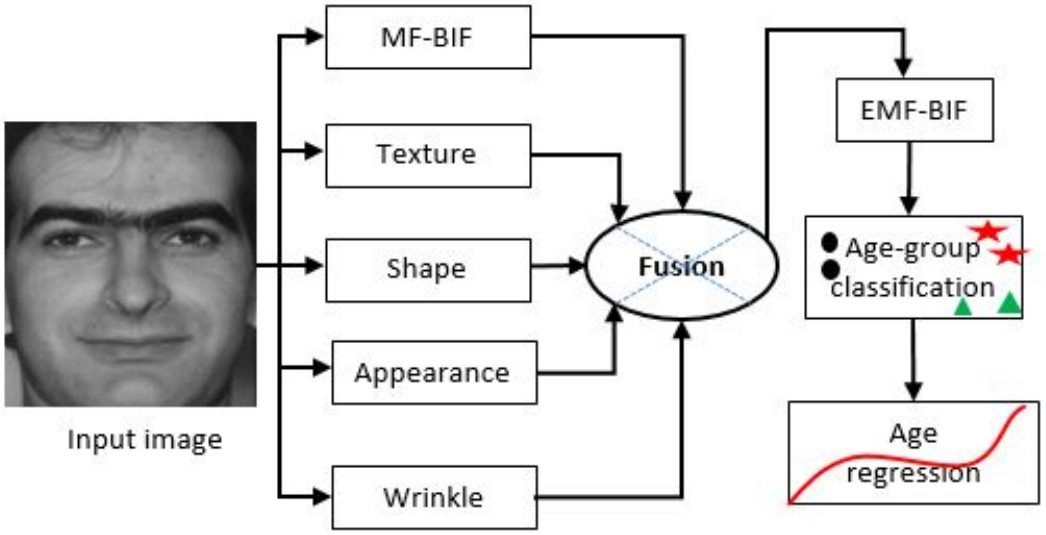


FIGURE 3.9: Hierarchical age estimation using EMF-BIF

it with texture, wrinkle, shape and appearance features to obtain EMF-BIF features. The EMF-BIF features is used for age-group classification using ANN followed by within-group age regression using SVR. Age regression is performed on a narrow age-range within each group.

As shown in Figure 3.10, this study also sought to find out how gender discrimination improves age estimation accuracies. It has been found that ageing patterns vary between males and females [9] [94] [92]. Investigation into differences in ageing between males and females is necessary [91]. This prompted us to incorporate gender estimation into the age estimation process to try and improve age estimation accuracies [287]. Images are first classified as male or female. In each gender-group, age-group classification is performed to group images into particular age-groups of 10 years. This is followed by within-group (and within gender) age regression for exact age estimation.

One model is trained for gender classification and two models for age-group classification; one for males and the other for females. For each gender, seven regression models are trained, one for each age-group. Regression models are trained to learn ageing patterns within the respective age-group.

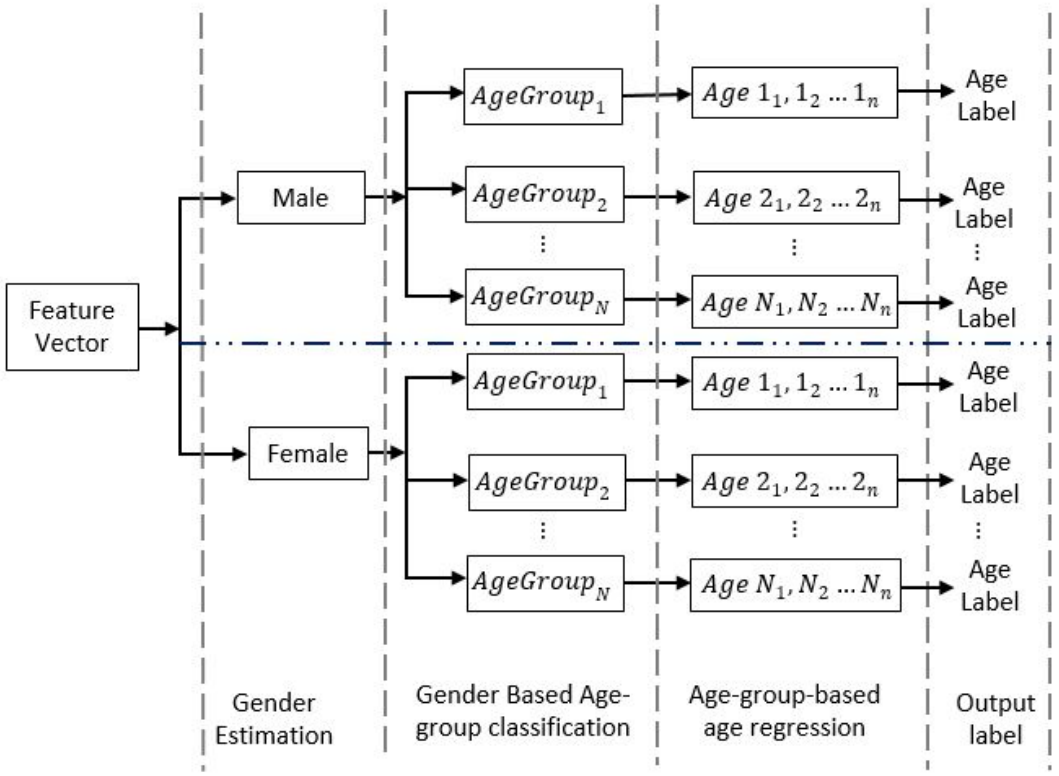


FIGURE 3.10: Gender and age-group based age estimation using fused features

At prediction, gender of the subject in the probe image is first determined. If gender is male, male age-group estimation model is used for age-group estimation, otherwise female age-group estimation model is used. The age-group in which the probe image falls is determined and a regression model trained to estimate actual age within that particular age-group is used for age estimation. Algorithm 2 summarizes the steps followed for age estimation while Figure 3.1 depict the hierarchical age prediction.

3.8 Feature extraction

In this section, feature extraction techniques used are described. Texture features were extracted with spatial LBP, appearance features were extracted with LDA and Gabor filters were used for wrinkle feature extraction while 2D landmark points were used as shape features. Multi-frequency BIF (MF-BIF) features were extracted using Gabor filters at different frequencies, orientations and scales. Local statistical BIF (LS-BIF) features were extracted using local pooling within feature selector. Two extensions of Local Directional Patterns (LDP) are proposed and used for texture feature extraction.

3.8.1 Local binary patterns feature extraction

The image is divided into 9 blocks. Histogram of each block is computed and concatenated to make a texture feature vector for age estimation. Figure 3.11 and 3.12 show our approach of holistic and landmark texture feature extraction respectively. LBP assigns a code to each pixel in an image by comparing it to its neighbors in a

Algorithm 2: Age prediction process

Input : Probe feature vector **F**
Output: Predicted age **p**

Determine **Gender** : Estimate with **Gender** model
if Gender ←**Male then**
  Model ← $Male_{Model}$: Load male age-group model for age-group estimation
else
  Model ← $Female_{Model}$: Load female age-group model for age-group estimation
Determine **Group** from loaded age-group **Model**
if Group ← **1 then**
  Regressor ← $\mathbf{SVR_1}$
else if Group ← **2 then**
  Regressor ← $\mathbf{SVR_2}$
⋮
else if Group ← $\mathbf{N-1}$ **then**
  Regressor ← $\mathbf{SVR_{N-1}}$
else
  Regressor ← $\mathbf{SVR_N}$
Estimate age using **Regressor**

particular radius. Local Binary Patterns code is defined as

$$LBP_{P,R}\left(x_c, y_c\right) = \sum_{n=0}^{N-1} 2^n s\left(g_n - g_c\right)$$
$$\tau\left(x\right) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{3.22}$$

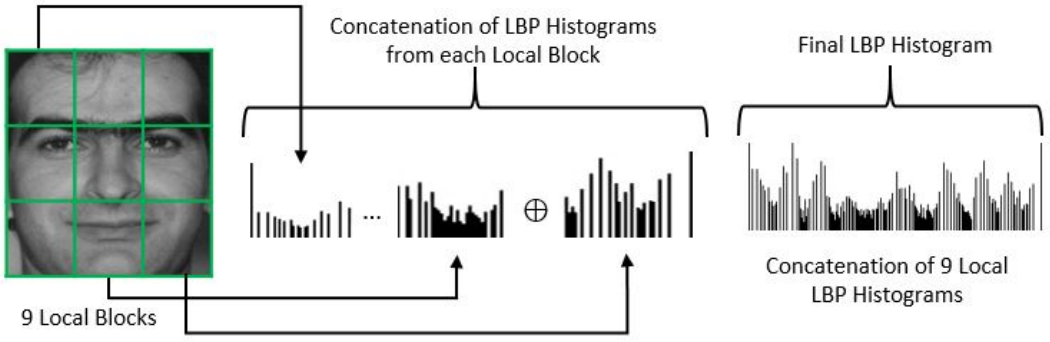


FIGURE 3.11: Spatial LBP texture feature extraction

where N is number of neighbouring pixels, R is distance of neighbouring pixel from center pixel, g_c is gray-value of center pixel, g_n for $n = 0, 1, 2, \ldots N-1$ correspond the n^{th} to value of neighbouring pixel on circular symmetric neighbourhood of distance $R > 0$ and the function $\tau\left(x\right)$ is a threshold function that generates a binary bit for a particular pixel. Concatenating all 8 bits gives a binary number. Resulting binary number is converted to a decimal and assigned to centre pixel as its LBP code.

As stated, the face image is split into 9 local blocks. Local Binary Patterns code

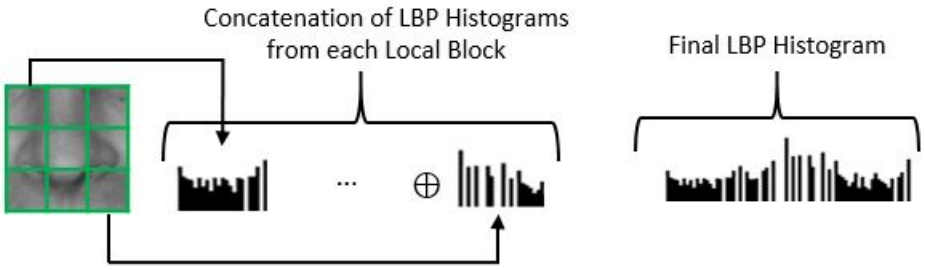
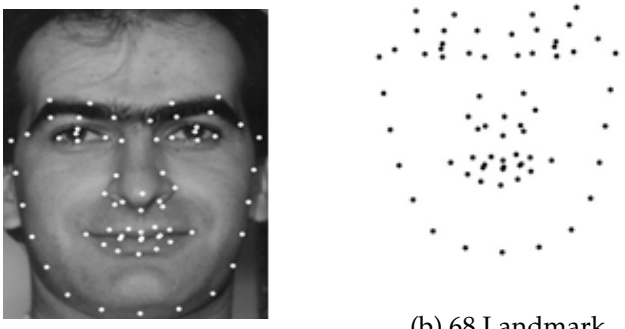


FIGURE 3.12: Landmark texture feature extraction

is calculated with $R = 1, P = 8$ and histogram with $2^8 = 256$ bins for each region is generated. Thereafter the local histograms are concatenated into one feature vector of size $9 \times 2^8 = 2304$, which is used as texture feature vector for age estimation. Same approach is used to extract texture features of facial landmarks (eyes, nose, mouth, cheeks, nose-bridge and forehead). Concatenating local histograms maintains the spatial relationship of the facial texture. Linear Discriminant Analysis is used to reduce dimensionality of the resultant texture feature.

3.8.2 Shape feature extraction

The 68 landmark points in Face and Gesture Recognition Network (FG-NET) ageing database and ratios of distances between fiducial landmarks to represent global facial shape. These points can be determined by using appropriate 2D landmarking algorithm like one proposed in [110]. Figure 3.13 show labelled face image and corresponding 68 landmark points used to label the face.



(a) Labeled face

(b) 68 Landmark points

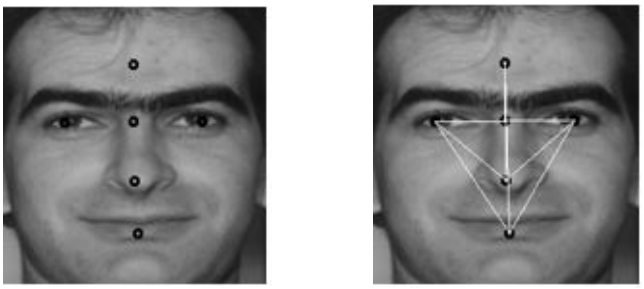
FIGURE 3.13: Landmark points used in facial shape

A landmark point p is represented by its x and y coordinates that show its location on the face. For each image, the shape feature vector is created by concatenating each point coordinates to a vector as

$$\vec{S} \leftarrow [x_1 \oplus y_1 \oplus x_2 \oplus y_2 \cdots \oplus x_n \oplus y_n] \tag{3.23}$$

where x_i and y_i are coordinates of point i and $\oplus$ is a concatenation operator.

Landmark localization technique proposed in [282] is used to locate left eye (LE), right eye (RE), nose (NO), mouth (MO), nose-bridge (NB) and forehead (FH) fiducial landmarks. Figure 3.14 shows landmark centroids and distances between them. For each fiducial landmark, a bounding rectangle (circle for NB) is returned. Land-



(a) Centroids (b) Distances

FIGURE 3.14: Landmark centroids and distances between them

mark centroids[3] are then determined using spatial and geometric information of the bounding rectangle or circle. For instance, the centroids of the Left-Eye (LE) $LE_c\,(x, y)$ is determined as

$$LE_c\,(x, y) = \left(LE.TL_x + \frac{W}{2}, LE.TL_y + \frac{H}{2}\right) \tag{3.24}$$

where $LE.TL_x$ and $LE.TL_y$ are coordinates of top-left corner of the LE bounding rectangle, W and H are the width and height of the rectangle respectively. Ten distances were calculated between these landmarks from which 7 ratios were derived, 4 for horizontal distances and 3 for vertical distances. These ratios were calculated as shown in equation 3.7 to 3.13.

For each image, the ratio feature vector is created concatenating ratios in section 3.5 as

$$\vec{R} \leftarrow [r_1 \oplus r_2 \oplus \ldots r_n] \tag{3.25}$$

where r_i is a particular ratio and $\oplus$ is a concatenation operator. The final shape feature vector is obtained by concatenating $\vec{S}$ and $\vec{R}$ as

$$F \leftarrow \vec{S} \oplus \vec{R} \tag{3.26}$$

LDA is applied on the $\vec{F}$ to reduce its dimensionality before being used for age estimation.

3.8.3 Appearance feature extraction

The detected face is resized to 100×100 then flattened into a 1×10000 row matrix. Linear Discriminant Analysis is used for appearance feature extraction. Face images are first projected to PCA subspace to reduce the dimensionality of the input image data from 10000 to $N - c$ where N is the number of samples and c is the number of age classes to be estimated. This is done before applying LDA to ensure that S_w does not become singular as proposed in [127] [128]. Linear Discriminant Analysis on face images projected on PCA space. For each image in all ages, with-class scatter matrix is defined as

$$S_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} \left(x_i^j - \mu_j\right) \left(x_i^j - \mu_j\right)^T \tag{3.27}$$

where x_i^j is the i_{th} image of age j, μ_j is the mean of age j, c is the number of ages to be estimated and N_j is the number of images in age j. Between-class scatter matrix

[3]The centre of mass of a geometric object of uniform density.

is defined as

$$S_b = \sum_{j=1}^{c} (\mu_j - \mu)(\mu_j - \mu)^T \tag{3.28}$$

where μ is the mean of all ages.

Linear Discriminant Analysis main objective is to maximize between-class scatter matrix while minimizing within-class scatter matrix. One way of doing this is maximizing the ratio $\frac{\det|S_b|}{\det|S_w|}$. Given that S_w is nonsingular, it has been proven [125] that this ratio is maximized when column vectors of projection matrix are the eigenvectors of $S_w^{-1}S_b$. S_w maximum rank is $N - c$ with N samples and c classes. This therefore requires $N = t + c$ samples to guarantee that S_w does not become singular, where t is the dimensionality of input data. The number of samples N is almost always smaller than t, making the scatter matrix S_w singular. To solve this problem Belhumeour *et al.* [127] and Swets and Weng [128] propose projecting input data to PCA subspace, to reduce dimensionality to $N - c$, or less, before applying LDA. Principal and Component Analysis and LDA are widely used appearance feature extraction methods in pattern recognition [129]. Consequently, LDA is adopted for extraction of global face appearance features for age-group estimation.

The input data is projected to PCA subspace as recommended before applying LDA. Thereafter, LDA and PCA eigenvectors are combined using GEMM[4][288] operation and use the combined eigenvectors to project each of the face data. These projections are taken as appearance feature vector for age estimation.

3.8.4 Wrinkle feature extraction

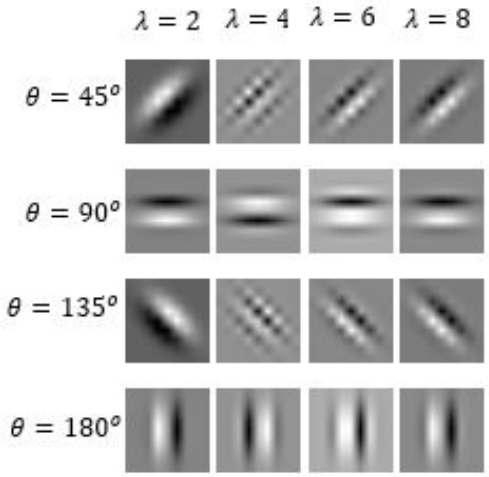
Wrinkles features are extracted using Gabor filter. Originally introduce by Denis Gabor in 1946 [69], Gabor filters has been extensively used for wrinkle, edge and texture feature extraction due to their capability of determining orientation and magnitude of wrinkles [8]. Wrinkles significantly influence facial appearance and can improve age estimation accuracies [74]. Since wrinkles appear as edge-like components with high frequency, Gabor edge analysis technique has been commonly used for wrinkle features extraction. Though, edges in a face image also consist of noise such as beards, mustache, hairs and shadows. To reduce the effect of this noise, Choi *et al.* [8] propose the use of predominant orientation of wrinkles to be considered in wrinkle feature extraction.

A bank of Gabor filters of different scales, frequencies and orientations is used to extract wrinkle features. A 2D spatial domain Gabor is defined as

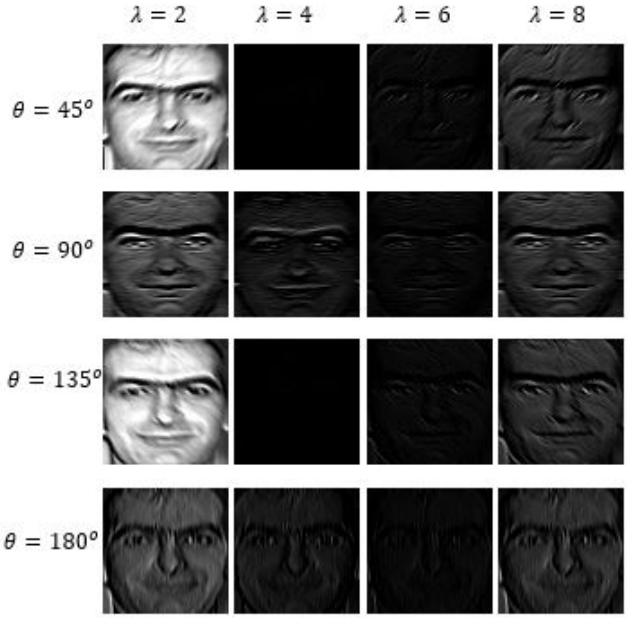
$$G(x, y) = \exp\left(-\frac{X^2 + \gamma^2 Y^2}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} X\right) \tag{3.29}$$

where $X = x\cos\theta + y\sin\theta$ and $Y = -x\sin\theta + y\cos\theta$ are angle of rotations of Gabor filters, θ varies from 0 to π, γ and σ are aspect ratio and standard deviation of the Gaussian envelope respectively and λ is the wavelength and determines spatial frequency $1/\lambda$. Figure 3.15a shows a bank of Gabor filters at 4 orientations and 4 frequencies while Figure 3.15b shows Gabor responses after convolving image in Figure 3.3d with filters in 3.15a. Mean and standard deviation of Gabor filter responses magnitude are used to determine wrinkle features. Given a Gabor filter bank of size b, for each image, b Gabor filter responses with N rows and M columns.

[4]Generalized Matrix Multiplication (GEMM) in Basic Linear Algebra Subprograms (BLAS) Level 3



(a) Gabor filters



(b) Gabor responses

FIGURE 3.15: Gabor filters and responses from 4 orientations and 4 frequencies

The mean is defined as

$$\mu = \frac{1}{N \times M} \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} x\left(i, j\right) \tag{3.30}$$

and the standard deviation is defined as

$$\sigma = \sqrt{\frac{1}{N \times M} \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} \left[\left(x\left(i, j\right)\right) - \mu\right]^2} \tag{3.31}$$

The mean and standard deviation are concatenated to form a wrinkle feature with $2 \times b$ dimensionality. This wrinkle feature is used for representing face for age estimation.

3.8.5 Significant orientation response local directional patterns

Objective five sought to define a variant of LDP operator that considers edge orientation and magnitude. This section describes Significant Orientation Response LDP (SOR-LDP) technique.

Significant Orientation Response LDP (SOR-LDP) is an eight-bit binary code assigned to every pixel in an image. It detects edge-features at four different orientations and is computed by taking the maximum response of each of the two responses for each orientation. Orientations of these edges are defined as horizontal (H_{0^o}), vertical (V_{90^o}), left-diagonal (LD_{45^o}) and right-diagonal (RD_{135^o}) and are defined as

$$H_{0^o} = \max\{|5S_0 - 3T_0|, |5S_4 - 3T_4|\} \tag{3.32}$$

$$V_{90^o} = \max\{|5S_6 - 3T_6|, |5S_2 - 3T_2|\} \tag{3.33}$$

$$LD_{45^o} = \max\{|5S_3 - 3T_3|, |5S_7 - 3T_7|\} \tag{3.34}$$

$$RD_{135^o} = \max\{|5S_1 - 3T_1|, |5S_5 - 3T_5|\} \tag{3.35}$$

where S and T are as defined in Equation 2.11.

A 3×3 image region shown in Figure 3.17(a) is convolved with Kirsch masks shown in Figure 2.15. The resultant response is compared for each orientation. The bit corresponding to the maximum response is set to 1 while the other bit is set to zero. This way, each directional response is considered for encoding image gradient. The response code shown in Figure 3.17(e) is calculated as

$$\begin{aligned} \text{SOR-LDP}_p &= \sum_{i=0}^{3} \tau\left(|n_i(p)| - |n_{i+4}(p)|\right) \times 2^i \\ &+ \tau\left(|n_{i+4}(p)| - |n_i(p)|\right) \times 2^{i+4} \end{aligned} \tag{3.36}$$

where $n_i(p)$ is the i^{th} neighbor of pixel p as shown in Figure 3.16 and function τ is defined as

$$\tau(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ \\ 0 & \text{otherwise} \end{cases} \tag{3.37}$$

$n_3(p)$	$n_2(p)$	$n_1(p)$
$n_4(p)$	p	$n_0(p)$
$n_5(p)$	$n_6(p)$	$n_7(p)$

FIGURE 3.16: 8-neighbors of a pixel p

The resultant bits are concatenated and the 8-bit binary string is converted to decimal and assigned to the center pixel as its SOR-LDP code. This approach is relevant where there is significance in encoding edges that originate from one point and stretch to multiple directions, like wrinkle lines. With this approach, all 8-directional responses are considered in determining image gradient. Figure 3.17 shows process of encoding an image using this approach. The same image region is used in Figure 2.16 for comparison purposes.

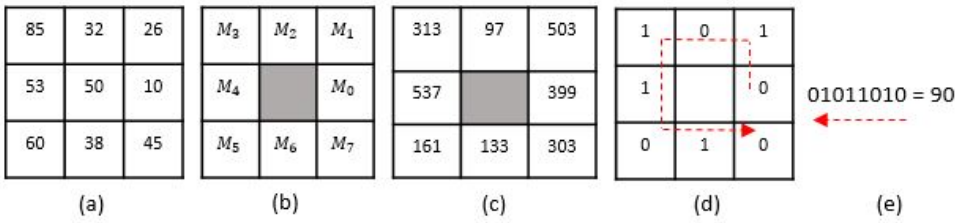


FIGURE 3.17: Process of encoding an image with SOR-LDP operator (a) Image region. (b) Kirsch masks as presented in Figure 2.15. (c) Result of convolving each pixel in (a) with 8 Kirsch masks. (d) Pick significant orientation responses for each direction, set there corresponding bit to 1 and the rest to 0 (e) Resultant LDP code

As seen in Figure 3.17 (e), the resultant LDP code is 90 as compared to 19 in Figure 2.16 (e). Figure 2.16 shows that the most significant edges were dominantly towards north-east and west. It can be seen that original LDP failed to capture that there was also edges towards south and north-west. As shown in Figure 3.17, it is evident that there are significant edge responses towards north-east, west, north-west and south of the reference pixel. Unlike LDP with $k = 4$, there is surety that this approach will encode edge responses in 4 different orientations. Figure 3.18 shows the LDP process with $k = 4$.

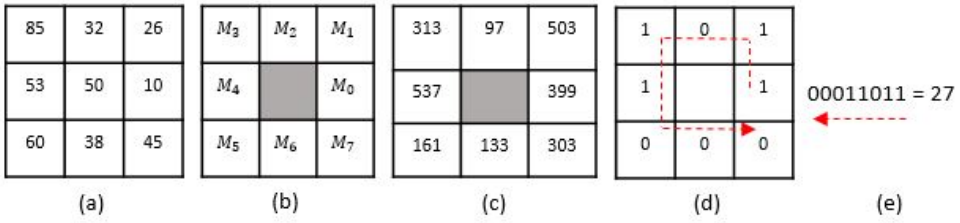


FIGURE 3.18: Process of encoding an image with LDP operator with $k = 4$ (a) Image region. (b) Kirsch masks as presented in Figure 2.15. (c) Result of convolving each pixel in (a) with 8 Kirsch masks. (d) Pick $k = 4$ significant responses, set there corresponding bit to 1 and the rest to 0 (e) Resultant LDP code

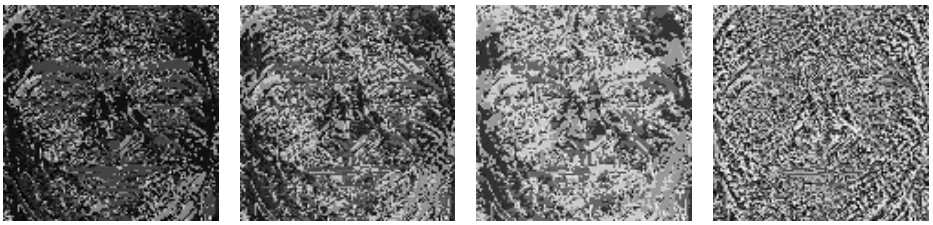
A histogram H_i with C_4^8 bins can be used to represent the input image of size $M \times N$ as

$$H_i = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f\left(LDP_{(n,m)}, i\right) \tag{3.38}$$

$$f\left(p, i\right) = \begin{cases} 1 & \text{if } p = i \\ \\ 0 & \text{otherwise} \end{cases} \tag{3.39}$$

where $f\left(p, i\right)$ is a logical function that compares if the SOR-LDP code at location $p\left(m, n\right)$ of the SOR-LDP encoded image is equal to current SOR-LDP pattern i.

As can be seen in Figure 3.18, LDP with $k = 4$ fails to capture the south (vertical) edge response. The original LDP operator has a risk of missing significant responses in some orientations with respect to reference pixel. Figure 3.19 shows LDP and SOR-LDP encoded responses of image shown in Figure 3.3d. Experimental results of age estimation using SOR-LDP are presented in 4.5.



(a) LDP k=2 (b) LDP k=3 (c) LDP k=4 (d) SOR-LDP

FIGURE 3.19: Input image encoded with original LDP and SOR-LDP. (3.19a) Original LDP k=2, (3.19b) Original LDP k=3, (3.19c) Original LDP k=4 and (3.19d) SOR-LDP encoded image

3.8.6 Local ternary directional patterns

This section describes the method used to achieve objective six which was to extend LDP to a ternary directional pattern based on pixel differential values and an adaptive threshold for encoding image gradient.

Local Ternary Pattern uses a static user defined threshold τ for all images in a dataset or for all experimental datasets making it not invariant to pixel value transformations. It is not practically easy to select an optimum value for τ in real application domains. The value of τ should be adaptive to different image conditions and datasets. LDP only considers top k directional responses and disregard the rest of $8-k$ responses in encoding image gradient. Furthermore, LDP does not consider gray level of current reference pixel when calculating the image gradient. Presence of an edge is depicted by sharp difference between a pixel and its neighbors [137]. LDP *encodes image gradient* without considering the central pixel thereby "capturing" an image edge even where there is none. This results into possible lose of discriminative information. In this section, LTDP operator that consider central reference pixel and all directional responses in encoding image gradient is proposed. Local Ternary Directional Patterns operator uses an adaptive τ that depends on the directional responses of the image region.

Local Ternary Directional Patterns computes eight directional response using Kirsch masks. Given a 3×3 image region, LTDP first determines the differences in pixel intensities between central pixel and its neighboring pixels. The absolute magnitude of the difference is set as the edge difference of the respective pixel as

$$P_{i,j} = |P_{i,j} - P_c| \tag{3.40}$$

where $P_{i,j}$ is the pixel value at index (i,j) and P_c is the pixel value of the central pixel. Figure 3.20 shows differential values and their corresponding directional responses.
The responses are then normalized before being used to generate LTDP code. Min-max normalization is done as

$$x_i^{norm} = \frac{x_i - min}{max - min} \tag{3.41}$$

where x_i is the absolute value of respective responses for $i = 0, 1, 2, \ldots 7$, min and max are minimum and maximum responses respectively and x_i^{norm} is the normalized value of x_i. The normalized responses are in the range of 0.0 and 1.0 which signify the probability of an edge from the central reference pixel stretching towards respective direction.

Threshold τ is set to ± 0.1667 deviation from 0.50 value. The value 0.5 is selected as offset reference value for τ because it shows equal chance of there being an edge

85	32	26
53	50	10
60	38	45

(a) Original image

35	18	24
3	0	40
10	12	5

(b) Differential values

7	175	215
57	0	111
241	225	15

(c) Directional responses

FIGURE 3.20: Resultant LDP codes from the LTDP code (a) Positive LDP code and (b) Negative LDP code

or not. The value of τ is chosen to ensure the probability space is divided into 3 equal segments, one for each ternary bit. If the normalized response value is greater or equal to $0.5+\tau$, its corresponding bit is set to $+1$, if the normalized response value is less or equal to $0.5-\tau$, its corresponding bit is set to -1, and the corresponding bit is set to 0 if the normalized response is between $0.5-\tau$ and $0.5+\tau$ as

$$f(x_i) = \begin{cases} 1 & \text{if } x_i^{norm} \geq 0.50 + \tau \\ 0 & \text{if } 0.50 - \tau < x_i^{norm} < 0.50 + \tau \\ -1 & \text{if } x_i^{norm} \leq 0.50 - \tau \end{cases} \tag{3.42}$$

Figure 3.21 shows the process of encoding an image with the proposed LTDP operator.

0.0	0.72	0.89
0.21	0	0.44
1	0.93	0.03

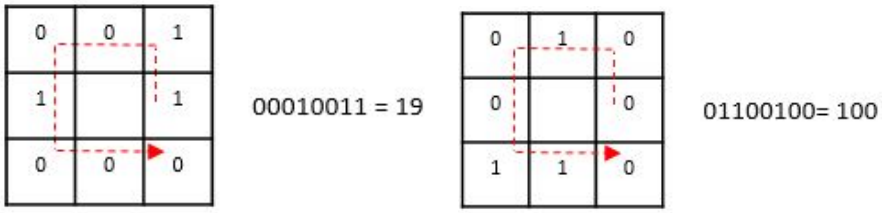
(a) Normalization of responses in Figure 3.20c

-1	1	1
-1	0	0
1	1	-1

(b) Assigning LTDP code at $\tau = 0.5 \pm 0.1667$

FIGURE 3.21: Process of encoding an image with LTDP operator

The presence of an edge towards a particular direction is signified by not only significant differential directional response towards that direction but also significant differential directional response of one of its neighboring direction. A differential directional response is significant if its value d is greater than $\bar{m} = 0.5 \times m + \tau$ where m is the maximum differential directional response of the local region. Differential directional responses closer to $\bar{m}$ are coded as being invariant relative to central pixel hence there corresponding bit set to 0. The differential directional response further away below $\bar{m} = 0.5 \times m - \tau$ are coded as having a negative image gradient hence there corresponding bit set to -1 and those further away above $\bar{m}$ are coded as having positive image gradient hence there corresponding bit set to 1. Each LTDP is split into its corresponding negative and positive segments as shown in Figure 3.22.



(a) Positive LDP code (b) Negative LDP code

FIGURE 3.22: Resultant LDP codes from the LTDP code

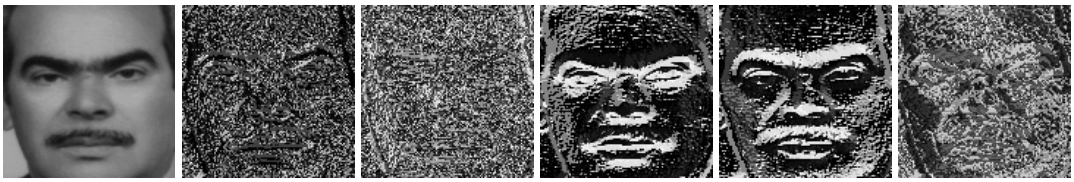
These codes are converted to decimal and assigned to corresponding central pixel of positive and negative LTDP encoded images respectively. A histogram is generated for both negative and positive LTDP encoded images as

$$H_i = \sum_{x,y} I\left(f\left(x,y\right)=i\right), i = 0,1,2\ldots,2^p-1 \tag{3.43}$$

where p is the number of patterns that can be encoded by the LDP operator (positive and negative) and

$$I(a) = \begin{cases} 1, & \text{if a is TRUE} \\ 0 & \text{otherwise} \end{cases} \tag{3.44}$$

Face region was detected from an input image using Haar-cascade face detection classifier [277]. The face is then cropped, converted to gray scale and resized to 120×120 pixels. The gray scale face image is smoothened using Gaussian filter as shown in Equation 3.5. The face is then encoded using LTDP operator. Figure 3.23 shows image encoded with LTDP, LTP and LDP operators.



(a) (b) (c) (d) (e) (f)

FIGURE 3.23: Image encoded with LTDP, LTP and LDP operators. (a) Input image (b) Resultant positive LTDP image (c) Resultant negative LTDP image (d) Resultant positive LTP image $\tau = \pm 3$ (e) Resultant negative LTP image $\tau = \pm 3$ (f) Resultant LDP image *k=3*

A histogram is generated for each of these images. The positive and negative histograms are concatenated and used as a feature vector for age estimation. The dimensionality of the resultant feature vector is reduced using LDA.

The resultant positive and negative histograms are concatenated and used as LTDP feature for pattern recognition. The histograms can be trimmed down by taking only uniform patterns into respective bins and put the rest of non-uniform patterns into one bin. A pattern is uniform if it contains utmost 2 transitions from 0 to 1 or vice versa. For *n-bit* patterns, the total number of uniform patterns is

$$P_U = n(n-1)+2 \tag{3.45}$$

where n is the number of bits used to represent the patters. LTDP generates $8(8-$

$1) + 2 = 58$ uniform patterns for both negative and positive LTDP encoded images. The resultant histogram could have 59 bins with 58 bins storing uniform patterns while the 59^{th} bin storing all non-uniform patterns. These two histograms are concatenated to form final LTDP feature vector. Experimental results of age estimation using LTDP are presented in section 4.7. The proposed LTDP technique is published in [289].

3.8.7 Directional Histogram of Oriented Gradients

This section describes proposed directional HOG texture feature descriptor. This technique was not extensively investigated and only preliminary results are reported in section 4.6.

Limitation of HOG is that it calculates image gradient using vertical and horizontal edges only. While this approach achieves better results in object detection, more orientations image gradients are needed to improve texture description using HOG. Directional Histogram of Oriented Gradients (DHOG) is proposed in this study. The proposed approach encodes image gradient at four orientations. The image gradient is calculated at $0^o, 45^o, 90^o$ and 135^o unlike HOG operator which calculates image gradient at 0^o and 90^o only.

Directional HOG uses 1D centred derivative mask $[-1, 0, +1]$ and its rotations at $45^o, 90^o$ and 135^o for vertical (g_x), left diagonal (g_{ld}), horizontal (g_y) and right diagonal (g_{rd}) directions respectively to calculate image gradient of 64×128 image divided into blocks of 16×16 with 28×8 cells making up a block. These filters are shown in Figure 3.24 while Figure 3.25 shows horizontal, vertical, left and right diagonal image gradients and the corresponding magnitudes and angles.

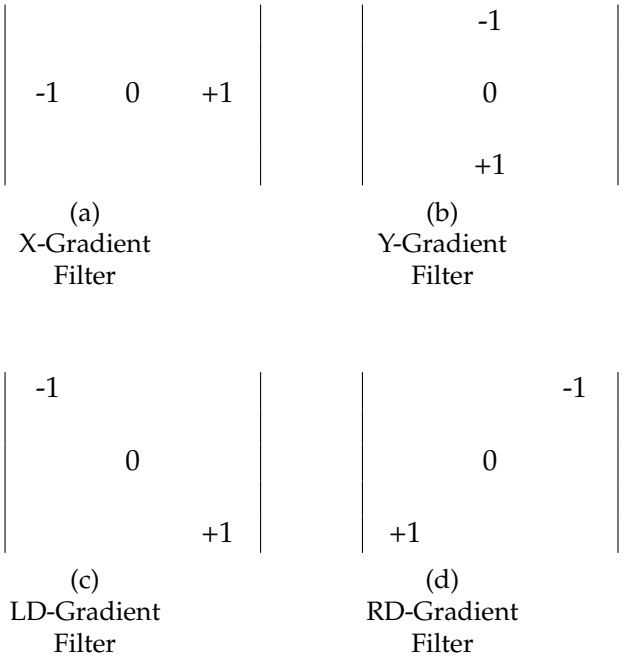


FIGURE 3.24: Filters used for image gradients calculations

Magnitude g_{xy} and angle θ_{xy} are calculated using x and y gradient responses as shown in equation 2.9. Using left-diagonal and right-diagonal gradient responses,

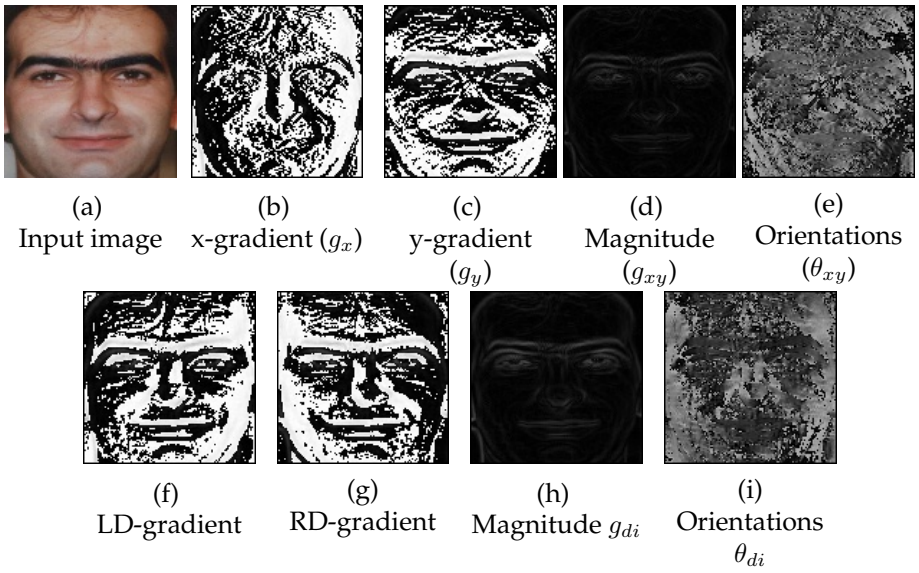


FIGURE 3.25: Image gradients visualizations for x, y, right and left diagonal directions and there corresponding magnitudes and orientations

second magnitude g_{di} and angle θ_{di} are calculated as

$$g_{di} = \sqrt{g_{ld}^2 + g_{rd}^2}$$
$$\theta_{di} = \arctan \frac{g_{ld}}{g_{rd}} \tag{3.46}$$

Given an image region shown in Figure 3.26a, four image gradients shown in Figure 3.26b are calculated. From these gradients, two orientation shown in Figure 3.26c

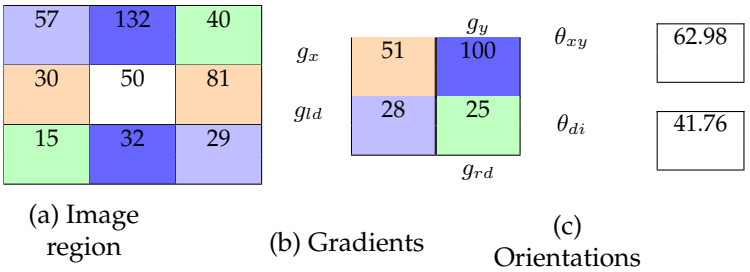


FIGURE 3.26: Sample DHOG generation process

are calculated for current reference pixel. The mean of the two angles is found and entered to the respective bin of DHOG. This is repeated on the whole image region and resultant DHOG is used as feature descriptor for age estimation. Preliminary results of age estimation using DHOG are presented in section 4.6.

3.8.8 Multi-Frequency BIF extraction

This section describes the technique used to achieve objective seven of the study. This objective sought to find out effect of frequency selectivity on the performance of BIF in age estimation.

Multi-Frequency BIF (MF-BIF) is extracted from Gaussian filter smoothened image. Biologically Inspired Features are derivatives of primates feed-forward model of visual object recognition pipeline [148].

The visual model of primates contains alternating layers of simple (S) and complex (C) cell units. Complexity of these cells increase as layers advance from primary visual cortex (V1) to inferior temporal cortex (IT). In primary visual cortex, S units use a bell-shaped tuning function to combine input intensities to increase scale and orientation selectivity.

S1 Units (Simple Cells): They represent the receptive field in primary visual cortex (V1) [72] which has basic attributes of multi-orientation, multi-frequency and multi-scale selection [73]. Gabor filters are used for modelling of cortical simple cell receptive fields. 2D spatial domain Gabor is defined as

$$G(x,y) = \exp\left(-\frac{X^2+\gamma^2Y^2}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda}X\right) \tag{3.47}$$

where $X = x\cos\theta + y\sin\theta$ and $Y = -x\sin\theta + y\cos\theta$ are angle of rotations of Gabor filters, θ varies from 0 to π, γ and σ are aspect ratio and standard deviation of the Gaussian envelope respectively and λ is the wavelength and determines spatial frequency $1/\lambda$. After applying Gabor filter, each pixel of the filtered image is represented as mean of neighbouring pixels in the $s \times s$ neighbourhood, where s is the size of Gabor filter used in that particular scale. Given Gabor response at scale S of size s, the mean gray value P_i at index i is calculated as

$$P_i = \frac{1}{N}\sum_{i=1}^{s\times s} S_i^s \tag{3.48}$$

where S_i^s is the gray value at index i of the response image at scale s and N is the number of pixels in $s \times s$ neighbourhood. The ratio σ/λ determines spatial frequency bandwidth of the filter. As proposed by De Valois *et al.* [290], narrowly tuned simple cells with half-response spatial frequency bandwidth of approximately one octave provide input for higher processing stages. This half-response spatial frequency bandwidth value correspond to 0.56 of the ratio σ/λ [291]. Spatial frequency can be varied by varying σ along both spatial dimensions [63]. Given that σ and λ are dependent, as in [291], λ is considered as a free parameter and is used to vary spatial frequency $1/\lambda$.

Unlike most previous studies that use only multiorientation and multi-scale selection of features, a technique that incorporate multi-frequency feature selection is proposed. This is because useful discriminating features are extracted using Gabor filters with different orientation and frequencies [64]. Consequently, previous studies [63] [64] suggest that spatial frequency processing is done in primary visual cortex. Spatial-frequency analysis extracts discriminative features that are more robust to distortions [154]. Daugman [65] found that visual system in primate's extracts information both in 2D spatial and frequency domains, and Shapley [66] proved that spatial frequency analysis helps the brain understand an image. This motivates inclusion of frequency parameter of Gabor filter in BIF extraction since it suggests that cortex performs some sort of spatial Fourier analysis. As shown in Figure 3.27, multifrequency Gabor response to an input image is different at same orientation. Gabor filter bank of 10 scales ranging from 3×3 to 21×21 pixels is used to model $S1$ visual response. Filters in four orientations $(0^o, 45^o, 90^o, 135^o)$, and four frequencies (wavelength of 4, 6, 8 and 10 pixels) are generated for each scale. This approach generates 160 Gabor filters $(4orientations \times 4frequencies \times 10scales)$. This way, $S1$ features will consist of multi-scale, multi-frequency and multi-orientation feature selection which provide multi-orientation, multi-frequency and multi-scale representation of

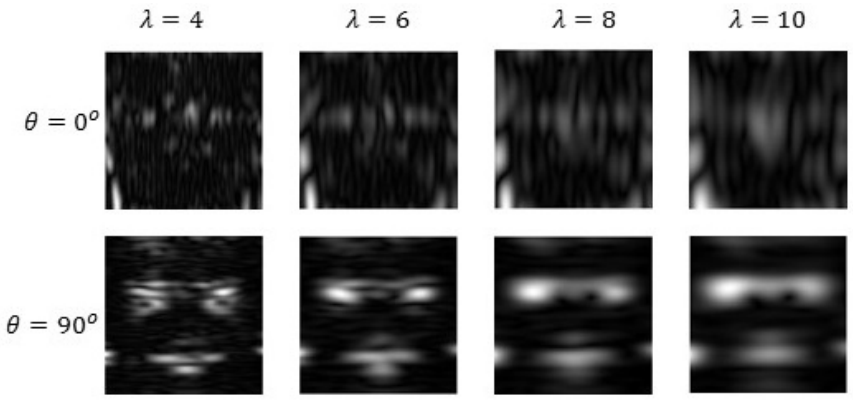


FIGURE 3.27: Gabor responses of image in Figure 3.3d at different frequencies ($f = 1/\lambda$) for $f = 0.25, 0.17, 0, 125, 0.1$.

input image. Once the image is convolved with Gabor bank of multi-frequency filters, Multi-Frequency BIF (MF-BIF) are extracted in $C1$ units.

C1 Units – Cortical Complex Cells: These units receive responses from S1 units and perform linear feature integration. C1 units represent complex cells that are shift invariant. Lampl *et al.* [155] proposed that spatial integration of complex cell in visual cortex that can be described by a series of pooling operations.

Riesenhuber and Poggio [70] demonstrated merits of using MAX pooling operator compared to SUM while Guo *et al.* [42] showed that standard deviation (STD) pooling operator outperforms MAX operator. Cai *et al.* [73] improved on STD by using a cell grid of 4×4 in normalization. The MAX operator returns maximum values at each index i of the two consecutive scale features. Given a feature at scale S_x and scale S_{x+1}, the maximum value F_i at index i is given by

$$F_i = \begin{cases} S_x^i & \text{if } S_x^i \geq S_{x+1}^i \\ \\ S_{x+1}^i & \text{if } S_{x+1}^i > S_x \end{cases} \tag{3.49}$$

where S_x^i and S_{x+1}^i are the filtered values at the position i of features from scale x and $x + 1$ respectively.

STD operator defined in [42] as

$$STD = \sqrt{\frac{1}{n_s \times n_s} \sum_{i=1}^{n_s \times n_s} (F_i - \overline{F})^2} \tag{3.50}$$

where maximum value at index i between two consecutive S1 scales is represented by F_i, $\overline{F}$ is the mean of filtered values within $n_s \times n + s$ neighbourhood is used. Given two $N \times N$ features at scales S_x and S_{x+1}, STD operator with $n_s \times n_s$ grid returns $\lfloor N/n_s \rfloor \times \lfloor N/n_s \rfloor$ features. STD operator captures local texture and wrinkle variations which are significant for subtle age estimation.

Based on [42] and [70], MAX and STD pooling operators are used to extract C1 features. MAX operations are done on S1 units for each frequency to get orientation-based features. MAX operator is applied on orientation-based features to obtain scale-based features. For every two consecutive scales, s_i and s_{i+1} in S^n , where n is

number of scales used, STD operation is performed $\beta = n/2$ features are obtained. These β features are concatenated and used for age estimation. This technique is named MF-BIF and is summarized in Algorithm 3, where

- n scales: $(s_i)_{i=1,2,...,n}$
- d orientations: $(\theta_j)_{j=1,2,...,d}$
- t frequencies:$(\lambda_k)_{k=1,2,...,t}$

are considered. MAX and STD operators are used to generate C1 features which

Algorithm 3: MF-BIF Extraction

input : S1 vector X of $n \times t \times d$ Gabor responses for n scales, t orientations and d frequencies
output: **MF-BIF** features vector

for $s \leftarrow s_1$ **to** s_n **do**
  for $\theta \leftarrow \theta_1$ **to** θ_t **do**
    for $\lambda \leftarrow \lambda_1$ **to** λ_d **do**
      $a_\lambda \leftarrow MAX(X_\lambda)$:*Feature at frequency* λ;
    $a_\mu \leftarrow AVG(a_\lambda)$:*Solve equation 3.48*;
    $a_\theta \leftarrow MAX(a_\mu)$:*Feature at orientation* θ;
  $c_s \leftarrow MAX(a_\theta)$:*Feature at scale* s;
  $S \leftarrow c_s$:*Push feature at scale* s *to vector* S;
for $s \leftarrow s_{n=1}$ **to** $s_{n=\lfloor n/2 \rfloor}$ **do**
  $c_0 \leftarrow MAX(S_n, S_{n+1})$:*Solve equation 3.49*;
  $c_s \leftarrow STD(c_0)$:*Solve equation 3.50*;
  $C \leftarrow c_s$
MF-BIF $\leftarrow C_1 \oplus C_2 \oplus C_3 \oplus \cdots \oplus C_n$: *Concatenate features at all scales*

are concatenated and used as feature representation of the input image. This process result into a very large dimensional feature vector. Dimensionality of the resultant feature vector is reduced by Principle Component Analysis (PCA) before classification. Experimental results on age estimation using MF-BIF are presented in section 4.8. The proposed MF-BIF technique is published in [292].

3.8.9 Local Statistical Biologically Inspired Features (LS-BIF)

This section describes LS-BIF feature extraction technique for age estimation. This technique is used to achieve objective eight of the study. This technique mimics visual image processing in mammal's visual cortex. Visual cortex contain alternating layers of simple (S) and complex (C) cells whose complexity increase as layers advance from primary visual cortex (V1) to inferior temporal cortex (IT). The S units use a bell-shaped tuning function to combine input intensities to increase scale, frequency and orientation selectivity. Pooling in C units introduce gradual invariance to scale, rotation and translation.

S1 Units - Simple Cells: They represent the receptive field in primary visual cortex (V1) [72] which has basic attributes of multi-orientation, multi-frequency and multi-scale selection [73]. The S1 units in LS-BIF modelling was done with a bank of Gabor filters of multiple frequencies, orientations and scales as described in section 3.8.8. Unlike in MF-BIF modelling, illumination and contrast normalization was introduced in LS-BIF modelling using center-surround divisive normalization.

Center-surround divisive normalization: Illumination and contrast normalization is performed both in the retina and visual cortex system [293]. For instance, rods, cones and an adjustable pupil are used by the retina to perform light intensity normalization. After light normalization, neurons in the retina transmit light-normalized signal into the visual cortex where contrast normalization takes place [293]. Consequently, center-surround divisive normalization is performed for illumination and contrast normalization. Illumination normalization is done by dividing every pixel in 3×3 neighbourhood with average of the neighbouring 8 pixels while contrast normalization is obtained by dividing every pixel in the same neighbourhood with standard deviation of 8 neighbouring pixels as

$$\begin{aligned} P_i &= \frac{P_i}{\frac{1}{n \times n} \sum_{j=1}^{n \times n} P_j} \quad illumination\ normalization \\ P_i &= \frac{P_i}{\sqrt{\frac{1}{n \times n} \sum_{j=1}^{n \times n} (P_j - \mu)^2}} \quad contrast\ normalization \end{aligned} \tag{3.51}$$

where P_i is pixel value at index i, P_j is the pixel value at index j in the local $n \times n$ neighborhood, μ is the mean of values in the local neighborhood.

C1 Units – Cortical Complex Cells: Shift invariant complex cells in visual cortex are represented by C1 units. These units receive responses from S1 units and perform hierarchy of pooling operations [155] to linear feature integration. Riesenhuber and Poggio [70] demonstrated merits of MAX pooling to SUM pooling while Guo *et al.* [42] showed that standard deviation (STD) pooling operator outperforms MAX operator. Cai *et al.* [73] improved on STD by using a cell grid of 4×4 in normalization. Based on [42], STD pooling operator is redefined for extraction of C1 features. This operator is redefined to operate within a local neighborhood of $n \times n$ unlike applying it on whole S1 features across scales or orientations as used in [42]. Another operator AVG_{local} is also defined to perform pooling in $n \times n$ neighborhood. The redefined STD operator (STD_{local}) and AVG_{local} are combined together with MAX [70] to extract C1 features.

MAX [70] pooling operator is applied to S1 units in different frequencies (within same scale and orientation) to obtain frequency tolerant features. The MAX_{global} operator returns maximum values at each index i of two consecutive frequency features. Given a feature at frequency λ_t and λ_{t+1}, the maximum value P_i at index i is given by

$$P_i = \begin{cases} \lambda_t^i & \text{if } \lambda_t^i \geq \lambda_{t+1}^i \\ \lambda_{t+1}^i & otherwise \end{cases} \tag{3.52}$$

where λ_t^i and λ_{t+1}^i are filtered values at index i of features in frequency t and $t+1$ respectively, within the same scale and orientation bands. Figures 3.28 and 3.29 show Difference between pooling across and within selectors. For simplicity, MAX pooling in 3×3 grids is shown. The same pooling approach is used for STD and AVG pooling. Given $N \times N$ feature vector with frequency selectivity, AVG_{local} and STD_{local} pooling in $n \times n$ local grid are applied in all orientations within a scale band, where n is the size of the Gabor filter used in that particular scale. AVG_{local} is defined as

$$AVG_{local} = \frac{1}{T} \sum_{i=1}^{n \times n} P_i \tag{3.53}$$

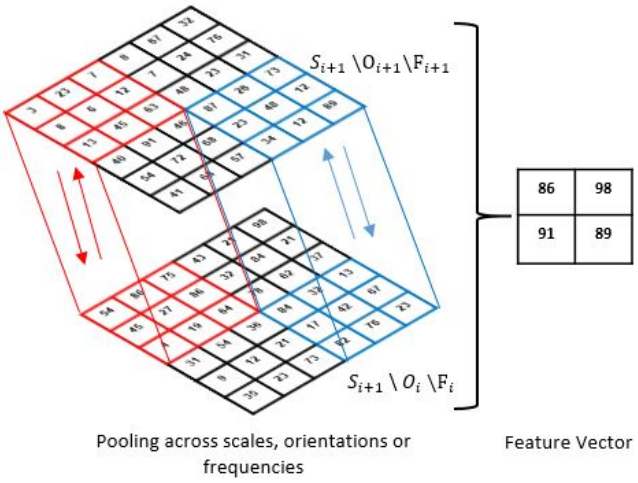


FIGURE 3.28: Pooling across scales, orientations or frequencies. This is the pooling approach used in previous studies

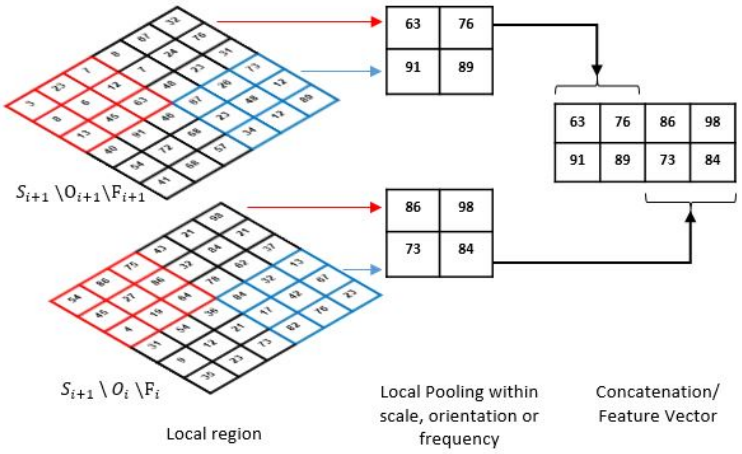


FIGURE 3.29: Pooling within scale, orientation or frequency. This is the proposed pooling in this study.

where T is the number of pixel values in the $n \times n$ neighborhood and P_i is a pixel value at index i. STD_{local} is defined as

$$STD_{local} = \sqrt{\frac{1}{n \times n} \sum_{i=1}^{n \times n} (P_i - \mu)^2} \tag{3.54}$$

where P_i is a pixel value at index i and μ is the mean of values in $n \times n$ region as defined in equation 3.53.

For each orientation, two $\lfloor N/n \times N/n \rfloor$ vectors are obtained, one with mean (μ) values and the other with standard deviation (σ) values. The two vectors are concatenated to obtain $F_d = \lfloor N/n \rfloor \times N$ feature vector at orientation d. F_d is pushed to a vector $\vec{F^D}$ such that $\vec{F^D} \leftarrow F_1, F_2, F_3 \ldots F_d$ where d is the number of orientations within a scale.

Given a vector $\vec{F^D}$ of orientation based features, AVG_{global} pooling operator is applied to all orientation-based features within the same scale. AVG_{global} is defined

as

$$AVG_{global} = \frac{1}{D}\sum_{i=1}^{D} M_i \tag{3.55}$$

where D is the number of orientations and M_i is feature vector at orientation i. This results into single feature vector for each scale. MAX [42] operator is applied for scale selectivity. MAX returns the maximum value at each index i of two consecutive scale features. Given a feature at scale S_x and scale S_{x+1}, the maximum value F_i at index i is given by

$$F_i = \begin{cases} S_x^i & \text{if } S_x^i \geq S_{x+1}^i \\ \\ S_{x+1}^i & \text{if } S_{x+1}^i > S_x \end{cases} \tag{3.56}$$

where S_x^i and S_{x+1}^i are the filtered values at position i of features from scale x and $x+1$ respectively. For z scales, MAX [42] operator returns $\lfloor z/2 \rfloor$ features. These $\lfloor z/2 \rfloor$ are concatenated and used an image representation for age estimation. This technique is named local statistical BIF (LS-BIF) and is summarized in Algorithm 4, where scales, orientations and frequencies are considered as in MF-BIF described in section 3.8.8. Experimental results on age estimation using LS-BIF are presented in

Algorithm 4: LS-BIF Extraction

Input : S1 vector X of $n \times t \times d$ Gabor responses for n scales, t orientations and d frequencies

Output: **LS-BIF** vector

for $s \leftarrow s_{n=1}$ **to** $s_{n=n}$ **do**
  for $\theta \leftarrow \theta_{d=1}$ **to** $\theta_{d=d}$ **do**
    for $\lambda \leftarrow \lambda_{t=1}$ **to** $\lambda_{t=\lfloor t/2 \rfloor}$ **do**
      $a_\lambda \leftarrow MAX_{global}(X_\lambda^t, X_\lambda^{t+1})$: equation 3.52;
    $\theta_d^\mu \leftarrow AVG_{local}(a_\lambda)$:*equation 3.53*;
    $\theta_d^\sigma \leftarrow STD_{local}(a_\lambda)$:*equation 3.54*;
    $\theta_d \leftarrow \theta_d^\mu \oplus \theta_d^\sigma$: Feature at orientation d ;
    $\mathbf{Y} \leftarrow \theta_d$: Push feature at orientation d to vector $\mathbf{Y}$
  $c_n \leftarrow AVG_{global}(\mathbf{Y})$:*equation 3.55. Feature at scale* n;
  $\mathbf{S} \leftarrow c_n$:*Push feature at scale* n *to vector* $\mathbf{S}$;
for $s \leftarrow s_{n=1}$ **to** $s_{n=\lfloor n/2 \rfloor}$ **do**
  $c_i \leftarrow MAX_{global}(S_n, S_{n+1})$:*equation 3.56*;
  $C \leftarrow c_i$
LS-BIF $\leftarrow \mathbf{C_0} \oplus \mathbf{C_1} \oplus \mathbf{C_3} \cdots \oplus \mathbf{C}_{\lfloor \mathbf{n/2} \rfloor - \mathbf{1}} \oplus \mathbf{C}_{\lfloor \mathbf{n/2} \rfloor}$

section 4.9. The proposed LS-BIF technique is published in [294].

3.8.10 Enhance MF-BIF extraction

This section describes how enhanced MF-BIF was achieved. The enhanced feature vector was used for age estimation with gender and age-group estimation. Enhanced Multi-Frequency BIF (EMF-BIF) is obtained by fusing MF-BIF with shape, texture, appearance and wrinkle features. Given shape feature vector $\vec{S} = [s_1, s_2, \ldots s_k]$, texture feature vector $\vec{T} = [t_1, t_2, \ldots t_n]$, appearance feature vector $\vec{A} = [a_1, a_2, \ldots a_m]$, wrinkle feature vector $\vec{W} = [w_1, w_2, \ldots w_p]$ and MF-BIF feature vector $\vec{B} = [b_1, b_2, \ldots b_q]$, where k, m, n, p and q are dimensionality of respective feature vectors. First z-score

normalization of each of the feature is performed as

$$f^i_{norm} = \frac{f_i - \mu_i}{\sigma_i} \tag{3.57}$$

where f_i is a facial ageing feature, μ_i and σ_i are the mean and standard deviation of this feature, and f^i_{norm} is the normalized feature of feature f_i.

For each face image, there are 5 normalized features. These features are concatenated to obtain EMF-BIF as

$$\mathbf{EMF\_BIF} = \vec{A} \oplus \vec{S} \oplus \vec{T} \oplus \vec{W} \oplus \vec{B} \tag{3.58}$$

The resultant **EMF_BIF** has the dimensionality $k + m + n + p + q$. LDA is used to reduce the dimensionality of **EMF_BIF** before using it for age estimation. Experimental results on age estimation using EMF-BIF are presented in section 4.10.

3.9 Information fusion

Information fusion is performed in two stages. For age group estimation, landmark wrinkle feature is fused with texture feature to obtain a rich feature vector for each component. A classifier (SVM or ANN) is trained to perform landmark-based age estimation. Z-score normalization is done as shown in equation 3.57 before fusing these features. A classifier is trained on each fused feature for learning ageing pattern based on respective landmark features. N classifiers are trained, where N is equal to number of landmarks used.

In total $N + 1$ classifiers are trained [5] to perform age-group estimation. During prediction, decisions from these classifiers are fused using majority vote rule and a label with highest number of votes is picked as the predicted age-group of the probe image.

Holistic shape, texture, appearance, wrinkle and MF-BIF features are fused to obtain EMF-BIF feature vector for age estimation. Hierarchical age estimation based on EMF-BIF feature is done by performing age-group classification followed by within-group exact age regression.

3.10 Classification and regression

Back propagation [295] Multi-Layer Perceptron (MLP) [296] Artificial Neural Network (ANN) and Support Vector Machines (SVM) [297] are used for classification and Support Vector Regression (SVR) for regression. These classifiers were selected due to their significant success in pattern recognition in recent years [298].

Number of layers and neurons[6] in each layer in the MLP are empirically determined. Different values are tried experimentally with input layer having neurons equivalent to dimensionality of the input feature vector and output layer having neurons equivalent to number of classes (ages/age-groups) to be classified. For age-group estimation, two topologies were use. For the technique proposed in 3.6, the ANN had 100 neurons at the input layer, two hidden layers of 120 neurons each and output layer with 4 neurons. For age-group used prior to age estimation, the ANN was configured to have 100 neurons in the input layer, 120 neurons in each of the two

[5] One extra classifier was trained to perform age estimation based on holistic facial shape and appearance.

[6] Fewer neurons in the hidden layers achieve data compression while many neurons may over represent the data hence improve performance given enough data.

hidden layers and 7 neurons in the output layer. The neurons in the hidden layers and number of hidden layers were determined empirically and parameters that posted better accuracies were selected. RBF kernel was used for SVM with parameters $\gamma = 0.12$ and $C = 100$. Different values of γ and C were tried and the selected values are those that posted better validation accuracies.

Artificial Neural Network are natively structured to handle non-categorical data, both as input or output. "One-hot" encoding is used to represent the output, since output data (labels) are categorical. If input data item is a member of class k, the entry at output k is set to 1.0 and other slots are filled with 0.0. Algorithm 5 is used to achieve this.

Algorithm 5: Change MLP output to continuous representation

Input : Training Data **data** of n rows, Labels **labels** of c classes
Output: Response Matrix **responses** of size $n \times c$

for $x \leftarrow 0$ **to** $x \leftarrow n - 1$ **do**
  for $y \leftarrow 0$ **to** $y \leftarrow c - 1$ **do**
    $\mathbf{responses}[x][y] \leftarrow 0.0$: *fill whole matrix with zeros*
for $i \leftarrow 0$ **to** $i \leftarrow n - 1$ **do**
  $b \leftarrow \mathbf{labels}\,[i]$
  $\mathbf{responses}[i][b] \leftarrow 1.0$: *use $b - a$ if your labels start at a*
Stop

One-versus-one soft-margin SVM [240] with radial basis function (RBF) kernel defined as

$$K\left(x_i, x_j\right) = e^{-\gamma|x_i - x_j|^2} \tag{3.59}$$

is used for classification. Different pairs of (C, γ) parameters are tried on small subsets of the training data[7] and parameters that return better cross-validation results are used to train a model on the whole training set. The parameter γ is the inverse of the radius of the samples selected as support vectors while C is the determines number of support vectors to be chosen[8].

Age is estimated using Support Vector Regression (SVR) [158] with a RBF kernel. As in classification, different pairs of (C, ε) are tried and parameters that give lower cross validation accuracy are chosen. The parameter ε determine number of support vectors used to construct a regression function by controlling the width of ε-*insensitive* zone.

3.11 Validation and evaluation protocols

3.11.1 Validation protocol

The dataset is randomly split into k subsets for *k-fold* cross-validation for selection of SVM/SVR and MLP parameters. $k\frac{(k-1)}{2}$ cross-validations are performed where for each iteration, a model is trained on $k - 1$ subsets and leave out one subset for testing. Although LOO is almost unbiased, it may give unreliable estimates due

[7] This process is also known as grid-search optimization of C and γ

[8] When γ is too large, radius of the area of influence of support vectors only includes the support vectors and no regularization is done with C to prevent over-fitting. When γ is too small, the model is constrained and cannot learn data distribution

to its high variance [196]. Nevertheless, we used LOO validation protocol for all experiments in this study.

3.11.2 Evaluation protocol

Mean Absolute Error (MAE) and Cumulative Scores (CS) to evaluate age and age-group estimation respectively. Mean Absolute Error (MAE) is the average of all absolute errors between estimated and is formally defined as

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|a_i - \bar{a}_i| \tag{3.60}$$

where N is total number of test images, a_i is the ground truth age of image i, and $\bar{a}_i$ the estimated age of image i. MAE for every age is calculated as

$$MAE_k = \frac{1}{n}\sum_{i=1}^{n}|a_k - \bar{a}_i| \tag{3.61}$$

where $\bar{a}_i$ is the estimated age for image i of age a_k and n is the number of test images belonging to age a_k. This gives age specific performance of age estimation technique. Overall MAE is found by summing all the MAE for all ages tested and divide by total number of test images in each age as

$$MAE_{TOTAL} = \sum_{i=1}^{k}\frac{(MAE_i \times n_i)}{N} \tag{3.62}$$

where $N = n_1 + n_2 + \cdots + n_k$. The overall accuracy of the estimator is given by cumulative score (CS) [39] [49] which is defined as

$$CS\,(x) = \frac{N_{e\leq x}}{N} \times 100\% \tag{3.63}$$

where $N_{e\leq x}$ is the number of images on which the age estimation technique makes an absolute error no higher than x years error tolerance and N is the total number of test images.

CS is used as an indicator of accuracy of age-group estimator [5]. In age group estimation, the age-group label represents a range of ages, hence the cumulative scores are compared at error level 0, i.e. the percentage of exactly correct age-group estimation. Therefore, the CS equation becomes

$$CS\,(x) = \frac{n_x}{N_x} \times 100\% \tag{3.64}$$

where n_x is the number of test images correctly recognized as belonging to age-group x and N_x is the total number of test images in age-group x.

3.12 Experimental database

FG-NET ageing database is used to evaluate the proposed age estimation approach. It is a publicly available ageing dataset that has 1002 images of 82 subjects aged between 0 and 69 years. Images have wide variation in illumination, color and expression. Some images have poor quality since they were scanned. Table 3.2 show distribution of images in FG-NET ageing dataset in 10-year age groups.
The limitation of this dataset is that it has few images in adult age groups (above 30

TABLE 3.2: Distribution of images in FG-NET dataset in 10-years age range

Age-group	No. of images	% Distribution
0-9	371	37.02%
10-19	339	33.83%
20-29	144	14.37%
30-39	70	6.99%
40-49	46	4.59%
50-59	15	1.50%
60-69	8	0.78%
Total	**1002**	**100%**

years). Studies that use this dataset for age-group estimation like [239] [243], [237] [236] tend to divide it into variable age groups of ranges similar to *0-5, 6 - 12, 13-24* and *25-69*. This is because age-group estimation accuracy reduces drastically with increase in number of groups due to reduction in number of images per group [239]. In this study, the age-groups were formulated as shown in Table 3.3 for age-group estimation. This grouping is almost similar to one used in [239] and [237].

TABLE 3.3: Distribution of images in random size age ranges

Age-group	Number of images	% Distribution
0-6	274	27.3%
7-14	271	27.0%
15-24	249	24.9%
25-69	208	20.8%
Total	**1002**	**100%**

Human expert was used to manually label gender of each subject in FG-NET based on the facial appearance. Labeling was done by appending *M* to the image label of males and *F* to the image labels of females. Table 3.4 show the distribution of images across group and gender in FG-NET dataset as labeled by a human.

TABLE 3.4: Image distribution across gender and groups of 10 years

Age-group	No. of images	
	Male	**Female**
0-9	206	166
10-19	174	164
20-29	80	64
30-39	33	46
40-49	27	19
50-59	7	8
60-69	5	3
Total	**532**	**470**

The grouping shown in Table 3.4 is used for hierarchical age estimation with age-group and gender discrimination.

3.13 Summary

This chapter presented hierarchical age estimation model adopted together with feature extraction techniques and experimental dataset. Age estimation process consists on gender estimation, age-group classification for each gender followed by age regression for each age-group. Ageing feature extraction techniques used are described. The chapter concludes by presenting experimental dataset used for age estimation.

Chapter 4

Results and Discussion

This chapter present and discuss results of the proposed hierarchical age estimation using enhanced facial and facial component features. In this age estimation approach, the first step is localization of facial landmarks like eyes, mouth, nose, cheeks, nose-bridge and forehead. Thereafter, facial ageing features are extracted from both holistic face and the localized landmarks. These features are fused and used for age and age-group estimation. Distance between centroids of mouth, nose, nose-bridge, eyes and forehead are measured across age and used to model regression functions that show how fiducial landmarks drift from each other across age. Ageing features from the face and facial landmarks are used for hierarchical age estimation modelling that consist of age-group classification followed by within-group age regression.

4.1 Landmark localization

This section presents results for the experiments described in section 3.4.

Face, fiducial and secondary landmarks are automatically detected. Figure 4.1 shows a sample output of a face image that was loaded, face region detected, detected and regions of interest (landmarks)cropped.

(a) (b) (c)

FIGURE 4.1: (a) Gray-scale input image (b) Detected and cropped face (c) Cropped Landmarks

Using Haar-cascade face detection classifier [277][278], 946 faces were detected out of 1002 images in the experimental dataset, which correspond to 94.41% accuracy. This detection rate could be attributed to adverse image conditions of some images in the dataset. Landmark detection was performed in these 946 detected faces and results are shown in Table 4.1.

As shown in Table 4.1, there was a 94.41% accuracy on the face detection since 946 faces were detected from the 1002 face images. The results show that 1718 eyes were detected from the possible 1892 of which 866 are $right\ eyes$ and 852 are $left\ eyes$. The lower rate of eye detection is attributed to poor image condition of some images in FG-NET-AD dataset, pose and face orientation. The same results as for the eyes were observed for the cheeks since spatial information and dimensions of the eyes were used to detect the cheeks. Therefore, for every true eye there was a true cheek

TABLE 4.1: Regions of interest detection accuracy

ROI	**True Detection**	**Total**	Accuracy
Face	946	1002	94.41%
FH	845	946	89.32%
RE	866	946	91.54%
LE	852	946	90.06%
RC	866	946	91.54%
LC	852	946	90.06%
NO	862	946	91.12%
NB	849	946	89.75%
MO	857	946	90.60%

and consequentially for every false eye there was a false cheek. Using spatial and geometric information of eyes, 849 nose-bridge regions were detected.

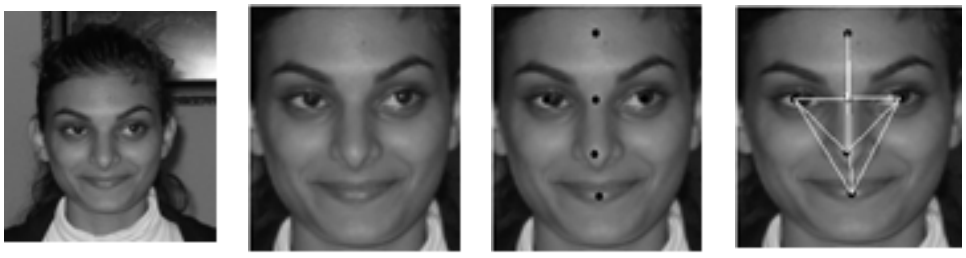
Nose and mouth were detected using Haar-cascade classifier proposed in [280]. To the best of our knowledge, there is no previous study that has used these classifiers for nose and mouth detection. The results obtained are promising as 862 (91.12%) of the 946 possible noses were accurately detected and 857 (90.60%) of the possible 946 mouths were detected. Performance of mouth classifier is adversely affected by beards, image orientation, and pose among other image conditions. However, it was observed that mouth classifier is robust to slight expression around the mouth region. There were 845 (89.32%) correctly detected foreheads out of the possible 946 foreheads.

The average time taken for face and regions of interest detection is 0.55 of a second. The speed and accuracy of landmark detection show that this technique could be used in a real time environment for detection of landmarks for further face computing. In the proposed approach, a simple tool is used making it fast and efficient for embedded systems with lower computing power.

4.2 Landmark displacement across age

This section presents results of experiments described in section 3.5.

Euclidean distances between centroids of facial landmark were measured across age and their ratios used to model a landmark displacement across age. Figure 4.2 shows landmark centroids and distances between them that were used for landmark displacement across age modelling.



(a) (b) (c) (d)

FIGURE 4.2: Landmark centroids localization and distances used. (a) Input image (b) Detected face (c) Landmark centroids and (d) Euclidean distances between landmarks

As shown in Figure 4.2(a), colored image is first converted to single channel gray-scale image. A 2D Gaussian filter is then applied to the image to reduce noise and enhance image details. The face in Figure 4.2(b) is detected using Haar-cascade classifier. The detected face region is cropped and landmark detection is done in the cropped face. Landmarks are detected and their respective centroids, as shown in Figure 4.2(c), are located. Euclidean distances between landmarks is then calculated as shown in Figure 4.2(d).

There were 946 (94.41%) faces detected from the 1002 images in the dataset. Some of the images in the dataset have adverse conditions as a result of scanning and this could have affected the detection rate. A total of 1732 (91.54%) eyes were accurately detected from the possible 1892. Lower rate of eye detection could be attributed to adverse image conditions, pose and orientation of images in experimental dataset. There were 845 (89.32%) foreheads accurately derived from spatial and geometric information of eyes. Using Haar-cascade for nose and mouth detection, 862 (91.12%) of noses and 857 (90.60%) of mouths were detected respectively from the possible 946. The centroids of these landmarks were derived and used for calculating distances between landmarks across age in order to establish landmark displacement as one ages.

As shown in Figure 4.2(c), the centroids used are those of the mouth (MO), nose (NO), nose-bridge (NB), left-eye (LE), right-eye (RE) and forehead (FH). Accuracies of these centroid detection is similar to that of respective landmark detection presented in Table 4.1. Distances between these centroids were measured and used to calculate 7 ratios shown in equations 3.7 to 3.13. Figure 4.3 shows the trend of these ratios across age.

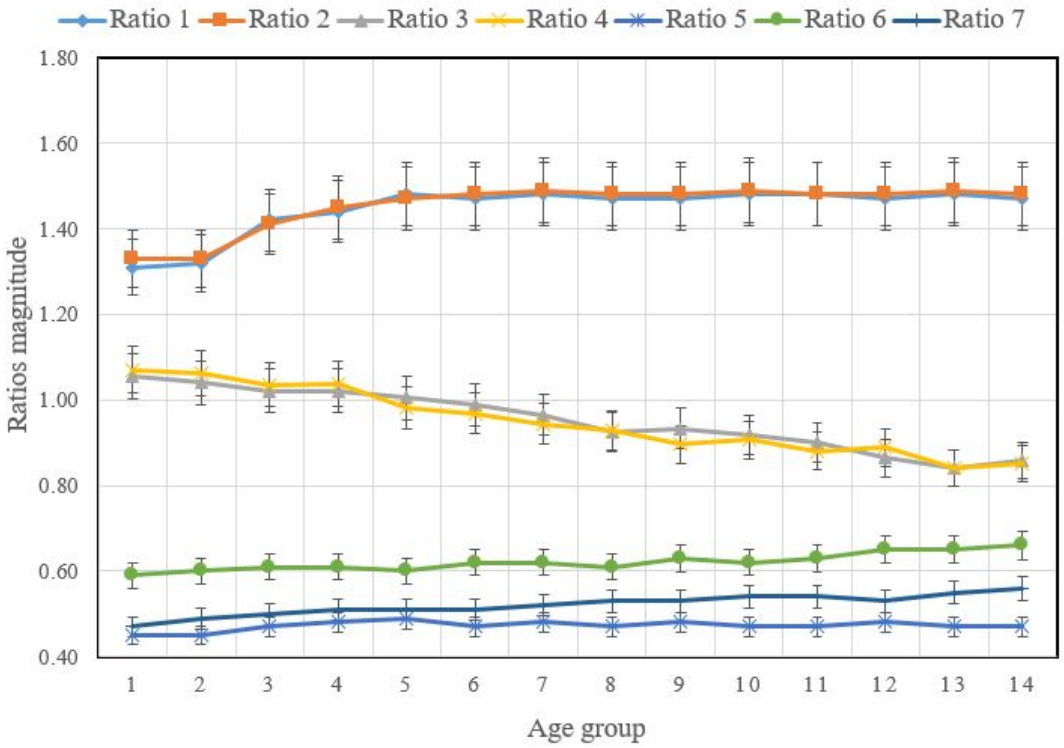


FIGURE 4.3: Trend of ratios of distance between landmarks across age

Ratios 1 and 2 track the displacement of eyes from the nose across age. As shown in Figure 4.3, ratios 1 and 2 remains relatively unchanged between 0-4 years. The

ratio changes as one moves to age-group 2 of 5-9 years. This could be due to craniofacial development that leads to change in general facial shape. The ratio is seen to significantly increase between 5 years and 20 years. In this age bracket, there is significant increase in craniofacial development where the forehead expands and slopes back hence creating space on the cranium. The increase in the ratios signify that change in distance between the eyes is higher compared to change in distance between respective eyes and nose. This shows that, at this stage, eyes tend to drift away from each other more than they drift from the nose. This signifies that change in horizontal size (width) is higher in this age bracket compared to change in vertical facial height. As the face grow, space is created on the cranium and eyes drift to occupy these space.

Ratios 3 and 4 track the displacement of eyes from the mouth across age. There is decrease in these ratios across age. This signifies that change in distance between the eyes is lower compared to change in distance between eyes and mouth. This shows that, as one age, the mouth tends to drift away from the eyes more than the eyes drift from each other. The face tend to increase in height as one age as its shape changes from circular to vertical oval shape. This causes the mouth to drift away from the eyes, especially during craniofacial development stage and in early adult age brackets.

Ratio 5 tracks the displacement of nose from forehead. This ratio slightly increases between 5-9 years. This signifies that in this age-group, there is slight displacement of the nose from forehead. This reveal some slight growth of the face in the vertical direction. This ratio remains relatively constant across the age. This shows that the growth in upper region of the face remains relatively constant across age. The same observation is seen with ratio 6 that tracks the displacement of mouth from the forehead. Ratio 7 slightly increases across age. This shows that change in distance between the mouth and nose is slightly higher compared to change in distance between mouth and midpoint of eyes. This signifies that there is slight displacement of the nose from the mouth as one age.

From the experiments, it was found that the nose centroid is located at a position $0.69 \times D_{eye\_eye}$, approximately 45^o degrees south-west of the right eye and south-east of left eye. The mouth centroid was found to be $1.06 \times D_{eye\_eye}$ from the eyes in the direction of about 60^o south-west of right eye and south-east of left eye. These findings slightly defer with that reported in [74] where they found that the nose is located at $0.60 \times D_{eye\_eye}$.

The relative position of the mouth centroid from the eye was found to be almost similar to $1.07 \times D_{eye\_eye}$ reported in [74]. The limitation of the findings in [74] is that they never investigated the direction of a particular landmark in respect to eyes. Intuitively, when specifying a point relative to a reference point, you must specify magnitude and direction of the vector between the point and reference point. Here, D_{nose_eye} and D_{mouth_eye} are vectors that show relative position of nose and mouth with respect to eye, θ and β are direction of these vectors respectively.

4.2.1 Regression analysis of ratios

Ratios of distances between facial landmarks were analysed to determine the effect of age on landmark displacement. Table 4.2 shows the results of regression analysis of these ratios against age.
As shown in Table 4.2, it was found that the regression explains over 52.2% of the variance in ration 1 and 2. Regression results show that for every five years, ratio 1

TABLE 4.2: Regression analysis of distance ratios between facial landmarks across age

Ratio	$\mathbf{R^2}$	**Coefficients**		**Sig.**
		Constant	**Age**	
1	0.522	1.371	0.010	0.001
2	0.579	1.376	0.010	0.002
3	0.974	1.080	-0.017	0.001
4	0.960	1.087	-0.018	0.001
5	0.123	0.465	0.001	0.218
6	0.853	0.587	0.005	0.001
7	0.914	0.476	0.006	0.001

or ratio 2 will change by 0.01 (1.0%). This can be represented as

$$Ratio\ 1|2 = 0.01x + 1.371 \tag{4.1}$$

where x represent five year-increase in age. This therefore means that as one grow, eyes tend to slightly move away from the tip of the nose. Change in ratio is found to be minimal and this signifies that change in distance between eyes and change in distance between eyes and nose tip is almost constant. Therefore, it can be concluded that the displacement of the nose from the eyes and the displacement of the eyes from each other is almost constant as one grows.

Ratio 3 and 4 show a negative regression with age with regression explaining 96.0% of the variance in these ratios. This means generally that increase in age by 5 years would decrease ratio 3 and 4 by 0.017 (1.7%). Linear relationship between ration 3 and 4 with age could be represented as

$$Ratio\ 3|4 = 0.017x + 1.080 \tag{4.2}$$

where x represent five year increase in age. This signify that, as one grow, the change in distance between eyes and mouth is greater than change in distance between eyes. This show that the mouth tend to be displaced more from the eyes than the eyes get displaced from each other as one grow. This could mean that there is difference in vertical and horizontal growth of face with age.

Ratio 5, 6, and 7 show displacement of nose from the forehead, mouth from the forehead and mouth from the nose. Regression results for displacement of nose from head were not significant. There is very minimal age introduced displacement between mouth and forehead with change in ratio 6 being less than 1% for every five-year increase in age with regression explaining 85.3% of the variance in ratio 6. Relationship between ratio 6 and age could be represented as

$$Ratio\ 6 = 0.005x + 0.587 \tag{4.3}$$

where x represent five-year increase in age. This signify that, as one grow, the change in distance between forehead and mouth is greater than change in distance between mouth and middle of eyes. This means that mouth is slightly displaced from middle of eyes and consequently from the forehead as one age. Displacement of the mouth from the nose as one age could be represented as

$$Ratio\ 7 = 0.006x + 0.476 \tag{4.4}$$

where x represent five-year increase in age. With this regression explaining 91.4% of the variance in ratio 6, it could be observed that the mouth gets slightly displaced

from the nose as one age. One increase in age group would cause 0.6% change in ratio 7. This signify that change in distance between nose and mouth is slightly greater than change in distance between mouth and middle of eyes. It can the concluded that there is slight displacement of the mouth from the nose as one age.

4.2.2 Landmark displacement across age summary

The results of analysis show that as one grows, facial landmarks get displaced from each other. The mouth is more displaced from the eyes than is any other landmark displaced from the other. It was also found that ageing does not contribute much displacement between mouth and nose and between mouth and forehead. Displacement of eyes from each other was lower than displacement of eyes from the nose and the mouth. Results show that the mouth is slightly displaced from the nose as one age. Findings show that nose is located at a distance slightly higher than half of distance between eyes, 45^o south of eyes and mouth is located at a distance almost equal to distance between eyes, about 63^o degrees south of eyes. It is evident that there are other factors that contribute to displacement of landmarks as one grows. This could be extrinsic factors like diet or intrinsic factors like genes. This study recommends an extensive analysis of factors affecting displacement of facial landmarks in humans.

4.3 Age-group estimation

This section presents results for the experiment described in section 3.6.

The dataset is split into four age groups as shown in Table 3.3 since age group estimation reduces drastically with increase in number of groups due to reduction in number of images per group [239]. Age grouping used is almost similar to grouping in [237] and [239]. Age-group classification was performed on individual fused feature vectors to determine their performance.

Experimental results for age-group estimation based on global shape and appearance features is shown in Table 4.3.

TABLE 4.3: Cumulative scores of age estimation using holistic shape and appearance features

Age-group	**Shape + Appearance**	
	SVM(%)	**ANN(%)**
1	78.2	94.5
2	75.1	75.8
3	77.4	80.6
4	80.3	82.7

It was observed that age-group estimation accuracy is higher in age-group 1 and 4. In lower age-groups, facial shape change significantly while appearance remains relatively constant. Shape features play a major role in classifying images in these age group.

In adulthood, facial shape remains slightly constant as there is significant change in appearance. In these age-groups, appearance features are vital in discriminating between age-group. Fusing of these two features contribute to better accuracy across all age groups. Table 4.4 shows the cumulative scores for age-group estimation based on forehead texture and wrinkle fused features.

Forehead Gabor and LBP features can discriminate between age-groups 1 and 2 by

TABLE 4.4: Cumulative scores of age estimation using forehead texture and wrinkle features

Age-group	Texture + Wrinkle	
	SVM(%)	ANN(%)
1	45.2	51.5
2	64.3	53.7
3	72.4	74.2
4	78.5	82.6

accuracies in the range of 45% – 60%. As one grow into age-group 3, the forehead texture become distinct from age-group 1 and 2. Accuracies improve to slightly around 80% in age-group 4. This could be attributed to introduction of heavy wrinkle lines on the forehead in age group 4. The performance of these features could also be adversely affected by hair style that covers part of the forehead.

As shown in Table 4.5, LE and RE LBP and Gabor features performs quite well in age-group estimation.

TABLE 4.5: Cumulative scores of age estimation using eyes texture and wrinkle features

Age-group	Texture + Wrinkle			
	LE		RE	
	SVM(%)	ANN(%)	SVM(%)	ANN(%)
1	54.2	52.3	55.7	53.1
2	66.0	64.4	65.8	65.5
3	73.1	75.7	74.2	76.7
4	76.3	79.6	76.8	78.4

Age-group discrimination using LE texture and wrinkle features achieved accuracy of 50% – 70% in age-groups 1 and 2. This could be attributed to negligible texture variations in these age groups. Wrinkles are also rarely found around eye regions in this age group. Accuracy of between 70% - 80% was achieved in age group 3 and 4. Accuracy improvement in last two age groups could be attributed to significant texture variations around the eye region and introduction of wrinkles as one grow.

As shown in Table 4.6, LBP and Gabor features from LC and RC performed slightly lower compared to eye features in age-groups 1 and 2. As one move to age-group 3 and 4, age spots, wrinkles and general texture variations start appearing in cheek bone areas. This leads to improvement in LC and RC features performance from 50

TABLE 4.6: Cumulative scores of age estimation using cheeks texture and wrinkle features

Age-group	Texture + Wrinkle			
	LC		RC	
	SVM(%)	ANN(%)	SVM(%)	ANN(%)
1	48.4	53.2	50.3	53.7
2	62.5	60.0	63.1	61.0
3	78.1	76.8	79.2	76.2
4	81.4	83.2	80.8	81.6

– 63% in age-groups 1 and 2 to the range of 75 – 84% in age-group 3 and 4.

As shown in Table 4.7, performance on age-group 4 was highest with accuracy of 75% - 84% on NB texture and wrinkle features. This could be attributed to introduction of heavy wrinkle lines and significant texture variations around NB in adults and old age. In old age, freckles start appearing around NB due to over-exposure to UV rays from the sun.

TABLE 4.7: Cumulative scores of age estimation using cheeks texture and wrinkle features

Age-group	**Texture + Wrinkle**			
	NB		**NO**	
	SVM(%)	**ANN(%)**	**SVM(%)**	**ANN(%)**
1	55.4	53.2	57.2	65.8
2	57.8	58.1	72.1	69.6
3	61.5	69.0	75.5	73.3
4	75.1	83.6	78.2	80.8

Local Binary Patterns and Gabor features from NB region in age groups 1 and 2 were slightly low with accuracies in the range of 50% - 60%. This is due to relatively constant skin texture and lack of wrinkle features in this region in these age groups. In age group 3, accuracies improve to the range of 60% - 70%. Similarities between subjects in age-group 3 and those in age groups 2 and 4 could have contributed to this lower accuracy. It was observed that NB region provides rich texture and wrinkle information for distinguishing age group 3 from age group 4. Wrinkle and texture feature from the NO region performed slightly better than features from NB region for age group 1 and 2.

Performance in age-group 4 had highest accuracy of 78% - 81% with age-group 2 and 3 having accuracies of between 70% and 75%. Similar to other regions, age-group 1 had lowest accuracy of between 55% and 60%. Performance in age-groups 3 and 4 could have been affected by partial occlusion/nose shadow and presence and absence of mustache.

Table 4.8 shows the performance of texture and wrinkle features from MO region. Age groups 1 and 2 had accuracies of between 67% - 71%. Performance of features

TABLE 4.8: Cumulative scores of age estimation using mouth texture and wrinkle features

Age-group	**Texture + Wrinkle**	
	SVM(%)	**ANN(%)**
1	67.2	69.3
2	70.5	68.8
3	59.5	58.7
4	62.4	63.2

from MO was lowest in age-groups 3 and 4 with accuracy of 58% - 64%. This could be attributed to facial expression like smiling and laughing that extremely influence LBP and Gabor features extracted from MO region. Another factor that could have affected this accuracy is mustache and beards around the mouth.

Two classifiers are combined to make an ensemble of 9 classifiers; 1 for each landmark and 1 for global fused feature. Using majority voting strategy, decision from these classifiers are fused in order to get single ensemble output for each test sample. Figure 4.4 shows the overall performance of age-group estimation using 68 2D Landmark Points, LDA, LBP, Gabor filter, SVM and ANN. Figure 4.4 shows overall performance of feature and decision level fusion in age group estimation. Using

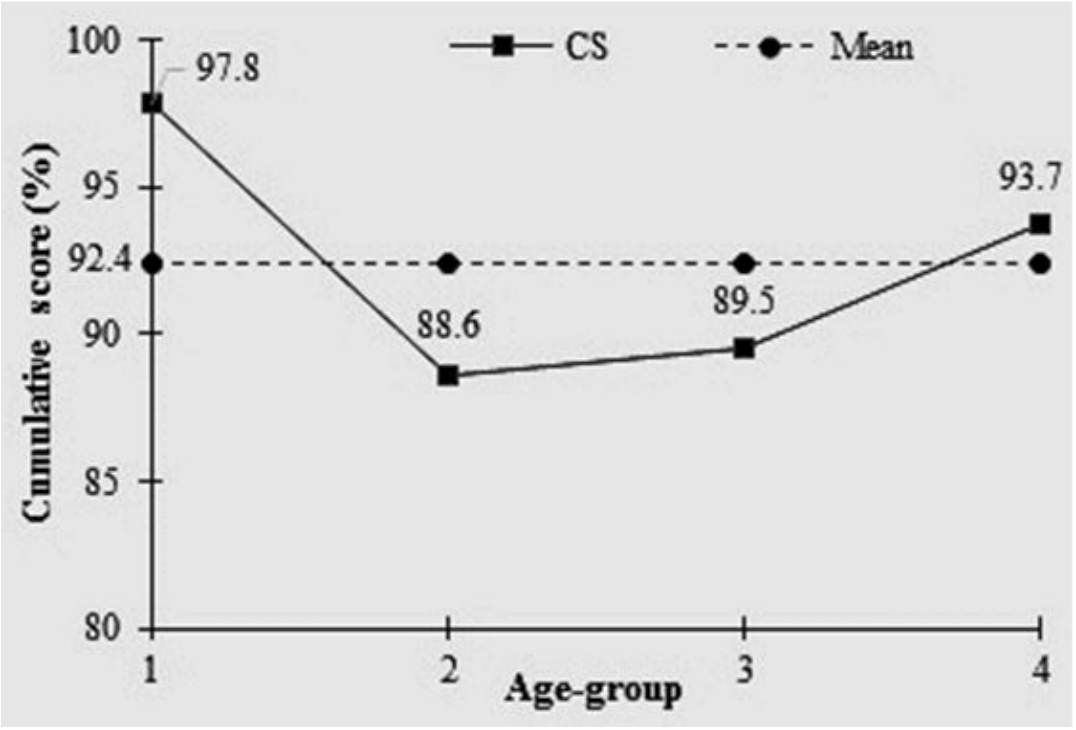


FIGURE 4.4: Cumulative score for age-group estimation based on feature and decision level fusion of global shape and appearance and local wrinkle and texture feature

hybrid of global and local facial features with feature fusion and decision fusion significantly improve age-group estimation. The overall performance of age-group 1 had highest accuracy of 97.8%. Age-group 2 recorded overall performance of 88.6% while age-group 3 overall performance was 89.5%. Age-group 4 had performance of 93.7%. The general performance of this approach is 92.4%.

Performance of age-group 4 could be attributed to significant contribution from appearance, texture and wrinkle features. In this age-group, appearance, texture and wrinkles vary significantly as compared to subjects in lower age-groups. Lower performance of age-group 2 could be attributed to similarity of subjects in this age-group with subjects in age group 1 and 3.

4.3.1 Component-based age-group estimation

This section present results of performance of each component in age-group estimation. Features from a particular component are fused and used to train ANN and SVM for age-group estimation. For eyes and cheeks, texture and wrinkle features from both components are fused to obtain one feature vector[1]. The resultant features are fused before training a classifier for age-group estimation based on the respective component.

Table 4.9 shows the cumulative scores of component-based age-group estimation using ANN and Table 4.10 using SVM.

As shown in Table 4.9, holistic facial shape and appearance performed better than each of the components with an average cumulative score of 83.4%. Combined

[1]LE texture is fused with RE texture to get texture feature $\mathbf{T_{eyes}}$. Left eye wrinkle is fused with RE wrinkle to obtain $\mathbf{W_{eyes}}$ wrinkle feature. Thereafter, $\mathbf{T_{eyes}}$ is fused with $\mathbf{W_{eyes}}$ to obtain final feature vector used for classification. Same approach is used for cheeks.

TABLE 4.9: Cumulative scores for component-based age estimation using ANN

Age-group	No. images	Holistic	Fore-head	Eyes	Cheeks	Nose	Mouth	Nose-bridge
0-6	274	94.5	51.5	53.1	53.9	65.8	69.3	53.2
7-14	271	75.8	53.7	65.6	61.4	69.6	68.8	58.1
15-24	249	80.6	74.2	77.2	76.9	73.3	58.7	69.0
25-69	208	82.7	82.6	80.3	83.5	80.8	63.2	83.6
Total	**1002**	**83.4**	**65.50**	**69.05**	**68.93**	**72.38**	**65.00**	**65.98**

holistic shape and appearance feature had highest cumulative score of 94.5% in age-group 0-6. This could be attributed to change in facial shape in this age-group due to craniofacial development. In age-group 7-14, the cumulative score dropped to 75.8%. Age-group 15-24 achieved cumulative score of 80.6% which is lower to 82.7% achieved in age-group 25-69 and 94.5% achieved in age-group 0-6. Lower cumulative score in age-groups 0-6 and 25-69 could be attributed to similarities between images in these age-groups with the lower and upper age-group.

The nose posted better average cumulative score of 72.38% compared to other facial landmarks. In age-group 0-6, nose achieved 65.8% compared to 53.1% achieved with eyes and 51.5% achieved with forehead in the same age-group. Artificial Neural Network achieved 69.6% with nose wrinkle and texture feature in age-group 7-14. The performance improved to 73.3% in age-group 15-24 and to 80.8% in age-group 25-69. This steady improvement of ANN performance with nose features could be attributed to change in wrinkles and texture of the nose as one grow. The nose is exposed to sun UV rays and its texture, skin tone and wrinkles change steadily with increase in age.

Eyes and cheeks posted almost same results with average cumulative scores of 69.05% and 68.93% respectively. Higher cumulative scores, 80.3% for eyes and 83.5% for cheeks, were achieved in age-group 25-69 because there is significant change in texture and wrinkle features in this age-group. Subjects in this age-group are exposed to external factors like UV rays, air pollution, smoke, dry wind that accelerates change in texture and wrinkles. Lower cumulative scores of about 53% were achieved in age-group 0-6 since in this age-group, texture of the eye and cheek regions remain relatively constant because this age-group is characterized with change in facial shape but facial appearance remains significantly constant. Another reason could be mostly subjects in this age-group are not exposed to extrinsic factors like air pollution, UV rays and wind that accelerates change in texture and wrinkles. Cumulative score increase with increase in age because as one age, there is changes in facial bony structure, lose of collagen and over-production of melanin leading to change in facial texture and wrinkle development. Intuitively, exposure to extrinsic factors that accelerates change in texture and wrinkles increase with increase in age.

There was an increase in cumulative scores across age when age-group estimation is performed with forehead texture and wrinkle features. Lower cumulative score of 51.5% was achieved in age-group 0-6 because subjects in this age-group hardly have any significant change in texture and wrinkle lines on the forehead. But as one age, exposure to UV rays, air pollution and other factors cause change in forehead texture. Appearance of wrinkle lines across the forehead in old age explains better performance of forehead features in age-groups 15-24 and 25-69. Forehead performance could have been also affected by different hair-lines and hair styles across age.

Age-group estimation with mouth texture and wrinkle features achieved 69.3% in age-group 0-6. The accuracy drops across age. This could be attributed to factors like facial expressions like smiling and laughing that adversely affect appearance of the mouth region given that images in the experimental dataset were not of neutral face. Another factor that affects performance of mouth texture and wrinkle features across age is presence of mustache and beards in some subjects within an age-group. Nose-bridge region cumulative score increased with increase in age. As one age, the region above the nose and between the eyes experience some significant change in texture and wrinkles. In some subjects, this region develop horizontal wrinkle lines in adult and old age and some have hair in this region joining the two eyebrows. This makes texture and wrinkle features discriminate people in these ages from ones in younger age-groups.

TABLE 4.10: Cumulative scores for component-based age estimation using SVM

Age-group	No. images	Holistic	Fore-head	Eyes	Cheeks	Nose	Mouth	Nose-bridge
0-6	274	78.2	45.2	56.1	51.3	57.2	67.2	55.4
7-14	271	75.1	64.3	66.3	63.4	72.1	70.5	57.8
15-24	249	77.4	72.4	74.2	79.9	75.5	59.5	61.5
25-69	208	80.3	78.5	77.0	81.8	78.2	62.4	75.1
Total	**1002**	**77.75**	**65.10**	**68.40**	**69.10**	**70.75**	**64.90**	**62.45**

As shown in table 4.10, same age-group estimation performance trend was observed when using SVM for age-group classification.

As it is evident from Table 4.9 and Table 4.10, ANN outperforms SVM in age-group estimation with ANN achieving better cumulative scores than SVM when trained on the same features.

4.3.2 Comparison with previous methods

It is a bit challenging to compare age-group estimation approaches due to differences in age-grouping in various approaches. This study compares performance in age-group estimation with similar age-groups in previous approaches.

As shown in Figure 4.4, fusing shape, appearance, texture and wrinkle features improved performance in middle age groups from 40% – 60% reported in [241] to the range 80% – 90%. Adult age-group performance was also improved to 93.7% as compared to 87.5% reported in [243]. Proposed technique achieved performance of 89.5% in teen age-group (15 – 24 years) as compared to teen age-group (12 – 21 years) performance of 87% reported in [243]. Age group 1 (0 – 6 years) performance improved to 97.8% as compared to 92.1% baby group (0 – 3 years) accuracy reported in [237]. Proposed approach achieved 93.7% accuracy in adult age group (25 – 69 years) compared to 90.1% reported in [237] for age group 20 -59 years. Proposed technique achieved performance of 89.5% in age group 3 (15 – 24 year) as compared to 75% for age range 16 – 30 years reported in [236].

Table 4.11 show a summary of comparison of overall performance of the proposed approach with previous methods in literature.

As shown in Table 4.11, proposed approach has improved overall age group estimation from 91.4 and 91.3% reported in [237] and [239] respectively to 92.4%. This shows that using both global and local features with feature fusion and decision fusion improves performance of age-group estimation.

TABLE 4.11: Age-group estimation performance comparison

Method	**Age grouping/Dataset**	**Accuracy (%)**
2D Landmark Points, Regression Vector Machines, Partial Least Squares, Nearest Neighbor, Naive Bayes, Fisher Linear Discriminant [234]	0 – 15, 15 - 30, 30+ FG-NET	70.0%
LBP, Minimum Distance, Nearest Neighbor [235]	$10 \pm 5, 20 \pm 530 \pm 5, 40 \pm 550 \pm 5, 60 \pm 5$ FERET	80.0%
Anthropometric Distance Ratios, Canny, Artificial Neural Network [74]	0-15, 16-30, 31-50, 51+ IFDB	86.6%
HOG, Probabilistic Neural Network [236]	0-15, 16-30, 31-50, 50+ IFDB	87.0%
2D Landmark Points, GOP, ROC, SVM [28]	0-3, 4-19, 20-59 FG-NET	91.4%
Gradient of Oriented Pyramids, SVM [239]	0-5, 6-12, 13-21, 22-69 FG-NET	91.3%
Proposed: Anthropometric Distance Ratios, LBP, 2D Landmark Points, Gabor, LDA, SVM, Artificial Neural Network	0-6, 7-14, 15-24, 25-69	92.4%

4.3.3 Age-group estimation summary

An age-group estimation method using feature fusion, decision fusion and ensemble of classifiers is proposed. Global shape features are extracted using landmark points and distance ratios between fiducial points. LDA is used to extract Global appearance features. After z-score normalization on each feature vector, global shape and appearance features are fused into one feature vector for training one of the classifiers. Face image is then divided into 8 regions of interest (ROI). ROI based texture and wrinkle features are extracted by LBP and Gabor filters respectively. For each ROI, these features are z-score normalized and then fused to make one feature vector.

Linear Discriminant Analysis is used to normalize dimensionality of features before z-score normalization and fusion. Fused features are used to train an ensemble made of ANN and SVM classifiers for age-group estimation. A total of 9 models are trained. Decision level fusion is adopted by use of majority voting scheme for final age-group label. Experimental results on widely used FG-NET ageing dataset show that the proposed approach achieves better accuracy in age-group estimation against previous methods. To improve exact age estimation, in-depth inquiry is necessary for performing age-group estimation with narrow uniform age groups (0 – 4, 5 – 9, 10 – 14...96 - 100). Further inquiry may be done to verify robustness of proposed method across various face ageing datasets. Further investigation is required to make age-group and age estimation robust to various dynamics like facial expression, cosmetics, illumination and differences in perceived ageing between males and

females.

4.4 Component-based age estimation

In this section, performance of each component in age estimation is presented. Features from a particular component are fused and used to train ANN and SVM for age estimation. For eyes and cheeks, texture and wrinkle features from both components are fused to obtain one feature vector. The resultant features are then fused before training a classifier for age estimation based on the respective component.

For each component, two classifiers are trained, SVM and ANN to perform exact age estimation based on fused texture and wrinkle features. Two classifiers are also trained, SVM and ANN, to perform age estimation based on fused holistic appearance and shape feature. Decisions from these two classifiers are fused to obtain final estimated age label. Table 4.12 show MAE of age estimation based on fused holistic and landmark features.

TABLE 4.12: Component-based age estimation performance using MAE

Age-group	No. images	Holistic	Fore-head	Eyes	Cheeks	Nose	Mouth	Nose-bridge
0-6	274	3.67	9.01	7.15	8.37	9.23	10.21	9.88
7-14	271	2.53	7.82	6.04	7.82	8.12	8.74	7.90
15-24	249	3.50	4.30	3.45	4.03	4.58	8.92	4.56
25-69	208	6.87	5.24	3.86	4.78	4.31	6.75	6.27
Total	**1002**	**5.97**	**7.10**	**5.62**	**5.96**	**6.23**	**7.02**	**6.87**

As shown in Table 4.12, holistic facial shape and appearance, eyes and cheeks wrinkle and texture achieves better MAE compared to other components.

Holistic shape and texture achieved MAE of 3.67 years in age-group 0-6 years and 2.53 years in age-group 7-14 years. The MAE improves in the 7-14 age-group because subjects in this age-group exhibit significant changes in facial shape due to craniofacial development. There is slight variation in facial appearance in these age-groups, but fusing appearance features with shape features results into a rich age discriminative feature. The performance deteriorates as one move to age-group 15-24 because of increase in the age range with individuals in at the top of this age-group (above 20 years) having significant change in facial appearance as compared to shape. Craniofacial development tend to stop as around 18 years where change in facial appearance start becoming more significant compared to change in shape. MAE 0f 6.87 years was achieved in age-group 25-69 years. This poor performance could have been contributed by wide range of the age group in addition to reduction of number of images as one move to senior ages in the experimental dataset[2].

Texture and wrinkle features from eyes and cheeks regions achieved MAE of 5.62 years and 5.96 years respectively. The MAE improves with age. This is because, as one ages, texture and wrinkles features around the eyes and the cheeks change significantly due to exposure of these regions to extrinsic factors like UV rays and air pollution. There is lose of collagen underneath the skin makes skin texture (tone) to change as one ages. Rearrangement and shifting of bony structure around the cheek-bone area makes the skin to become leathery and wrinkles start appearing

[2]There are only 69 images with the age of above 40 years with very few images per subject. These images are very few to learn ageing pattern of subjects

in this region in adult and old age. Development of wrinkle lines at the eye-corners also make eye features more age discriminative as one move up the age-group. Generally, the eye features perform better compared to cheek features because cheeks may have unstable appearance due to facial expressions, facial hair and occlusion.

Nose and nose-bridge regions achieved MAE of 6.23 years and 6.87 years respectively. The performance improves across age because as one age, the more the nose and nose-bridge region change in texture and appearance of wrinkle lines. The nose and nose-bridge are adversely affected by exposure to UV rays from the sun. This causes change in skin tone (texture) and in adult and old age, wrinkle lines start to appear at the nose-bridge region. Forehead features achieved MAE of 7.10. As observed in other landmarks, the performance improved with increase in age. This is because forehead texture and wrinkles vary with increase in age. Although forehead could be having more discriminative texture and wrinkle features, its performance could have been affected by varied hair lines and hair styles. Mouth achieved MAE of 7.02 years. Mouth performance could have been affected by expressions like smiling and laughing in addition to effect of mustache and beards in some individuals.

4.4.1 Component based age estimation summary

Eyes posted better MAE of 5.62 years compared to the rest of components used. Holistic based age estimation approach accuracies deteriorated as one move up the age group. This is because as one move from childhood to adulthood and old age, facial shape remains relatively unchanged as compared to facial texture. Although facial texture varies in adulthood and old age, holistic feature extraction approach does not capture more discriminative texture and facial appearance features. Local features descriptors achieve better accuracies as one move up the age groups. This is attributed to robustness of local feature descriptors in extracting subtle age discriminative features.

4.5 Age estimation using SOR-LDP

This section presents results for experiment described in section 3.8.5.

FG-NET ageing dataset is split into 7 age-groups of 10 years. The first age-group is $0-9$ years and last age-group is $60-69$ years. Table 1 show MAE in years achieved in each group and comparison with LBP, LDP and HOG.

TABLE 4.13: Age-group MAE (years) comparison using LBP, LDP and SOR-LDP

Group	**images**	**SOR-LDP [Proposed]**	**HOG** [130]	**LDP** [47]	**LBP** [44][299]
0-9	371	2.68	2.79	2.83	2.97
10-19	339	3.22	3.42	3.47	3.70
20-29	144	3.55	3.88	4.20	4.25
30-39	70	7.62	8.15	8.74	9.03
40-49	46	12.85	13.92	14.60	15.26
50-59	15	18.24	19.63	20.59	21.34
60-69	8	29.22	31.12	31.61	32.20
Total	**1002**	**4.05**	**4.64**	**5.12**	**5.97**
CS		**80.2%**	**77.2%**	**76.6%**	**70.4%**

As shown in Table 4.13, SOR-LDP achieved MAE of 2.68 years in age-group 0-9 compared to 2.97, 2.83 and 2.75 achieved by LBP LDP and HOG respectively recording an improvement of about 0.11 years in age estimation. In age-group 10-19, SOR-LDP achieved MAE of 3.22 posting age estimation improvement of 0.20 years, compared to 3.42 achieved by HOG, 3.47 achieved by LDP and 3.70 achieved by LBP in the same age-group. As a subject move up the age-groups, the number of images per group in the experimental dataset reduces drastically. With 144 images in age-group 20-29, SOR-LDP achieved MAE of 3.55 years while HOG achieved 3.88, LDP achieved 4.20 and LBP achieved 4.25 years. SOR-LDP improves age estimation in this age-group by 0.33 years as compared to HOG in the same age-group. SOR-LDP achieved MAE of 7.62 in age-group 30-39 as compared to 8.15 years achieved by HOG, 8.74 years achieved by LDP and 9.03 years achieved by LBP. SOR-LDP improved age estimation accuracies in this age-group by 0.53 years. In age-group 40-49, SOR-LDP achieved MAE of 12.85 years, which is an improvement of 1.07 years in age estimation accuracy compared to 13.92 years achieved by HOG, 14.60 years achieved by LDP and 15.26 years achieved by LBP. In age-group 50-59, SOR-LDP achieved 1.39 years improvement on age estimation accuracies by achieving MAE of 18.24 years as compared to 19.63 years achieved by HOG, 20.59 years achieved by LDP and 21.34 years achieved by LBP. SOR-LDP posted an improvement of 1.9 years on age estimation accuracies in age-group 60-69, as it achieved MAE of 29.12 years as compared to 31.12 achieved by HOG, 31.61 years achieved by LDP and 32.20 years achieved by LBP. SOR-LDP achieves overall MAE of 4.05 which is better compared to 4.64 achieved by HOG, 5.12 years achieved by LDP and 5.97 achieved by LBP. The MAE are poor in older age-groups because there are very few images in these age-groups making it hard to learn any ageing patterns.

As can be seen from Table 4.13, SOR-LDP improves age estimation accuracies by 1.07 years as compared to LDP and by 0.59 as compared to HOG. Age estimation improvements increase as a subject move up the age-groups for all the three techniques. This is because SOR-LDP, HOG, LDP and LBP are local texture descriptors and capture more discriminative texture features in older age-groups as compared to lower age-groups. As one move from age-group 0-9 year to 10-19 years, the most significant changes on the face is shape with very slight variations in texture. This explains why there is minimal improvements in these age-groups. In age-group 20-29, facial texture start varying with minimal variation in facial shape. This variation in texture explains the improvements in age estimation accuracies in age-groups 20-29 and above. In old age, age-groups 40-49 and above, facial wrinkles start appearing on some parts of the face like forehead, eye-corners, nose-bridge, cheeks and mouth-corners. The edge responses of these wrinkles are more significant in these age-groups compared to lower age-groups. This explains the significant improvements in age estimation accuracies in age-groups above 40 years. SOR-LDP robustness to orientation-based edge response improves general age estimation by 1.07 years as compared to LDP accuracy. SOR-LDP takes an average of $2.1 \times 10^6 \mu s$ to encode an image of size 120×120 pixels as compared to average of $3.4 \times 10^6 \mu s$ taken by LDP to encode image of same size. LBP has the fastest average execution time of $1.2 \times 10^5 \mu s$ measures of the same image size.

4.5.1 Summary of age estimation using SOR-LDP

Significant-orientation response local directional pattern (SOR-LDP) which is a variant of local directional pattern (LDP) operator is proposed. The proposed method divides an image into local regions and then calculates texture features by comparing gradient responses for each orientation. The texture description of a single local region describes texture of that region. Combining texture description for all the regions describes the global texture of the image. SOR-LDP assigns 8-bit binary code to a central pixel by comparing two responses for each orientation. The orientations

considered in this paper are $0^o, 45^o, 90^o$ and 135^o. The bit corresponding to the maximum response is set to 1 while the other bit is set to 0. The resultant binary string is converted to decimal and assigned to reference pixel as its SOR-LDP code. Unlike LDP which considers top k responses, SOR-LDP considers all 8-directional responses by pairing them into four distinct orientations. Therefore, all the responses play a role in determining the image gradient. Experimental results on FG-NET ageing dataset show that SOR-LDP out-performs LBP, LDP and HOG in age estimation. Experiments in this paper focused on using proposed method for age estimation. Performance of SOR-LDP in other application domains like face recognition, gender classification, satelite image classification among others may be investigated. Further research could be done to investigate robustness of SOR-LDP to random noise and its performance in other datasets. A comparison of SOR-LDP operator with LTP is also recommended. Although SOR-LDP improves performance of LDP, it is found to be slower compared to LBP. An optimization of LDP and SOR-LDP is required to make them faster for real time application.

4.6 Age estimation using DHOG

This section presents results for experiment discussed in section 3.8.7.

Table 4.14 shows results achieved for age estimation using HOG and DHOG features. Findings in Table 4.14 shows that DHOG outperforms HOG in age estimation. With DHOG encoding image gradients at four orientation, it encodes more age discriminative features across all age-groups. There is significant difference in MAEs achieved in each age-group apart form age-group 20-29.

TABLE 4.14: Age-group MAE comparison using DHOG and HOG

Group	**DHOG [Proposed]**	**HOG** [130]
0-9	2.56	2.79
10-19	2.78	3.42
20-29	3.32	3.88
30-39	7.40	8.15
40-49	12.67	13.92
50-59	18.05	19.63
60-69	29.34	31.12
Total	**3.97**	**4.64**
CS	**79.8%**	**77.2%**

The slight improvement of DHOG in this age-group could be attributed to similarities of individuals in this age-group and those in age-group 10-19 and 30-39. The overall accuracy of DHOG is lower compared to HOG since DHOG encodes more discriminative image gradients in all age-groups.

For every pixel in the image, DHOG considers all the eight neighbouring pixels in encoding the image gradient at that particular point, unlike HOG which only consider four pixels. It is evident from the findings that image gradients extracted are more discriminative when all eight neighbouring pixels of each reference pixel are considered. This results into a rich feature vector with more discriminative edges/-gradients introduced by the extra four neighbouring pixels considered. Preliminary results show that DHOG performs better in age estimation compared to HOG.

4.7 Age estimation using LTDP

This section presents results for experiment described in 3.8.6.

FG-NET ageing dataset is split into 7 age-groups of 10 years. The first age-group is $0-9$ years and last age-group is $60-69$ years. Table 1 shows MAE error achieved in each group for LTDP, LDP and LTP operators.

TABLE 4.15: Age-group MAE (years) comparison using LTDP, LDP and LTP

Group	**images**	**LTDP [Proposed]**	**LDP** [139]	**LTP** [48]
0-9	371	2.72	2.83	2.94
10-19	339	3.26	3.47	3.71
20-29	144	3.41	4.20	4.04
30-39	70	7.58	8.74	8.98
40-49	46	13.65	14.60	15.30
50-59	15	19.84	20.59	21.02
60-69	8	30.37	31.61	31.88
Total	**1002**	**4.35**	**5.12**	**5.74**
CS		**78.3%**	**76.6%**	**73.4%**

As shown in Table 4.15, LTDP achieved MAE of 2.72 years in age-group 0-9 compared to 2.83 and 2.94 achieved by LDP and LTP respectively. In age-group 10-19, LTDP achieved MAE of 3.26 compared to 3.47 achieved by LDP and 3.71 achieved by LTP in the same age-group. With 144 images in age-group 20-29, LTDP achieved MAE of 3.41 years compared to 4.20 and 4.04 achieved by LDP and LTP respectively. Performance of all three operators deteriorates drastically as from age-group 30-39 due to drastic decrease in number of images per group. Nevertheless, LTDP performed better than LDP and LTP in age-group 30-39 by achieving MAE of 7.58 years compared to 8.74 achieved by LDP and 8.98 achieved by LTP. The performance is poorest in age-group 60-69 because the dataset used has only 8 images in this age-group, which are not sufficient to learn any ageing pattern. LTDP achieved overall MAE of 4.35 which is superior relative to 5.12 achieved by LDP and 5.74 achieved by LTP.

It is evident from the experiments that LTDP encodes more discriminative local texture features compared to LDP and LTP. LTDP improves age estimation accuracies by MAE of 0.77 compared to LDP and by 1.42 compared to LTP. The accuracy of LTDP could be attributed to involvement of central pixel as well considering all eight directional responses in calculating image gradient. This shows that all the responses as well as central pixel are vital in achieving age discriminative local texture features.

4.8 Age estimation using MF-BIF

This section presents results for experiments described in section 3.8.8.

Accuracy of the proposed age estimation technique is evaluated using LOPO validation protocol on FG-NET ageing dataset. Each face image is cropped and MF-BIF features are extracted from global face are used for age-group classification using ANN followed by SVR exact age regression using MF-BIF. A tuning set, part of the training set, is used to select parameters for the classifier and regressor. RBF regression kernel width is set to be between 0.1 and 0.01 [158]. Classification and regression is combined in order to improve age estimation results as recommended in [49].

FG-NET ageing dataset is split into 7 age-groups of 10 years as shown in Table 3.2. The first age-group is 0 – 9 years and last age-group is 60 – 69 years. Table 4.16 shows the cumulative scores of age-group classification (estimation) using MF-BIF.

TABLE 4.16: Age-group estimation using MF-BIF cumulative score

Age-group	Number of images	Cumulative Score (CS %)
0-9	371	84.2
10-19	339	81.6
20-29	144	72.8
30-39	70	59.6
40-49	46	42.3
50-59	15	22.8
60-69	8	0.5
Total	**1002**	**80.3**

MF-BIF achieved cumulative score of 84.2% in age-group 0-9 and 81.6% in age-group 10-19. Age-group performance drops with increase in age because of reduction of number of images in higher age-groups. Age-group 20-29 achieved cumulative score of 72.8% while age-group 30-39 achieved 59.6%. Drop in performance in this age-group could be attributed to similarities of images in this age-group with images in 10-19 and 30-39 age-groups in addition to reduction of total number of images in the age-group. Cumulative score of 42.3% was achieved with 46 images in age-group 40-49. The worst cumulative score of 0.5% in age-group 60-69 is because the number of images in this age-group are too few to learn any ageing pattern.

The difference between these cumulative score and that reported in table 4.9 and Table 4.10 is attributed to differences in groupings and features used. The grouping used in Table 4.9 and Table 4.10 result into many images within an age-group. As it can seen from these results, number of images in a group affects the accuracy of age-group estimation with fewer images per group posting lower accuracies compared with groups with many images. The width of the group (range of years within a group) also affect accuracy of age group estimation because age-introduced variations are temporal and hence learning an ageing pattern over so many years results into lower accuracies. Overall age-group estimation accuracy using MF-BIF was 80.3%.

Table 4.17 show MAE error achieved in each group and comparison with previous studies, with same grouping, evaluated on FG-NET ageing dataset.

This technique achieved MAE of 4.02 years on FG-NET which is significantly lower than 4.77 years reported in [42], 4.43 years reported in [52], 4.60 years reported in [255], 4.32 years reported in [254]. MF-BIF achieved better results on FG-NET than the best earlier reported MAE of 4.32 in [254]. Better results than ours were reported in [2] (MAE 3.80) and [300] (MAE 3.17), however, these study used an enhanced BIF feature constituted by fusing BIF and Active Shape Model (ASM) features [300] or boosted BIF [2] . This is an indication that fusing MF-BIF with other features like shape and texture could further improve age estimation accuracy to MAEs lower than achieved by state-of-the-art techniques.

As shown in Table 4.17, MAE of all the studies significantly deteriorates as one move from lower to higher age-groups. This is because, FG-NET dataset has few images of adults (≥ 30 *years*) compared to teen and child images, (≤ 30 *years*). Age-group 60-69 achieves worse MAE because there are only 8 images in this group, which are too few to learn any ageing pattern.

TABLE 4.17: Age-group MAE comparison using MF-BIF

Age-group	**# imag**	**MF-BIF [Proposed]**	**BIF** [42]	**EBIF** [300]	**RUN** [54]
0-9	371	2.23	2.99	1.67	2.51
10-19	339	2.89	3.39	1.09	3.76
20-29	144	3.95	4.30	3.31	6.38
30-39	70	8.02	8.24	7.87	12.51
40-49	46	13.88	14.98	13.02	20.09
50-59	15	20.26	20.49	18.00	28.07
60-69	8	29.97	31.62	26.25	42.50
Total	**1002**	**4.02**	**4.77**	**3.17**[a]	**5.78**

[a]Used fused features(Fusion of ASM and BIF)

A comparison of the overall performance of MF-BIF and BIF in age estimation on FG-NET is shown in Table 4.18.

TABLE 4.18: Comparison of BIF age estimation performance on FGNET

Method	**MAE(years)**	**CS(%)**
Guo *et al.* [42]	4.77	≈69%
Guo and Mu (MORPH) [52]	4.43	—
Han *et al.* [255]	4.60	74.8%
Spizhevoi & Bovyrin [254]	4.32	74.7%
Han *et al.* [2]	3.80 [a]	78.0%
El Dib & El-Saban [300]	3.17[b]	—
Proposed Method (MF-BIF)	**4.02**	**76.2%**

[a]BIF features were selected using a boosting algorithm
[b]Used fused features(Fusion of ASM and BIF)

As shown in Table 4.18, MF-BIF performs better than best reported MAE of 4.32 years on FG-NET. This is an indication that using multiple frequencies results in a better BIF. Cumulative score improved from 74.8% in [255] to 76.2%. However, MF-BIF is outperformed by Enhanced BIF (EBF) [300] and boosted BIF [2]. This is an indication that fusing BIF with other features gives better results than MF-BIF. Since MFBIF outperforms BIF, enhancing and boosting MF-BIF could result in better accuracies than reported in [2] and [300].

4.8.1 Summary of age estimation using MF-BIF

MF-BIF is examined for age estimation. To extract MF-BIF, Gabor filters at multiple spatial frequencies, multiple orientation and multiple scales are used to build S1 features. MAX pooling is applied to S1 features at each frequency to obtain maximum responses X^f. Each pixel in X^f is then represented as a mean of the neighbouring pixels in the $s \times s$ neighbourhood (AVG pooling) to obtain X^o. MAX pooling is applied to X^o to get a single feature vector X^s for each scale. For each X^s for $s = 1, 2, 3, \ldots, n$ MAX pooling is applied to features in two consecutive scales, X^s and X^{s+1} to obtain X^p for $p = 1, 2, 3, \ldots n/2$ features. STD operator is then applied on each X^p feature. The resultant features are concatenated and used for age estimation. MF-BIF have been evaluated for age estimation and results show improved accuracies compared to normal BIF that use only multi-orientation and

multi-scale feature selection. However, enhanced BIF [300] and boosted BIF [2] outperforms MF-BIF. Motivated by performance of EBIF (Fusion of BIF and ASM) in [300], component based BIF in [255], and boosted BIF in [2]. Exploring enhancement of holistic and landmark-based MF-BIF by fusing MF-BIF with other features like texture, shape features and evaluate its performance in age estimation is recommended. Boosting algorithm could be applied on the enhanced MF-BIF to investigate its improvement in age estimation.

4.9 Age estimation using LS-BIF

This section presents results for experiment described in section 3.8.9.

FG-NET ageing dataset is split into 7 age-groups of 10 years. The first age-group is $0-9$ years and last age-group is $60-69$ years. Table 4.19 shows MAE error achieved in each group and comparison with previous studies.

TABLE 4.19: Age-group MAE comparison

Age-group	**# imag**	**LS-BIF** [**Proposed**]	**MF-BIF** [292]	**BIF** [42]	**RUN** [54]
0-9	371	2.08	2.23	2.99	2.51
10-19	339	2.45	2.89	3.39	3.76
20-29	144	3.73	3.95	4.30	6.38
30-39	70	7.81	8.02	8.24	12.51
40-49	46	13.54	13.88	14.98	20.09
50-59	15	19.87	20.26	20.49	28.07
60-69	8	29.76	29.97	31.62	42.50
Total	**1002**	**3.84**	**4.02**	**4.77**	**5.78**

LS-BIF achieved MAE of 3.84 years compared to 4.02 years achieved using MF-BIF in [292]. MAE achieved by LS-BIF is significantly lower compared to 4.77 years MAE achieved by BIF in [42], 4.43 years reported in [52], 4.60 years reported in [255], 4.32 years reported in [254]. Therefore, LS-BIF extracts robust age discriminative features compared to BIF and MF-BIF. This signifies that feature integration by pooling in a local neighborhood extracts robust age discriminative features compared to feature integration by global or holistic pooling across frequencies, scales and orientations. Further, using local statistical features (local mean and standard deviation) increases robustness of extracted ageing features compared to use of holistic maximum values, means and standard deviations pooled across different scales, orientations and frequencies.

Using enhanced BIF (enhancement done by applying boosting algorithm to BIF), Han *et al.* [2] achieved MAE of 3.80 years. Fusing BIF with Active Shape Model (ASM) features, El-Dib and El-Saban [300] achieved MAE of 3.17 years. It is noted that MAE achieved by LS-BIF is almost equal to MAE achieved by Boosted BIF. This is an indication that, boosting LS-BIF could achieve much better MAE than accuracy reported in [2]. Fusing LS-BIF with texture, shape and wrinkle features could see further improvement in age estimation accuracies.

As shown in Table 4.19, MAE of all the studies significantly deteriorates as one move from lower to higher age-groups. This is because, FG-NET dataset has few images of adults (*above 30 years*) compared to teen and child images, (*below 30 years*). Age-group 60-69 achieves worse MAE because there are only 8 images in this group, which are too few to learn any ageing pattern. A comparison of the overall performance of LS-BIF and BIF in age estimation on FG-NET is shown in Table 4.20.

TABLE 4.20: BIF vs LS-BIF performance on FG-NET

Method	MAE(years)	CS(%)
Guo et al.[42]	4.77	69%
Guo and Mu (MORPH) [52]	4.43	—
Han et al.[255]	4.60	74.8%
Spizhevoi & Bovyrin[254]	4.32	74.7%
Han et al. [2]	3.80	78.0%
El Dib & El-Saban [300]	3.17	—
Angulu *et al.* [292]	4.02	76.2
Proposed Method (LS-BIF)	**3.84**	**77.8%**

As shown in Table 4.20, LS-BIF performs better compared to MF-BIF reported in [292]. LS-BIF achieved CS of 77.8% compared to 76.2% achieved in [292], 69% achieved in [42], 74.8% achieved in [255] and 74.7% achieved in [254]. This is an indication that local pooling in C1 feature extraction achieves better age discriminative features compared to pooling whole image holistically. Further, local statistics of an image (mean and standard deviation) outperforms use of global maximums, standard deviations and means pooled over consecutive scales or orientations. LS-BIF is outperformed by Enhanced BIF [300] and boosted BIF [2]. Since LS-BIF performs better than BIF, it then follows that enhancing LS-BIF or boosting it will achieve better results compared to accuracies reported in [2] and [300].

4.9.1 Summary of age estimation using LS-BIF

Gabor filters at multiple spatial frequencies, orientation and scales are used to model S1 features. A global MAX operator is applied on consecutive frequency features to obtain $N \times N$ maximum frequency-based responses X^{λ}. AVG_{local} and STD_{local} operators are applied on X^{λ} in a $n \times n$ local grid, where $n << N$, to obtain two $\lfloor N/n \rfloor \times \lfloor N/n \rfloor$ features, one with locally pooled means X^{μ} and the other with locally pooled standard deviations X^{σ}. X^{μ} and X^{σ} are concatenated to form X^{t}, feature at orientation t and pushed to a vector. This process is repeated for all orientations within a scale and a vector X^{T} with features at different orientations within a scale is obtained. AVG_{global} operator is applied to features at different orientations within a scale to obtain X^{n}, feature at scale n. This process is repeated for all scales to obtain a vector X^{z} with features at all scales. MAX_{global} is applied to features in two consecutive scales X^{z} and X^{z+1} to obtain $\lfloor z/2 \rfloor$ features. These $\lfloor z/2 \rfloor$ features are concatenated and used as LS-BIF for age estimation. LS-BIF achieves better age estimation accuracies compared to MF-BIF [292] and BIF [42]. Therefore, LS-BIF extracts robust age discriminative features compared to BIF and MF-BIF. This signifies that feature integration by pooling in a local neighborhood extracts age discriminative features compared to feature integration by global (holistic) pooling. Further, using local statistical features (local mean and standard deviation) increases robustness of extracted ageing features compared to use of holistic maximum values, means and standard deviations across scales, orientations and frequencies. Motivated by performance of enhanced BIF [300] and boosted BIF [2], this study recommends enhancing LS-BIF by applying a boosting algorithm to concatenated features or fusing concatenated features with shape, texture, appearance and wrinkle features before age estimation.

4.10 Age estimation using EMF-BIF

This section presents results for experiment described in section 3.8.10.

Performance of EMF-BIF in age estimation is evaluated using ANN for age-group classification followed by exact age regression using SVR. Table 4.21 show the cumulative scores of age-group estimation using EMF-BIF.

TABLE 4.21: Age-group estimation using EMF-BIF cumulative score

Age-group	Number of images	Cumulative Score (CS %)
0-9	371	86.7
10-19	339	85.2
20-29	144	73.9
30-39	70	62.4
40-49	46	44.6
50-59	15	26.1
60-69	8	0.8
Total	**1002**	**88.4**

EMF-BIF achieved an overall age-group estimation accuracy of 88.4%. This accuracy is lower compared to 92.4% achieved using fused features and ensemble of classifiers as shown in Figure 4.4. EMF-BIF achieved an accuracy of 86.7% in age-group 0-9 and 85.2% in age-group 10-19. This performance could be attributed change in facial shape in this age-groups due to craniofacial development. With 144 images in age-group 20-29, EMF-BIF achieved an accuracy of 73.9%. Although the number of images in this age-group has reduced drastically, EMF-BIF still achieved a relatively better accuracy because fusion of shape, appearance and texture features with MF-BIF features makes EMF-BIF more age discriminative. The performance deteriorates as one move up the age groups because of drastic decrease in number of images in a group. Age-group 60-69 still achieves worst accuracy 0f 0.8% because it has only 8 images which are far too few to learn any ageing pattern within this group.

Table 4.22 show the MAE for age estimation using EMF-BIF features and ANN for age-group classification followed by within-group SVR exact age regression.

TABLE 4.22: Age-group MAE comparison using EMF-BIF

Age-group	# imag	EMF-BIF [Proposed]	MF-BIF [292]	BIF [42]	EBIF [300]	RUN [54]
0-9	371	1.52	2.23	2.99	1.67	2.51
10-19	339	1.10	2.89	3.39	1.09	3.76
20-29	144	2.98	3.95	4.30	3.31	6.38
30-39	70	6.35	8.02	8.24	7.87	12.51
40-49	46	12.05	13.88	14.98	13.02	20.09
50-59	15	17.15	20.26	20.49	18.00	28.07
60-69	8	25.10	29.97	31.62	26.25	42.50
Total	**1002**	**3.05**	**4.02**	**4.77**	**3.17**	**5.78**

Generally, EMF-BIF achieved better MAE as compared to what was achieved by EBIF [300] and MF-BIF [292]. EMF-BIF achieves MAE of 1.52 years in age group 0-9 as compared to 1.67 years achieved by EBIF and 2.23 achieved by MF-BIF. There was also improvement in age-group 10-19 with EMF-BIF achieving MAE of 1.10 as

compared to 2.89 years achieved by MF-BIF. EMF-BIF also performed better in age-group 20-29 with MAE of 2.98 as compared to 3.95 achieved by MF-BIF and 3.31 achieved by EBIF. Performance of all the techniques deteriorates as one move up the age-groups because of drastic reduction in number of images per group in adult and old age age-groups in FG-NET dataset. EMF-BIF achieved a general MAE of 3.05 years as compared to 3.17 years achieved by EBIF and 4.02 years achieved by MF-BIF. This is an indication that fusing MF-BIF with shape, texture, appearance and wrinkles results into a richer age discriminative feature as compared to fusing BIF with only shape (ASM) features as done in [300]. This is because, in adult and old ages, facial shape tend to remain slightly unchanged and hence enhancing BIF features with appearance, wrinkle and texture features could be more discriminative for age estimation.

In younger age-groups, 0-9 and 10-19, there is significant change in facial shape while facial appearance (texture and wrinkle) remain relatively constant. Fusing MF-BIF with shape features makes it more age discriminative in these age-groups. As subjects grow to age-groups 20-29 and above, there is very negligible change in facial shape therefore shape features may not give MF-BIF significant age discriminative edge. In these age-groups, facial texture and tone start to change. Therefore, fusing MF-BIF with appearance and texture features give age discriminative edge in these age-groups. Adult and old age are more characterized by appearance of wrinkle lines in some regions of the face like forehead, eye-corners, mouth-corners and cheek-bone area. Texture and wrinkle features enhance discriminative power of MF-BIF in these age-groups.

4.11 Age estimation with gender discrimination

It has been found that male and females age differently. Therefore, age estimation accuracies could be improved by preceding it with gender estimation [287]. Effect of gender on age estimation is investigated. A gender estimation model based on fused EMF-BIF is build. After estimating ones gender, age-group classification is performed followed by exact age regression. An accuracy of 87.3% is achieved for gender estimation with accuracy of 86.21% accuracy for males and 87.3% for females. Table 4.23 shows cumulative scores of age-group estimation with gender discrimination.

TABLE 4.23: Cumulative scores for age-group estimation using EMF-BIF with gender discrimination

Age-group	No. of images		Cumulative Score (%)		Accuracy (%)
	Male	Female	Male	Female	
0-9	206	166	88.2	85.1	86.7
10-19	174	164	87.9	86.5	87.2
20-29	80	64	77.6	73.4	75.1
30-39	33	46	60.4	64.8	62.6
40-49	27	19	48.7	41.9	45.3
50-59	7	8	27.0	25.2	26.1
60-69	5	3	0.9	0.6	0.8
Total	**532**	**470**	**90.1**	**88.9**	**89.5**

Age-group estimation with gender discrimination improved estimation accuracies in some age-groups. Accuracy did not change in age-group 0-9 because facial appearance remain relatively constant in this age-group both in males and females. Change in facial shape in both sexes result into improved accuracies in age-groups

10-19 where an improvement from 85.2% to 87.2% is observed. Accuracy improved from 73.9% to 75.1% in age-group 20-29. This could be attributed to slight difference in facial appearance between males and females in this age-group. There was minimal improvements in accuracies in age-groups 30-39 through 60-69. This could be attributed to few images of males and females in these age groups.

Table 4.24 show MAE for age estimation using EMF-BIF with gender discrimination. EMF-BIF age estimation with gender estimation achieved MAE of 1.02 years in

TABLE 4.24: MAE for age estimation using EMF-BIF with age-group and gender discrimination

Age-group	**No. of images**		**MAE** (Years)	
	Male	**Female**	**Male**	**Female**
0-9	206	166	1.02	1.54
10-19	174	164	1.04	1.06
20-29	80	64	2.87	2.97
30-39	33	46	6.41	5.68
40-49	27	19	11.38	12.58
50-59	7	8	17.16	17.10
60-69	5	3	24.4	25.30
Total	**532**	**470**	**2.52**	**2.56**

age-group 0-9 for males and 1.54 years for females. These MAE are lower compared to 1.52 achieved in the same age-group in Table 4.22 without gender discrimination. With gender discrimination, EMF-BIF achieved MAE of 1.04 years for males and 1.06 years for females in age-group 10-19. Age-group 20-29 achieved MAE of 2.87 years for males and 2.97 for females. The performance of EMF-BIF with gender discrimination deteriorates with increase in age because of drastic reduction in number of images per gender per group as one move up the age-groups.

Table 4.25 show the comparison in performance of EMF-BIF with gender and age-group discrimination to performance of EMF-BIF with only group discrimination and other age-estimation techniques.

TABLE 4.25: Age-group MAE comparison using EMF-BIF with age-group gender discrimination

Age-group	**Images**	**EMF-BIF**[a] **[Proposed]**	**EMF-BIF**[b] **[Proposed]**	**MF-BIF**[b] **[Proposed]**	**BIF** [42]	**EBIF**[c] [300]	**RUN** [54]
0-9	371	1.28	1.52	2.23	2.99	1.67	2.51
10-19	339	1.05	1.10	2.89	3.39	1.09	3.76
20-29	144	2.92	2.98	3.95	4.30	3.31	6.38
30-39	70	6.04	6.35	8.02	8.24	7.87	12.51
40-49	46	11.98	12.05	13.88	14.98	13.02	20.09
50-59	15	17.13	17.15	20.26	20.49	18.00	28.07
60-69	8	24.85	25.10	29.97	31.62	26.25	42.50
Total	**1002**	**2.54**	**3.05**	**4.02**	**4.77**	**3.17**	**5.78**

[a]With gender and age group discrimination
[b]With age-group discrimination
[c]Used fused features(Fusion of ASM and BIF)

EMF-BIF with gender and age-group discrimination achieved MAE of 1.28 in age-group 0-9 as compared to 1.52 years without gender discrimination and 2.23 years using MF-BIF and 1.67 years using EBIF. In age-group 10-19, EMF-BIF with gender and age-group discrimination achieved MAE of 1.05 years which is better compared to 1.05 achieved by EMF-BIF without gender discrimination and 2.89 achieved by MF-BIF and 1.09 years achieved by EBIF. MAE of 2.92 years was achieved in age-group 20-29 using EMF-BIF with gender and age-group discrimination as compared to 2.98 without gender discrimination and 3.95 years with MF-BIF and 3.31 years with EBIF. In all the techniques, the performance deteriorates as one move up the age-groups because of reduction in number of images with increase in age. Generally, EMF-BIF with gender and age-group discrimination achieves MAE of 2.54 years which is lower compared to 3.05 achieved without gender discrimination and 3.17 years achieved by EBIF.

A comparison of the overall performance of EMF-BIF, MF-BIF, EBIF and BIF in age estimation on FG-NET is shown in Table 4.26.

TABLE 4.26: Comparison of age estimation performance on FG-NET

Method	MAE(years)	CS(%)
BIF Guo *et al.* [42]	4.77	≈69%
BIF Guo and Mu (MORPH) [52]	4.43	—
BIF Han *et al.* [255]	4.60	74.8%
BIF Spizhevoi & Bovyrin [254]	4.32	74.7%
LBP Ojala *et al.* [46]	5.98	70.4%
LTP Tan *et al.* [48]	5.74	73.4%
LDP Jabid *et al.* [47]	5.12	76.6%
Boosted BIF Han *et al.* [2]	3.80 [a]	78.0%
EBIF El Dib & El-Saban [300]	3.17[b]	—
Proposed Methods		
SOR-LDP[c]	**4.05**	**80.2%**
LTDP[c]	**4.35**	**78.3%**
MF-BIF[c]	**4.02**	**76.2%**
LS-BIF[c]	**3.84**	**77.8%**
EMF-BIF[c]	**3.05**	**85.8%**
EMF-BIF[d]	**2.54**	**87.3%**

[a] BIF features were selected using a boosting algorithm
[b] Used fused features(Fusion of ASM and BIF)
[c] With age-group discrimination
[d] With gender and age-group discrimination

As shown in Table 4.26, EMF-BIF with gender and age-group estimation outperforms BIF, EMF-BIF without gender discrimination, EBIF and MF-BIF based techniques. EMF-BIF with gender discrimination achieves MAE of 2.54 years compared to 4.77 years achieved by BIF, 3.80 years achieved by boosted BIF, 3.17 years achieved by BIF fused with ASM features. The approach achieved better results compared to 3.05 years achieved with same features but without gender discrimination and 4.02 years achieved by MF-BIF features.

Results show that proposed local feature descriptors outperform LBP, LTP and LDP local feature descriptors. SOR-LDP achieves MAE of 4.05 compared to 5.98 achieved by LBP, 5.74 years achieved by LTP and 5.12 achieved by LDP. Although LTDP

achieves better results than LBP, LTP and LDP, it is outperformed by SOR-LDP operator. Figure 4.5 shows comparison of performance between proposed feature extractors and state-of-the-art.

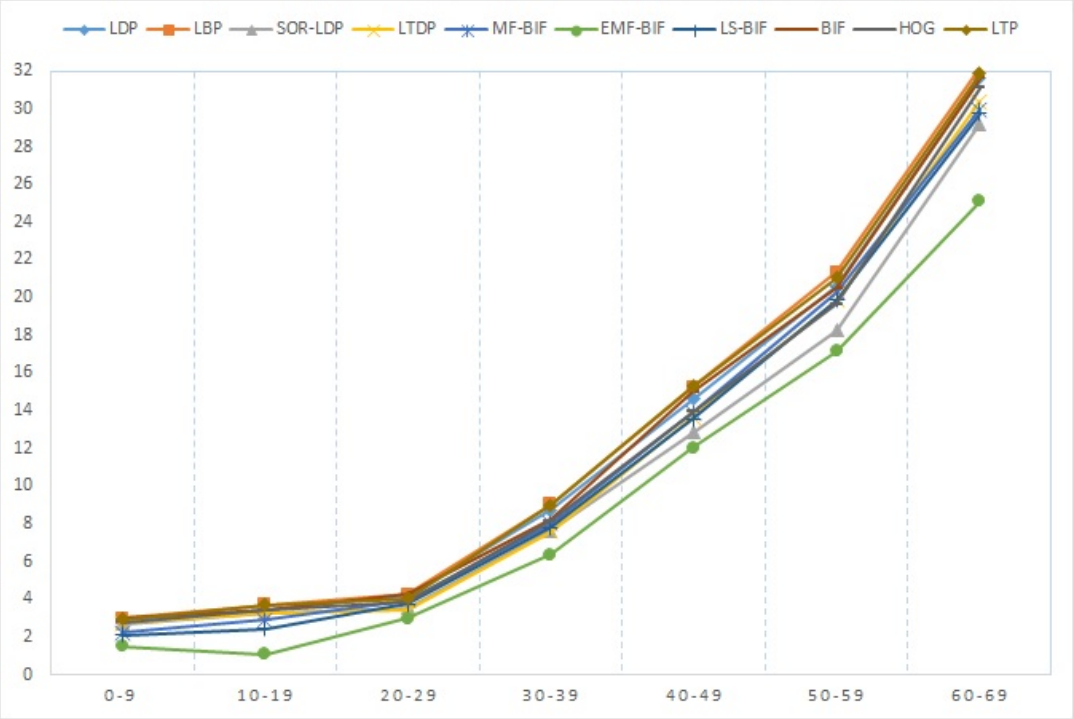


FIGURE 4.5: Comparison of age estimation performance between proposed techniques and state-of-the-art techniques

As shown in Figure 4.5, proposed feature extractors outperform state-of-the-art feature extraction techniques. It is also evident that enhancing ageing features by fusion improve age estimation accuracies across all age groups.

4.12 Summary

In this chapter, results of landmark localization, landmark displacement, age-group and age estimation are presented and discussed. The proposed landmark localization approach locates eyes, nose, cheeks, nose-bridge, mouth and forehead with highest accuracies of above 90%. These landmarks are located within 0.55 seconds in an image with a single face. Results of facial landmark displacement across age are also presented. It was found that as one age, eyes, nose and mouth drift away from each other. The mouth is more displaced from the eyes than is any other landmark displaced from the other. Displacement of eyes from each other was found to be lower compared to displacement of eyes from nose and mouth. This signify that there is more vertical facial development compared to horizontal development. It was also found that the nose about 45^o south of eyes at a distance of about half the distance between eyes. The mouth is found to be located at a distance almost equal to distance between eyes, about 63^o south of eyes.

Fusing appearance and shape features from whole face and wrinkle and texture features from facial components has significantly improved age-group and age estimation accuracies. Age-group estimation results are higher in 0-6 and 25-69 years age groups. This is attributed to more discriminative shape features in lower age-group and more discriminative texture, appearance and wrinkle features in adult and old age-group. Similarity of faces in age-group 7-14 years to those in 0-6 and

15-24 years resultant into lower age-group estimation accuracies in age-group 7-14 years. The same effect was observed in age-group 15-24 years.

Component-based age-group estimation results show that holistic shape and appearance features perform better compared to texture and wrinkle features extracted from facial components. Among components, nose features posted better results compared to other components. Results show that ANN performs better in age estimation compared to SVM.

SOR-LDP, LS-BIF, MF-BIF and LTDP features posted better results in age estimation compared to LBP, LTP, HOG and LBP operators. This shows that considering all directional responses in image gradient computation improves discriminative power of the extracted features. MF-BIF achieved lower MAE compared to BIF. Consequently, incorporating frequency selectivity if BIF feature extraction results into discriminative features compared to use of orientation and scale selectivity only. With LS-BIF achieving lower MAE than MF-BIF and BIF, it can be concluded that pooling within feature selectors achieves better age discriminative features. Results also show that fusing several features results into a discriminative feature vector that further improves age estimation accuracies.

Hierarchical approach to age estimation also improved age estimation results. Performing gender classification, within-gender age-group classification and within-group age regression saw improvements in age estimation performance.

Chapter 5

Conclusion and Future Work

This study presents a hierarchical age estimation model that improves age estimation accuracies using age-group and gender discrimination. The model was build using a rich ageing feature vector constituted of shape, wrinkle, texture, appearance and Biologically Inspired Features (BIF). Texture features were extracted using spatial LBP, LDP, SOR-LDP, and LTDP operators. Ratios of distances between fiducial facial landmarks and 2D landmark points constituted shape feature vector. Gabor filters were used for wrinkle feature extraction where mean and standard deviation of Gabor responses were concatenated to form wrinkle feature. Facial appearance features were extracted using LDA. Using Gabor filters at multiple scales, frequencies and orientations, two BIF feature extraction approaches are proposed. Multi-Frequency BIF (MF-BIF), which has better performance compared to BIF features extracted with scale and orientation selectivity only and Local Statical BIF (LS-BIF) which outperforms both original BIF and MF-BIF. Various features were fused and used for hierarchical age estimation that consist of gender and age-group classification using Back-propagation Multilayer Perceptron (MLP) Artificial Neural Network (ANN) followed by with-in gender and within-group age regression with Support Vector Regressor (SVR).

Age-group estimation using holistic and landmark features is also investigated. Holistic appearance and shape features were used alongside landmark texture and wrinkle features for age estimation using ensemble of classifiers that consisted MLP ANN and Support Vector Machines. Evaluating this approach to age-group estimation on FG-NET ageing dataset, it was found that fusion of both holistic and landmark features together with decision fusion improve age-group estimation accuracies.

This study also proposes a fast landmark localization approach for subsequent facial computing. Using Haar-cascade classifiers, geometric and spatial information of fiducial landmarks like eyes, landmark localization approach that detect secondary facial landmarks like cheeks, nose-bridge and forehead with accuracy of above 90% at a speed of about 0.5 of second is proposed. This technique does not use any machine learning tool hence making it more applicable in plug-and-play scenarios with its speed and accuracy making it ideal for real-time landmark-based facial processing systems.

Two different variants of LDP are proposed and their performance in age estimation investigated. The first variant of LDP considers all directional responses by grouping them into four orientations with respect to central reference pixel. The second variant extends LDP to a ternary pattern where all directional responses are also considered in encoding the image gradient. New algorithms for extraction of biologically inspired features with frequency selectivity and local pooling are also proposed.

This study also sought to find out techniques, algorithms and approaches commonly used for age estimation by reviewing recent studies. Multi-disciplinary literature was analysed in order to find out factors affecting ageing and how this information could be used to improve age estimation accuracies.

In subsequent subsections, summary and conclusion of each of the main contributions of this study is presented.

5.1 State-of-the-art techniques and facial ageing

A comprehensive survey of various techniques and approaches used for age estimation have been presented in this thesis. There are a number of face image representation techniques used in age estimation. Active Appearance Model presents a parametric modelling of facial shape and texture which makes it more appropriate for both young and old ages. Anthropometric representation only capture changes in facial shape and can only be reliable for young faces that are still undergoing facial development. When a dataset has images at almost every age, AGES can model the ageing pattern more accurately compared to other representations. However, datasets hardly have sequential images of subjects at all ages. In this scenario, age manifold is convenient for modelling ageing patterns. Appearance based models extract facial features like texture and wrinkle and use them to model ageing patterns. Commonly used feature extraction techniques are LBP, Gabor filters, BIF, LDA and PCA. To the best of our knowledge, no study has used LDP for extraction of texture features for learning ageing patterns.

Human face undergo both shape and appearance changes as one grows. During early years, the more dominant changes are facial shape due to craniofacial development. Facial landmarks like eyes, nose and mouth drift to occupy space created on the cranium. This changes the distances between these facial landmarks. In adulthood and old age, there is minimum changes in shape and more significant changes in facial appearance. Loss of collagen due to exposure to UV rays lead to reduction in production of melanin which cause uneven skin tone. Restructuring of facial bony structure among other intrinsic and extrinsic factors cause appearance of wrinkle lines in old age. Ageing is affected by intrinsic factors like genetic makeup as well as extrinsic factors like weather, use of cosmetics, pollution, lifestyle (diet and cigarette smoking).

Age estimation can be approached as a classification or regression problem or a hybrid of classification and regression. Most of the studies in the literature have approached age estimation as either a classification or regression problem with few studies embracing the hybrid approach. Of the classification approaches, majority of them were estimating age-group rather than exact age. Most of the studies that used regression approach were learning ageing patterns with intention of estimating exact age of a subject. Hybrid approach has seen improvements in the accuracies of age estimation systems.

Age estimation can be approached as either age-group estimation or exact age estimation. Age-group estimation approximates range of ages within which a subjects age belong. Exact age estimation approximates single value that correspond to a subject age. Both age and age-group estimation can be classification-based, regression-based or hybrid of both classification and regression. Choice between regression and classification may be guided by face image representation and size and age distribution of the dataset. For big datasets with sequential age-labels, both classification and regression can be used while for datasets with only age-group labels or significantly missing images at some ages, classification-based approach may be more appropriate. Both classification and regression can be combined in a hierarchical manner. In this hybrid approach, often classification is used for age-group

estimation followed by exact age estimation within the age-group using regression techniques.

There are a number of evaluation protocols used to evaluate age estimation techniques. The most commonly used ones are cross-validation and LOPO. The choice between cross-validation and LOPO could be guided by size of the dataset and number of subjects in the dataset. LOPO could be more appropriate for a large dataset of few subjects with many images per subject. Cross-validation would be appropriate for a dataset with many individuals with few images per subject. Performance of age estimation technique is measure in Mean Absolute Error (MAE) and Cumulative Score (CS). MAE is appropriate when training set has a lot of missing ages while CS is used when training dataset has samples at almost every age. Overall performance of the system is represented by CS. In practice, both MAE and CS are used because different techniques, datasets may be biased for evaluation.

5.2 Feature extraction techniques

This study proposes two improvements of LDP operator and introduction of frequency selectivity and local statics pooling in BIF extraction. Feature fusion strategy to improve discriminative power of extracted features is also proposed. A landmark localization technique and investigation of landmark displacement across age are also presented in this study.

5.2.1 Significant orientation response LDP

Significant orientation response LDP (SOR-LDP) convolves a 3×3 image region with Kirsch masks to obtain eight directional responses. These directional responses are grouped into four pairs with each pair consisting of two responses at a particular orientation with respect to central reference pixel. The values in each pair are compared and the bit corresponding to the maximum value in the pair is set to 1 and the one corresponding to the minimum value is set to 0. The resultant binary string is converted to decimal and assigned to central pixel as its code. Experimental results show that SOR-LDP outperforms LBP, LDP and HOG in age estimation.

5.2.2 Multi-frequency biologically inspired features

Multi-Frequency BIF (MF-BIF) is extracted by first convolving an image with a bank of Gabor filters are different scales, orientations and frequencies. Mean, maximum and standard deviation pooling operators are used to extract MF-BIF. Experimental results show that MF-BIF achieves better accuracies in age estimation compared to BIF extracted with scale and orientation selectivity only.

5.2.3 Local statistical biologically inspired features

Gabor filters at multiple spatial frequencies, orientation and scales are used to model S1 features. A global maximum pooling operator is applied on consecutive frequency features to obtain frequency-based responses. Local mean and standard deviation pooling operators are used in local neighbourhood to extract LS-BIF features. Experimental results show that LS-BIF achieves lower MAE in age estimation compared to MF-BIF and BIF.

5.2.4 Local ternary directional patterns

Local ternary directional patterns (LTDP) calculates the difference in neighbouring pixel values in a 3×3 image region with the central pixel. These difference in pixel values are convolved with Kirsch masks to obtain eight directional responses. The

normalized responses are used as probability of an edge occurring in a particular direction. The probability space is split into three subspaces and used to generate LTDP codes. Experimental results on age estimation show that LTDP performs better compared to LDP, LTP and LBP operators.

5.2.5 Enhanced MF-BIF

Multi frequency BIF features were enhanced by fusing them with shape, appearance, wrinkle and texture features. Wrinkle features are extracted using a bank of Gabor filters at different orientations and scales. Means and standard deviations of each Gabor filter responses are concatenated and used as wrinkle features. Appearance features are extracted using LDA on data projected in a PCA subspace. PCA eigenvectors are combined with LDA eigenvectors to constitute appearance feature vector. Texture features are extracted using spatial LBP. 2D landmark points and ratios of distances between fiducial landmarks constituted the shape feature vector. Experimental results show that fusing MF-BIF with other features improves its age discriminative power.

5.3 Landmark localization

A two stage fast facial landmark localization approach was developed. In the first step, fiducial landmarks like eyes, nose and mouth are detected using different Haar-cascade classifier. In the second phase, secondary landmarks like cheek-bone area, forehead and nose-bridge are located using geometric and spatial information of fiducial landmarks. Not incorporating machine learning makes the proposed technique fast without compromising accuracy. The proposed technique is appropriate for real time landmark localization for subsequent image processing tasks.

5.4 Landmark displacement across age

The results of analysis show that as one grow, facial landmarks get displaced from each other. The mouth is more displaced from the eyes than is any other landmark displaced from the other. It was also found that ageing does not contribute much displacement between mouth and nose and between mouth and forehead. Distance between facial landmarks can be used for age and age group estimation purposes as it is evident these distances vary across age. It is evident that there are other factors that contribute to displacement of landmarks as one grow. This could be extrinsic factors like diet or intrinsic factors like genes. This study recommends an extensive analysis of factors affecting displacement of facial landmarks in humans.

5.5 Hierarchical age estimation

This section present conclusions drawn from age group and age estimation results.

5.5.1 Age group estimation

5.5.1.1 Holistic age-group estimation

Holistic shape and appearance features outperforms texture and wrinkle features from facial components in age-group estimation. Age-group estimation accuracies deteriorates as one moves up the age groups due to minimal changes in facial shape in adult and old age. However, variation in texture and wrinkle provides a discriminative edge in these ages. Combining shape, appearance, texture and wrinkle features could help improve age-group estimation accuracies.

5.5.1.2 Component-based age-group estimation

Texture and wrinkle features extracted from facial components were fused and used for age-group estimation. Features from the nose posted better accuracies compared to the rest of components. Features from the eyes and cheeks were discriminative compared to features from the nose-bridge, mouth and forehead. This means that nose poses age discriminative features than other facial components. This could be as result of the nose being less affected by facial expressions as compared to eyes, mouth and forehead. Findings show that facial ageing does not uniformly affect all components.

5.5.1.3 Hybrid age-group estimation

Hybrid age-group estimation was done by use of both appearance and shape features extracted from the whole face and wrinkle and texture features extracted from facial components. Fusing shape, appearance, texture and wrinkle features improve age estimation accuracies.

5.5.2 Age estimation with age-group discrimination

5.5.2.1 Holistic age estimation

Findings show that holistic feature based age estimation approach achieves better results in childhood age-groups as compared to adult and old age age-groups. This is attributed to discriminative shape features during craniofacial development. In old age there is significant change in facial appearance with negligible change in facial shape. Holistic feature extraction techniques are not able to capture subtle age discriminative features in adulthood and old age.

5.5.2.2 Component based age estimation

Texture and wrinkle features from eyes posted superior results compared to other components, followed by wrinkle and texture features from cheek region. Texture and wrinkle features from the nose region achieved higher accuracies compared to same features from nose-bridge and mouth regions. Mouth features could have been affected by facial expressions like smiles while nose features could be affected by self occlusion and shadows. Forehead features posted highest MAE compared to other components. Forehead performance could have been affected by hairline and differences image profile. Texture and wrinkle features from eye and cheek regions are more discriminative for age estimation compared to other components. These results show that ageing does not uniformly affect all components of the face. Some components may exhibit age introduced variations while other components may remain relatively age invariant. Findings show that fusion of texture and wrinkle features from different facial components and use of ensemble of classifiers improves age estimation accuracies compared to holistic features extracted from the whole face.

5.5.2.3 Age estimation with SOR-LDP

SOR-LDP features were extracted from the whole face and used for age estimation. Experimental results show that SOR-LDP achieves lower MAE compared to LDP and LBP operators. This signify that, considering all directional responses in encoding image gradient result into discriminative features compared to considering only a subset of the directional responses. Furthermore, age estimation using image gradient features is more accurate compared to LBP.

5.5.2.4 Age estimation with LTDP

Local Ternary Directional Patterns achieves lower MAE compared to LDP and LTP operators. However, SOR-LDP achieved better results compared to LTDP. Consequently, this emphasizes that considering all directional responses in image gradient calculations encodes more discriminative features compared to considering a subset of directional responses.

5.5.2.5 Age estimation with DHOG

It is concluded that DHOG extracts more discriminative image gradient features compared to HOG. This study recommends investigation of DHOG performance in pedestrian or object detection. A further investigation to determine the best was to combine the two histograms of DHOG is also recommended.

5.5.2.6 Age estimation with MF-BIF

Findings show that using BIF features with frequency, scale and orientation selectivity achieves better results compared to use of orientation and scale selectivity only. This signify that, incorporation of frequency selectivity in BIF extraction improves discriminative power of BIF features extracted.

5.5.2.7 Age estimation with LS-BIF

Local statistical BIF achieves better age estimation accuracies compared to MF-BIF and BIF. Therefore, LS-BIF extracts robust age discriminative features compared to BIF and MF-BIF. This signifies that feature integration by pooling in a local neighbourhood within same scale or orientation extracts age discriminative features compared to feature integration by global (holistic) pooling. Further, using local statistical features (local mean and standard deviation) increases robustness of extracted ageing features compared to use of holistic maximum values, means and standard deviations across scales, orientations and frequencies.

5.5.2.8 Age estimation with EMF-BIF

Age estimation was performed using enhanced MF-BIF. Experimental results shows that fusing MF-BIF with shape, texture, appearance and wrinkle features improves it's discriminative ability. This shows that, fusing MF-BIF with more than one feature improves its discriminative power.

5.5.3 Age estimation with age-group and gender discrimination

Enhanced MF-BIF with gender and age-group estimation outperforms BIF, EMF-BIF without gender discrimination, EBIF and MF-BIF based techniques. The approach achieved better results compared to results achieved without gender discrimination. This is an indication that performing gender profiling prior to age estimation improves accuracies.

5.6 Recommendations for further research

There are a number of promising future directions for age estimation. Following are some of future research directions that may see improvement in age estimation performance

- **Fusion**-Feature and decision fusion for age estimation has not been extensively investigated. Fusing shape, wrinkle and texture features may result into rich feature set that can distinguish faces in different ages or age-groups.

Decisions from multiple classifiers or regressor could also be fused to see how they impact age estimation performance.

- **Multi-instance-Component based**–Facial landmarks can be extracted and considered as an instance for age estimation. Which parts of the face age faster and how? A face can be broken down into its components (eyes, forehead, nose, nose-bridge, mouth and cheeks) and ageing investigation done on each component. Both geometric, anthropometric and appearance face modeling can be used on each component.

- **Ethnic based**-Faces of subjects from different ethnic groups age differently. Incorporating ethnic parameters as in [166] improves age estimation performance. This approach has not been fully investigated due to lack of large datasets with images from different ethnic groups like African, Asian, and Caucasian.

- **Life style**-one's lifestyle affects how the face age. Faces of individuals of the same age but with different life style will appear different. Research has shown that smoking has an influence in facial ageing [83] [84] [85] [86] [75]. It may be interesting to investigate ageing and age estimation among a smoking population and how it compares to non-smoking population. Taister et al [83] asserts that exposure to drug and psychological stress affects skin texture and color making skin complexion spotted and blemished. Drug use and stress could also be investigated to determine their effect on age estimation.

- **Environment**-Taister *et al.* [83] found that general exposure to wind and arid air influence facial ageing. Arid environment and wind dehydrates the skin leading to wrinkle formation. An investigation of age estimation in populations in different environments is an interesting direction for further research.

- **Databases**-A large multi-racial database is needed for effective investigation of ageing in different ethnic groups and gender. Collecting a large database with well distributed age labels is essential. Web image collection is efficient way of achieving this [270] [177].

- **Profile face ageing**-how does non-frontal parts of the face age? How to estimate age from non-frontal face images? Investigations to answer these two questions could be necessary; though is based on availability of such databases (non-frontal face images). 3D face modelling could be vital in investigating profile face ageing and age estimation.

- **Multi-sensor**-image collection from multiple imaging sensors could be appropriate for mitigating degrading factors from uncontrollable and personalized attributes. Fusion could be done on the image features for age estimation.

- **CNN**- Convolutional Neural Network (CNN) have posted better results compared to traditional approaches of pattern recognition. There is emerging interest in investigating performance of CNN in age estimation domain.

- **Feature enhancement**- This study recommends further investigation into enhancement using boosting or feature selection machine learning techniques. Machine learning boosting and feature selection could be applied to fused feature or to individual features before fusion.

Bibliography

[1] P. A. George and G. J. Hole, "Factors influencing the accuracy of age estimates of unfamiliar faces", *Perception*, vol. 24, no. 9, pp. 1059–1073, 1995.

[2] H Han, X. O. CLiu, and A. K Jain, "Demographic estimation from face images:human vs machine performance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, pp. 1148–1161, 6 2015.

[3] Y Fu, G Guo, and T Huang, "Age synthesis and estimation via faces: A survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1955–1976, 11 2010.

[4] J. Naval, M. Coma, R. Valls, J. M. Mas, A. Pujol, M. A. Herranz, and V. Alonso, "Methods for diagnosing perceived age on the basis of an ensemble of phenotypic features", *Clinical, Cosmetic and Investigational Dermatology*, p. 133, 2014. DOI: 10.2147/CCID.S52257.

[5] X Geng, Z Zhau, and K Smith-miles, "Automatic age estimation based on facial aging patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 2234–2240, 12 2007.

[6] J. Twigg and S. Majima, "Consumption and the constitution of age: Expenditure patterns on clothing, hair and cosmetics among postwar baby boomers", *Journal of Aging Studies*, vol. 30, no. 1, pp. 23–32, 2014.

[7] U Park, Y Tong, and A. K. Jain, "Age invariant face recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 947–954, 5 2010.

[8] S. E Choi, Y. J Lee, J. L. S, R. P. K, and J Kim, "Age estimation using hierarchical classfier based on global and local features", *Pattern Recognition*, vol. 44, pp. 1262–1281, 2011.

[9] A. M Alberta, K Ricanek, and E Pattersonb, "A review of the literature on the aging adult skull and face: Implications for forensic science research and applications", *Forensic Science International*, vol. 172, pp. 1–9, 1 2007.

[10] R. L. Bowen and C. S. Atwood, "Living and dying for sex", *Gerontology*, vol. 50, pp. 265–290, 5 2004.

[11] A. Dillin, D. E Gottschling, and T. Nystrom, "The good and the bad of being connected: The integrons of aging", *Current Opinion in Cell Biology*, vol. 26, pp. 107–112, 2014.

[12] A. R. Valiente, A. Trinidad, J. R. G. Berrocal, C. Górriz, and R. R. Camacho, "Extended high–frequency (9–20 khz) audiometry reference thresholds in 645 healthy subjects", *International Journal of Audiology*, vol. 53, pp. 531–545, 8 2014.

[13] S. A. Thurstan, N. K. Gibbs, A. K. Langton, C. E. M. Griffiths, R. E. B. Watson, and M. J. Sherratt, "Chemical consequences of cutaneous photoageing", *Chemistry Central Journal*, vol. 6, pp. 1–7, 34 2012.

[14] T Wu, P Turaga, and R Chellappa, "Age estimation and face verification across aging using landmarks", *IEEE Transactions on Information Forensics Security*, vol. 7, pp. 1780–1788, 6 2012.

[15] B. F. Klare, M. J. Burge, J. C. Klontz, R. W. V. Bruegge, and A. K. Jain, "Face recognition performance: Role of demographic information", *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1789–1801, Dec. 2012. DOI: 10.1109/TIFS.2012.2214212.

[16] R Alley, *Social and Applied Aspects of Perceiving Faces*. New Jersey: Lawrence Erlbaum Associates, Inc, 1998.

[17] A Gallagher and T Chen, "Estimating age, gender and identity using first name priors", in *Proceedings of IEEE Conference on Computer Vision and Pattern recognition*, 2008.

[18] ——, "Understanding images of groups of people", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[19] D. Berry and L. McArthur, "Perceiving character in caces: The impact of age-related craniofacial changes on social perception", *Psychological Bull*, vol. 100, no. 1, pp. 3–18, 1986.

[20] E Drakaki, C Dessinioti, and C. V. Antoniou, "Air pollution and the skin", *Environmental Science*, vol. 2, no. 11, pp. 1–6, 2014.

[21] D. Zoe and M. D Draelos, "Aging in a polluted world", *Journal of Cosmetic Dermatology*, vol. 13, p. 85, 2014.

[22] A Vierkotter, T Schikowski, U Ranft, D Sugiri, M Matsui, U Kramer, and J Krutmann, "Airborne particle exposure and extrinsic skin aging", *Journal of Investigative Dermatology*, vol. 130, no. 12, pp. 2719–2726, 2010.

[23] C. Guinot, D. J.-M. Malvy, and L. Ambroisine, "Relative contribution of intrinsic vs extrinsic factors to skin aging as determined by a validated skin age score", *Archives of Dermatology*, vol. 138, no. 11, pp. 1454 –1460, 2002.

[24] M. Yaar and B. A. Gilchrest, "Photoageing: Mechanism, prevention and therapy", *The British Journal of Dermatology*, vol. 157, no. 5, 874–887, 2007.

[25] G. Lemperle, R. E. Holmes, S. R. Cohen, S. M. Lemperle, and S. Fagien, "A classifcation of facial wrinkles", *Plastic and Reconstructive Surgery*, vol. 108, no. 6, 1735–1752, 2001.

[26] FG-NET, *Face and gesture recognition working group*, 2002. [Online]. Available: http://www-prima.inrialpes.fr/FGnet/.

[27] N. Ramanathan, R. Chellappa, and S. Biswas, "Computational methods for modeling facial aging: A survey", *Journal of Visual Languages and Computing*, vol. 20, pp. 131–144, 2009.

[28] M. J. Raval and P. Shankar, "Age invariant face recognition using artificial neural network", *International Journal of Advance Engineering and Research Development*, vol. 2, pp. 121–128, 6 2015.

[29] A. Sonu, K. Sushil, and K. Sanjay, "A novel idea for age invariant face recognition", *International Journal of Innovative Research in Science Engineering and Technology*, vol. 3, pp. 15 618–15 624, 8 2014.

[30] S. N. Jyothi and M. Indiramma, "Stable local feature based age invariant face recognition", *International Journal of Application or Innovation in Engineering & Management*, vol. 2, pp. 366–371, 12 2013.

[31] S. Jinli, C. Xilin, S. Shiguang, G. Wen, and D. Qionghai, "A concatenational graph evolution aging model", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 2083–2096, 11 2012.

[32] Z Song, B Ni, D Guo, T Sim, and S Yan, "Learning universal multi-view age estimator using video context", in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 241–248.

[33] A Lanitis, C Draganova, and C Christodoulou, "Comparing different classifiers for automatic age estimation", *IEEE Transactions on Man, Systems and Cybernatics*, vol. 34, pp. 621–628, 1 2004.

[34] E Patterson, A Sethuram, M Albert, K Ricanek, and M King, "Aspects of age variation in facial morphology affecting biometrics", in *Proceedings of 1ST IEEE Conference on Biometrics, Theory and Application Systems*, 2007, pp. 1–6.

[35] K Ricanek and E Boone, "The effect of normal adult aging on standard pca face recognition accuracy rates", in *Proceedings of IEEE International Joint Conference on Neural networks*, 2005, pp. 2018–2023.

[36] X Geng, Z.-H Zhou, Y Zhang, G Li, and H Dai, "Learning from facial aging patterns for automatic age estimation", in *Proceedings of ACM Conference on Multimedia*, 2006, pp. 307–316.

[37] F Gao and H Ai, "Face age classification on consumer images with gabor feature and fuzzy lda method: Lecture notes in computer science", in *Proceedings of 3rd International Conference on Advances in Biometrics*, 2009, pp. 132–141.

[38] T. F. Cootes, G. J. Edwards, and C. J Taylor, "Active appearance models", *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, pp. 681–685, 6 2001.

[39] Y Fu and H. T. S., "Human age estimation with regression on discriminative aging manifold", *IEEE Transaction on Multimedia*, vol. 10, pp. 578–584, 4 2008.

[40] Y Fu, Y Xu, and H. T. S, "Estimating human ages by manifold analysis of face pictures and regression on aging features", in *Proceedings of IEEE Conference Multimedia and Expo*, 2007, pp. 1383–1386.

[41] Y Kwon and N Lobo, "Age classification from facial images", *Computer Vision and Image Understanding*, vol. 74, pp. 1–21, 1 1999.

[42] G Guo, G Mu, Y Fu, and T. S Huang, "Human age estimation using bio inspired features", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 112–119.

[43] S Yan, M Liu, and T. S Huang, "Extracting age information from local spatially flexible patches", in *Proceedings of IEEE Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 737–740.

[44] T Ojala, M Pietikäinen, and T Mäenpää, "A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification", in *Proceedings of 2nd ICAPR*, 2001, pp. 397–406.

[45] T Ojala, M Pietikainen, and T Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, 7 2002.

[46] T Ojala, M Pietikainen, and D Harwood, "A comparative study of texture measures with classification based on featured distribution", *Pattern Recognition*, vol. 29, pp. 51–59, 1 1996.

[47] T. Jabid, M. Kabir, and O. Chae, "Local directional pattern (LDP) for face recognition", in *Proceedings of 2010 Digest of Technical Papers International Conference on Consumer Electronics (ICCE)*, IEEE, 2010. DOI: 10.1109/ICCE.2010.5418801.

[48] X Tan and B Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions", *IEEE Transactions on Image Processing*, vol. 19, pp. 1635–1650, 6 2010.

[49] G Guo, Y Fu, C Dyer, and T Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression", *IEEE Transactions on Image Processing*, vol. 17, pp. 1178–1188, 7 2008.

[50] K Ueki, T. Hayashida, and T Kobayashi, "Subspace-based age group classification using facial images under various lighting conditions", in *Proceedings of IEEE Conf. Automatic Face and Gesture Recognition*, 2006, pp. 43–48.

[51] S Yan, X Zhou, M Liu, M Hasegawa-Johnson, and T. S Huang, "Regression from patch-kernel", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[52] G Guo and G Mu, "Simulteneous dimensionality reduction and human age estimation via kernel partial least squares regression", in *Proceedings of IEEE Conference on Computer Vision and Pattern recognition*, 2011, pp. 657–664.

[53] S Yan, H Wang, T. S Huang, and X Tang, "Ranking with uncertain labels", in *Proceedings of IEEE Conference on Multimedia and Expo*, 2007, pp. 96–99.

[54] S Yan, H Wang, X Tang, and T. S Huang, "Learning auto-structured regressor from uncertain non-negative labels", in *Proceedings of IEEE Conference on Computer Vision*, 2007.

[55] Y Zhang and D.-Y Yeung, "Multi-task warped gaussian process for personalized age estimation", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 2622–2629.

[56] S. K Zhou, B Georgescu, X. S Zhou, and D Comaniciu, "Image based regression using boosting method", in *Proceedings of IEEE International Conference on Computer Vision*, 2005, pp. 541–548.

[57] G Guo, Y Fu, T. Huang, and C Dyer, "Locally adjusted robust regression for human age estimation", in *Proceedings of IEEE Workshop on Applications of Computer Vision*, 2008.

[58] G Guo, y Fu, T. S Huang, and C Dyer, "A probabilistic fusion approach to human age prediction", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition-Semantic Learning and Applications in Multimedia Workshop*, 2008, pp. 1–6.

[59] K.-Y Chang, C.-S Chen, and Y.-P Hung, "Ordinal hyperplane ranker with cost sensitivities for age estimation", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 585–592.

[60] Y Ma, T Xiong, Y Zou, and K Wang, "Person-specific age estimation under ranking framework", in *Proceedings of 1st ACM International Conference on Multimedia Retrieval*, 2011.

[61] P Yang, L Zhong, and D Metaxas, "Ranking model for facial age estimation", in *Proceedings of 20th International Conference on Pattern Recognition*, 2010, pp. 3404–3407.

[62] N Ramanathan and R Chellappa, "Face verification across age progression", *IEEE Transactions in Image Processing*, vol. 15, pp. 3349–3361, 11 2006.

[63] M. R Turner, "Texture discrimination by gabor functions", *Biological Cybernetics*, vol. 55, pp. 71–82, 1986.

[64] M Haghighat, S Zonouz, and M Abdel-Mottaleb, "Identification using encrypted biometrics", in *Proceedings of Computer Analysis of Images and Patterns*, 2013, pp. 440–448.

[65] J. G Daugman, "Spatial visual channels in the forier plane", *Vision Res*, vol. 24, pp. 891–910, 1984.

[66] R Shapley, "Spatial frequency analysis in the visual system", *Annual Reviews in Neuroscience*, vol. 8, pp. 547–583, 1985.

[67] G Panis, A Lanitis, N Tsapatsoulis, and T. F Cootes, "Overview ofreserach on facial ageing using the fg-net ageing database", *IET Biometrics*, vol. 5, pp. 37–46, 2 2016.

[68] T. F Cootes, C. J Taylor, D. H Cooper, and J Graham, "Active shape models-their training and application", *Computer Vision and Image Understanding*, vol. 61, pp. 38–59, 1 1995.

[69] D Gabor, "Theory of communication", *Journal of the Institute of Electrical Engineers*, vol. 93, pp. 429–457, 1946.

[70] M Riesenhuber and T Poggio, "Hierarchical models of object recogntion in cortex", *Nature Neuroscience*, vol. 2, pp. 1019–1025, 11 1999.

[71] T Serre, L Wolf, S Bilechi, M Riesenhuber, and T Poggio, "Robust object recognition with cortex-like mechanism", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 411–426, 3 2007.

[72] G. D. J, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters", *Journal of the Optical Society of Ameica*, vol. 2, pp. 1160–1169, 7 1985.

[73] B Cai, X Xu, and X Xing, "BIT: Bioologically inspired tracker", *IEEE Transactions on Image Processing*, vol. 25, pp. 1327–1339, 3 2016.

[74] M. D. M and A Bastanfard, "A new algorithm for age recognition from facial images", *Signal Processing*, vol. 90, pp. 2431–2444, 2010.

[75] P. O'Hare, A. B Fleischer, R. B D'Agostino, S. R Feldman, M. A Hinds, A. A Rassette, A Mcmichael, and P. M Williford, "Tobacco smoking contributes little to facial wrinkling", *Journal of the European Academy of Dermatology and Venereology*, vol. 12, pp. 133–139, 2 1999.

[76] A. J. O'Toole, T. Price, T. Vetter, J. C. Bartlett, and V. Blanz, "3d shape and 2d surface textures of human faces: The role of "averages" in attractiveness and age", *Image Vision Comput.*, vol. 18, pp. 9–19, 1999.

[77] P. Turaga, S. Biswas, and R. Chellappa, "The role of geometry in age estimation", in *Proceedings ofInternational Conference on Acoustics Speech and Signal Processing (ICASSP)*, IEEE, 2010. DOI: 10.1109/ICASSP.2010.5495292.

[78] L. A Zebrowitz, *Reading Faces: Window to the Soul*. Westview Press, 1997.

[79] L. S Mark, J. B Pittenger, H Hines, C Carello, R. E Shaw, and J. T Todd, "Wrinkling and head shape as coordinated sources for age-level information", *Perception and Psychophysics*, vol. 27, pp. 117–124, 2 1980.

[80] N Ramanathan and R Chellappa, "Face verification across age progresion", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2005, pp. 462–469.

[81] M. S Zimbler, M. S Kokosa, and J. R Thomas, "Anatomy and pathophysiology of facial aging", *Facial Plastic Surgery Clinics of North America*, vol. 9, pp. 179–187, 2 2001.

[82] I Macleod and B Hill, *Heads and Tales: Reconstructing Faces*. Edinburg: National Museums of Scotland, 2001.

[83] M. A Taister, S. D Holliday, and H. I. M Borman, "Comments in facial aging in law enforcement investigation", *Forensic Science Communication*, vol. 2, 2000.

[84] F. G. Fedok, "The aging face", *Fcaial Plastic Surgery*, vol. 12, pp. 107–115, 1996.

[85] I. L. W. C Harvey, "Is skin ageing in the elderly caused by sun exposure or smoking?", *The British Journal of Dermatology*, vol. 147, pp. 1187–1191, 6 2002.

[86] H. B. López, J. Tercedor, J. M. Ródenas, L. F.R. M. Simón, and O. S. Serrano, "Skin aging and smoking", *Revista Clinica Espanola*, pp. 147–149, 1995.

[87] R. B Shaw, E. B Katzel, P. F Koltz, D. M Kahn, J. A Girotto, and H. N Langstein, "Aging of the mandible and its aesthetic implications", *Plastic Reconstruction Surgery*, vol. 125, pp. 332–342, 1 2010.

[88] M Situm, M Buljan, V Cavka, V Bulat, I Krolo, and L. L. Mihić, "Skin changes in the elderly people–how strong is the influence of the uv radiation on skin aging?", *Collegium anthropologicum*, vol. 34, pp. 9–13, 2 2010.

[89] R Neave, *Age changes to the face in adulthood*. New York: Oxford Press, 1998, pp. 225–234.

[90] S Coleman and R Grover, "The anatomy of the aging face: Volume loss and changes in 3-dimensional topography", *Aesthetic surgery journal / the American Society for Aesthetic Plastic surgery*, vol. 26, pp. 4–9, 1S 2006.

[91] P. L Leong, "Aging changes in the male face", *Facial Plastic Surgery clinics of North America*, vol. 16, pp. 277–279, 3 2008.

[92] K Sveikata, I Balciuniene, and J Tutkuviene, "Factors influencing face aging. literature review", *Stomatologija, Baltic Dental and Maxillofacial Journal*, vol. 13, pp. 113–116, 4 2011.

[93] E. C Paes, H. J Teepen, W. A Koop, and M Kon, "Perioral wrinkles: Histologic differences between men and women", *Aesthetic Surgery Journal*, vol. 29, pp. 467–472, 6 2009.

[94] D. A Gunn, H Rexbye, C. E Griffiths, P. G Murray, A Fereday, S. D Catt, C. C Tomlin, B. H Strongitharm, D. I Perrett, M Catt, A. E Mayes, A. G Messenger, M. R Green, F van der Ouderaa, J. W Vaupel, and K. Christensen, "Why some women look young for their age", *PloS One*, vol. 4, 12 2009.

[95] C Chen, A Dantcheva, and A Ross, "Impact of facial cosmetics on automatic gender and age estimation algorithms", in *Proceedings of 9th International Conference on Computer Vision Theory and Applications*, Lisbon, Portugal, 2014, pp. 182–190.

[96] G Guo and X Wang, "A study on human age estimation under facial expression changes", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2547–2553.

[97] D. T Nguyen, S. R Cho, K. Y Shin, J. W Bang, and K. R Park, "Comparative study of human age estimation with or without preclassification of gender and facial expression", *The Scientific World Journal*, pp. 1–15, 2014.

[98] M Minear and D Park, "A lifespan database of adult facial stimuli", *Behaviour Research Methods, Instruments and Computers*, vol. 36, pp. 630–633, 4 2004.

[99] N Ebner, M Riediger, and U Lindenberger, "Faces-a database of facial expressiona in young, middle-aged, and older women and men: Development and validation", *Behaviour Research Methods*, vol. 42, pp. 351–362, 1 2010.

[100] M. C Voelkle and N. C Ebner, "Let me guess how old you are: Effects of age, gender, and facial expression on perceptions of age", *Psychology and Aging*, vol. 27, pp. 265–277, 2 2012.

[101] B Bruyer and J.-C Scailquin, "Person recognition and ageing: The cognitive status of addresses—an empirical question", *International Journal of Psychology*, vol. 29, pp. 351–366, 3 1994.

[102] A. K Jain, S. Dass, and K Nandakumar, "Soft biometrics traits for personal recognition systems", in *Proceedings of International Conference on Biometric Authentication*, 2004, pp. 731–738.

[103] L Farkas, *Anthropometry of the Head and Face*. Raven Press, 1994.

[104] K. Bush and O. Antonyshyn, "3-dimensional facial anthropometry using a laser-surface scanner–validation of the technique", *Plastic and reconstructive surgery*, vol. 98, no. 2, 226–235, Aug. 1996.

[105] J. Kolar and E. Salter., *Craniofacial Anthropometry: Practical Measurement of the Head and Face for Clinical, Surgical and Research Use*. Charles C. Thomas Publisher LTD, 1996.

[106] T. R Alley, *Social and Applied Aspects of Perceiving Faces*. Lawrence Erlbaum Assoc., 1988.

[107] T. Todd, S. Leonard, R. Shaw, and J. Pittenger, "The perception of human growth", *Scientific American*, vol. 242, no. 2, pp. 132–144, 1980.

[108] N Ramanathan and R Chellappa, "Modeling age progression in young faces", in *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*, 2006, pp. 384–394.

[109] M Kass, A Witkin, and D terzopoulos, "Snakes: Active contour models", in *Proceedings of 1st IEEE International Conference on Computer Vision*, 1987, pp. 259–268.

[110] N Duta, A. K Jain, and M. P Dubuisson-Jolly, "Automatic construction of 2d shape model", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 433–446, 5 2001.

[111] G Edwards, A Lanitis, C Taylor, and T Cootes, "Statistical models of face images-improving specificity", *Image and Vision Computing*, vol. 16, pp. 203–211, 1998.

[112] A Lanitis, J Taylor, and T. F Cootes, "Toward automatic simulation of aging effects on face images", *IEEE Transactions on pattern Analysis and Machine Intelligence*, vol. 24, pp. 442–455, 4 2002.

[113] K Scherbaum, M Sunkel, H.-P. Seidel, and V Blanz, "Prediction of individual non-linear aging trajectories of faces", in *Proceedings of Annual Conference of European Association on Computer Graphics*, vol. 26, 2007, pp. 285–294.

[114] L Huang, J Lu, and Y.-P. Tan, "Multi-manifold metric learning for face recognition based on image sets", *Journal of Visual Communication and Image Representation*, vol. 25, no. 7, pp. 1774–1783, 2014.

[115] D Beymer and T Poggio, "Image representations for visual learning", *Science*, vol. 272, pp. 1905–1909, 1996.

[116] J Hayashi, M Yasumoto, H Ito, and H Koshimizu, "A method for estimating and modeling age and gender using facial image processing", in *Proceedings of the 7th International Conference on Virtual Systems and Multimedia*, 2001.

[117] J Hayashi, M Yasumoto, H Ito, Y Niwa, and H Koshimizu, "Age and gender estimation from facial image processing", in *Proceedings of SICE Annual Conference*, 2002, pp. 13–18.

[118] J. P. P, H Moon, S. A Rizvi, and P. J Rauss, "The feret evaluation methodology for face recognition algorithms", *IEEE Transactions on pattern Analysis and Machine Intelligence*, vol. 22, pp. 1090–1104, 10 2000.

[119] Z Yang and H Ai, "Demographic classification with local binary patternsns", in *Proceedings of International Conference on Biometrics*, 2007, pp. 464–473.

[120] F Gao and H Ai, "Face age classification on consumer images with gabor feature and fuzzy lda method", in *Proceedings of International Conference on Advances in Biometrics*, 2009, pp. 132–141.

[121] B. S Manjunathi and W. Y Ma, "Texture features for browsing, retrieval of image data", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 837–842, 8 1996.

[122] 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, pp. 183–192, 3 2001.

[123] H Takimoto, Y Mitsukura, M Fukumi, and N Akamatsu, "Robust gender and age estimation under varying facial poses", *Electronics and Communications in Japan*, vol. 91, pp. 32–40, 7 2008.

[124] H. G Jung and J Kim, "Constructing a pedestrian recognition system with a public open database, without the necessity of re-training: An experimental study", *Pattern Analysis and Applications*, vol. 13, pp. 223–233, 2 2010.

[125] R. A Fisher, "The statistical utilization of multiple measurements", *Annals of Eugenics*, vol. 8, pp. 376–386, 1938.

[126] K Fukunaga, *Introduction to Statistical Pattern recognition*, 2nd. Academic Press, 1990.

[127] P. N. Belhumeour, J. P Hespanda, and D. J Kriegman, "Eigenfaces vs fisherfaces:recognition using class specific linear projection", *IEEE Transactions on pattern Analysis and Machine Intelligence*, vol. 19, pp. 711–720, 7 1997.

[128] D. L Swets and J. J Weng, "Using discriminant eigenfeatures for image retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 71–86, 8 1996.

[129] A. M Martinez and A. C Kak, "Pca versus lda", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 228–233, 2 2001.

[130] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", in *in Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010.

[131] A. Satpathy, X. Jiang, and H.-L. Eng, "Extended histogram of gradients feature for human detection", in *in Proceedings of 2010 17th IEEE Conference on Image Processing (ICIP)*, 2010. DOI: 10.1109/ICIP.2010.5650070.

[132] R. A. Kirsch, "Computer determination of the constituent structure of biological images", *Computers and Biomedical Research*, vol. 4, no. 3, pp. 315–328, 1971. DOI: 10.1016/0010-4809(71)90034-6.

[133] T Maenpaa and M Pietikainen, *Texture analysis with local binary patterns: Handbook of Pattern Recognition and Computer Vision*. London: World Scientific, 2005.

[134] T Ahonen, A Hadid, and M Pietikainen, "Face recognition with local binary patterns", in *Proceedings of European Conference on Compututer Vision*, 2004, 469–481.

[135] J. M. S. Prewitt, "Object enhancement and extraction", in *Picture Processing and Psychopictorics*, Academic Press, 1970.

[136] I Sobel and F. G, "A 3 x 3 isotropic gradient operator for image processing", in *Presented at the Stanford Artificial Intelligence Project (SAIL)*, 1968.

[137] W. K. Pratt, *Digital Image Processing*. New York: Wiley, 1978.

[138] S.-W. Lee, "Off-line recognition of totally unconstrained handwritten numerals using multilayer cluster neural network", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, pp. 648–652, 1996. DOI: 10.1109/34.506416.

[139] T. Jabid, M. H. Kabir, and O. Chae, "Gender classification using local directional pattern (LDP)", in *Proceedings of 2010 20th International Conference on Pattern Recognition*, IEEE, 2010. DOI: 10.1109/ICPR.2010.373.

[140] R. C Gonzalez and R. E Woods, *Digital Image Processing*, 3rd. Pearson, 2008.

[141] M. H. R, K Shanmugam, and I Dinstein, "Texture features for image classification", *IEEE Transactions on system, Man and Cybernetics*, vol. 3, pp. 610–621, 6 1973.

[142] M Unser, "Sum and difference histograms for texture classification", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 118–125, 1 1986.

[143] N Zulpe and V Pawar, "Glcm texture features for brain tumor classification", *International Journal of Computer Science Issues*, vol. 3, no. 3, 2012.

[144] F. R. Siquira, W. R. Schwatz, and H Pedrini, "Multi-scale gray level co-occurrence matrices for texture description", *Neurocomputing*, vol. 120, pp. 336–345, 2013.

[145] L. K Soh and C Tsatsoulis, "Texture analysis of sar see ice imagery using grey level co-occrence matrices", *IEEE Transactiona on Geoscience and Remote Sensing*, vol. 37, pp. 780–795, 2 1999.

[146] A Edelman, T. A Arias, and S. T Smith, "The geometry of algorithms with orthogonality constrains", *SIAM Journal on Matrix Analysis and Applications*, vol. 20, pp. 303–353, 2 1999.

[147] H Karcher, "Riemannian center of mass and mollifier smoothing", *Communications on Pure and Applied Mathematics*, vol. 30, pp. 509–541, 5 1977.

[148] T Serre and M Riesenhuber, "Realistic modeling of simple and complex cell tuning in the hmax model and implications for invariant object recognition in cortex", Massachusetts Institute of Tech Cambridge Computer Science Artificial Intelligence Lab DTIC Washington DC USA Tech. Rep. AI-MEMO-2004-017, Tech. Rep., 2004.

[149] B Ma, Y Su, and F Jurie, "Covariance descriptor based on bio-inspired features for person re-identification and face verification", *Image and Vision Computing*, vol. 32, pp. 379–390, 6-7 2014.

[150] Y Huang, K Huang, D Tao, T Tan, and X Li, "Enhanced biologically inspired model for object recognition", *IEEE Transactions in Systems, man and Cybernatics*, vol. 41, pp. 1668–1680, 6 2011.

[151] D Song and D Tao, "Biologically inspired feature manifold for scene classification", *IEEE Transactions on Image Processing*, vol. 19, pp. 174–184, 1 2010.

[152] S Marcelja, "Mathematical description of the responses of simple cortical cells", *Journal of Optical Society of America*, vol. 70, pp. 1297–1300, 11 1980.

[153] G Guo and G Mu, "Joint estimation of age, gender and ethnicity:cca vs pls", in *Proceedings of IEEE Conference on Face and Gesture Recognition*, 2013, pp. 1–6.

[154] L Shen and L Bai, "A review on gabor wavelets for face recognition", *Pattern Analysis and Application*, vol. 9, pp. 273–292, 2006.

[155] I Lampl, D Ferster, T Poggio, and M Riesenhuber, "Intracellular measurements of spatial integration and the max operation in complex cells of the cat primary visual cortex", *journal of Neurophysiology*, vol. 92, pp. 2704–2713, 5 2004.

[156] T. Serre, L. Wolf, and T. Poggio, Eds., *Object recognition with features inspired by visual cortex*, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[157] J. Mutch and D. Lowe, "Object class recognition and localization using sparse features with limited receptive fields", in *Prooceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 11–18.

[158] V. Vapnik, *Statistical Learning Theory*. New York: John Wiley, 1998.

[159] I Huerta, C Fernandez, C Segura, J Hernando, and A Prati, "A deep analysis on age estimation", *Pattern recognition Letters*, vol. 68, pp. 239–249, 2015.

[160] H Bay, T Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features", *Computer Vision-ECCV*, vol. 3951, pp. 404–417, 2006.

[161] B Triggs and N Dalal, "Histograms of oriented gradients for human detection", in *Proceedings of IEEE on Compter Vision and Pattern Recognition*, 2005, pp. 886–893.

[162] Z Hu, Y Wen, J Wang, M Wang, R Hong, and S Yan, "Facial age estimation with age difference", *IEEE Transactions on Image Processing*, pp. 1–13, 99 2016.

[163] D. T Nguyen, S. R Cho, and K. R Park, "Age estimation-based soft biometrics considering optical blurring based on symmetrical sub-blocks for mlbp", *Symmetry*, pp. 1882–1913, 2015.

[164] J Lu and Y Tan, "Ordinary preserving manifold analysis for human age and head pose estimation", *IEEE Transactions on Human-Machine System*, vol. 43, pp. 249–258, 2 2013.

[165] O. F. W Onifade and D. J Akinyemi, "A groupwise age ranking framework for human age estimation", *International Journal of Image, Graphics and Signal Processing*, vol. 5, pp. 1–12, 2015.

[166] J. D Akinyemi and O. F. W Onifade, "An ethnic-specific age group ranking approach to facial age estimation using raw pixel features", in *Proceedings of IEEE Symposium on Technologies for Homeland Security*, 2016, pp. 1–6.

[167] A Gunay and V. V Nabiyev, "Facial age estimation based on decision level fusion of amm, lbp and gabor features", *International Journal of Adanced Computer Science and Applications*, vol. 6, pp. 19–26, 8 2015.

[168] K Ricanek and T Tesafaye, "Morph: A longitudinal image database of normal adult age-progression", in *Proceedings of IEEE 7th International Conference on Automatic Face and Gesture Recognition*, Southampton, 2006, pp. 341–345.

[169] P. Viola and M. Jones, "Robust real-time object detection", *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[170] K Ueki, T Hayashinda, and T Koboyashi, "Subspace-based age group classification using facial images under various lighting conditions", in *Proceedings of IEEE Conference on Automatic Face and Gesture Recognition*, 2006, pp. 43–48.

[171] Y Fu and N Zheng, "M-face: An appearance-based photorealistic model for multiple facial attributes rendering", *IEEE Transactions Circuits and Systems for Video Technology*, vol. 16, pp. 830–842, 7 2006.

[172] D. M Burt and D. I Perrett, "Perrception of age in adult caucasian male faces: Computer graphic manipulation of shape and colour information", in *Proceedings of Royal Society of London Series B Biological Sciences*, vol. 259, 1995, pp. 137–143.

[173] J Suo, T Wu, S Zhu, S Shan, X Chen, and W Gao, "Design sparse features for age estimation using hierarchical face model", in *Proceedings of IEEE Conference in Automatic Face and Gesture Recognition*, 2008.

[174] A Bastanfard, M. A Nik, and M. M Dehshibi, "Iranian face database with age, pose and expression", in *Proceedings of International Conference on Machine Vision*, 2007, pp. 50–55.

[175] A Gallagher and T Chen, "Understanding images of groups of people", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[176] B Ni, Z Song, and S Yan, "Web image mining towards universal age estimator", in *Proceedings of ACM International Conference on Multimedia*, Beijing, 2009.

[177] ——, "Web image and video mining towards universal and robust age estimator", *IEEE Transactions on Multimedia*, vol. 13, pp. 1217–1229, 6 2011.

[178] S. P Kyaw, J.-G. Wang, and E. K Teoh, "Web image mining for facial age estimation", in *Proceedings of IEEE International Conference on Information, Communication and Signal Processing*, 2013.

[179] V Blanz and T Vetter, "A morphable model for the synthesis of 3d faces", in *Proceedings of ACM Conference SIGGRAPH*, 1999, pp. 187–194.

[180] A Ross and A Jain, "Information fusion in biometrics", *Pattern Recognition Letters*, vol. 24, pp. 2115–2125, 2003.

[181] G. M. U, S Samanta, S Das, and P. R Chowdhury, "A survey of decision fusion and feature fusion strategies for pattern classification", *IETE Technical Review*, vol. 27, pp. 293–307, 4 2010.

[182] l Hong and A. K. Jain, "Integrating faces and fingerprints for personal identification", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 2, pp. 1295–1307, 1998.

[183] M Budka and B Gabrys, "Density-preserving sampling: Robust and efficient alternative to cross-validation for error estimation", *IEEE Transactions on Neural Network and Learning Systems*, vol. 24, pp. 22–34, 1 2013.

[184] S Weiss and C Kulikowski, *Computer Systems that Learn*. San Fransisco: Morgan Kaulmann, 1991.

[185] Y Bengio and Y Grandvalet, "No unbiased estimator of the variance of k-fold cross validation", *Journal of Machine Learning Research*, vol. 5, pp. 1089–1105, 2004.

[186] A Jain, R Duin, and J Mao, "Statistical pattern recognition: A review", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 4–37, 1 2000.

[187] R. R Picard and R. D Cook, "Cross validation of regression models", *Journal of American Statistical Association*, vol. 79, pp. 575–583, 1984.

[188] R Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection", in *Proceedings of International Joint Conference on Artificial Intelligence*, vol. 14, 1995, pp. 1137–1143.

[189] O Kiline and I Uysal, "Source-aware partitioning for robust cross-validation", in *Proceedings of 14th International Conference on Machine Learning and Applications*, 2015.

[190] R Duda, P Hart, and D Stork, *Pattern Classification*, 2nd. New York: Wiley, 2001.

[191] M Stone, "Cross-validatory choice and assessment of statistical predictions", *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 36, pp. 111–147, 2 1974.

[192] T Hastic, R Tibshirani, J Friedman, and J Franklin, "The elements of statistical learning: Data mining inferences and prediction", *Mathematical intelligencer*, vol. 27, pp. 83–85, 2005.

[193] P Rafaeilzadeh, L Tang, and H Liu, "Cross-validation", *Encyclopedia of Database Systems*, pp. 532–538, 2009.

[194] S Geisser, "The predictive sample reuse method with applications", *Journal of the American Statistical Association*, vol. 70, pp. 320–328, 350 1975.

[195] D. M Allen, "The relationship between variable selection and data augmentation and a method for prediction", *Technometrics*, vol. 16, pp. 125–127, 1 1974.

[196] B Efron, "Estimating the error rate of a prediction rule: Improvement on cross-validation", *Journal of the American Statistical Association*, vol. 78, pp. 316–331, 382 1983.

[197] J Shao, "Linear model selection by cross-validation", *Journal of the American Statistical association*, vol. 88, pp. 486–494, 422 1993.

[198] A Elisseeff and M Pontil, *Leave-one-out error and stability of learning algorithms with applications*. IOS Press, 2003.

[199] A Celisse and S Robin, "Nonparametric density estimation by exact leave-p-out cross-validation", *Computational Statistics and Data Analysis*, vol. 52, pp. 2350–2368, 2008.

[200] S Arlot, "A survey of cross-validation procedures for model selection", *Statistical Surveys*, vol. 4, pp. 40–79, 2010.

[201] B Efron and R Tibshirani, "Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy", *statistical Science*, vol. 1, pp. 54–77, 1 1986.

[202] ——, *An introduction to the bootstrap*. London: Chapman & Hall, 1993.

[203] S Salzberg, "On comparing classifiers: Pitfalls to avoid and a recommended approach", *Data Mining and Knowledge Discovery*, vol. 1, pp. 317–328, 1997.

[204] O Celiktutan, S Ulukaya, and B Sankur, "A comparative study of face landmarking techniques", *EURASIP Journal on Image and Video Processing*, pp. 1–27, 2013.

[205] E Chutorian and M Trivedi, "Head pose estimation in computer vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 607–626, 2009.

[206] Y Cohn, J Tian, and T Kanade, "Recognizing action units for facial expression analysis", *IEEE Transactions in Pattern Analysis and Macine Intelligence*, vol. 23, pp. 97–114, 2 2001.

[207] M Pantic and L. J. M Rothkrantz, "Automatic analysis of facial expressions: The state of the art", *IEEE Transactions in Pattern Analysis and Machine Itelligence*, vol. 22, pp. 1424–1445, 12 2000.

[208] S Ioannou, G Caridakis, K Karpouzis, and S Kollias, "Robust feature detection for facial expression recognition", *Journal of Image Video Processing*, vol. 2007, pp. 5–5, 2 2007.

[209] N Pears, T Heseltine, and M Romero, "From 3d point clouds to pose-normalised depth maps", *International Journal of Computer Vision*, vol. 89, pp. 152–176, 2 2010.

[210] L Wiskott, J. M Fellous, N Kruger, and C Malsburg, "Face recognition by elastic bunch graph", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, pp. 775–779, 1997.

[211] P Campadelli, R Lanzarotti, and C Savazzi, "A feature-based face recognition system", in *Proceedings of International Conference on Image Analysis and Processing*, Mantova, Italy, 2003, pp. 68–73.

[212] A Lanitis, C. J Taylor, and T. F Cootes, "Automatic face idenitification system using flexible appearance models", *Image Vision Computing*, vol. 13, pp. 393–401, 5 1995.

[213] F Dornaika and F Davoine, "Online appearance-based face and facial feature tracking", in *Proceedings of International Conference on Pattern Recognition*, Washington, DC, USA, 2004, pp. 814–817.

[214] J Cohn, A Zlochower, J. J. J Lien, and T Kanade, "Feature-point tracking by optical flow discriminates subtle differences in facial expression", in *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 1998, pp. 396–401.

[215] D. G Lowe, "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2 2004.

[216] P Viola and M Jones, "Rapid object detection using a boosted cascade of simple features", in *Proceedings of IEEE International Conference on Computer Vision and Pattern (CVPR)*, kauai, HI, USA, 2001, pp. 8–14.

[217] H Hatem, Z Beiji, R Majeed, M Lutf, and W. J, "Face detection and pose estimation based on evaluating facial feature selection", *International Journa of Hybride Information Technology*, vol. 8, pp. 109–120, 2 2015.

[218] X Zhu and D Ramanan, "Face detection, pose estimation, and landmark localization in the wild", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, USA, 2012, pp. 16–21.

[219] Y Yan, X Zhang, Z Lei, and S. Z Li, "Real-time high performance deformable model for face detection in the wild", in *Proceedings of IEEE International Conference on Biometrics*, Madrid, Spain, 2013, pp. 4–7.

[220] H Pan, Y Zhu, and L Xia, "Efficient and accurate face detection using heterogeneous feature descriptors and feature selection", *Journal of Computer Vision and Image Understanding*, vol. 117, pp. 12–28, 1 2013.

[221] T Mita, T Kaneko, and O Hori, "Joint haar-like features for face detection", in *Proceedings of International Conference on Computer Vision 2005*, Beijing, 2005.

[222] B. A Efraty, M Papadakis, A Profitt, S Shah, and I. A Kakadiaries, "Facial component-landmark detection", in *Proceedings of IEEE International Conference on Automatic Face & Gesture Recognition Workshops*, Santa Barbara, CA, 2011.

[223] C Hakan, B Triggs, and V Franc, "Face and landmark detection by using cascade of classifiers", in *Proceedings of 10TH IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, Shanghai, China: IEEE, 2013, pp. 1–7.

[224] S Baskan, M. M Bulut, and V. Atalay, "Projection-based method for segmentation of human faces and its evaluation", *Pattern Recognition Letter*, vol. 23, pp. 1623–1629, 14 2002.

[225] S. Arca, P Campadelli, and R Lanzarotti, "A face recognition system based on automatically determined facial fiducial points", *Pattern Recognition*, vol. 39, pp. 432–443, 2006.

[226] B Sankur and H. C Akakın, "Robust 2d/3d face landmarking", in *Proceedings of Conference on 3DTV*, Kos, Greece, 2007, pp. 1–4.

[227] H Dibeklioglu, A. A Salah, and T Gevers, "A statistical method for 2-d facial landmarking", *IEEE Transactions on Image Processing*, vol. 21, pp. 844–858, 2 2012.

[228] Y Kwon and N Lobo, "Age classification from facial images", in *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*, 1994, pp. 762–767.

[229] M. Gonzalez-Ulloa and E. Flores, "Senility of the face: Basic study to understand its causes and effects", *Plastic and Reconstructive Surgery*, vol. 36, pp. 239–246, 1965.

[230] F. Yun, G Guodong, and S. H. Thomas, "Age synthesis and estimation via faces: A survey.", *IEEE Transactions on Pettern Analysis and Machine Intelligence*, vol. 32, pp. 1955–1976, 11 2010.

[231] L. G Farkas, M. J Katic, and C. R Forrest, "International anthropometric study of facial morphology in various ethnic groups/races", *Journal of Craniofacila Surgery*, pp. 615–646, 2005.

[232] R Tiwari, A Shukla, C Prakash, D Sharma, R Kumar, and S Sharma, "Face recognition using morphological methods", in *Proceedings of IEEE International Advance Computing Conference*, 2009.

[233] N Ramanathan, R Chellapa, and W Biswas, "Age progression in human faces: A survey", *Journal of Visual Languages and Computing*, vol. 15, pp. 3349–3361, 2009.

[234] P Thukral, K Mitra, and R Chellappa, "A hierarchical approach for human age estimation", in *Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing*, 2012, pp. 1529–1532.

[235] A Gunay and V. V Nabiyev, "Automatic age classification with LBP", in *Proceedings of 23rd International Symposium of Computer and Information Sciences*, 2008, pp. 1–4.

[236] M. A Hajizadeh and Ebrahimnezhad, "Classification of age groups from facial image using histogram of oriented gradient", in *Proceedings of 7th Iranian Machine Vision and Image Processing Conference*, 2011, pp. 1–5.

[237] K.-H Liu, S Yan, and J. C.-C Kuo, "Age group classification via structured fusion of uncertainty-driven shape features and selected surface features", in *Proceeding of IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2014, pp. 445–452.

[238] H Ling, S Soatto, N Ramanathan, and D Jacobs, "Face verification acoss age progression using discriminanative methods", *IEEE Transactions on Information Forensics and Security*, vol. 5, pp. 82–91, 1 2010.

[239] K.-H Liu, S Yan, and J. C.-C Kuo, "Age estimation via grouping and decision fusion", *IEEE Transactions on Information Forensics and Security*, vol. 10, pp. 2408–2423, 11 2015.

[240] C.-C Chang and C.-J Lin, "Libsvm: A library for support vector machines", *ACM Transactions on Intelligent Systems and Technology*, vol. 2, Article No. 27, 3 2011.

[241] P Sai, J Wang, and E Teoh, "Facial age range estimation with extreme learning machines", *Neurocomputing*, vol. 149, pp. 364–372, 2015.

[242] G.-B Huang, Q.-Y Zhu, and C.-K Siew, "Extreme learning machines: Theory and applications", *Neurocomputing*, vol. 70, pp. 489–501, 1-3 2006.

[243] J Wang, W Yau, and H. L. Wang, "Age categorization via ecoc with fused gabor and lbp features", in *Proceedings of IEEE Applications of Computer Vision (WACV)*, 2009, pp. 313–318.

[244] G Guo, G Mu, Y Fu, C Dyer, and T. S Huang, "A study on automatic age estimation using a large database", in *Proceedings of IEEE Conference on Computer Vision*, 2009.

[245] S Yan, H Wang, Y Fu, J Yan, X Tang, and T. S Huang, "Synchronized submanifold embedding for person independent pose estimation and beyond", *IEEE Transactions on Image Processing*, vol. 18, pp. 202–210, 1 2009.

[246] J Suo, F Min, S Zhu, S Shan, and X Chen, "A multi-resolution dynamic model for face aging simulation", in *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*, 2007.

[247] W Chao, J Liu, and J Ding, "Facial age estimation based on label-sensitive learning and age-oriented regression", *Pattern Recognition*, vol. 46, pp. 628–641, 3 2013.

[248] K Luu, T. Bui, and C. Y Suen, "Kernel spectral regression of perceived age from hybrid facial features", in *Proceedings of IEEE Conference on Automatic Face & Gesture Recognition and Workshops*, 2011.

[249] K Luu, K Seshadri, M Savvides, T Bui, and C Suen, "Contourlet appearance model for facial age estimation", in *Proceedings of IJCB*, 2011, pp. 1–8.

[250] C.-T. Lin, D.-L. Li, J.-H. Lai, M.-F. Han, and J.-Y. Chang, "Automatic age estimation system for face images", *International Journal of Advanced Robotic Systems*, vol. 9, pp. 1–9, 216 2012.

[251] A. K Choober, "Improving automatic age estimation algorithms using an efficient ensemble technique", *International Journal of Machine Learning and Computing*, vol. 2, pp. 118–122, 2 2012.

[252] A Hadid and M Pietikanen, "Demographic classification from face videos using manifold learning", *Neuracomputing*, vol. 100, pp. 197–205, 2013.

[253] X Geng, C Yin, and Z.-H Zhou, "Facial age estimation by learning from label distribution", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 2401–2412, 10 2013.

[254] A. S Spizhevoi and A. V Bovyrin, "Estimating human age using bio-inspired features and the ranking method", *Pattern Recognition and Image Analysis*, vol. 25, pp. 547–552, 3 2015.

[255] H Han and A. K. O. CJain, "Age estimation from face images: Human vs. machine performance", in *Proceedings of IAPR International Conference on Biometrics*, 2013.

[256] Y Sun, X Wang, and X Tang, "Deep convolutional network cascade for facial point detection", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3476–3483.

[257] Y Taigman, M Yang, M Ranzato, and L Wolf, "Deepface: Closing the gap to human level performance in face verification", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701–1708.

[258] M Yang, S Zhu, F Lv, and K Yu, "Correspondence driven adaptation for human profile recognition", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 505–512.

[259] X Wang, R Guo, and C Kambhamettu, "Deeply-learned feature for age estimation", in *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 534–541.

[260] Z Niu, M Zhou, L Wang, X Gao, and G Hua, "Ordinal regression with multiple output CNN for age estimation", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4920–4928.

[261] J.-C. Chen, A Kumar, R Ranjan, V. M Patel, A Alavi, and R Chellappa, "A cascaded convolutional neural network for age estimation of unconstrained faces", in *Proceedings of IEEE Conference on Biometrics, Theory, Applications and Systems*, 2016.

[262] Z Huo, X Yang, C Xing, Y Zhou, P Hou, J Lv, and X Geng, "Deep age distribution learning for apparent age estimation", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 722–729.

[263] R. C Malli, M Aygun, and H. K Ekenel, "Apparent age estimation using ensemble of deep learning models", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 714–721.

[264] B Hebda and T Kryjak, "A compact deep convolutional neural network architecture for video based age and gender estimation", in *Proceedings of IEEE Conference on Computer Science and Information Systems*, 2016, pp. 787–790.

[265] T Kanno, M Akiba, Y Teramachi, H Nagahashi, and T Agui, "Classification of age-group based on facial images of young males using neural networks", *IEICE Transactions on Information Systems*, vol. E84-D, pp. 1094–1101, 8 2001.

[266] R Iga, K Izumi, H Hayashi, G Fukano, and T Ohtani, "A gender and age estimation system from face images", in *Proceedings of SICE Annual Conference*, 2003, pp. 756–761.

[267] H Takimoto, Y Mitsukura, M Fukumi, and N Akamatsu, "A design of gender and age estimation system based on facial knowledge", in *Proceedings of ICASE International Joint Conference*, 2006, pp. 3883–3886.

[268] H Takimoto, T Kuwano, Y Mitsukura, H Fukai, and M Fukumi, "Appearance-age feature extration from facial image based on age perception", in *Proceedings of SICE Annual Conference*, 2007, pp. 2813–2818.

[269] X Zhuang, X Zhou, M Hasegawa-Johnson, and T. S Huang, "Face age estimation using patch-based hidden markov model supervectors", in *Proceedings of International Conference on Pattern Recognition*, 2008.

[270] B Ni, Z Song, and S Yan, "Web image mining towards universal age estimator", in *Proceedings of ACM International Conference on Multimedia*, 2009.

[271] B Xiao, X Yang, and Y Xu, "Learning distance metric for regression by semidefinite programming with application to human age estimation", in *Proceedings of ACM Conference on Multimedia*, 2009.

[272] H Han and A. K. Jain, "Age, gender and race estimation from unconstrained face images", Michigan State University, Tech. Rep. MSU-CSE-14-5, 2014.

[273] C Li, Q Liu, W Dong, X Zhu, J Liu, and H Lu, "Human age estimation based on locality and ordinal information", *IEEE Transactions on Cybernetics*, vol. 45, pp. 2522–2534, 11 2015.

[274] X. Yang, B.-B. Gao, C. Xing, Z.-W. Huo, X.-S. Wei, Y. Zhou, J. Wu, and X. Geng, "Deep label distribution learning for apparent age estimation", in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, IEEE, 2015. DOI: 10.1109/ICCVW.2015.53.

[275] R Angulu, J.-R. Tapamo, and O. A Adewumi, "Age estimation via face images: A survey", *EURASIP Journal of Image and Video Processing*, 2018.

[276] C. Solomon and T. Breckon, *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. Wiley, 2010. DOI: 10.1002/9780470689776.

[277] R Lienhart, "Stump-based 20x20 gentle adaboost frontal face detector", *Intel Corporation*, 2000.

[278] R Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2002, pp. 900–903.

[279] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Li, "Learning multi-scale block local binary patterns for face recognition", *Advances in Biometrics*, 828–837, 2008.

[280] M Castrillon-santana, O Deniz-Suarez, T. M. Hernandez, and C. G. A, "Real-time detection of multiple faces at different resolutions in video streams", *Journal of Visual Communication and Image Representation*, vol. 18, pp. 130–140, 2 2007.

[281] M Castrillon-santana, O Deniz-Suarez, L Anton-Canalis, and J Lorenzo-Navarro, "Face and facial feature detection evaluation", in *Proceedings of Third International Conference on Computer Vision Theory and Applications*, 2008.

[282] R. Angulu, J.-R. Tapamo, and A. O. Adewumi, "Landmark localization approach for facial computing", in *Proceedings of IEEE 2017 Conference on Information Communication Technology and Society (ICTAS)*, IEEE, Mar. 2017, pp. 1–6. DOI: 10.1109/ICTAS.2017.7920664.

[283] ——, "Frontal face landmark displacement across age", in *IEEE Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech)*, IEEE, 2017.

[284] A Ross and R Govindarajan, "Feature level fusion in biometric systems", in *Proceedings of the Biometric Consortium Conference (BCC)*, 2004.

[285] J. V Kittler, M Hatef, R. P. W Duin, and J Matas, "On combining classifiers", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 226–239, 1998.

[286] R Angulu, J.-R. Tapamo, and O. A Adewumi, "Age-group estimation using feature and decision level fusion", *The Computer Journal*, 2018.

[287] N. S Lakshmiprabha, J Bhattacharya, and S Majumder, "Age estimation using gender information", in *Communications in Computer and Information Science*, Springer Berlin Heidelberg, 2011, pp. 211–216. DOI: 10.1007/978-3-642-22786-8_26.

[288] J. J. Dongarra, J. D. Croz, S. Hammarling, and I. S. Duff, "A set of level 3 basic linear algebra subprograms", *ACM Transactions on Mathematical Software*, vol. 16, no. 1, pp. 1–17, 1990. DOI: 10.1145/77626.79170.

[289] R. Angulu, J. R. Tapamo, and A. O. Adewumi, "Image and video technology", in, M. Paul, C. Hitoshi, and Q. Huang, Eds. Springer International Publishing, Mar. 2018, ch. Age Estimatiomation with Local Ternary Directional Patterns, pp. 421–434, ISBN: 978-3-319-75785-8.

[290] R. L. D. Valois, D. G Albrecht, and L. G Thorell, "Spatial frequency selectivity of cells in macaque visual cortex", *Vision Research*, vol. 22, pp. 545–559, 1982.

[291] P Kruizinga and N Petkov, "Nonlinear operator for oriented texture", *IEEE Transactions on Image Processing*, vol. 8, pp. 1395–1407, 10 1999.

[292] R. Angulu, J.-R. Tapamo, and A. O. Adewumi, "Human age estimation using multi-frequency biologically inspired features (MF-BIF)", in *Proceedings of 2017 IEEE AFRICON Conference*, Cape Town, South Africa: IEEE, Sep. 2017, pp. 26–31.

[293] C. M and H. D. J., "Normalization as a canonical neural computation", *Nature Reviews Neuroscience*, vol. 13, no. 1, pp. 51–62, 2011. DOI: 10.1038/nrn3136.

[294] R Angulu, J.-R. Tapamo, and O. A Adewumi, "Age estimation with local statistical biologically inspired features", in *Proceedings of 2017 IEEE International Conference on Computational Science and Computational Intelligence (CSCI)*, 2017.

[295] D. E Rumelhart, G. E. Hinton, and R. J Williams, "Parallel distributed processing:explorations in the microstructures of cognition", in. MIT Press, 1988, ch. Learning internal representations by error propagation, pp. 318–362.

[296] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York:Springer-Verlag, 2007.

[297] V Vapnik, *The nature of statistical learning theory*. New York: Springer-Verlag, 1995.

[298] A. Kaehler and G. Bradski, *Learnin OpenCV 3: Computer Vision in C++ with the OpenCV Library*, D. Schanafelt, Ed. O'Reilly Media, 2016.

[299] T Ahonen, A Hadid, and M Pietikainen, "Face description with local binary patterns: Aplication to face recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 2037–2041, 12 2006.

[300] M. Y El-Dib and M. El-Saban, "Human age estimation using enhanced bio-inspired features (ebif)", in *Proceedings of 17th IEEE International Conference on Image Processing*, 2010, pp. 1589–1592.