PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
Ministry of Higher Education and Scientific Research

Larbi Tebessi University – Tebessa
Faculty of Exact Sciences and Sciences of Nature and Life
Département of Mathematics and Computer Science

Thesis submitted in fulfillment of the requirements for the degree of :

MASTER

in

Domain : Mathematics and Computer Science
Field: Computer Science
Speciality: Networks and cyber security

Theme:

Automatic Facial Age Estimation

Presented by :

Rouili Mohamed

Before the jury composed of :

Mr. Med Cherif Nait Hamoud	MAA	University Larbi Tébessi	Chair
Mr. Mohamed Aouine	MAA	University Larbi Tébessi	Examiner
Mr. M.Y Haouam	MCB	University Larbi Tébessi	Supervisor
Mr. L. Laimeche	MCA	University Larbi Tébessi	Co-Supervisor

On the date of : 14/09/2020

Abstract

In recent times, human facial age estimation attracted a lot of attention in the fields of computer vision and pattern recognition because of its significant applications in many domains such as biometrics, law enforcement, health care, security control and surveillance. Numerous approaches were offered and studied by researchers concerning this problem. Nevertheless, age estimation systems that can be globally used for every situation have not been reached yet. In our thesis, we presented an in-depth analysis of facial age estimation. In the proposed solution, three feature extraction algorithms are evaluated, including a Hand-craft technique (LBP), a learned hand-craft technique (CBFD), and a deep learning technique (DCTNet). A classification phase using MLP and SVM classifiers is implemented in order to test the extracted features. Through multiple experiments, our proposed DCTNet achieved 3.84 MAE (mean absolute error) for exact age estimation using the MORPH-II facial image database. In this regard, the proposed method provided a very respectable performance compared to existing state-of-the-art methods.

Keywords : Age Estimation ; Feature Extraction ; Classification ; DCTnet; CBFD; Facial Images

Résumé

Dans ces derniers temps, l'estimation de l'âge à partir du visage humain a attiré beaucoup d'attention dans les domaines de la vision par ordinateur et de la reconnaissance de formes en raison de ses applications importantes dans de nombreux domaines tels que la biométrie, l'application de la loi, la santé, le contrôle de sécurité et la surveillance. De nombreuses approches ont été proposées et étudiées par les chercheurs concernant ce problème mais la performance parfaitement précise n'est pas encore atteinte. Dans ce mémoire, nous avons présenté une analyse approfondie de l'estimation de l'âge du visage. Dans la solution proposée, trois algorithmes d'extraction de caractéristiques sont évalués, dont une technique dite Hand-craft (LBP), une technique Hand-craft d'apprentissage (CBFD) et une technique d'apprentissage profond (DCTNet). Une phase de classification utilisant des classificateurs MLP et SVM est implémentée afin de tester les caractéristiques extraites. Grâce à de multiples expérimentations, notre DCTNet proposé a atteint une MAE (erreur absolue moyenne) de 3,84 pour l'estimation exacte de l'âge, en utilisant la base de données d'images faciales MORPH-II. À cet égard, la méthode proposée a fourni une performance très respectable par rapport aux méthodes de pointe existantes.

Les mots clés : estimation de l'âge ; Extraction des caractéristiques ; Classification ; DCTnet; CBFD; Images du visage.

ملخص

في الآونة الأخيرة، جذب تقدير العمر من وجه الانسان الكثير من الاهتمام في مجالات الرؤية الحاسوبية والتعرف على الأنماط نظرًا لتطبيقاته المهمة في العديد من المجالات مثل القياسات الحيوية، القانون ،الرعاية الصحية والمراقبة الأمنية . تم تقديم العديد من الأساليب ودراستها من قبل الباحثين فيما يتعلق بهذه المشكلة، ومع ذلك، لم يتم الوصول إلى الأداء الدقيق تمامًا بعد. في أطروحتنا، قدمنا تحليلًا معمقًا لتقدير عمر الوجه. في الحل المقترح، يتم تقييم ثلاث خوارزميات لاستخراج الميزات، بما في ذلك تقنية Hand-craft (LBP) ، وتقنية deep learning (DCTNet) وتقنية learned hand-craft (CBFD). يتم تنفيذ مرحلة التصنيف باستخدام مصنفات MLP و SVM من أجل اختبار الميزات المستخرجة. من خلال تجارب متعددة، حقق DCTNet المقترح MAE 3.84 (متوسط الخطأ المطلق) لتقدير العمر الدقيق باستخدام قاعدة بيانات صور الوجه MORPH-II. في هذا الصدد، قدمت الطريقة المقترحة أداءً محترمًا للغاية مقارنةً بأحدث الأساليب الحالية.

الكلمات الرئيسية: تقدير العمر؛ استخراج الميزات ؛ التصنيف ؛ DCTNet ؛ CBFD ؛ صور الوجه

Acknowledgments

First and foremost I offer my sincerest gratitude to my supervisors, Dr. Lakhdar Laimeche and Dr. Mohamed Yassine Haouam, who have supported me throughout my thesis with their patience and knowledge whilst allowing me the room to work in my own way. I attribute the level of my Masters degree to their encouragement and effort and without them this thesis, too, would not have been completed or written. One simply could not wish for a better or friendlier supervisors. They gave me the opportunity to benefit from their wide knowledge and experience. Their ideas and fashion of thinking contributed considerably to this thesis.

Also , I would like to thank Dr. Abdallah Meraoumia , Dr. Isam Bendib and Dr. Djeddi Chaouki as well as anyone who helped me directly or indirectly with their advice, encoragment and support.

Without the support and encouragement from my family, this thesis would have been impossible. I especially want to thank my parents for encouraging me to strive for excellence and advising me to pursue what matters most to me.

Contents

Abstract ………………………………………………………………………………... I
Resumé……………………………………………………………………………….. II
ملخص………………………………………………………………………………... III
List of Figures…………………………………………………………………….. VI
List of Tables…………………………………………………………………… III
Abbreviations……………………………………………………………………... IX

GENERAL INTRODUCTION……………………………………………………….. 1

CHAPTER 1 : FACIAL AGE ESTIMATION…………………….……………........ 3
1.1 Introduction …………………………………………………….……..…..... 3
1.2 Age in theory ……………………………………………….……………...... 3
1.2.1 Factor influencing face aging……………………...……………….. 4
1.2.2 Facial aging stages…………………………………….…………... 5
1.2.3 Application of age estimation……………………..………………… 7
1.3 Age estimation system functionality……………………..…..………….. 9
1.3.1 Processing ……………………………………………….…………….. 10
1.3.2 Feature extraction………………………………………………….. 11
1.3.3 Classification……………………………………………………... 12
1.3.4 Decision………………………………………………………………... 12
1.4 Evaluation protocols………………………………………………………… 13
1.5 Facial aging databases……………………………………….................... 13
1.6 Conclusion……………………………………………………………………… 16

CHAPTER 2: BACKGROUND AND LITTERATURE REVIEW……………..… 17
2.1 Introduction…………………………………………………………...……... 17
2.2 Handcraft based methods……………………………………………………… 18
2.2.1 Anthropometric based methods …………………………...………… 18
2.2.2 Texture based methods ……………………………………………….. 19
2.2.3 Active appearance models ……………………………..……………...... 20

2.2.4 Aging pattern subspace (AGES) …………….……………………….. 21
2.2.5 Age manifold …………………………………………….………... 23
2.2.6 Multi-feature fusion ……………………………...…………... 24
2.2.7 Summary ……………………………………………………….……... 24
2.3 Deep learning based methods ………………………………………………. 25
2.3.1 Training from scratch ………………………………………………... 28
2.3.2 Transfer learning ……………………………………...…………... 28
2.3.3 Feature extraction …………………………………..…………….. 30
2.3.4 Summary …………………………………………..…………….…. 31
2.4 Conclusion …………………………………………………….…………… 31

CHAPTER 3 : PROPOSED METHODS AND EXPERIMENTS RESULTS 33
3.1 Introduction ………………………………………………………………….. 33
3.2 Methodology and proposed methods ……………………………...…….. 33
3.2.1 Preprocessing…………………………………………………………….. 34
3.2.2 Feature extraction………………………………………………………... 35
3.2.3 Classification……………………………………………………………… 43
3.3 Experiments and discussions ………………………………………………….. 44
3.3.1 Image Database……………………………………………………….…... 44
3.3.2 Age estimation evaluation metrics……………………………………..... 45
3.4 Proposed age estimation methods evaluation ………………………………... 46
3.4.1 LBP based age estimation……………………….……………………… 46
3.4.2 CBFD based age extimation……………….…………………………….. 47
3.4.3 DCT-NET based age estimation……………………….………………… 48
3.5 Comparison with state-of-the-art age estimation …………………..………… 49
3.6 Conclusion ………………………………………………..…………….. 50

GENERAL CONCLUSION…..………………………………………………………. 51
Bibliography……………………………………………………………………………. 52

List of Figures

Figure

1.1	Craniofacial growth (shape change) on a human face with age progression...	6
1.2	Skin aging with age progression………………………………………………..	6
1.3	Main parts of a pattern recognition system…………………………………..	9
1.4	Faces detection from images………………………………………………….	10
1.5	Face rotation……………………………………………………………….	11
1.6	Cropped images…………………………………………………………….	11
2.1	Overview of age estimation models and algorithms………………………..	17
2.2	Age pattern vectorization…………………………………………………….	22
3.1	The general block diagram of the proposed Age estimation methods……….	34
3.2	The 68 landmarks of sample images………………………………………….	35
3.3	Cropping and resizing process……………………………………………….	35
3.4	The original image (left) processed by the LBP operator (right)…………….	36
3.5	LBP based face description…………………………………………………..	36
3.6	LBP Mechanism…………………………………………………………...	37
3.7	Learning the handcrafted codebook for feature encoding…………………....	38
3.8	Learned LBP based feature extraction task…………………………………..	40
3.9	Block diagram of the DCTNet……………………………………………….	41

List of Tables

Table

1.1	Summary of face image databases…………………………………………	15
2.1	Summary of hardcraft based methods……………………………………..	26
2.2	Summary of deep learning based methods…………………………………	32
3.1	Number of images by gender and race……………………………………	45
3.2	LBP based age estimation results…………………………………………..	46
3.3	CBFD Age estimation results using MLP and SVM classifiers……………...	47
3.4	DCTNet Age estimation results using MLP and SVM classifiers……………	48
3.5	DCTNet Age estimation results using MLP and SVM classifiers……………	50

Abbreviations

DCTNet:	Discrete Cosine Transform Network
LBP:	Local Binary Patterns
CBFD:	Compact Binary Face Descriptor
DCT:	Discrete Cosine Transform
PCA:	Principale Component Analysis
HOG:	Histogram Of Gradients
BSIF:	Binarized Statistical Image Features
LPQ:	Local Phase Quantization
CA-LBFL:	Contexte Aware Local Binary Feature Learning
CNN:	Convolutional Neural Network
PCANet :	Principale Component Analysis Network
VGG :	Visual Geometry Group
SVM :	Support Vector Machine
KNN :	K-Nearest Neighbors
MLP :	Multi Layer Perceptron
SVR :	Support Vector Regression
rCCA :	Regularized Canonical Correlation Analysis
MAE :	Mean Absolute Error
CS :	Cumulative Score

GENERAL INTRODUCTION

The human face holds important amount of information and attributes such as expression, gender and age. The majority of people are able to easily recognize human traits like emotional states, where they can tell if the person is happy, sad or angry and determine their gender. Most people also have the ability to say from looking at someone if they are a child, a teenager, an adult or an old person. However, knowing a person's exact age just by looking at old or recent pictures for them is often a bigger challenge. Developing automatic facial systems that estimate people's age from their facial images is a challenging task for computer vision due to changes in facial appearance, which are mainly caused by internal and external factors.

Automatic age estimation systems have been growing rapidly in recent years due to their important modules and beneficial uses for many computer vision applications including human-computer interaction, security systems, and visual surveillance. For example, automatic age estimation is currently being used by hotels, airports, bus stations, casinos, government buildings, universities, hospitals, cinemas, etc. to increase the level of security and facing any possible threats or deficiencies. Besides the security applications, age estimation techniques are utilized also in health care systems, information retrieval, academic studies and researches, and Electronic Customer Relationship Management (ECRM) systems, where customers are distributed to different age groups like children, teenagers, adults and senior adults. Furthermore, gathering some customer's daily life information like activities, habits, traditions, priorities etc. may help the corporations to classify products and services depending on their gender or age groups, which lead to increase their incomes and earn more money.

Several age estimation methods have been proposed for different applications. Despite the advantages of these methods, they suffer from several limitations due to several challenges encountered when attempting to develop them. These limitations are explained more in the following points:

☞ The age estimation problem is particularly challenging as age depends on many factors, some of them are visual and many others are non-visual such as ethnic background, living style, working environment, health condition and social life. Changes in the appearance of a face are also attributed to shape (e.g. weight loss/gain) and changes in texture (e.g. wrinkles, scar, facial structure, skin color etc.) as age progresses. Besides biological factors, factors such as ethnicity, habits, etc., and external factors such as glasses, facial hair, changes in pose and expression, etc. often contribute to physical changes in the face. In particular, aging can be accelerated by smoking, genetic predisposition, emotional stress, disease processes, dramatic weight changes, and exposure to extreme climates.

- ☞ The visual features that can help in estimating age such as people's facial features are affected by pose, lighting and imaging conditions.
- ☞ Males and females may have different general discriminative features displayed in images due to the different extent in using makeup, accessories and cosmetic surgeries which increase the negative influence of individual differences. The appearance of many female face images can make a person looks lot younger than they actually.
- ☞ The difficulty of acquiring large-scale databases, which covers enough age range with chronological face aging images, makes the estimation tasks more difficult to achieve.

This study sought to achieve the following objectives:

- ☞ To analyze previous studies in age estimation: factors affecting ageing, algorithms used for age estimation, results achieved, age estimation challenges, image representation techniques for age modelling, age estimation evaluation protocols.
- ☞ To develop a model for age estimation based on the types of feature extraction methods including the Handcraft feature (LBP), Learning Handcraft feature (CBFD), and deep learning feature (DCTNet).
- ☞ To find out the type of feature method effect on the performance in age estimation systems.

In our thesis, we have chosen to organize our study around three main chapters:

- ☞ The first chapter defines basic concepts in age theory, outlines the general structure of an age estimation system and presents a selection of the most prominent face image datasets who are commonly utilized in age estimation tasks.
- ☞ The second chapter provides background information on the state-of-the-art methods in age estimation.
- ☞ The last chapter explains the core components of our age estimation system. An in-depth view is presented on the LBP, CBFD and DCTNet to perform the face features.

Finally, a general conclusion with the targeted perspectives that we will consider is given at the end of this thesis.

Chapter 1.

FACIAL AGE ESTIMATION

Chapter 1

FACIAL AGE ESTIMATION

1.1 Introduction

The appearance of the human face is impacted by several factors; this is why each and every single one of us has a unique face and aging pattern. Aging involves both variations in soft tissues and bony structure on the face. As a person gets older, his/her face changes completely. Hence, these variations that the age introduces can be learned and employed for age estimation. The face delivers noticeable evidence about a person's identity, age, gender, mood and ethnicity. The data rendered by the face has brought considerable attention within the face image processing study field. Automatic age estimation from facial images has predominantly attracted vast research interest thanks to its many application areas. It is pattern recognition system that consists of labeling the face with a precise age or age group automatically. In this chapter we dive into the theory of the human age and define the general structure of age estimation systems as well as the face image datasets who were developed specifically to be exploited in age estimation research.

1.2 Age in Theory

Aging is a procedure that can't be stopped or inverted. The facial features of a human change as time passes by which reflects major variations in appearance. The signs of age progression that are displayed on faces are uncontrollable and personalized such as graying hair, dropping muscles, loose skin and wrinkles. The aging signs depend on many external factors such as life style and degree of stress. For instance, smoking causes several facial characteristics changes. A 30 years old person who smokes a box of cigarettes each day will look like 42 years old instead.

Compared with other facial characteristics such as identity, expression and gender, aging effects display three unique characteristics:

- The aging progress is uncontrollable. No one can advance or delay aging at will since the aging process is slow and can't be reversed under any circumstances.

- Different people age in different ways and each has unique aging variations. The aging variation of each person is determined by his/her genes as well as many external factors, such as health, lifestyle, weather conditions, etc.
- The aging variations are temporal data. The aging progress must obey the order of time. The state of the face at a certain age will impact all older faces, but not younger ones.
- All of these characteristics contribute to the complications of automatic age estimation.

1.2.1 Factors influencing face aging

Many factors have a heavy impact on facial aging from natural, psychological and environmental factors to the person's lifestyle and occupation. These factors can be split into two main categories, intrinsic factors and extrinsic ones. Extrinsic factors are those that are external to the human body like environmental and occupation factors while intrinsic are internal factors like bone structure and genetic influence which occur naturally over time. In childhood, facial changes are mainly caused by craniofacial development which leads to changes in facial shape due to growth, modeling, and deposition of bony tissues in the face. This leads to changes in height and shape of the face [1]. The back sloping of the forehead creates space on the cranium which gets occupied by facial landmarks which causes changes in facial shape during childhood.

On the other hand, during adulthood, the aging of the face is demonstrated primarily in texture variations which are the result of various factors.

- Arid environment and wind dehydrate the skin leading to wrinkle formation.

- Air pollution has also been found to affect aging by accelerating wrinkle development. Research on air pollution and aging has shown that city dwellers that are prone to polluted air from industries develop deeper wrinkles than individuals who are not exposed to pollution.

- Smoking has negligible effect on wrinkling compared to UV rays and photoaging effects [2]. However, it's worth noting that smoking affects the skin's collagen production which leads to wrinkles appearing around the mouth.

- The ultraviolet (UV) rays affect the skin's pigmentation heavily as they dry and destroy the skin cells which hasten wrinkle development around the most prominent face areas like the eyes along with multiple variations in photoaging like skin wrinkling, elastosis and irregular pigmentation.

- The gravity's force reduces the skin's elasticity thus augmenting the skin wrinkling. Internally, changes in bone structure and subsequent variations in musculature cause skin wrinkling. Aging was also found to be different between males and females with female faces tending to age faster compared to male faces [3]. Aging in males and females share many common characteristics, but there are some differences.

- Cosmetics have a heavy impact on age estimation since they hide the wrinkles and age spots all around the face which makes the person look a lot younger than what they actually are in most cases. It was also proven that facial expression has its impact. Some facial expressions like smiling, frowning, surprise, and laughing may introduce wrinkle-like lines on some regions of the face like the forehead, cheek bone area, mouth region, and nose-bridge regions all of which could be registered all wrinkles during age estimation.

1.2.2 Facial Aging Stages

Aging affects the human facial appearance significantly as it's an irreversible natural process that cannot be controlled. Moreover, although age progression affects facial appearance of different people differently, biological or anthropometric studies suggests that based on some common features, facial aging can be roughly divided into two stages, birth-to-adulthood and adulthood-to-Senior**.**

1.2.2.1 Changes in the texture and shape of the face during birth-to-adulthood stage

During birth-to-adulthood, usually bone growth takes place that causes major changes in the facial shape as shown by the six prototype images of Figure 1.1.

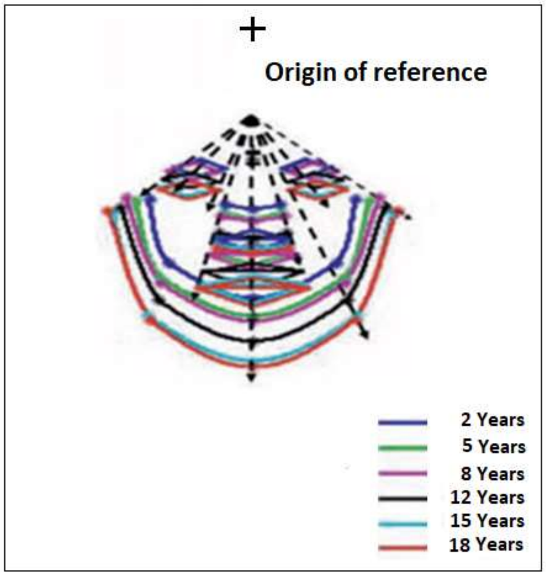

Figure 1.1: Craniofacial growth (shape change) on a human face with age progression, originally shown in [4]

The most prominent changes in the texture and shape of the face during this stage of aging consist of:

- The forehead slopes back shrinking and creating space on the cranium.
- Facial features such as eyes, nose, mouth and ears grow their areas.
- Cheeks expand their areas and chin becomes more bulging.
- Skin texture does not change much but facial hair become denser.

1.2.2.2 Changes in the texture and shape of the face during adulthood-to-senior stage

During the adulthood-to-senior stage the most perceptible age-related deformations are associated with texture changes (see Figure 1.2).

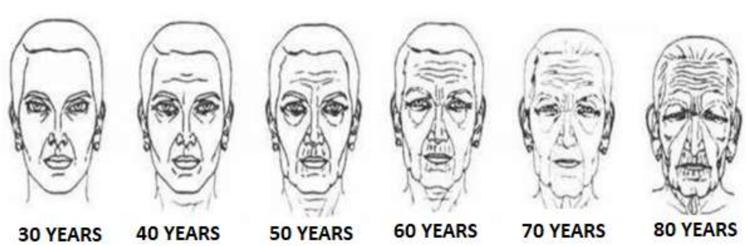

Figure 1.2: Skin aging with age progression [5]

The most prominent changes in the texture and shape of the face during this stage of aging consist of:

- The face skin thins out and becomes leathery with less flexibility, its tone becomes darker.
- The gradual appearance of blemishes and wrinkles due to aging.
- Muscle motion generates dynamic wrinkles and folds which become more projecting.
- Cheeks start dropping, double chin and lower eyelid bags appear.
- Even though craniofacial growth is minimal during this aging period, there is still considerable change in the facial geometry from adulthood to senior age. In female especially, faces alter from an upside-down triangle form to a rectangular one.

1.2.3 Applications of age estimation

In the last few decades researchers came up with a plethora of techniques for the extraction and use of these facial characteristics in areas such as age estimation, gender classification, expression recognition and human identification. Among these, human age estimation, whose objective is to determine the specific age or age range of a person based on a given facial image, is a challenging yet attractive topic due to its roots in numerous real-life applications such as:

- **Law Enforcement**:

  A system that is armed with decent age estimation unit can be very practical in filtering out the possible suspects in a more precise and effective way a database using the estimated age of the input mugshot.

- **Biometrics**:

  Multi-cue identification. Age estimation is a type of biometrics that can be employed to complement the primary biometric features to improve the performance of a primary hard biometrics system for example: recognize or identify faces after a gap of several years.

- **Security Control and Surveillance**:

  Security control and surveillance monitoring issues are becoming more and more crucial in daily life. For example, accurate age estimation can stop under aged teenagers from entering unauthorized places, also prevent smokers among them from buying cigarettes from vending machines and block children from accessing adult websites or restricted videos.

- **Health Care**:

  Age estimation can also be helpful in health care systems, like a robotic nurse or an intelligent intensive care unit, for customized services. For example, an adapted Avatar will be designated automatically from the custom-built Avatar database to communicate with patients from various age groups and specific preferences.

- **Forensic Art**:

  Human age is one of the principal artistic techniques in forensic art and here we find that the face age synthesis is the most useful technique then the age estimation like predicting a missing person face after few years.

- **Employment**:

  Some government employments like the 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 employees should retire after reaching a particular age. Age estimation systems could also play a significant role in finding if one has reached retirement age.

- **Human-computer interaction (HCI):**

  The system can adjust the contents presented to a user based on his/her age. For example, a smart shopping cart can be designed to provide recommendations according to the age of the customer.

- **Electronic customer relationship management**:

Satisfy preference of all ages. Age can be helpful for companies or advertisers for planning the best strategy like which group of ages exactly is interested to such product or witch group age is considered as the biggest client number.

These few example application areas not only highlight the contribution of face image automatic age estimation systems to real world but also inspire the need for more research work that can produce state-of-the-art systems to accurately estimate human age.

1.3 Age estimation system functionality

The general structure of age estimation systems involves three main parts which are: the pre-processing step, features extraction and classification (see Figure 1.3). Pre-processing and data collecting part is related to choose an appropriate database, enhancing samples and isolating unnecessary ones. The enhancing comes in the shape of image modifications that can be done like image resizing, image rotation, segmentation, modification of the illuminations, or any other corrections may enhance the appearance of images. As for features extraction part, it is related to finding critical features, define pattern classes and representation, and reduce the data by measuring these features. The classification part is where a classifier divides the features space into regions and assigns a pattern to a category.

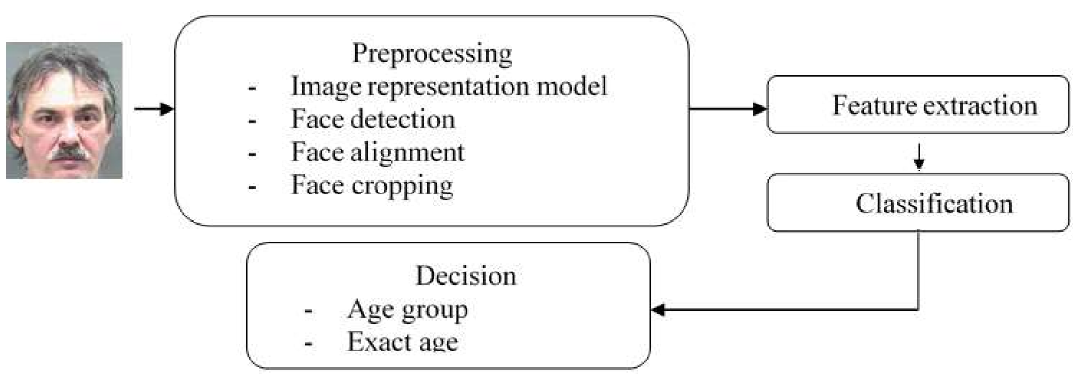

Figure 1.3: Main parts of a pattern recognition system

1.3.1 Preprocessing

Solving age estimation problems require overcoming some main difficulties, such as differing image dimensions and qualities, image background noise, varying levels of luminosity, choosing an appropriate database for each problem, and employing sufficient number of images in each experiment. Therefore, some important pre-processing techniques should be applied in order to keep the essential information only and therefore prepare the image for the next step. There are several types of image quality processing and enhancement, such as converting color to grayscale, face detection, eyes rotation and image cropping.

1.3.1.1 Image grayscale conversion

An acquired image usually belongs to the RGB space; it is converted to grayscale in order to reduce the influence of inconsistent colors. Also, the computational cost in terms of time and memory usage is reduced.

1.3.1.2 Face detection

Face detection in the image is an essential and crucial step. It involves searching an image for the position of faces and extracting them as a set of images to facilitate further processing. A face is considered correctly detected if the extracted image size does not exceed 20% of the actual size of the facial region, and it contains primarily the eyes, nose and mouth. There are multiple effective ways to detect and outline the faces from images like the usage of predefined shape predictors, like the 68 landmarks shape predictor, another way would be to use specialized algorithms like the VJ (Viola et Jones) [6] face detection algorithm (see Figure 1.4).

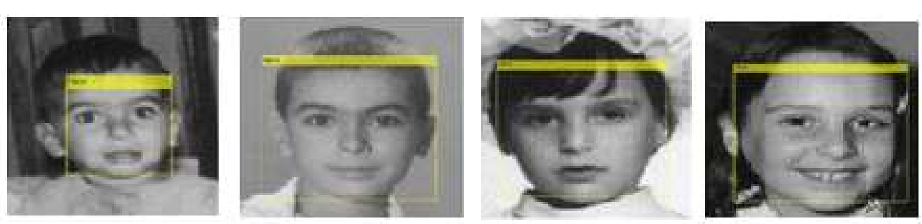

Figure 1.4: Faces detection from images

1.3.1.3 Face Rotation

Face rotation is one of the most important pre-processing stages in image-based age estimation. There are various methods to correct the alignment of the eyes, the first step consists of detecting them using specific detection algorithms or shape predictors for facial landmarks localization, after that the face is rotated by a calculated angle (see Figure 1.5).

Figure 1.5: Face rotation

1.3.1.4 Image cropping

Image cropping consists of keeping the maximum intrinsic variations in the face, and removing other information such as background, hair, shirt collars, ears, etc. in order to improve the performance of the age estimation system (see Figure1.6).

Figure 1.6: Cropped images

1.3.2 Feature extraction

In an age estimation system and after having acquired the images and preprocessed them, the feature extraction is performed. Feature extraction is important factor for the success of the age prediction process, and should be able to extract more information even under difficult conditions, such as: bad lighting, noise and redundant data. The feature extraction techniques are split into classical handcrafted methods like the DCT[7], PCA[8], HOG[9], BSIF[10], LBP[11,12,74], LPQ[13] etc., learning handcraft features like CA-LBFL[14], CBFD[15], etc., and the more advanced deep learning methods like the different CNNs[16,17], DCTNet[18], PCANet[19,20], VGG-16[21,22,23] etc.

1.3.3 Classification

When the feature vector is obtained, a classifier (SVM[24][25], KNN[26], MLP-C[27]…) or regressor (SVR[28], rCCA[29]…) is used in training stage on the available databases for age estimation. Using classification-based methods, the subjects can be classified into a real age label or an age group. Else, the age of a person could be estimated as a numerical value using regression-based methods. Lately, ranking-based approaches have attracted a lot of researchers in computer vision and have been used in age estimation. The hybrid-based methods can be used to combine the benefits of the classification approach with regression. Lastly, a soft classification algorithm that can be labeled as a middle case between classification and regression is being used by different models to enhance the performance of age estimation systems [30].

1.3.4 Decision

The decision in age estimation systems is based on two prediction values including exact age and/or age group.

1.3.4.1 Exact age

Exact age estimation approaches estimate a single label (value) that represent the age of a face image which is relatively difficult when compared with age-group based age estimation since the margin of error is very small. Estimation based on exact age also requires a very large and balanced aging dataset for optimal results. The main advantage of exact age estimation is that it aims to estimate the precise actual age of the person rather than approximate its range which means that highly performance age estimation systems which are based on exact age estimation are more accurate than age group-based systems with the same performance.

1.3.4.2 Age group

Due to the difficulty of achieving accurate exact age estimation, many researchers adopt the age group-based age estimation instead and split their datasets into age groups which is relatively easier since the estimated age just needs to be classified in his respective age range. The age groups range also affects performance since getting good results in an age estimation based on groups of 5 years for e.g. is harder than achieving the same performance in estimation conducted on groups of 10 years range since it gives a larger error margin.

1.4 Evaluation Protocols

This section concerns the evaluation of the system's performance in terms of accuracy. Notably, there are many techniques used to evaluate performance of pattern recognition systems such as Leave-One-Out, Cross-validation, k-folds Cross-validation, Mean Absolute Error (MAE), and Cumulative Score (CS). In this section, we present only cross-validation technique, while the Mean Absolute Error (MAE) and Cumulative Score (CS) used in the proposed age estimation methods are detailed in the following chapters.

Cross-Validation is the procedure of judging how the outcomes of a statistical analysis adapt to an independent database. In an estimation system, a model is usually trained using a dataset of known data (training set), and a dataset of first seen data to test the model (testing set). The aim of cross-validation is to define a dataset to test the model during the training stage, to limit issues like over fitting, give an idea on how the model will generalize to an independent dataset.

There are two kinds of cross-validation, exhaustive and non-exhaustive cross-validation. Exhaustive cross-validation techniques are methods divide the original sample into a training and a validation set through learning and testing on all possible ways, these approaches include Leave-p-out cross-validation and Leave-one-out cross-validation. On the other hand, non-exhaustive cross-validation techniques do not calculate all the ways of dividing the original sample. The Holdout method and k-fold cross-validation are two examples of non-exhaustive cross validation.

1.5 Facial aging databases

The most common and available database that are used in age estimation are presented in this section. Accurate age estimation requires large, balanced and labeled database. Assembling such databases is very difficult and the collection of multiple images of one person at different ages is even harder.

FG-NET [31] is made up of 1002 images of 82 persons and was created with the aim of studying real age. The images are both gray scaled and colored. The ages range from 0 to years. Each person has an average of 12 pictures. The database provides also annotation for different human races. There exists a diversity of head poses and some facial expression as well as some illumination on the images. Moreover, the database provides 68 landmark points that are useful for facial shape modeling. The dataset is available online.

MORPH [32] is face dataset made up of 2 albums. Album1 holds 1724 images of 515 persons. The Album 2 (MORPH-II) dataset on the other hand has 55,134 images of 13,618 persons gathered over a 4-year time span. The ages range from 16-77 years. This database is predominantly employed to evaluate diverse deep learning techniques on the age estimation task. The database has labels for age, gender, race, birth date, and the date it was taken. Even eye coordinates can be obtained. There is a larger set that is commercially available gathered over a longer time. It offers additional data of the person like his height and weight.

PAL [33] dataset is made up of 575 images, it also contains four age groups for adults and older that demonstrate the age throughout the person's life. The ages range from 18-70+. The first age group ranging from 18-29 has 218 images, the second one ranging from 30-49 holds 76 images, the third ranging from 50-69 contains 123 images, and the last group ranging from 70 and older has 158 images, the PAL database can be found online.

Iranian Face Database IFDB [34] is made up of 3600 images of 616 persons. All of the images are colored. This database can be utilized to study various face related tasks like age classification and race detection. The images are split into 787 males and 129 females. The ages range from 2 to 85. The images have multiple variations like facial expression, the pose of the head, without glasses. The dataset can be retrieved online.

YGA dataset [35], contains 1600 Asian people with a total of 8,000 images. The images are high-resolution and colored. 800 of the images are females and 800 are males. The ages range from 0-93 years. There are 5 images per person with a label of the approximate age on average. There is a huge diversity of illumination and facial expression. A face detector was used to crop the faces.

WebFace [36] is made up of 77,021 images of 219,892 faces. The ages range from 1 to 80 years. The database was gathered from Flickr and Google Image. The images were cropped to 240 × 240 size and normalized.

CACD [37] is a very large dataset that was established for face recognition and age retrieval tasks and collected from Internet. This database has 163,446 images belong to 2,000 celebrities. The ages range from 16-62 years. The dataset is available online.

IMDB-WIKI [23] is by far the biggest publicly available database; it was developed for age estimation tasks. This dataset has facial images labeled with gender and age annotations. It is made of 523,051 images that belong to 20,284 different people. The ages range from 0-100 years. The database was formed thanks to the Wikipedia. All the images have metadata about the date when they were captured. The dataset is available online.

WIT-DB database [39] The Waseda human-computer interaction technology database has 12,008 face images of 2500 females and about 14,214 images of 3000 males all from japan, with age ranging between 3 and 85 years. The ages are split into 11 age groups. The database has wide variations in the illumination. Face images are cropped and resized to 32 × 32 grayscale patches.
Table 1.1 summarizes some of the most used aging databases for different age ranges. Some of these databases have annotations with real age while others have age group labels.

Database Name	**Images**	**Subjects**	**Age range**	**Age type**
FG-NET	1002	82	0-69	Real Age
PAL	575	575	18-93	4 Age groups
Iranian Face Database IFDB	3600	616	2-85	Real Age
YGA	8000	1600	0-93	Real Age
MORPH-II	55134	13618	16-77	Real Age
WebFace	77021	219892	1-80	Real Age
CACD	163446	2000	16-62	Real Age
IMDB-WIKI	523051	20284	0-100	Real Age
WIT-DB	26222	5500	3-85	11 Age groups

Table 1.1: Summary of face image databases

1.6 Conclusion

In this chapter, we defined the aging theory, outlined the facial aging steps, and presented the different intrinsic and extrinsic factors that influence facial aging as well as the various real-world application areas of age estimation. We then presented the general architecture of an age estimation system followed by its evaluation protocols. Finally, we summarized some of the most utilized and available databases for age estimation.

In the next chapter, we take a look on different image representation approaches for age estimation and some of techniques and approaches of age estimation.

Chapter 2.

BACKGROUND AND LITERATURE REVIEW

Chapter 2

BACKGROUND AND LITERATURE REVIEW

2.1 Introduction

Age estimation models are mainly based on either handcrafted algorithms or deep learning technology. The feature extraction and selection stage represent the main difference between the two as it is accomplished manually for handcrafted models while being achieved automatically with no human meddling in the case of deep learning models. It's also worth mentioning that not all handcraft-based methods and deep learning methods represent facial images and learn aging patterns in the same way as both have multiple unique techniques and approaches split into specific categories (see Figure 2.1) .In this chapter we offer an overview on the different state-of-the-art methods and approaches used by researchers and outline the achieved results and advances in the age estimation studies.

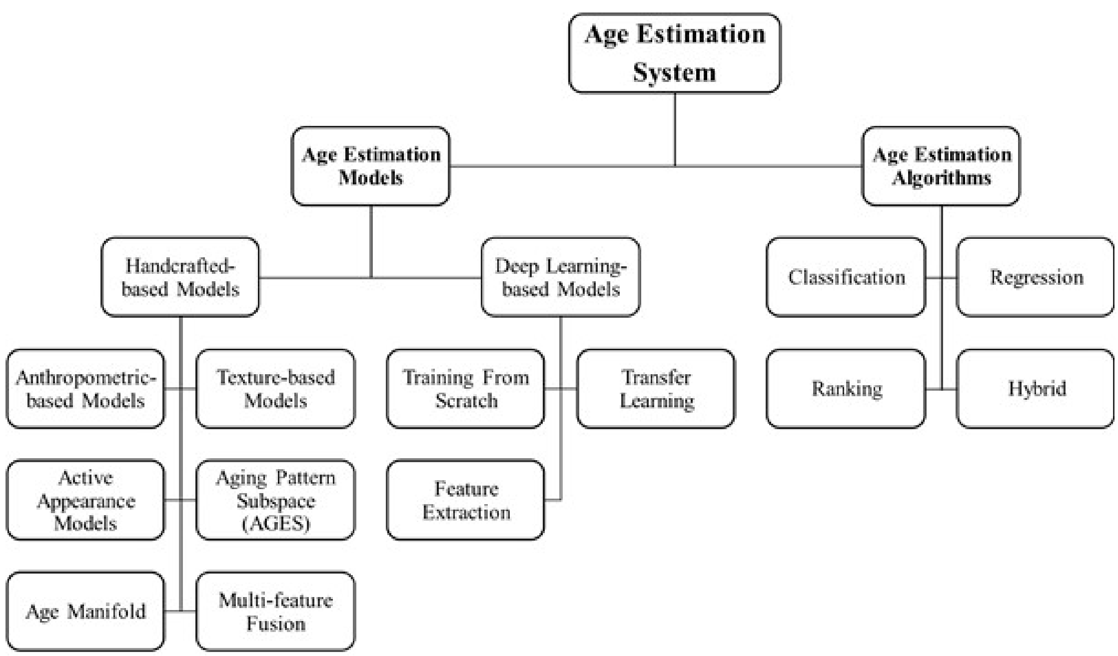

Figure 2.1: Overview of age estimation models and algorithms

2.2 Handcraft based methods

Knowledge by hand and human interference is required in order to use the handcraft-based methods in extracting the features from images as it's a manual process achieved through the usage of specified algorithms by experienced researchers or academics. The handcraft-based or traditional techniques for feature extraction can be mainly split up on 6 main categories of image representation for age estimation:

2.2.1 Anthropometric based methods

Anthropometry-based features are a set of distances and ratios which are calculated between fiducial (landmark) points in a frontal normalized face. The idea is to use these distances in order to describe the topological differences between faces of different ages. Basically, the anthropometric features derive the geometric dimensions of the skull based on the provided face image.

Kwon and Lobo [39] presented the first work on age estimation, their system evolved around the idea of classifying the images into 3 age groups depending on the wrinkle patterns and their geometrics. They differentiated between adults and children by computing 6 distance ratios and used snakelets to detect curves and extract the wrinkle patterns from skin areas. Their system's performance achieved 100% accuracy on a private database consisting only of 47 images that was employed in the tests.

Ueki et al. [40] presented a 2-Dimensional Linear Discriminant Analysis (2DLDA) model to uncover the most discriminant vectors that related to age-group based classification. The performance of their system displayed a 50% CS for males and 43% CS for females. They used the Waseda human computer Interaction Technology DataBase (WIT-DB) for their experiments.

Dehshibi and Bastanfard [41] computed the geometric ratios from facial images in order to extract the texture and associated features of facial components which were then employed to classify the images into 4 age groups. Essentially, to measure the facial geometric from face images, these should be in a frontal view, since the computing these ratios can be very sensitive. It's worth noting that this method is considering only the geometric features which might be unfitting for adults and old people since the appearance of the skin is the clear feature that characterizes the aging material

Their system's performance achieved 86% accuracy on the Interactive Fiction DataBase (IFDB) that was used in the experiments.

2.2.2 Texture based methods

Many studies on age estimation from face images are based on extraction of both holistic and local texture information. The most straightforward way to achieve that is to directly use the pixel intensities. Raw pixels contain a lot of redundant information which can be removed using dimensionality reduction methods. Many texture descriptors emerged and achieved interesting results in age estimation and thus became widely used by researchers.

One of the most significant texture descriptors is the BIF that was studied by Guo *et al.* [10]. The interesting thing about this method is that the new set of aging patterns can efficiently handle minor alternations like scale changes translations and rotations which is a very good advantage. The BIF descriptor as displayed as being very performant when it comes to age estimation by multiple studies. In Guo *et al.* experiments they opted for a regression-based age estimation with their best achieved result being a 4.77 MAE on the FG-NET database.

Active Shape Models (ASM) were also employed by Dib and ElSaban [42] who tried to extend BIF by integrating facial features with automatic initialization to boost the performance of age estimation. They experimented on both the MORPH and FG-NET databases and achieved considerable results of 3.17 (FG-NET) and 4.11 (MORPH) MAE.

Choi *et al.* [11] proposed age estimation approach using hierarchical classifiers with local and global facial features. Gabor filters were used for wrinkles while LBP was used for skin feature extraction. They classified face images into age groups using support vector machines. The disadvantage of this method is that it only depends on a single classifier which makes it susceptible to errors. 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. Their system's performance achieved a 4.7 MAE and 73% CS on the FG-NET database and a MAE of 4.3 along with a 70% CS on the PAL database.

A handful of other traditional methods rather than LBP and BIF were also tested for age estimation like the boosting model that was introduced by Zhou *et al.* [43] for regression-based age estimation which was trained using Haar-like descriptor. This model's performance achieved a 5.81 MAE on the FG-NET database.

Yan *et al.* [44] employed the SPF (spatially flexible patch) feature descriptor for local feature extraction of local patches and their specific positions from the face, other aspects such as head pose and occlusion could also be effectively handled. They experimented on the YGA database. The performance of their system displayed a 7.82 MAE and 75% CS for males and 8.53 MAE along with 70% CS for females.

Hajizadeh and Ebrahimnezhad [45] used HOG for feature extraction and classified the images into 4 age groups using a probabilistic neural network. Their system's performance achieved 87.025% accuracy on the IFDB dataset that was used in the experiments.

On another hand, the scattering transform descriptor which is used to scatter the Gabor coefficients was evaluated for feature extraction by Chang and Chen [46] using a ranking based age estimation system. Results from their experiments showed that scattering transform is more general than conventional BIF and very performant on face-based age estimation. They tested on both the FG-NET and MORPH2 databases achieving 4.7 and 3.82 MAE scores respectively.

2.2.3 Active appearance models

The appearance features are based both on shape and texture data. A typical method for extracting appearance features is Active Appearance Models (AAM) which was initially proposed for image coding. Using the training dataset, AAM separately applies PCA to learn a statistical shape model and an intensity model of face images.

Lanitis et al. [47] adapted Active Appearance Model (AAM) for age estimation problem by proposing ageing function. They defined age as a function that relates the model-based representation of each subject to the actual age; age = f (b), where age is the real estimated age of a subject, b consists of 50 AAM-learned-parameters feature vector, f is the ageing function. They conducted their tests on a private dataset of 500 images of 60 persons of which 45 had images of 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. They achieved an MAE of 4.3 years.

By adopting a ranking based age estimation system, the age features were displayed by Yan *et al.* [49] from a low up to high rank levels with indefinite labels. They experimented on the YGA and FG-NET databases. The performance of their model was estimated at a 5.33 MAE on FG-NET and a 79% CS along with a 6.95 MAE on the YGA database.

On a regression model, Chen *et al.* [50] applied cumulative attribute space for learning aging features with the aim of overcoming the issue of imbalanced and scarce information. The associated features were extracted using AAM and tested on the FG-NET and MORPH datasets. They achieved MAE of 4.67 years with a CS of 74.5% on FG-NET ageing dataset and 5.88 years with a
CS of 57.9% on the MORPH dataset.

Feng *et al.* [51] on their ranking based age estimation system proposed another AAM approach consisting of making the ranks of each of the age labels by estimating their significance to the facial image. They reported MAE of 4.35 years on FG-NET, 4.59 years on MORPH and 6.03 years on the Webface dataset.

It’s safe to say that when comparing anthropometric methods with AAM, the AAM technique shows a strong advantage since it can manage any age while considering both texture and geometric models. However, it’s worth noting that the loss of wrinkle data and some skin areas due to dimensionality reduction along with the intensive computations and the need of large number of images to learn the features that are related to the shape and appearance are considered as important inconvenients with this method.

2.2.4 Aging pattern subspace (AGES)

Geng et al. [52] [53] proposed an automatic age estimation method dubbed AGES (Aging pattern subspace) that uses the appearance of face images with the main concept of being the utilization of the images which are belonging to one person and define the related aging pattern for that subject.

The definition of aging pattern is basically a sequence of facial images ordered by time. The pattern’s images must all belong to the same individual. This aging pattern is called a complete pattern if images at all ages for an individual are available or else it is called an incomplete pattern.

AGES learn the subspace representation of a person's images when modeling a sequence of an individual's aging face to deal with lacking ages. To estimate age, test image is positioned at each possible location in the aging pattern to find a point that can best reconstruct it. Aging subspace that minimizes reconstruction error determines age of the test image. The vectorization of the aging pattern with lacking images in the aging pattern vector (labeled with *m*) is portrayed in figure 2.2. Available face images in the pattern (ages 2, 5, and 8) are placed at their respective positions and ages at which images are not available if their positions are left blank.

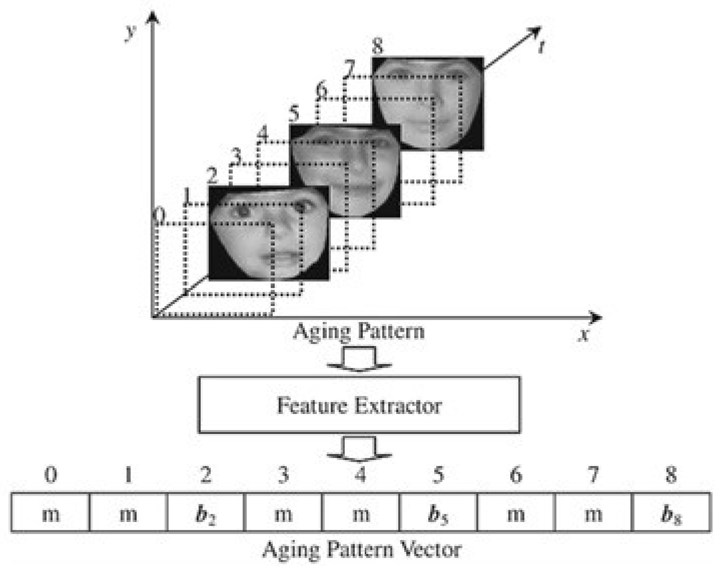

Figure 2.2: Age pattern vectorization

Following the vectorization, face images at ages 2, 5, and 8 are represented by feature vectors b_2, b_5, and b_8, correspondingly. Representing aging pattern with AGES guarantees that label *age* (*I*) and *id* (*I*) are implemented in the data where every pattern indicates an ID and very single age is fixed at a specific time-ordered location in the aging pattern.

Testing the AGES on FG-NET, 200 AAM parameters were employed to characterize very image for age estimation. Geng et al. reported a MAE of 6.77 years with their regression-based model and later on, they tested a classification-based model using the MORPH database and reported a MAE of 8.83 years.

2.2.5 Age manifold

The aging pattern can be commonly learned as a trend for several subjects at different ages rather than the process of finding a specific pattern for each person, like the previously discussed AGES approach. The nature of the aging pattern for a specific age indicates that many faces could be considered to represent the age. Moreover, many images may be found for one person at a single age or related to a range of ages. This means that portraying facial aging with this kind of learning proves to be more flexible than AGES which is why a method dubbed age manifold was offered for the learning of the common pattern among various ages of multiple faces.

Fu and Huang [54] defined the formulation of manifold through Conformal Embedding Analysis (CEA) to portray a low-dimensional subspace. The final statistical age was estimated using multiple linear regression (MLR). They achieved a MAE of 5-6 years on the YGA dataset.

Redundant information that can be eliminated using dimensionality reduction approaches is contained in raw pixels. Since age estimation is fairly complex process, Guo *et al.* [55] observed that using unsupervised methods to reduce the dimensionality like PCA, or locally linear embedding are not appropriate to discriminative subspaces. Instead, they opted for Orthogonal Locality Preserving Projections (OLPP) which is a supervised manifold learning technique. First, they predicted ages by a regression function. After, in order to match the right values within a bound the ages get adjusted locally. They experimented on the YGA and FG-NET databases. The MAE of their model was estimated at 5.07 years on FG-NET database. On the YGA database, they reported a CS of 81% for females and 83% for males along with MAEs of 5.30 years for males and 5.25 years for females.

Depending on 3-D head pose, Yan *et al.* [56] offered a framework through looking for submanifold embedding using data of the subject identity. They reported a MAE of 5.21 years with their regression-based model on the FG-NET database.

An impressive age estimation system based on discovering the low-dimensional manifold was offered by Cai *et al*. [57]. The associated features were extracted using dual Histogram LBP and tested on the FG-NET and MORPH datasets. They achieved MAE of 4.64 years with a CS of 72.1% on FG-NET ageing dataset and 4.66 years with a CS of 62.2% on the MORPH dataset.

2.2.6 Multi-feature fusion

Many researchers have applied the data fusion approach to enhance the performance of their systems. It revolves around the concept of fusing a handful of feature extraction techniques in one system.

Huerta *et al.* [58] proposed a classification-based system which fused appearance and local texture descriptors HOG, SURF and AAM, that gave very decent age estimation results when experimented on the MORPH and FRGC datasets. They achieved MAE of 4.25 years with a CS of 71.17% on MORPH dataset and 4.17 years with a CS of 76.24% on the FRGC dataset.

In a study conducted by Liu *et al* [59] ordinal information along with the geometrical information that assures the discriminative capability of the designated features and their smoothness on the manifold were combined. They reported MAE of 1.3 years on FG-NET dataset and 8.2 years on the FACES dataset.

A fusion of the AAM for global feature extraction with texture-based methods mainly LBP, GW and LPQ to extract the local features were employed by Pontes *et al.* [60] to offer a flexible hierarchical structure as an approach for age estimation. Their system was a hybrid of multi-class classification and regression with SVM used to classify the age into age groups while SVR estimated the final exact age. They achieved MAE of 4.50 years on FG-NET ageing dataset and 5.85 years on the MORPH2 dataset.

2.2.7 Summary

The main advantage of using hand-crafted methods for feature extraction and face representation is the molding of less complex systems. On the hand it’s worth noting that the usage of these techniques requires knowledge by hand and they also sometimes experience loss of critical and important data hence why there was an urge to replace the hand-crafted features with more powerful feature learning techniques. While the traditional techniques may not be compatible with the process of aging because of the loss of important data during the process of feature extraction and selection. In the best scenario, age image representation should portray adequate age data. Table 2.1 summarizes the multiple techniques and methods that were evaluated on different datasets. Hand-crafted techniques were used to extract the relevant aging related features from facial images. Multiple attempts and experiments were conducted with the aim to improve the performance of age estimation system either for age group classification or estimating the real age. We also noticed that the fusion approach or hybrid approach can improve the performance of age estimation systems when compared with other approaches. Fusing the multiple feature approaches in one model has proven to be the most effective technique to get more robust and accurate system while a more complex system could be generated.

2.3 Deep learning-based methods

The usage of deep learning networks in age estimation systems has seen a great rise in recently published works where they were either using it the feature extraction or the age estimation (classification/regression/ranking) stage. Another method is employed by deep learning-based systems (CNN's in particular) when compared with the different hand-craft based methods that have been discussed previously.

While implementing an age estimation system and choosing the optimal deep learning network structure for it, multiple key factors should be considered. These factors are:

1) The age estimation algorithm (classification, regression, ranking or hybrid).
2) The associated loss function of the training algorithm that measures the inaccuracy of predictions.
3) The deep network architecture's depth in layers (Shallow < two layers or Deep > two layers or very Deep >10 layers) [61].
4) The pre-training methodology: (training the network from scratch on specific task or already pre-trained on general task). The main goal of training process is to fine-tune the network's parameters such as the weights and biases which will allow the trained network to predict the classes from the databases [62].

The three main deep learning approaches are: Training from scratch, transfer learning and feature extraction.

Feature Model	Ref.	Database & Testing protocol	Descriptor	Age estimation Algorithm	Performance		
					MAE	CS	Accuracy
Anthropometric	Kwon and Lobo 1999 [39]	**Private dataset**: 15 images for test	Snakelets	Group classification	N/A	N/A	100%
	Ueki et al. 2006 [40]	**WIT-DB** :2-fold cross-validation	(2DLDA)	Group classification	N/A	50% (M), 43%(F)	N/A
	Dehshibi and Bastanfard 2010.[41]	**IFDB** : 298 for training 200 or testing	Combined Projection Function (CPF)	Group classification	N/A	N/A	86.64%
Texture	Guo *et al. 2009* [10]	**FG-NET**: LOPO	BIF	Regression	4.77	89%	N/A
	Dib and ElSaban 2010 [42]	**FG-NET**, **MORPH**	ASM	Hybrid	3.17 FG-NET, 4.11 MORPH	N/A	N/A
	Choi *et al.* [11]	**FG-NET**, **PAL**	LBP	Group classification	4.7 FG-NET, 4.3 PAL	73% FG-NET, 70% PAL	N/A
	Zhou *et al. 2005* [43]	**FG-NET**: 5-fold cross-validation	Haar-like	Regression	5.81	N/A	N/A
	Yan *et al.2008* [44]	**YGA**:1000 for training , 3000 for testing	SPF	Regression	7.82(M), 8.53(F)	75%(M), 70%(F)	N/A
	Hajizadeh and Ebrahimnezhad 2011[45]	**IFDB**: 246 for training,131 for testing	HOG	Group classification	N/A	N/A	87.025%
	Chang and Chen 2015 [46]	**FG-NET**, **MORPH-II** 80% for training 20% for testing	Scattering transform	Ranking	4.7 FG-NET, 3.82 MORPH	N/A	N/A
Active appearance	Lanitis et al. 2002 [47]	**Private**: 500 for training, 65 for testing	AAM	Regression	4.3	N/A	N/A
	Lanitis et al. 2004 [48]	**Private**: 50% for training, 50% for testing	AAM	Classification	3.82-5.58	N/A	N/A
	Yan *et al. 2007* [49]	**FG-NET**: LOPO **YGA**: 4-fold cross validation	AAM	Ranking	5.33 FG-NET,6.95(M), YGA 6.95(F)	79% (M), 78%(F) YGA	N/A
	Chen *et al. 2013* [50]	**FG-NET**:LOPO	AAM	Regression	4.67 FG-NET, 5.88 MORPH-II	74.5%,FG-NET,57.9%	N/A

		MORPH-II:80% for training, 20% for testing				MORPH	
	Feng *et al. 2016* [51]	**FG-NET**:LOPO **MORPH-II**:10-fold cross validation **WebFace**:4/5 for training and 1/5 for testing over 10 trails	AAM	Ranking	4.35 FG-NET, 4.59 MORPH-II, 6.03 WebFace	N/A	N/A
AGES	Geng et al. 2006 [52]	**FG-NET**:LOPO	AAM	Regression	6.77	N/A	81%
	Geng et al. 2007 [53]	**FG-NET**:LOPO, **MORPH**	AAM	Classification	8.83	N/A	70%
Age Manifold	Fu and Huang 2008 [54]	**YGA** : 50% for training, 50% for testing	CEA	Quadratic Regression	5-6 years	N/A	N/A
	Guo *et al. 2008* [55]	**FG-NET**:LOPO **YGA** : 4-fold cross validation	OLPP	Regression	5.07FG-NET ,5.30(M), YGA 5.25(F)	83% (M), 81%(F) YGA	N/A
	Yan *et al. 2009* [56]	**FG-NET**:LOPO	SSE	Regression	5.21	N/A	N/A
	Cai *et al.* 2016[57].	**FG-NET**:LOPO **MORPH-II**:80% for training, 20% for testing	Dual Histogram LBP	Gaussian Process Regression	4.64 FG-NET, 4.66 MORPH	72.1%,FG-NET,62.2% MORPH	N/A
Multi-Feature Fusion	Huerta *et al.2015* [58]	**MORPH,FRGC**: 4-fold cross validation	HOG,LBP, SURF	Classification	4.25 MORPH, 4.17 FRGC	N/A	76.24%FRGC, 71.17% MORPH
	Liu *et al 2015*[59]	**FG-NET**:LOPO **MORPH-II**:divided into 3 sets: 2 for training , 1 for testing	LBP,HOG, BIF	Hybrid	2.81 FG-NET, 2.97 MORPH-II	N/A	N/A
	Pontes *et al. 2016* [60]	**FG-NET**:LOPO **MORPH-II**:75% for training, 25% for testing	AAM,LBP, LPQ,GW	Hybrid	4.50 FG-NET, 5.86 MORPH-II	N/A	N/A

Table 2.1: Summary of handcraft-based method

2.3.1 Training from scratch

Training the network from scratch on large labeled data to allow the feature learning can be very practical when having a large number of classes as output. The training from scratch approach is not very popular since in most cases it requires very large amounts of data and the training takes very long times to complete. Yang *et al.* [63] took the initiative to estimate the age using CNN. They created the deep network and trained it from scratch. They implemented a recognition system to recognize the face from videos and images then extract the corresponding human profile. Using regression-based age estimation, they reported their system's performance at a MAE of 4.88 years on the FG-NET database.

On another study, Xing *et al.* [21] implemented a multi-task deep model to estimate the age based on race and gender classification using VGG-16 CNNs and a hybrid approach. Their system was trained from scratch and tested on the MORPH-II and Webface datasets. They reported a MAE of 2.96 years on MORPH-II and 5.75 years on Webface.

Using a regression-based system, Wan *et al.* [22] implemented five cascaded deep VGG-16 CNN networks that were all trained from scratch to extract race and gender related features and use them for the enhancement of the aging features. They held experiments on the MORPH-II and CACD datasets. They reported a MAE of 2.93 years on MORPH-II and 5.22 years on the CACD database.

Chen *et al.* [64] implemented a system based on a ranking CNNs network that was trained via the age's ordinal information as labels. They conducted multiple tests on the MORPH-II and FG-NET datasets. They reported a MAE of 2.96 years on MORPH-II and 4.93 years on FG-NET.

2.3.2 Transfer learning

The concept of transfer learning revolves around the idea of tuning already trained models by transferring the knowledge from a pretrained network and then train the network on an alternative or related task. It requires less amount of data as well as less computation time. This approach is used very frequently in age estimation systems.

Using a large and labeled facial dataset for the process of estimating the accurate age or classifying the images into age groups is mandatory. Even in the case of unlabeled images which are present in some databases, the age difference between a couple of images of the same person can still be easily extracted. Hu *et al.* [65] proposed a learning classification-based system that relies on the age difference. In order to benefit from the weakly labeled data, they employed GoogLeNet CNNs along with Kullback-Leibler diverge to locate the age difference information between image pair. They conducted their experiments on the MORPH-II and FG-NET datasets. They reported a MAE of 2.78 years on MORPH-II and 2.8 years on FG-NET.

In order to reduce the age estimation error on validation data to a minimum, it is important to learn models from training data. Also, the age estimators must work on 3 aspects that involve the extracted features and feature representation. The age estimation process assumes that the data being used for training, testing, and validating have the same characteristic according to the distribution and the conditions where the images are captured.

The age estimation system should be trained, fine-tuned or validated under unconstrained environment for it to be more reliable. Rothe *et al.* [23] enhanced their system's performance by fine-tuning the VGG-16 CNNs over unconstrained IMDB-WIKI database and then tested the regression-based system on the MORPH-II, FG-NET and CACD databases. They reported a MAE of 2.68 years on MORPH-II, 3.09 years on FG-NET and 6.521 on the CACD dataset.

Li *et al.* [66] proposed a novel method that uses AlexNet for feature extraction with cumulative hidden layer which has the main benefit of overcoming the issue of sample imbalance by learning indirectly via the faces with neighboring ages. An extra layer was added as an improvement to their regression-based age estimation system in order to make a comparative ranking with the objective of facilitating the learning of aging features and thus enhancing the overall performance. Through their experiments they reported a MAE of 3.06 years on MORPH-II dataset and 6.04 years on Webface dataset.

A multi-path CNN is proposed by Liu *et al.* [67] to extract the features at different scales. To obtain the joint fine-tuning for the networks, the objective loss function was deployed on the top of the multi-path CNNs. The system was made of a VGG-16 Face Net that was pretrained with two shallow CNNs for fine-grain feature extraction. A normalization of the output feature vector to a single vector was carried out then it was used for age estimation. The ordinal ages were splitting into discrete groups, to deeply learn the feature transformations across these groups. When testing their system, they reported a MAE of 3.25 years on MORPH-II and 3.93 years on FG-NET dataset.

On another research, the Residual networks of Residual networks (RoR) models were combined by Zhang *et al.* [68], they employed RoRmodel that was already pre-trained on ImageNet, they then did the fine-tuning to extract more features using the IMDB-WIKI-101 dataset and tested their system on the Adience database, Zhang *et al.* [69] also implemented the DEX technique [51] for exact age estimation on another study. They improved their age estimation system by extracting the fine-grained features using the attention mechanism. They managed to achieve a MAE of 2.36 years on MORPH-II and 2.39 years on FG-NET.

Liu et al. (2020) [70], developed a lightweight CNN network (ShuffeNetV2), based on the mixed attention mechanism (MA-SFV2) that transforms the output layer, that shapes the age estimation as a classification problem (that classify age as a separate label), regression problem (that rank the age of the human face having a particular order) and distribution learning (that consider the age correlation between adjacent ages). The model includes image pre-processing that reduces the effect of noise vectors and a data augmentation method like filtering, sharpening, histogram enhancement, etc., that increase the image size and alleviate the overfitting of the network, it combines classification, regression and distributed learnings algorithms for the age estimation task. The achieved experimental results are a MAE of 2.68 years on MORPH-II along with 3.81 years on the FG-NET dataset.

2.3.3 Feature extraction

Another approach which is less commonly used that we adopted in our own work is the employment of deep learning networks for feature extraction. After the feature extraction phase the obtained feature vector is then loaded into the classifiers/regressors for the age estimation phase.

Wang *et al.* [71] used this approach for extracting facial age-related feature where instead of obtaining feature maps at the top layer, they were acquired in different layers. They also implemented a manifold learning algorithm for an improved system performance. Many SVM for classification and SVR for regression structures were evaluated using a single layer Deep Learned Aging pattern (DLA). They conducted their experiments on the MORPH-II and FG-NET datasets. They reported a MAE of 4.77 years on MORPH-II and 4.26 years on FG-NET.

Another model proposed by Duan *et al.* [72] utilized a deep learning network for feature extraction and then fed the output to Extreme Machine Learning (ELM) for age group classification then to regress the final value of the age via ELM regressor. The final obtained result was a MAE of 2.61 years on the MORPH-II database.

2.3.4 Summary

The different deep learning methods were evaluated using different datasets and using different architecture of CNNs. From these age estimation algorithms, many used the regression to estimate the real age and achieved good results. While some others opted for a classification to estimate the final age and improved the performance of the age estimation system [65]. The advantage of combining the regression with classification in one hybrid model such as in [21] can be clearly seen. This signifies that the enhancement of the performance does not depend on one factor but rather a mixture between several ones to obtain a good and robust system. The different discussed deep learning-based techniques and methods are summarized in Table 2.2.

2.4 Conclusion

In this chapter, we went through some of the most important papers in age estimation and discussed handcraft-based methods and deep learning-based methods. Handcraft-based methods have the advantage of needing a small database and a model of low complexity to display decent performance. However, multiple key features and information may be lost to of lack knowledge in extracting the most relevant aging features. The deep learning technology gained a lot of the researchers' attention in the age estimation field. A main factor that has a straight effect on the performance of deep network is related to whether the model is pre-trained or has been trained from scratch for the age estimation task on large databases. A good idea that adds a decent performance boost to the age estimation system is the usage of model that was pre-trained on a field related to age estimation such as training on face recognition. Nevertheless, deep learning methods require massive data amounts and resources to achieve promising results. Unbalanced and limited datasets may have a negative effect on the performance of an age estimation system. Therefore, data augmentation could be a practical solution. Furthermore, the system performance can be further enhanced by usage of data fusion of different biometrics with age to obtain more useful and robust systems. In the next chapter we present our proposed age estimation system and the different methods we used in its implementation, we also display our achieved age estimation results and compare them with the performance of the reviewed state-of-the-art studies.

Deep Learning	Ref.	Database & Testing protocol	Model	Age Estimation Algorithm	MAE
Training from Scratch	Yang et al. 2011. [63]	**FG-NET** : LOPO	CNNs with correspondence driven adaptation	Regression	4.88
	Xing et al. 2017. [21]	**MORPH-II**: divided into 3 sets, 2 for training and 1 for testing. **WebFace**	Multi-task and fusion using VGG-16 CNNs	Hybrid	**2.96** Morph-II **5.75** WebFace
	Wan et al. 2018. [22]	**MORPH-II**: divided into 3 sets, 2 for training and 1 for testing. **CACD**	Data Fusion using Five cascaded VGG-16 CNNs	Regression	**2.93, 3.30** MORPH-II, 5.22 CACD
	Chen et al. 2018. [64]	**MORPH-II, FG-NET**: 80% for training, 20% for testing	Basic CNNs for each age group using Ranking-CNNs	Group Classification	**2.96** MORPH_II **4.13** FG-NET
Transfer Learning	Hu et al. 2016. [65]	**FG-NET**: LOPO **MORPH-II**: divided into 3 sets, 2 for training and 1 for testing.	Age Difference Using GoogLeNet CNN	Soft classification	**2.78** MORPH-II **2.8** FG-NET
	Rothe et al. 2016. [23]	**MORPH-II**: 80% for training, 20% for testing **FG-NET**: LOPO **CACD**: 145,275 images for training, 7600 for testing	VGG-16 CNNs with DEX method	Regression	**2.68** MORPH-II **3.09** FG-NET **6.521** CACD
	Li et al. 2017. [66]	**MORPH-II**: divided into 3 sets, 2 for training and 1 for testing. **WebFace**	AlexNet CNNs with D2C	Regression	**3.06** MORPH-II **6.04** WebFace
	Liu et al. 2016. [67]	**FG-NET**: LOPO **MORPH-II**: divided into 3 sets, 2 for training and 1 for testing	Group-Aware + OHranker using VGG16 CNN + two shallow CNNs	Ranking	**3.93** FG-NET **3.25** MORPH-II
	Zhang et al. 2018. [68]	**MORPH-II**:80% for training, 20% for testing **FG-NET** : LOPO	Fine-Grained with attention mechanism using ResNets + RoR model	Regression	**2.36 MORPH**-II **2.39** FG-NET
	Liu et al. (2020). [70]	**MORPH-II**:80% for training, 20% for testing **FG-NET**	ShuffleNetV2	Hybrid	**2.68** MORPH-II **3.81** FG-NET
Fearture Extraction	Wang et al. 2015. [71]	**MORPH-II**: 80% for training, 20% for testing **FG-NET** : LOPO	DLA using Single layer CNN+RNN	Regression	**4.77 MORPH**-II **4.26** FG-NET
	Duan et al. 2018. [72]	**MORPH-II**: 35440 images for training, 8860 for validation,8860 for testing.	Data Fusion+ELM using ThreeCNNs (AgeNet,GenderNet,Race-Net)	Hybrid	**2.61**

Table 2.2: Summary of deep learning-based methods

Chapter 3.

PROPOSED METHODS AND EXPERIMENTAL RESULTS

Chapter 3

PROPOSED METHODS AND EXPERIMENTAL RESULTS

3.1 Introduction

Age estimation system is considered as a pattern recognition system which includes several linked steps: image preprocessing, feature extraction step and classification step. One of the most critical problems in pattern recognition is the selection of suitable characteristics for a compact and optimal representation of the image. For that, many approaches for feature extraction process have been suggested in the literature. In general, the proposed techniques can be divided into three different categories: Hand-craft techniques (e.g., LBP, DCT, HOG…), Learned hand-craft techniques (e.g., CA-LBFL, CBFD…), and deep learning techniques (e.g., PCANet, CNN…).

In this work, we have used three feature extraction methods including a Hand-craft technique (LBP), a learned hand-craft technique (CBFD), and a deep learning technique (DCTNet). The main objective of this study is to obtain the feature vectors that can effectively represent the discriminating characteristics of the different faces. MORPH II database is then used to assess the performance of the proposed age estimation schemes. Finally, we analyze the results and compare them to multiple state of the art age estimation researches.

3.2 Methodology and proposed methods

The general block diagram of our proposed age estimation methods, presented in Fig. 3.1, is composed of three steps. The image preprocessing step which comprises face detection and image cropping. The second step consists to extract a set of features for age prediction using LBP, CBFD, and DCTNet in each proposed method, respectively. Finally, a classification step using MLP and SVM classifiers is used for testing the constructed features.

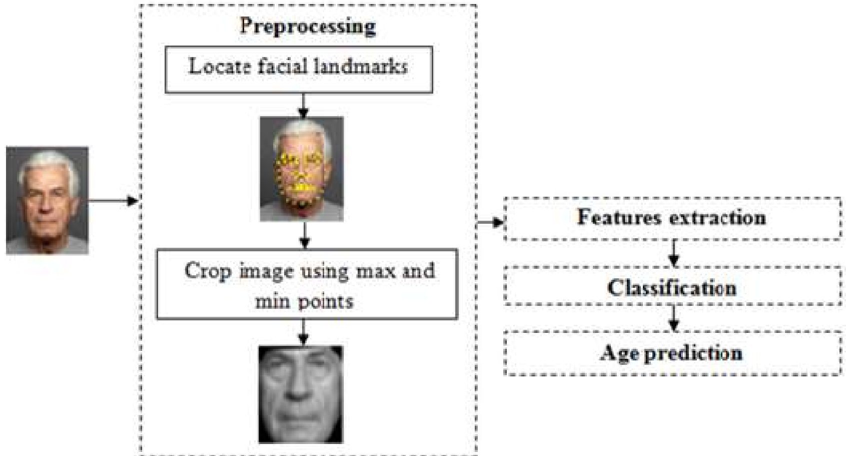

Figure 3.1: The general block diagram of the proposed Age estimation methods

3.2.1. Preprocessing

Information contained in the image can be used to determine 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, a preprocessing step is applied to prepare our test images for training as it greatly boosts our model's accuracy by enhancing image information. In our proposed age estimation methods, the preprocessing step involves two steps including face detection, cropping and resizing.

3.2.1.1. Face detection

Face detection is a fundamental step in main face related applications. Landmarks detection algorithms are used to detect and localize key points on a face. In this step, we aim at accurately localizing the facial region to extract features only from the relevant parts of the input image. We have used in our proposed methods a 68-landmark shape detector [73] for automatic localization of facial landmark points as shown in figure 3.2. The purpose of this analysis is to determine the locations of the most important aging features using the whole face area.

Figure 3.2: The 68 landmarks of sample images

3.2.1.2. Cropping and resizing

The 68 landmarks algorithm is used to capture the signatures of the aging process: corners of the eyes, mouth and nose, cheeks, chin, top of the nose and forehead, totaling eleven skin areas that are highly correlated with aging. In our study, to accurately localizing the facial region, two landmarks were used from the top of the eyebrows and the cheeks as reference for the width and the height of the rectangle. It was preferable to define some rectangles slightly larger and with some areas overlapping others in order to cover a wider facial region. Once the facial region is defined, the input image is cropped to the area covered by the rectangle landmark points as shown in figure 3.3.

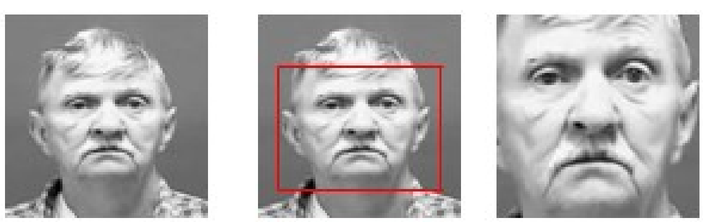

Figure 3.3: Cropping and resizing process

3.2.2. Feature Extraction

Feature extraction is a crucial module in any pattern recognition application since good classification results are directly related to the uniqueness and variability of the extracted features used to distinguish between different patterns. In our proposed methods, the LBP handcraft technique, learned handcraft technique CBFD, and DCTNet deep learning are used in each proposed method to extract the feature vector.

3.2.2.1. Local Binary Pattern (LBP)

The Local Binary Patterns which was presented by Ojala et al. [74] in 2002 is an efficient and powerful texture descriptor that is widely used in image processing to remove lighting effects (see Fig. 3.4), and computer vision areas as a feature and histogram representation (see figure 3.5).

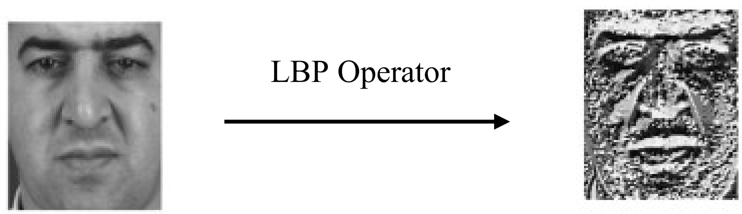

Figure 3.4: The original image (left) processed by the LBP operator (right)

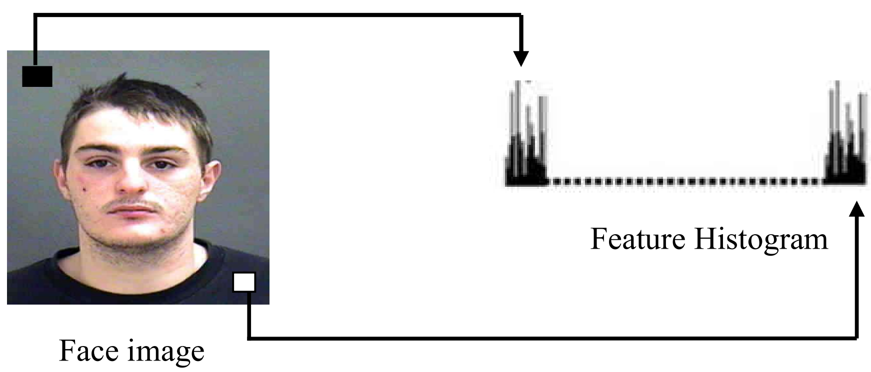

Figure 3.5: LBP based face description.

The main LBP mechanism is that the input image is divided into ‘N×N’ local regions, each local region is composed of a 3×3 neighborhood of each pixel by the central pixel’s value. Then, type of binary pattern will be assigned a label to each pixel according to its intensity value, where the distribution of these binary patterns in each block represents the results with an 8-bits integer, where the calculation of these patterns are used as a feature representation.

In general, by assuming a pixel at(Xc, Yc), the LBP results can be calculated as the following formula:

$$LBP(X_c, Y_c) = \sum_{n=0}^{N-1} s(i_n - i_c)\, 2^n \qquad (3.1)$$

Where i_n is a neighbor pixel of the N pixels around the central pixel i_c.

To explain the LBP mechanism (see Fig. 3.6), we assume that the input image has been divided by LBP algorithm into 100 blocks, and it is now processing the block number 29. This block is divided into '3×3' neighborhood pixels (9 cells), and then encoded each pixel by using its intensity value. By threshold from the value of the central pixel (which is 65 in above example), LBP ordering all surrounding pixels within the block from the upper-left corner down to the right one depending on whether have bigger or smaller intensity value than the central pixel (bigger value = 1; smaller = 0). Finally, we will get an 8-bit binary number (it is "11000011" in above example), which can be converted to a decimal number later, e.g. (11000011)2 = (195)10, where the calculation of these numbers represents a one-dimensional array of patterns.

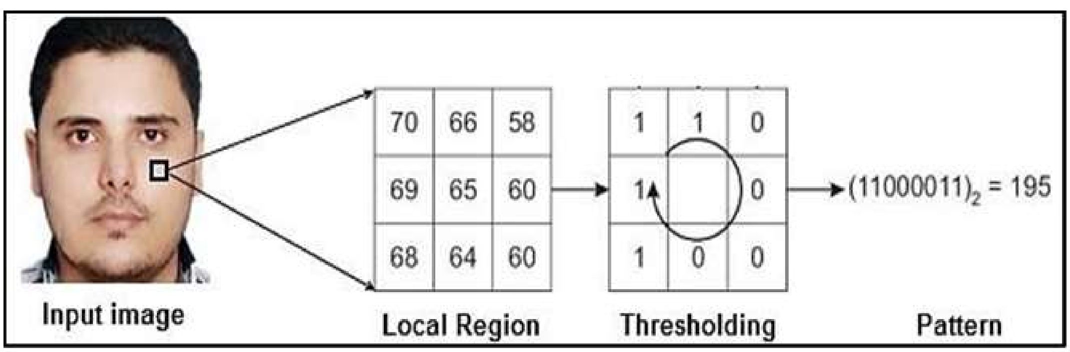

Figure 3.6. LBP Mechanism

3.2.2.2. CBFD image feature learning

The growing interest in the use of images in machine vision applications, several methods are being used to analyze and therefore understand the content of images. One of the most effective methods is the LBP descriptor. So, given the two advantage of the LBP based texture image analysis method (speed and low memory occupation), Jiwen Lu et al. in [75] propose an efficient learned hand-craft feature extraction method called the Compact Binary Face Descriptor (CBFD). In this technique, authors tried to improve the performances of by including a learning phase to become rather intelligent and thus to enhance his strength and eliminate its weaknesses. The training phase and image feature extraction are summarized as follows:

A. Training phase

This method relies on the robustness of the binary codes versus the local changes of image texture. The compact binary codes are learned directly from raw pixels for the image representation. Indeed, to ensure better classification results, CBFD features must be constructed from a set of image samples provided from the same context. To better understand the training phase (construction of the projection matrix (CBFD feature) and the codebook, Fig. 3.7 illustrates the details of the training phase flowchart for MORPH-II images.

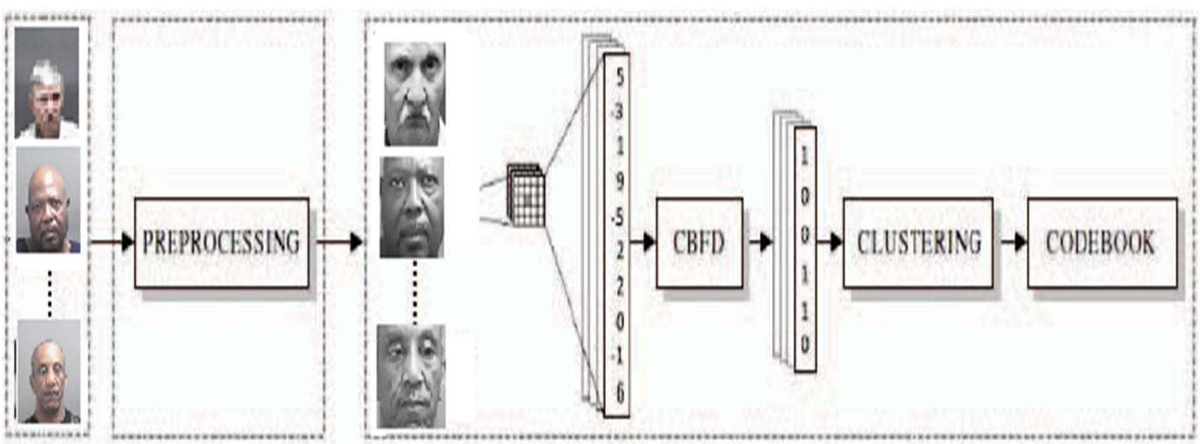

Figure 3.7: Learning the handcrafted codebook for feature encoding.

1) *CBFD feature learning*: The training vectors are obtained by using the relationship of each pixel with their neighborhoods. Indeed, the image is analyzed by a rectangular window of size $(2R + 1) \times (2R + 1)$, where R is a positive integer, centered on each pixel.

Let $X = [x_1, x_2, \cdots, x_n]$ be the set of training vectors (called Pixel Difference Vector (PDV)) obtained by measuring the difference between the central pixel and the neighboring pixels within the predefined window. The size of each vector is equal to $(2R + 1) \cdot (2R + 1) - 1$ or the

PDV between the central pixel and itself ($PDV_0 = 0$) is excluded. The CBFD feature (projection matrix) aims to learn K hash functions ($\{w_k\}_{k=1 \dots K}$) in order to quantize each vector $x_n | n = 1 \dots N$ into a binary vector $b_n = [b_{n1}, \cdots, b_{nK}]^T$, as follow:

$$b_{kn} = 0.5 \times (sgn(w_k^T x_n) + 1) \qquad (3.2)$$

Where sgn (v) equals to 1 if v ≥ 0 and -1 otherwise.
So first we have to build the projection matrix W which includes all the hash functions w_k:

$$W = [w_1, w_2, \dots, w_K] \qquad (3.3)$$

The W matrix is initialized with the K first eigenvectors of the covariance matrix ($C = XX^T$). Then, an optimization task is applied to minimize the objective function $J(w_k)$, which is defined as follows:

$$\min J(w_k) = J_1(w_k) + \lambda_1 J_2(w_k) + \lambda_2 J_3(w_k) \qquad (3.4)$$

The two predefined parameters λ_1 and λ_2 are chosen to balance the effects of different terms. The terms J_1,J_2, and J_3 are chosen for ensuring that: i) the variance of the learned binary codes are maximized, ii) the quantization loss between the original feature and the encoded binary codes is minimized and iii) feature bins in the learned binary codes evenly distribute as much as possible (for more details, see [75]). After the convergence of the optimization algorithm, a projection matrix W (CBFD features) is obtained containing all the hash functions and making it possible to transform a PDV vector into a binary vector of size K bits.

2) *Codebook learning:* The purpose of the codebook is to reduce the number of binary vectors found for each image. Indeed, all the training vectors (PDVs) are projected in the matrix W, and then the k-means clustering algorithm is applied to all the binary vectors obtained. Finally, the centroid of the classes obtained represents the codebook.

B. MORPH-II feature extraction

The feature extraction task (see Fig. 3.8) is based on the projection matrix W (CBFD feature) and the codebook obtained during the training phase.

However, after the formulation of all the PDV vectors ($X = \{x_n\} n = 1 \cdot\cdot N$) of the image, their binary correspondences are achieved through their projection in the matrix W:

$$V_b = 0.5 \times (sgn(W^T X) + 1) \qquad (3.5)$$

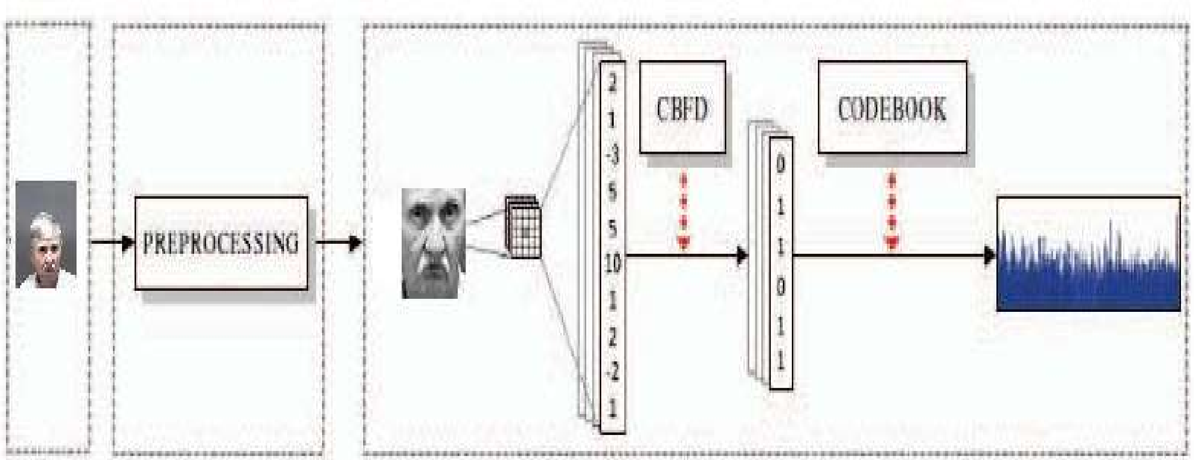

Figure 3.8: Learned LBP based feature extraction task.

Then, each binary vector is replaced by the closest vector coordinate in the codebook (bin). Finally, using the different coordinates, a histogram is formed to represent the entire image feature. To find discriminative feature vectors, the raw image is segmented into several regions, and each region is analyzed as an image (each region has its own CBFD features (W) and its codebook).

For each region, a histogram (H_s) is created, and finally, the concatenation of all the histograms makes it possible to obtain a precise vector to represent the whole of the image.

$$V = [H_1, H_2, \cdot\cdot\cdot, H_M] \qquad (3.6)$$

Where, M denotes the number of regions. In our experimental results, we try to choose the number of regions that gives the best accuracy of the biometric system.

3.2.2.3. DCTNet framework

DCTNet [76] is a simple deep learning baseline for image classification which comprises only the very basic data processing components: (1) cascaded Discrete Cosine Transform (DCT), (2) binary hashing, and (3) histograms. Thus, the block diagram of DCTNet algorithm presented in Figure 3.9 can be summarized as follows:

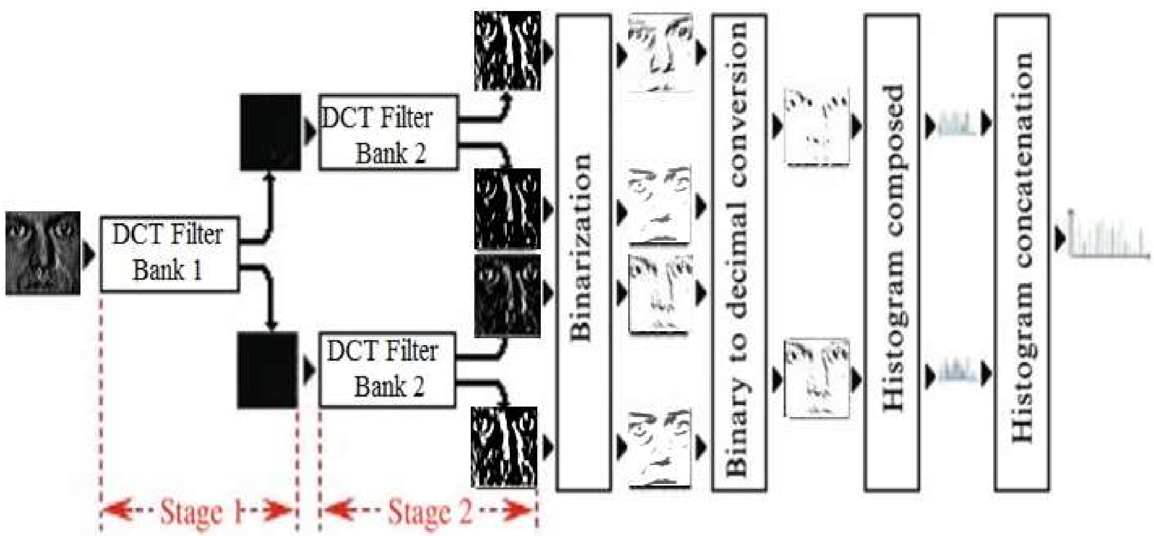

Figure 3.9: Block diagram of the DCTNet.

(1) *DCT Filter bank:* As shown in Fig. 3.9, the DCT filter bank comprises two stages of filter bank convolutions. In the first stage, the filter banks are estimated by performing Discrete Cosine Transform (DCT) over a set of vectors where each vector represents the $k_1 \times k_2$ points around each pixel. For each vector we take the mean of the entries and then subtract the mean from each entry of it. Then, we perform DCT over these vectors and retain the principal components W size of $(k_1 . k_2 \times L_{S1})$ where L_{S1} is the primary eigenvectors.

Next, each discrete cosine transform (column of W) is a filter and may be converted to $k_1 \times k_2$ kernel which is convolved with the input image. So,

$$I_\ell(x,y) = H_\ell(x,y) * I(x,y), \text{ where } 1 \le \ell \le L_{S1} \qquad (3.7)$$

Where * denotes the discrete convolution and $I^{\prime}$ is the resulting filtered image using the H_ℓ filter. So, using the L_{S1} columns of W we take each input image I and convert it into L_{S1} output images.

The second stage is constructed by iterating the algorithm from the first stage (Filter bank convulsions) over each of the output images. For each output image I_ℓ we take the vector (points around each pixel), the mean of the entries and then subtract the mean from each entry of the vector is computed. The vectors produced are then concatenated together and we estimate another DCT filter bank (with L_{S1} filters). Finally, each obtained filter is convolved with I_ℓ to produce a new image.

Hence, using all output images of the first stage and by repeating convulsion process for each filter, $L_{S1}.L_{S2}$ output images can be produced.

(2) *Binary hashing*: In this step, the $L_{S1}.L_{S2}$ output images obtained in second layer are converted into binary format by using a Heaviside step function who's their value is one for positive entries and zero otherwise.

$$I^B_{\ell,m}(i,j) = \begin{cases} 1 & if\ I_{\ell,m}(i,j) \ge 0 \\ 0 & otherwise \end{cases} \qquad (3.8)$$

where $I^B_{\ell,m}$ is a binary image. In addition, around each pixel, we view the vector of L_{S2} binary bits as a decimal number. This converts the L_{S2} outputs into a single integer-valued (image).

$$I^D_\ell(i,j) = \sum_{m=1}^{L_{S2}} 2^{m-1}\, I^B_{\ell,m}(i,j) \qquad (3.9)$$

(3) *Histogram composition:* Each hashed image (I_ℓ^D) is partitioned into N_B blocks and a histogram of the decimal values in each block (B) is computed. These blocks can be non-overlapping (disjoint) or they can be overlapping (depending on the application). Thus, the feature of I_ℓ^D is obtained by concatenating all B histograms such as:

$$\mathcal{V}_\ell^{hist} = [B_1^{hist}, B_2^{hist}, \ldots, B_{N_B}^{hist}] \qquad (3.10)$$

Finally, after this encoding process, the feature vector of the input image I is then defined as:

$$\mathcal{V}_\ell^{hist} = [\mathcal{V}_1^{hist}, \mathcal{V}_2^{his}\ , \ldots, \mathcal{V}_{L_{S1}}^{hist}] \qquad (3.11)$$

In conclusion, the hyper-parameters of the DCTNet include the filter size$(k_1,\ k_2)$, the number of filters in each stage L_{Si}, the number of stages (Ns), and the block size for local histograms in the output layer (B).

3.2.3. Classification

Age estimation can be treated as a classification problem, when each age is considered as a class label. In our experiments, we have used both SVM and MLP classifiers for age estimation on the MORPH-II database.

3.2.3.1. *Support Vector Machine (SVM)*

SVM was developed by Cortes and Vapnik in 1995 [24] and has extensively been used as a popular and powerful supervised learning tool for general pattern recognition applications. In addition, it gives promising and excellent performance on the range of machine learning and many other fields by applying it to different classification problems like our age estimation, data separation, regression, and density estimation.

SVM classifier has many advantages, which gives high performance even with small number of images in training set, not sensitive to the number of dimensions, which gives it promising performance with any images size, and the ability to minimize both empirical and structural risk, which leads to better generalization for data classification even when the number of train set is high. SVM has been chosen since it proven advantages in handling large scale classification tasks with good generalization performance. Additionally, it has demonstrated superior results in various classification and pattern recognition problems.

3.2.3.2. Multilayer perceptron (MLP)

Multilayer perceptron is a type of Artificial Neural Network (ANN) that belongs to the class of supervised neural classifiers. ANN is inspired by the function of the human brain. The MLP consists of perceptron that are organized in layers: an input layer, one or more hidden layers, and the output layer. Every perceptron in a particular layer is usually connected to every perceptron in the layer above and below. These connections carry weights. Each perceptron calculates the sum of the weighted inputs, and feeds it into its activation function, regularly a sigmoid one. The result is then passed on to the next layer. The output layer has, for example, the same number of perceptron as there are classes, and the perceptron with the highest activation will be consider the classification of the input sample. Training is achieved by successively feeding all training samples into the network, and comparing the output with the true class label.

Furthermore, MLP has also been chosen as second classifier to provide a large generalization about classification techniques especially with a big number of training samples and a high number of input variables.

3.3. Experiments and discussions

3.3.1. Image Database

MORPH database is one of the largest publicly available longitudinal face databases [32]. Multiple versions of MORPH have been released, but for our study, we use the 2008 MORPH-II non-commercial release. MORPH-II dataset includes 55,134 mugshots with longitudinal spans taken between 2003 and late 2007. For each image, the following metadata is included: subject ID number, picture number, date of birth, date of arrest, race, gender, age, time since last arrest, and image filename (see Table 3.1).

	Black	White	Asian	Hispanic	Other	Total
Male	36832	7961	141	1667	44	46645
Female	5757	2598	13	102	19	8489
Total	42589	10559	154	1769	63	55134

Table 3.1: Number of images by gender and race.

Because of its size, longitudinal span, and inclusion of relevant metadata, MORPH-II dataset is widely utilized in the field of computer vision and pattern recognition, including a variety of race, gender, and age face imaging tasks.

3.3.2. Age estimation evaluation metrics

The performance evaluation of age estimation system can be described by two essential metrics, i.e., Mean Absolute Error (MAE) and Cumulative Score (CS). These metrics are the most common and most used for age estimation evaluation.

3.3.2.1. Mean Absolute Error (MAE)

MAE is used in many image processing applications and is considered a reference model for evaluating the quality of the processed image. The MAE is defined as follows:

$$MAE = \frac{1}{n_s}\sum_{i=0}^{n_{s-1}}|e_i| \qquad (3.12)$$

where n_s is the number of samples and e_i is the error of the i^{th} instance, i.e. $e_i = |\hat{y}_i - y_i|$, where y_i is the real label and $\hat{y}_i$is the predicted label. This metric tells the average number of years that the prediction is wrong.

3.3.2.2. Cumulative Score (CS)

The Cumulative Score (CS) is defined as the percentage of test images such that the absolute error is not higher than a threshold t (in years).

$$CS(t) = (1 - \frac{1}{n_s}\sum_{i=0}^{n_{s-1}} h(|\hat{y}_i - y_i| - t)).100 \qquad (3.13)$$

$$h(x) = \begin{cases} 1, & if\ x \geq 0 \\ 0, & otherwise \end{cases} \qquad (3.14)$$

where y_i is the age label of the i^{th} test image and $\hat{y}_i$ is the age prediction of the *ith* test image.

3.4. Proposed age estimation methods evaluation

This section is concerned with evaluating the methods' performance in terms of MAE and CS metrics; we present our experimentations and analysis of our obtained results. We divide this section into three sub-sections for our trio of used feature extraction methods. Each proposed scheme was tested with two classifiers of choice (SVM and MLP).
We note that, the SVM and MLP classifiers are trained on 43998 images from the MORPH-II database using the holdout technique (80:20).

3.4.1. LBP based age estimation

The proposed age estimation method, that uses LBP features, results are in general acceptable. The MAE is reported at 6.44 along with a cumulative score showing that 56.67% of estimations have less than 5 year deviation from true age. In contrast, the MLP classifier achieved an MAE of 13.97 along with a cumulative score of 30.38% which is fairly poor results compared to the SVM classifier.
Table 3.2 summarizes the MAE and the CS obtained by the LBP features extractor using SVM and MLP classifiers (highest accuracy is highlighted in bold).

Method	Dataset	Training images	Testing images	MAE	CS (5)
LBP + SVM	MORPH II	43998	11000	**6.44**	**56.67**
LBP+ MLP	MORPH II	43998	11000	13.97	30.38

Table 3.2: LBP based age estimation results

3.4.2 CBFD based age estimation

The Compact binary facial descriptor feature learning technique has a set of multiple parameters preset to meet our specific needs. However, some parameters need some testing and tuning which is why conducted a series of experiments to find the best parameters of CBFD which help our age estimation system perform the best possible with this feature extraction method. The different parameters of CBFD algorithm are the number of vertical regions, horizontal regions, window size, dictionary size and projection matrix size. In our experiments, we have considered only the number of vertical regions, horizontal regions and the window size. To have better parameters (optimal parameters), we tested the CBFD algorithm with different horizontal and vertical regions sizes, for example 1×1, 2 ×4, 3×4,…, with different windows size and with the two classifiers MPL and SVM. The goal of the variation in the size these parameters is to have better results as a function of training time, recognition time, feature vector size and the error rate.

Windows Size (R)	Number of Vertical & Horizontal regions (N,M)	MPL		SVM	
		MAE	CS (5)	MAE	CS (5)
3	(1,1)	10.72	23.89	11.72	19.22
3	(2,4)	9.01	43.05	8.40	45.05
3	(3,3)	8.65	46.53	8.45	45.65
5	(3,3)	5.99	63.02	6.13	60.86
7	(2,2)	5.13	68.34	**4.91**	**70.05**
7	(2,4)	7.66	52.45	7.36	51.83
7	(3,4)	5.63	63.27	5.93	60.86
7	(3,3)	5.45	64.12	5.99	61.55

Table 3.3: CBFD Age estimation results using MLP and SVM classifiers

Table 3.3 shows that the age estimation based on CBFD achieved an age prediction MAE of 4.91 with a cumulative score of 70.05% with a number of vertical and horizontal region of (2, 2) and windows size equal to 7 when using SVM classifier. The estimations have less than 5 year deviation from true age, this represents the best achieved score with this method which is slightly better than the results obtained by the MLP classifier at a 5.13 MAE along with a cumulative score of 68.34%.

3.4.3 DCT-NET based Age estimation

To evaluate the efficiency of our proposed method, a series of experiments were carried out in order to select the best parameters of DCTNet. Their objective is to study the impact of each parameter (number of filters, filter size, and block overlap ratio) on the age estimation accuracy. In our experiments, the DCTNet is tested with two and three layers at each time. For DCTNet with two layers, we have used multiple numbers of filters from (2, 2) to (4, 6) along with filter sizes ranging from (3, 3) up to (19, 19). Contrary, for the DCTNet with three layers, we have concluded that the performance of the system, in terms of MAE, gets gradually better the higher when the number of filters reach (4, 4). Thus, the number of filter was fixed to (4, 4, 4) and we have used different filter sizes ranging from (3, 3, 3) up to (19, 19, 19).

Number of Convolution Layers	Number of Filters	Filter Sizes	MLP		SVM	
			MAE	CS (5)	MAE	CS (5)
2	(2, 2)	(3, 5)	6.00	65.46	6.16	63.56
2	(2, 4)	(17, 19)	7.36	54.74	7.25	53.92
2	(4, 4)	(3, 5)	4.67	75.31	5.40	64.03
2	(4, 4)	(3, 3)	5.62	59.44	5.28	65.89
2	(4, 4)	(17, 19)	5.61	67.80	5.34	62.25
2	(4, 6)	(3, 5)	4.83	72.19	4.74	69.25
3	(4, 4, 4)	(3, 5, 7)	4.18	76.65	4.29	71.63
3	(4, 4, 4)	(5, 5, 5)	4.61	72.98	4.60	70.09
3	(4, 4, 4)	(19, 19, 19)	5.57	65.06	5.24	63.98
3	**(4, 4, 4)**	**(9, 9, 9)**	**3.84**	**81.37**	4.04	78.89

Table 3.4: DCTNet Age estimation results using MLP and SVM classifiers

From table 3.4 we can learn that the DCTNet deep feature extraction method with three layers and especially with a number of filters equal to (4, 4, 4) gives an age prediction MAE of 3.84 along with a cumulative score showing that 81.37% of estimations have less than 5 year deviation from true age. This score is slightly better than the results obtained by the SVM classifier at a 4.04 MAE along with a cumulative score of 78.89%. This result represents the best achieved score out of all our experiments with all our proposed methods for age estimation.

3.5. Comparison with state-of-the-art age estimation

We have quantitatively compared the performance of our proposed age estimation methods based on LBP, CBFD, and DCTNet feature extraction techniques against the recent state-of the-art approaches. To enable fair comparison, we have selected approaches that used the same image datasets MORPH-II and with the same evaluation protocol.

First, the proposed age estimation method based on ***handcraft feature*** (LBP), we can learn from table 3.5 that the proposed method achieve an acceptable MAE equal to **6.44** and which is competitive compared to the state-of-the art methods based on handcraft feature [51, 57, 60, 58, 50], which gives an MAE equal to **4.95**, **4.66**, **5.86**, **5.88**, **4.25**, respectively. However, our proposed method is less accurate than the approaches based deep learning feature [70, 72, 68].

On the other hand, the proposed age estimation method based on ***DCTNet deep features***, we can assert that the proposed age estimation method outperforms and improves the state-of-the art methods based on handcraft feature [51,57, 71, 50, 58, 60], which gives an MAE equal to **3.84** compared to their obtained MAE which equal to **5.88**, **4.25**, **4.77**, **5.86**, **4.59**, **4.66**, respectively. Nevertheless, our proposed method is slightly less accurate than the approaches based deep learning feature [70, 72, 68] and it is competitive with the proposed method in [66] which give an MAE equal to **3.06**.

Finally, the proposed age estimation method based on ***learning handcraft feature*** (CBFD) improves the proposed approaches proposed by Chen *et al.* [50] and Pontes *et al.* [60], which give an MAE equal to **4.91** compared to their obtained MAE which equal to **5.88** and **5.86**, respectively.

However, it is slightly less accurate than the approaches presented in [51, 57, 71, 58] but it is better in term of the length of their feature vector.

Publication	**Approach**	**MAE**
Liu et al. (2020) [70]	ShuffleNetV2	2.68
Duan et al. 2018 [72]	Data Fusion + ELM using Three CNNs (AgeNet, GenderNet, Race-Net)	2.61
Zhang et al. 2018 [68]	Fine-Grained with attention mechanism using ResNets + RoR model	2.36
Li et al. 2017 [66]	AlexNet CNNs with D2C	3.06
Feng *et al. 2016* [51]	AAM	4.59
Cai *et al*. 2016[57].	Dual Histogram LBP	4.66
Pontes *et al. 2016* [60]	AMM+LBP+LPQ+GW	5.86
Wang et al. 2015 [71]	DLA using Single layer CNN+RNN	4.77
Huerta *et al.2015* [58]	HOG+LBP+SURF	4.25
Chen *et al. 2013* [50]	AMM	5.88
	DCTNet	**3.84**
Proposed methods	**CBFD**	**4.91**
	LBP	**6.44**

Table 3.5: MAE obtained with different state-of-the-art approaches on MORPH database

3.6. Conclusion

In this chapter we have proposed three age estimation methods using different types of feature extraction techniques including the handcraft technique: LBP, the learning handcraft feature technique: CBFD, and the deep learning features: DCTNet. The main objective in this work consists on choosing the best features that give an accurate age estimation prediction in terms of MAE and CS. Each proposed scheme was tested with two classifiers of choice (SVM and MLP) trained on 43998 images from the MORPH-II database using the holdout technique (80:20). The achieved results presented in this work validate the robustness and the effectiveness of the deep learning methods over the handcraft and learning handcraft feature extraction techniques with a very high age prediction (MAE = 3.84 and CS = 81.37) achieved with DCTNet.

GENERAL CONLUSION

In our work, we inspected an image processing field problem, the estimation of human age, an extensive study of this problem has been done. However, solving this issue is no easy task since prediction methods suffer from some complications in analyzing data, and making these data generalized, thus applying it to everyone is not simple.

We began this thesis by introducing the challenges about the estimation of age from facial images, discussed the motivations that encouraged us to study this field and outlined our objectives. Moreover, we investigated the theory of age and drew the steps of aging and the factors that impact it, we mentioned the various real-world application areas of age prediction, we also presented the general structure of age estimation systems with its different stages and methods used in each one as well as the datasets that are commonly used for the age estimation task. Next, we presented an overview of several researches and advances that have been done in the age prediction realm and discussed their achieved results and chosen approach. Finally, we presented our methodology and techniques of choice that consist of three age estimation methods for feature extraction LBP, CBFD and DCTNet which we tested using the SVM and MLP classifiers.

In general, the usage of the proposed methods, databases, and convenient number of images had a positive effect on the system's performance and results. Our results show that deep learning DCTNet features achieve superior precision compared to the handcrafted LBP technique and learned Handcraft CBFD technique. Another factor to point out is that the pre-processing of the facial images is a significant factor that improves the system's accuracy. Furthermore, our investigation confirms that the correct tuning of the algorithms' parameters has a strong effect on the prediction accuracy as displayed from our CBFD and DCTNet results tables in which we presented multiple obtained results with various parameters.

When compared to other similar studies, our work displays very respectable and significant results especially when taking into consideration the limitations of time and the hardware we used for our experiments. Thus, this work could be improved in the future when more computational power is available at our disposal which will allow us to experiment with stronger parameters that could lead to the achieving of better results.

Bibliography

[1] MA Taister, SD Holliday, HIM Borman, Comments in facial aging in law enforcement investigation. Forensic Sci. Commun. 2, 1–11 (2000).

[2] PM O'Hare, AB Fleischer, RB D'Agostino, SR Feldman, MA Hinds, AA Rassette, A Mc michael, PM Williford, Tobacco smoking contributes little to facial wrinkling. J. Eur. Acad. Dermatol. Venereol. 12, 133–139 (1999).

[3] AM Alberta, K Ricanek, E Pattersonb, A review of the literature on the aging adult skull and face: implications for forensic science research and applications. Forensic Sci. Int. 172, 1–9 (2007).

[4] A.J. O'Toole, T. Price, T. Vetter, J.C. Bartlett, and V. Blanz, "3D Shape and 2D Surface Textures of Human Faces: The Role of 'Averages' in Attractiveness and Age," Image and Vision Computing, vol. 18, pp. 9-19, (1999).

[5] A. Stone, The Aging Process of the Face & Techniques of Rejuvenation, http://www.aaronstonemd.com/Facial_Aging_ Rejuvenation.shtm, (2010).

[6] P. Viola and M. J. Jones, "Robust real-time face detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, (2004).

[7] A. Meraoumia. "Modèle de Markov cachée appliquée à la multi-biométrie ", Thèse de Doctorat (USTHB), 2014.

[8] I.Goodfellow, Y.Bengio, A.Courville. "Deep Learning ", Book at: http://www.deeplearningbook.org

[9] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 1, pp. 886893). IEEE.(2005).

[10] G. Guo, G. Mu, Y. Fu, and T. S. Huang, ''Human age estimation using bioinspired features,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 112–119.(2009).

[11] 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).

[12] J. Kittler, A. Hilton, M. Hamouz, and J. Illingworth. 3d assisted face recognition: a survey of 3d imaging, modelling and recognition approachest. In ECCV, pages 469–481, (2004).

[13] S. U. Hussain, T. Napoleon, F. Jurie, et al. Face recognition using ´ local quantized patterns. In BMVC, pages 1–12, (2012).

[14] Yueqi Duan, Jiwen Lu, Jiangjiang Feng, and Jie Zhou, Context-Aware Local Binary Feature Learning for Face Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 40(5):1139-1153.(2018).

[15] Lu J, Liong VE, Zhou X, Zhou J. Learning Compact Binary Face Descriptor for Face Recognition. *IEEE Trans Pattern Anal Mach Intell*. 37(10):2041-2056, (2015).

[16] J. Schmidhuber. ”Deep learning in neural networks: An overview ”, Neural Networks, Vol. 61, pp. 85-117, (2015).

[17] S. Chen, C. Zhang, and M. Dong, ‘‘Deep age estimation: From classification to ranking,’’ *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2209–2222, (2018).

[18] C. Jie Ng, A. B. Jin Teoh. ”DCTNet : A Simple Learning-free Approach for Face Recognition ”, IEEE Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), 16-19 Dec, (2015).

[19] A. Meraoumia, L. Laimeche, H. Bendjenna, S. Chitroub. ”Do We Have to Trust the Deep Learning Methods for Palmprints Identification? ”, Proceedings of the Mediterranean Conference on Pattern Recognition and Artificial Intelligence, Tebessa, Algeria, pp. 85-91, (2016).

[20] T. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma. ”PCANet: a simple deep learning baseline for image classification ”, IEEE Trans. Image Processing 24, 5017, (2015).

[21] J. Xing, K. Li, W. Hu, C. Yuan, and H. Ling, ‘‘Diagnosing deep learning models for high accuracy age estimation from a single image,’’ *Pattern Recognit.*, vol. 66, pp. 106–116, Jun. (2017).

[22] J. Wan, Z. Tan, Z. Lei, G. Guo, and S. Z. Li, ‘‘Auxiliary demographic information assisted age estimation with cascaded structure,’’ *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2531–2541, Sep. (2018).

[23] R. Rothe, R. Timofte, and L. Van Gool, ‘‘Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks,’’ *Int. J. Comput. Vis.*, vol. 126, no. 2, pp. 144–157, Apr. (2016).

[24] N. Vapnik, Vladimir. ”Support-vector networks ”, Machine Learning, Vol. 20(3), pp. 273-297, (1995).

[25] C. R. Huang, Y.-T. Chen, W.-Y. Chen, H.-C. Cheng, and B.-S. Sheu. ”Gastroesophageal reflux disease diagnosis using hierarchical heterogeneous descriptor fusion support vector machine ”, IEEE Trans. Biomed. Eng., Vol. 63, No. 3, pp. 588-599, (2016).

[26] S. B. Imandoust, M. Bolandraftar. ”Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background ”, Int. Journal of Engineering Research and Applications, Vol. 3(5), pp. 605-610, (2013).

[27] Yu, H., Samuels, D.C., Zhao, Y. *et al.* Architectures and accuracy of artificial neural network for disease classification from omics data. *BMC Genomics* **20,** 167 (2019).

[28] Liu, J., Ma, Y., Duan, L., Wang, F., & Liu, Y. *Hybrid constraint SVR for facial age estimation. Signal Processing, 94, 576–582.* (2014)

[29] Fernández, Carles & Huerta, Ivan & Prati, Andrea. A Comparative Evaluation of Regression Learning Algorithms for Facial Age Estimation. (2015).

[30] A. S. Al-Shannaq and L. A. Elrefaei, "Comprehensive Analysis of the Literature for Age Estimation From Facial Images," in *IEEE Access*, vol. 7, pp. 93229-93249, (2019).

[31] A. Lanitis and T. Cootes, *Fg-Net Aging Data Base*. Engomi, Cyprus: Cyprus College, (2002).

[32] K. Ricanek and T. Tesafaye, ‘‘MORPH: A longitudinal image database of normal adult age-progression,’’ in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Apr. 2006, pp. 341–345. (2006).

[33] M. Minear and D. C. Park, ‘‘A lifespan database of adult facial stimuli,’’ *Behavior Res. Methods, Instrum., Comput.*, vol. 36, no. 4, pp. 630–633, (2004).

[34]A. Bastanfard, M. A. Nik, and M. M. Dehshibi, ‘‘Iranian face database with age, pose and expression,’’ in *Proc. Int. Conf. Mach. Vis.*, Dec. 2007, pp. 50–55 (2007)..

[35] Y. Fu and T. S. Huang, ‘‘Human age estimation with regression on discriminative aging manifold,’’ *IEEE Trans. Multimedia*, vol. 10, no. 4, pp. 578–584, Jun. (2008).

[36] B. Ni, Z. Song, and S. Yan, ‘‘Web image and video mining towards universal and robust age estimator,’’ *IEEE Trans. Multimedia*, vol. 13, no. 6, pp. 1217–1229, Dec. (2011).

[*37] B.-C. Chen, C.-S. Chen, and W. H. Hsu, ‘‘Cross-age reference coding for age-invariant face recognition and retrieval,’’ in Proc. Eur. Conf. Comput.. (2014).*

[38]Agbo-Ajala, O., & Viriri, S. Deep learning approach for facial age classification: a survey of the state-of-the-art. Artificial Intelligence Review.(2020).

[39] Y. H. Kwon and N. da Vitoria Lobo, ‘‘Age classification from facial images,’’ *Comput. Vis. Image Understand.*, vol. 74, no. 1, pp. 1–21, (1999).

[40]: K. Ueki, T. Hayashida, and T. Kobayashi, ‘‘Subspace-based age-group classification using facial images under various lighting conditions,’’ in *Proc. 7th Int. Conf. Autom. Face Gesture Recognition.*, Apr. 2006, pp. 46–48. (2006).

[41] M. M. Dehshibi and A. Bastanfard, ‘‘A new algorithm for age recognition from facial images,’’ *Signal Process.*, vol. 90, no. 8, pp. 2431–2444, Aug. (2010).

[42]:M. Y. El Dib and M. El-saban, ‘‘Human age estimation using enhanced bio-inspired features (EBIF),’’ in *Proc. IEEE 17th Int. Conf. Image Process.*, Sep. 2010, pp. 1589–1592.(2010).

[43]K. Zhou, B. Georgescu, X. S. Zhou, and D. Comaniciu, ‘‘Image based regressionusingboostingmethod,’’in*Proc.10thIEEEInt.Conf.Comput. Vis.*, Oct. 2005, pp. 541–548. (2005).

[44] S. Yan, M. Liu, and T. S. Huang, ‘‘Extracting age information from local spatially flexible patches,’’ in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2008, pp. 737–740. (2008).

[45] M. A. Hajizadeh and H. Ebrahimnezhad, ‘‘Classification of age groups from facial image using histograms of oriented gradients,’’ in *Proc. 7th Iranian Conf. Mach. Vis. Image Process.*, Nov. 2011, pp. 1–5. (2011).

[46] K. Y. Chang and C. S. Chen, ‘‘A learning framework for age rank estimation based on face images with scattering transform,’’ *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 785–798, Mar. (2015).

[47] A Lanitis, J Taylor, TF Cootes, Toward automatic simulation of aging effects on face images. IEEE Trans. Pattern Anal. Mach. Intell. 24, 442–455 (2002).

[48] Lanitis A., Draganova C. and Christodoulou C. ‘Comparing Different Classifiers for Automatic Age Estimation’. IEEE Transactions On Systems Man and Cybernetics, Part B, 34(1), pp 621-629. (2004).

[49] S. Yan, H. Wang, T. S. Huang, Q. Yang, and X. Tang, ‘‘Ranking with uncertain labels,’’ in *Proc. ICME*, Jul. 2007, pp. 96–99. (2007).

[50] K. Chen, S. Gong, T. Xiang, and C. C. Loy, ‘‘Cumulative attribute space for age and crowd density estimation,’’ *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jul. 2013, pp. 2467–2474. (2013).

[51] S. Feng, C. Lang, J. Feng, T. Wang, and J. Luo, ‘‘Human facial age estimation by cost-sensitive label ranking and trace norm regularization,’’ *IEEE Trans. Multimedia*, vol. 19, no. 1, pp. 136–148, Jan. (2016).

[52] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai, ‘‘Learning from facial agingpatternsforautomaticageestimation,’’in*Proc.14thACMInt.Conf. Multimedia*, Jul. 2006, pp. 307–316. (2006).

[53] X. Geng, Z.-H. Zhou, and K. Smith-Miles, ‘‘Automatic age estimation based on facial aging patterns,’’ *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2234–2240, Dec. (2007).

[54] Fu and T. S. Huang, ''Human age estimation with regression on discriminative aging manifold,'' *IEEE Trans. Multimedia*, vol. 10, no. 4, pp. 578–584, Jun. (2008).

[55] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, ''Image-based human age estimation by manifold learning and locally adjusted robust regression,'' *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1178–1188, Jul. (2008).

[56] 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 Trans. Image Process.*, vol. 18, no. 1, pp. 202–210, Jan. (2009).

[57]L. Cai, L. Huang, and C. Liu, ''Age estimation based on improved discriminativeGaussianprocesslatentvariablemodel,''*MultimediaTools Appl.*, vol. 75, no. 19, pp. 11977–11994, Oct. (2016).

[58] I. Huerta, C. Fernández, C. Segura, J. Hernando, and A. Prati, ''A deep analysis on age estimation,'' *Pattern Recognit. Lett.*, vol. 68, pp. 239–249, Dec. (2015).

[59] K.-H. Liu, S. Yan, and C.-C. J. Kuo, ''Age estimation via grouping and decision fusion,'' *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 11, pp. 2408–2423, Nov. (2015).

[60] J. K. Pontes, A. S. Britto, C. Fookes, and A. L. Koerich, ''A flexible hierarchical approach for facial age estimation based on multiple features,'' *Pattern Recognit.*, vol. 54, pp. 34–51, Jun. (2016).

[61] A. Rosebrock, *Deep Learning for Computer Vision with Python: Starter Bundle*, 1st ed. Pyimagesearch, (2017).
https:// www.pyimagesearch.com/deep-learning-computer-vision-python-book/

[62] M. Duan, K. Li, C. Yang, and K. Li, ''A hybrid deep learning CNN–ELM for age and gender classification,'' *Neurocomputing*, vol. 275, pp. 448–461, Jan. (2017).

[63] M. Yang, S. Zhu, F. Lv, and K. Yu, ''Correspondence driven adaptation for human profile recognition,'' in *Proc. CVPR*, Jun. 2011, pp. 505–512. (2011).

[64] S. Chen, C. Zhang, and M. Dong, ''Deep age estimation: From classification to ranking,'' *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2209–2222, Aug. (2018).

[65] Z. Hu, Y. Wen, J. Wang, and M. Wang, ''Facial age estimation with age difference,'' *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3087–3097, Jul. (2016).

[66] K. Li, J. Xing, W. Hu, and S. J. Maybank, ''D2C: Deep cumulatively and comparatively learning for human age estimation,''*Pattern Recognit.*, vol. 66, Jun. 2016, pp. 95–105, (2017).

[67] H. Liu, J. Lu, J. Feng, and J. Zhou, ''Group-aware deep feature learning for facial age estimation,'' *Pattern Recognit.*, vol. 66, pp. 82–94, Jun. (2017).

[68] K. E. Zhang, C. E. Gao, L. Guo, M. Sun, and S. Member, ‘‘Age group and gender estimation in the wild with deep RoR architecture,’’ in *Proc. Chin. Conf. Comput. Vis. (CCCV)*, May 2017, pp. 1–9. (2017).

[69] K. Zhang, N. Liu, X. Yuan, X. Guo, C. Gao, and Z. Zhao, ‘‘Finegrained age estimation in the wild with attention LSTM networks,’’ 2018, arXiv:1805.10445 (2018).

[70] Liu X, Zou Y, Kuang H, Ma X. Face image age estimation based on data augmentation and lightweight convolutional neural network. Symmetry 12(1):146 (2020).

[71] X. Wang, R. Guo, and C. Kambhamettu, ‘‘Deeply-learned feature for age estimation,’’ in *Proc. IEEE Winter Conf. Appl. Comput. Vis. DeeplyLearned*, Jan. 2015, pp. 534–541. (2015).

[72] M.Duan,K.Li,andK.Li,‘‘AnensembleCNN2ELMforageestimation,’’ *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 3, pp. 758–772, Mar. (2018).

[73] http://dlib.net/face_landmark_detection.py.html

[74] Ojala, T., Pietikäinen, M. and Mäenpää, T. (2002), Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Trans. Pattern Analysis and Machine Intelligence 24(7): 971-987. (2002).

[75] Jiwen Lu, Venice Erin Liong, Xiuzhuang Zhou, and Jie Zhou, “Learning Compact Binary Face Descriptor for Face Recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), Vol. 37, Issue 10, pp. 2041-2056, (2015).

[76] C. Ng and A. Teoh, “DCTNet: A simple learning-free approach for face recognition,” arXiv preprint arXiv: 1507.02049, (2015).