الجمهـــــــــــــورية الجزائــــرية الديمـقراطيــــــة الشعبيـــــــــــة
République Algérienne Démocratique et populaire

Ministère de l'enseignement Supérieur et de la recherche scientifique
Université Mohamed Khider Biskra
Faculté des sciences et de la Technologie
Département de Génie Electrique
Réf : …………….

وزارة التعليم العالي والبحث العلمـي
جامعة محمد خيضر بسكرة
كلية العلوم والتكنولوجيا
قسم الهندسة الكهربائية
المرجع: ....

Thesis submitted for obtaining the degree of

Doctorate in: electronic engineering

Specialty (Option): Signals and communications

Extraction of soft biometric characteristic in the wild

Presented by:

ZIGHEM Mohammed-En-nadhir

Thesis defended publicly on: 12/09/2021

The jury members:

Pr. Megherbi Ahmed Chaouki	President	Professor	Univ Biskra
Pr. OUAFI Abdekrim	Supervisor	Professor	Univ Biskra
Dr. ZITOUNI Athmane	Co-supervisor	MCA	Univ Biskra
Pr. TALB-AHMED Abdelmalik	Examiner	Professor	Univ Polytechnic Hauts-de-France
Pr. BAARIR Zine-Eddine	Examiner	Professor	Univ Biskra
Pr. Ferdi Youcef	Examiner	Professor	Univ Constantine

Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work. The material has not been submitted, either in whole or in part, for a degree at this, or any other university.

Mohammed-En-nadhir Zighem

Abstract

This thesis deals with soft biometrics. Soft biometrics refers to the physical, behavioral or adhered human characteristics like age, gender, ethnicity etc. These characteristics can be much easier to extract (especially from distance) than traditional biometric traits (like fingerprints and iris) yet they can be very useful in many applications. The thesis gives an extensive analysis of the state of the art in the field of soft biometrics by discussing the existing methods, the reported results and the public databases. Importantly, the thesis describes a number of novel contributions including new methods for age estimation from face images. Extensive experimental results are reported. Some open issues and future directions are also highlighted.

Keywords: Age estimation, Demographic classification, Gender classification, Race classification, Features extraction, Face analysis.

This thesis deals with soft biometrics. Soft biometrics refers to the physical, behavioral or adhered human characteristics like age, gender, ethnicity etc. These characteristics can be much easier to extract (especially from distance) than traditional biometric traits (like fingerprints and iris) yet they can be very useful in many applications. The thesis gives an extensive analysis of the state of the art in the field of soft biometrics by discussing the existing methods, the reported results and the public databases. Importantly, the thesis describes a number of novel contributions including new methods for age estimation from face images. Extensive experimental results are reported. Some open issues and future directions are also highlighted.

Publications

The following publications have been generated while developing this thesis, and to an extent has guided the thesis into what it has become:

Two-stages based facial demographic attributes combination for age estimation. **Zighem, M.E.**, Ouafi, A., Zitouni, A., Ruichek, Y., Taleb-Ahmed, A.,2019. "Two-stages based facial demographic attributes combination for age estimation". Journal of Visual Communication and Image Representation 61, 236–249, 2019

Age estimation based on color facial texture.
Zighem, M.E., Ouafi, A., Bekhouche, SE., Benlamoudi, A., Taleb-Ahmed, A.,2017. "Age estimation based on color facial texture". CGE'10, 17-18 avril 2017, Ecole polytechnique militaire, Alger

Face Anti-Spoofing Combining MLLBP and MLBSIF.
Benlamoudi, A., **Zighem, M.E.**, Bougourzi, F., Bekhouche, SE., Ouafi, A., Taleb-Ahmed, A.,2017. CGE'10, 17-18 avril 2017, Ecole polytechnique militaire, Alger

A Comparative Study On Textures Descriptors In Facial Gender Classification. Bougourzi, F., A., Bekhouche, SE., **Zighem, M.E.**, Benlamoudi, A.,Ouafi, A., Taleb-Ahmed, A. CGE'10, 17-18 avril 2017, Ecole polytechnique militaire, Alger

Acknowledgements

This thesis is the results of a hard work done at the MOHAMED KHIDER UNIVERSITY of BISKRA from 2016 to 2021. The first merit goes to my supervisors Ouafi and xx who supervised the work and guided me during these years. The thesis would have never been realised without their valuable advises. The courses and knowledge that I received from all my teachers and mentors at the MOHAMED KHIDER UNIVERSITY of BISKRA were also very useful in conducting the experiments. This is very much appreciated. Some parts of thesis are conducted during a research visit at the University Valenciennes (France) under the guidance of Prof. Taleb-Ahmed Abdelmalik. I will never forget the precious help and support of Prof. TalebAhmed Abdelmalik. All the co-authors of my papers are also fully acknowledged. I hope that our collaboration will further strengthen after this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and motivation

Biometric systems recognize users based on their physiological and behavioral characteristics [1]. Unimodal biometric systems make use of a single biometric trait for user recognition. It is difficult to achieve very high recognition rates using only unimodal systems due to factors like noisy sensor data and non-universality and/or lack of distinctiveness of the chosen biometric traits. Multimodal biometric systems address some of these problems by combining evidence obtained from multiple sources [2]. A multimodal biometric system that utilizes a number of different biometric identifiers like face, fingerprint, hand-geometry, and iris can be more robust to noise and alleviate the problem of non-universality and lack of distinctiveness. Hence, such a system can achieve a higher recognition accuracy than unimodal systems. However, a multimodal system will require a longer verification time thereby causing inconvenience to the users.

Another alternative to improve the performance of biometric systems is to use soft biometric traits. It is indeed possible to improve the recognition performance of a biometric system without compromising on user-friendliness by utilizing ancillary information about the user like height, weight, age, gender, ethnicity, and eye color. We refer to these traits as soft biometric traits because they provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals (see Figure 1 for examples of soft biometric traits). The soft biometric traits can either be continuous or discrete. Traits such as gender, eye color, and ethnicity are discrete.

Primary biometric traits such as face, fingerprints, and iris can suffer from noisy sensor data, non-universality, and lack of distinctiveness. Further, in certain applications, these traits may fail to achieve high recognition rates. Multimodal biometric systems (Ross et al., 2006) can solve these problems by combining multiple biometric traits, resulting in a biometric signature that is robust and more distinctive.

Multimodal systems offer improved performance, but the time taken to verify users can drastically increase thereby causing inconvenience to the subjects and reducing the throughput of the system. Soft biometric traits have been investigated to solve this problem (Jain et al., 2004b).

1.2 Problematic

Extracting soft biometrics in controlled conditions is relatively easy but the task becomes very challenging when facing real-word conditions (i.e. in the wild). All the factors (illumination changes, pose, low quality images etc.) affecting general object recognition can be also be encountered in extracting soft biometrics. Very recently, when people are wearing masks due to Covid-19 pandemic, this situation makes the extraction of some soft biometrics traits even more challenging. There are also many open questions regarding the extraction and the use of soft biometrics. This includes:

- What soft biometrics trait can be extracted in a given situation?
- How to efficiently extract different soft biometrics in the wild?
- How to efficiently combine different soft biometrics trait?
- What is the relationship between different soft biometrics? For instance, how gender affects age estimation?
- How soft biometrics can be optimally used in real world applications?
- How the recent progress in deep learning can benefit the extraction of soft biometrics?

1.3 Contributions of the thesis

The thesis answers some of the above mentioned problems by proposing a new methodology for extracting soft biometrics in the wild. The main focus is on estimating human age from facial images. The thesis also gives an extensive analysis of the state of the art in the field of soft biometrics by discussing the existing methods, the reported results and the public databases. Extensive experimental results are reported on public databases. Some open issues and future directions are also highlighted.

1.4 Outline of the thesis

The thesis is organized as follows. Chapter 2 gives definitions and general overview of soft biometrics including the existing databases. Chapter 3 focuses on facial soft biometrics and presents the state of the art in extracting facial soft biometrics including deep learning methods. Chapter 4 presents the experimental analysis on age prediction from faces. Chapter 5 summarizes the thesis and highlights open issues and future directions.

Chapter 2

Overview on Soft Biometrics

2.1 Introduction

Soft biometrics are defined by Jain *et al.* [76] as *"characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals"*. In other words, soft biometrics are physical, behavioral or adhered human characteristics, classifiable in predefined human compliant categories. The aim of these characteristics is to distinguish between different individuals [27]. The history of soft biometrics dates back to 1896 when a French police officer Alphonse Bertillon [68] created the personal identification system (bertillonage) based on the anthropometric determinations and morphological characteristics of the person (See Fig 2.1).

Figure 2.1: An example of a real French police card of a criminal which was created as a part of bertillonage personal identification system. Bertillonage is often given as a first example of large collection of the soft biometrics data (Source: Google Images).

These characteristics have also been referred to as semantics [131], in reference to their semantic interpretation (e.g., describing a human face as "young male"). One advantage of soft biometric with respect to "primary" biometric systems (the goal of which is to identify a person) is that the former provides only a partial description of an individual preserving his/her privacy. This allows for the use of soft biometrics to collect anonymous statistics. For example, human age, gender and race, are particular cases of soft biometrics attributes which are often called *demographics*, because they are commonly used for population analysis [27].

In this chapter we give an overview of the benefits and characteristics of soft biometrics, their classifications, and the related advantages and limitations. In addition, we analyze the work already done on soft biometric attributes or systems that integrate soft biometric attributes.

2.2 Soft biometric attributes

Soft biometric attributes can be classified based on the modalities of face, body and accessory (see Table 2.1). Recognition and synthesis of facial attributes (gender, age, race, beard etc.) are an active research area involving the studies with various data modalities such as voice [109], iris [121], body [130], hand [135] and others.

However, among all soft biometric modalities, the face is the richest source that provides information about a person [75](see Fig 2.2). In addition, the development of social networks has dramatically increased the amount of face images (especially of celebrities) which are publicly uploaded in the Internet. Not only does it underpin the practical interest of creating automatic systems of face analysis, but it also offers an opportunity to train complex machine learning models (for example, deep CNNs), which was not possible even a decade ago. In the next subsections, We discuss soft biometric attributes based on the aforesaid classification of soft biometrics. We discuss facial soft biometrics attributes (Subsection 2.2.1), and body soft biometrics attributes (Subsection 2.2.2) and finally in the subsection 2.2.3, we discuss accessory attributes.

Soft biometric attribute	Face, Body, Accessory	Nature of value	Permanence	Distinctiveness	Subjective perception
Age	Face/Body	Continuous	Low/Medium	Medium	Medium
Skin color	Face	Continuous	Medium	Low	Medium
Hair color	Face	Continuous	Medium	Medium	Medium
Eye color	Face	Continuous	High	Medium	Medium
Beard	Face	Binary	Low/Medium	Low	Medium
Mustache	Face	Binary	Low/Medium	Low	Medium
Facial measures	Face	Continuous	High	Medium	Medium/High
Make–up	Face	Discrete	Low	Low	Medium
Race	Face	Discrete	High	Medium	Medium
Gender	Face/Body	Binary	High	Low	Low
Gait	Body	Continuous	Medium	Medium	High
Weight	Body	Continuous	Low/Medium	Medium	Medium
Height	Body	Continuous	Medium/High	Medium	Medium
Glasses	Accessory	Binary	Low/Medium	Low	Low
Clothes color	Accessory.	Discrete	Low	Medium	Medium

Table 2.1: Table of soft biometric attributes

2.2.1 Facial soft biometrics

Doubtless, an important number of soft biometric attributes can be extracted from face images and facial movements. This generally includes gender recognition (i.e. male vs. female), age categorization (e.g. child, youth, adult, middle age and elderly) and race [1] classification (e.g. Asian, Caucasian and African). These are often referred to as demographic attributes and are very useful for more affective human computer interaction (HCI) and smart environments in which the systems should adapt to the users whose behaviors and preferences are not only different at different ages but also specific to a given race and/or gender. Automatic demographic attributes extraction is also useful in many other applications such as content based image and video retrieval,

[1] In general English the term “race” and “ethnicity” are often used as though they were synonymous. However, they are related to biological and sociological factors respectively. Generally, race refers to a person’s physical appearance or characteristics, while ethnicity is more viewed as a culture concept, relating to nationality, rituals and cultural heritages, or even ideology. Since there are over 5,000 ethnic groups all over the world [13], the idea of “ethnicity recognition” seems to be both questionable and impractical from current computer vision and pattern recognition point of view. Therefore, considering the soft biometric characteristics of distinctive human population, we prefer to use “race” as more suitable category terminology in this thesis.

restricting access to certain areas based on gender and/or age, collecting demographic information in public places and counting the number of women entering a retail store and so on.

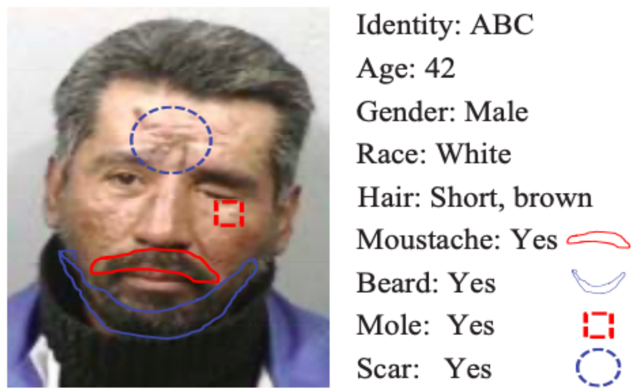


Figure 2.2: A wide variety of information can be gleaned from a face image, such as identity, age, gender, race, scars, marks and tattoos (SMT).

Actually, apart demographic attributes a human face can reveal a person's identity [150], mood [97] and many other details (Fig 2.2). Though there has been a great deal of progress in face analysis in the last years, facial soft biometric tasks have not been associated to that progress as most work has mainly focused on face detection and recognition problems. Consequently, the design of algorithms that are effective in discriminating between males and females, or classifying faces into different age and race categories is still challenging and remains an open area of research. Since the early publications in [110], [138], research on this class of soft biometrics has been embraced by the computer vision community.

2.2.1.1 Gender recognition

Human gender recognition is one of the fundamental tasks in the area of computer vision, which has recently gained a lot of attraction in research communities as well as industries due to its substantial role in a notable number of real time applications.

Particularly, in social interactions, different salutations and grammar rules are used for men and women. In the targeted advertisement, the billboard's contents can be visualized based on the demographics of pedestrians. The gender can be used as a key characteristic to perceive the shopping nature for the future of retail.

Gender recognition can be considered as a binary class problem that is only two classes are assumed to be present (male and female class). Human beings are diverse in nature. They differ in their physical appearance, personality, behaviour, thoughts and many more features. Even if two individuals are similar by physical appearance, there will be many distinguishing features separating them.

To solve the issue of gender recognition, many methods use physical appearance as input for classification. Physical appearance includes facial features like the eyes, nose, cheeks, lips, hair, forehead, ears, and the mid and lower body parts such as hands, legs, stomach area etc. Many research papers have facial features as an input to the classification problem. However, gender recognition is still a challenging task due to various changes in viewing angles, facial expressions, extreme poses, background, resolution variations, and face image appearance. It is more challenging in unconstrained imaging conditions. Indeed, an average female face is rounder than an average male one, while men often have more facial hair than women. Nevertheless, Loth and Iscan [103] showed that not a single face characteristic can be solely used to confidently recognize gender. Moreover, the difficulty of gender recognition can be increased by the presence of makeup (Fig 2.3) and facial accessories (eyeglasses, scarf, etc.). Gender recognition can be very challenging in unconstrained imaging conditions.

2.2.1.2 Age estimation

Human age is one of the crucial facial soft biometric attributes. Therefore, age is important for the facial analysis (e.g. age estimation and age classification). Typically,

Figure 2.3: Female to Male Makeup Transformation.

a human facial appearances are varying as age growth, such as the shape of facial contour, skin textures, shape of facial features, and so on. Generally, there are some general changes and resemblances in human facial aging process [4]. Therefore, we can always describe some global biological characteristics from the statistic view, such as craniofacial growth (shape change) from the birth to the adulthood [124], more protrusive chin as aging, smaller eyes as aging, growing wrinkle as aging, more dense mustache as mature, and skin aging from adulthood to agedness [43]. Fig 2.4 illustrates the common human facial changes in the human aging process.

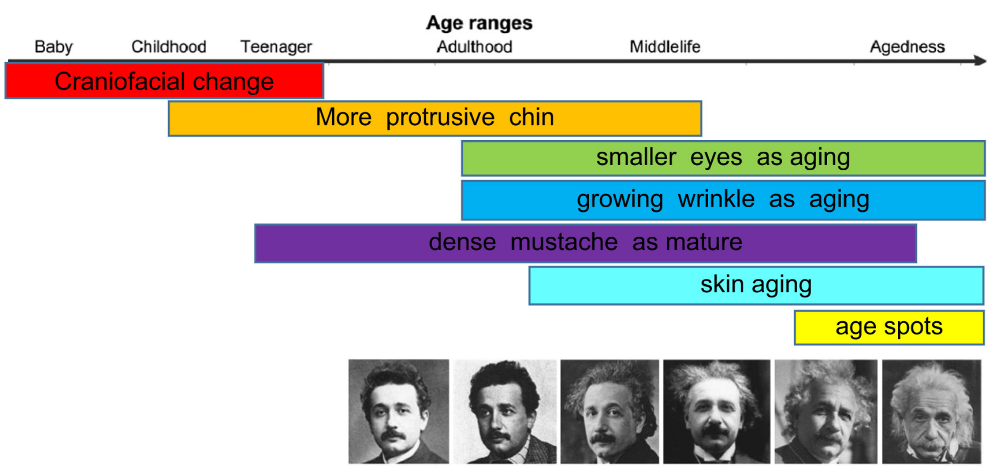


Figure 2.4: An illustration of human aging progress. At various stages of life, aging affects different face parts. [136]

In recent years, computer vision has made a great progress in various kinds of practical applications, e.g., bio-vision bionics, monitoring system, data analysis. Therefore, can a machine performs better human? Technology advances in computer science and engineering have given a positive answer to this question. There are two basic tasks in this field, computer based age progression and estimation, that are described as follows:

- *Age progression*: Re-render a face image aesthetically with natural aging and rejuvenating effects on the individual face.
- *Age estimation*: Label a face image automatically with the exact age (year) or the age group (year range) of the individual face.

To further understand the tasks, we can differentiate three concepts about human age:

- *Actual age*: The real age (cumulated years after birth) of an individual.
- *Apparent age*: The age information shown on the visual appearance.
- *Estimated age*: The individual age recognized by machine from the visual appearance.

The apparent age is typically consistent with the actual age. However, the variation is often inevitable due to the generic difference between different individuals and environmental/artificial factors. The estimated age is defined on the appearance age. The actual age is often defined as the ground truth.

2.2.1.3 Race classification

Besides gender and age, race is arguably the most prominent and dominant personal attribute, which can be demonstrated empirically by its omnirelevance with a series of social cognitive and perceptual tasks (attitude, biased view, stereotype, emotion, belief, etc.). Furthermore, it yields deep insights into how to conceptualize culture and socialization in relation to individual appearance attributes, including social categorization, association and communication. Therefore, the estimation of racial variance by descriptive approaches for practical purposes is indeed indispensable in both social and computer science. However, while race demarcation drives the intrinsically genetic variation structure of essential facial regions to gather more explicit appearance information, the core question emerges as the computational mechanism underlying this extraordinary complexity. This raises the following fundamental multi-disciplinary conundrum: How does a computer model and categorize a racial face?

To answer this fundamental question, numerous research consortium and scholars have developed intensive investigations from different angles. For example, psychologists have studied behavior correlations of race perception such as other race effect (ORE) and attention model [122], [34], [42], which show existence of racially-discriminative facial features such as eye corners or nose tip. Neurophysiologists have shown how race perception influences and regulates cognitive processes such as affection [134] and stereotype [11], [82]. Computational neuroscientists have built models to simulate and explain race perceptions (e.g., [92], [119]). Other cognitive experiments in [12] have also indicated the existence of racial features as a salient visual factor. Among the general public, the validity of racial categories is often taken for granted. Following these quantitative analysis, computer vision scholars have been motivated to tackle the inherent problem of racial demarcation by building computationally intelligent systems capable of categorizing races.

2.2.1.4 Relation between age, gender and race

It is often the case that a single facial feature can carry information about different soft biometric attributes. That is why attributes such as age, gender and race are often treated and categorized simultaneously [99], [52], an approach that is in line with the perceived correlation of these soft biometric attributes.

In addition to the correlation, exploring this intertwined nature of attributes can carry substantial advantages; for example, from a genetic point of view, understanding the interaction of race with aging allows for conclusions on race based differences in longevity and aging associated diseases [133]. This perceived correlation between these attributes has motivated additional work such as that in [53], which studied the influence of gender and race on age estimation process. The work concluded that age estimation can be impacted by the gender and race differences considerably. While joint treatment is not beneficial for gender and race classification and that gender and race can be estimated separately, due to the fact that features used to estimate gender are shared by all race groups and features used for race classification are present in both female and male faces.

2.2.1.5 Beard and mustache detection

Another typical face attribute is provided by the presence of beard or mustache, especially for men's categorization. Generally considered as an obstacle to the face recognition algorithms because of its high time variability, the study of beard and mustache detection is principally focused towards a detect-and-remove approach. As major example, in [84] authors proposed an algorithm for beard removal from images of people with facial hairs, the system is designed using the concept of structural similarity and coordinate transformations.

2.2.1.6 Evaluation Metrics

A. Gender classification

Gender recognition is a binary classification problem, and the accuracy of automatic gender classification algorithms is most often measured by a classification accuracy, which is simply defined as the ratio between the number of correct predictions N_c and the total number of predictions N:

$$Accuracy = \frac{N_c}{N} \tag{2.1}$$

The biggest downside of gender classification accuracy is that it does not reflect the partial prediction scores for male and female. Thus, sometimes the standard binary classification metrics such as True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) rates are used together with classification accuracy (see Table 2.2).

<table>
<tr><td></td><td></td><td colspan="2">Ground Truth</td></tr>
<tr><td></td><td></td><td>Positive</td><td>Negative</td></tr>
<tr><td rowspan="2">Predicted</td><td>Positive</td><td>True Positive</td><td>False Positive</td></tr>
<tr><td>Negative</td><td>False Negative</td><td>True Negative</td></tr>
</table>

Table 2.2: Confusion Matrix

Otherwise, ROC curve is often employed in addition to classification accuracy for binary classification problems. A simple way to summarize a ROC curve in one real value is the usage of an Area Under Curve (AUC) [65] which allows the direct comparison between a pair of ROC curves.

B. Age

Various evaluation metrics have been proposed for characterizing the performance of facial age analysis methods. In the following we briefly recall the ones commonly used.

1) Mean Absolute Error (MAE): is the most used evaluation metric for real age estimation. Given x_i the age estimated on the ith sample and y_i the corresponding true label, the MAE is the average error over the N test samples. Denoting with $e_i = |x_i - y_i|$ the error on the ith sample:

$$MAE = \frac{\sum_{i=1}^{N} e_i}{N} \tag{2.2}$$

2) Cumulative Score (CS): is used to evaluate the age estimator performance at different absolute error levels, where the age estimation error is lower than a threshold value. The CS is calculated by:

$$CS(k) = \frac{N_{e \leq k}}{N} * 100 \tag{2.3}$$

Where k is the threshold (years), N_e is the number of test images on which the age estimation makes an absolute error no higher than k and N is the total number of samples.

3) ϵ-error: is a metric specifically proposed for the Chalearn Looking at People challenge [37], [38] for apparent age estimation, where no ground truth is available; the age labels are computed by asking people to guess them by watching the images. Let be μ_i and ${\delta_i}^2$ respectively the average and the variance of the distribution of the estimates provided by the persons for the ith sample of the database, the classification ϵ_i on the ith sample, is calculated as:

$$\epsilon_i = 1 - e^{-\frac{(x_i-\mu_i)^2}{2\delta_i^2}} \tag{2.4}$$

The idea behind this index is to weight the estimation errors by taking into account the complexity of the sample under consideration as experienced by people. In particular, the error on the ith sample is normalized by the corresponding variance, according to (eq:2.4), samples with high variance give less contribution to the error.

4) Accuracy: is the evaluation metric generally adopted for age group classification, defined as the ratio between the number of correct classifications and the total number of test samples. This index is used in two forms: the *top-1* and the *1-off*. In the first case, a classification is considered correct if and only if exactly corresponds to the true age group. In the second case, the evaluation metric is more tolerant and considers correct also the classifications for age groups that are adjacent to the true age group.
Better performance refers to higher CS, higher accuracy, lower MAE, and lower ϵ-error values.

2.2.1.7 Databases

In recent years several databases have been collected for face analysis, In Table 2.3, we list different databases that have been used to evaluate face analysis tasks, together with the most relevant characteristics as the size, the label, the presence of additional annotations as identity, expression. An idea of the size and the popularity of the various database along the years, in terms of the number of methods using them, is given in Fig 2.5. A more detailed description of each database is reported in the next subsections.

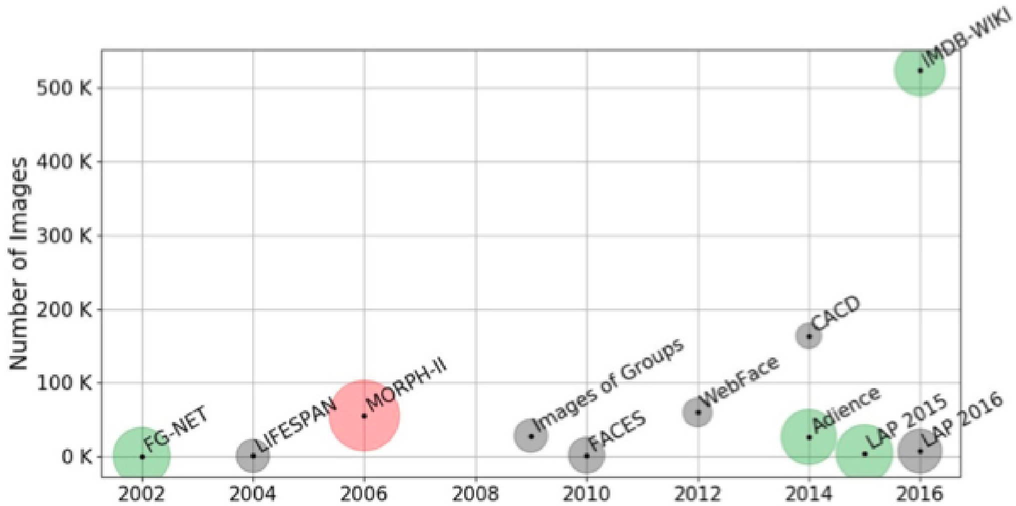


Figure 2.5: The size and popularity of the various database along the years.

Database	Year	# Images	# Subjects	Label			
				Gender	Age	Race	Other info
FG-NET	2002	1,002	82	✓	✓	N/A	Identity
PAL	2004	1,142	575	✓	✓	✓	Expression
MORPH-II	2006	55,608	13,673	✓	✓	✓	N/A
FACES	2010	2,052	171	N/A	✓	N/A	Expression
CACD	2014	163,446	2,000	✓	✓	N/A	N/A
PCSO	2014	1,500,000	100,012	✓	✓	✓	N/A
Adience	2014	26580	2,984	✓	✓	N/A	N/A
LFW+	2015	2,052	171	✓	✓	✓	N/A
IMDB-WIKI	2016	523,061	82,612	✓	✓	N/A	Identity
AgeDB	2017	16,488	568	✓	✓	N/A	Identity

Table 2.3: Publicly available databases for facial soft biometrics.

IMDB-WIKI: [128] is the largest available database annotated with age and gender labels. It comes from the fusion of the 460,723 face images of 20,284 people from IMDB, the famous web portal of the celebrities. Fig 2.6 illustrates distribution of age in IMDB-WIKI database. Being composed by pictures of celebrities, mostly looking towards the camera, the images do not present very challenging variations. It is important to mention that the authors do not assure the accuracy of the identities and of the age annotations.

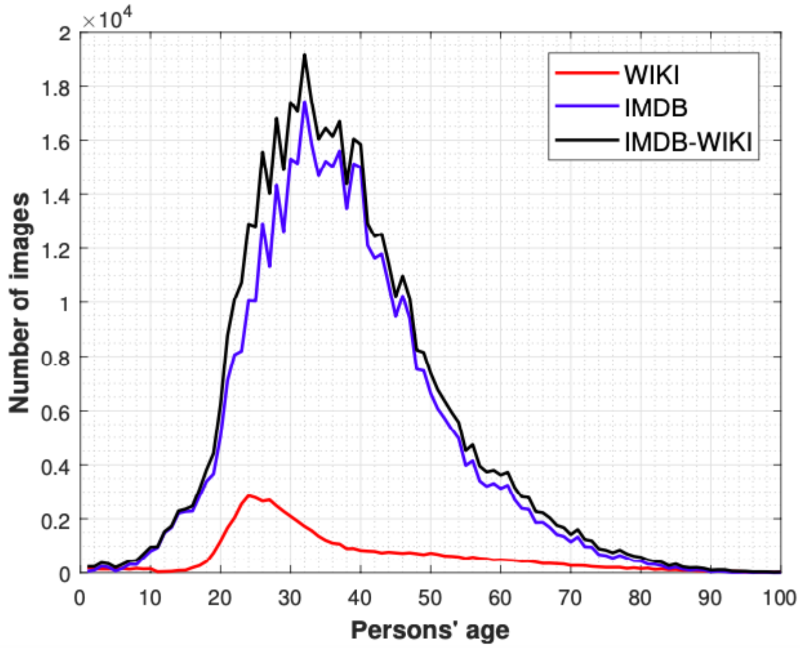


Figure 2.6: Distribution of age in IMDB-WIKI database.

Indeed, they took the images and the birth date from the personal profiles of the celebrities and assumed that the timestamp of the photo was correct. IMDB-WIKI is currently the most used database for pre-training networks for age estimation, which probably allows to achieve remarkable performance thanks to its size. Few examples of face images extracted from IMDB-WIKI databases are shown in Fig 2.7.

MORPH-II: [123] is collected by Face Aging Group at the University of

Figure 2.7: Some facial images from the IMDB-WIKI database.

North Carolina at Wilmington in controlled laboratory conditions, obtained with collaborative people that look towards the camera. So, the images do not have significant pose variations and the quality and image resolution is rather poor (between 200x240 and 400x480 pixels). Despite, this database is characterized by a high variability in terms of the age range, being [16-99 years], race, with a black/white ratio of about 4:1, and a male/female ratio of 5.5:1; it is created so as to preserve the distribution of age, gender and race of the whole database for each of its 5 folds.

Age range	0-20	21-30	31-40	41-50	51-60	61+	Total
Female	1081	2264	2939	1832	355	17	8488
Male	8372	13236	12690	9427	2678	243	64,646
Black	7571	12077	11848	8626	2263	177	42562
White	1351	2571	3368	2491	740	79	10600
Other	531	852	413	142	30	4	1972
Total	9453	15500	15629	11259	3033	260	55134

Table 2.4: The Age range, Gender, and Race distributions of subjects in the MORPH II database.

MORPH-II is highly recommended to be used in the evaluation of demographic estimation systems due to the high number of images in it and the distribution of the demographic attributes, this can be seen in Table 2.4 and Fig 2.8.

Figure 2.8: Some facial images from the MORPH-II database [123].

The Labeled Faces in the Wild (LFW+): is the most widely used database for studying the problem of unconstrained face recognition. Han et al [62], performed age, gender, and race estimation on a subset of LFW with 4,211 subjects (one image per subject), where the face images have relatively small pose variations. Since the label of race and real age are not available, they collected human (crowdsourced) estimates of age, gender and race of each face image by using the Amazon Mechanical Turk (MTurk) crowdsourcing service with three workers per task. A few examples of face images extracted from LFW+ databases are shown in Fig 2.9.

FG-NET: [94] has been proposed by the BeFIT for age estimation in uncontrolled real life conditions. In our opinion this database is not particularly challenging, since the people are aware they are in the field of view of the camera. In addition, the images are not equally distributed over age and only a few images of individuals older than 40 are available. Moreover, its limited size imposes the leave-one-person-out (LOPO) evaluation protocol.

The Productive Aging Lab Face (PAL) database: [112] contains totally 1,046 frontal face images from 580 subjects with two different expressions: neutral

Figure 2.9: Some facial images from the LFW+ [62].

and happy. The resolution of the faces is 640x480, and with age range from 18 to 93 years old.

Age range	0-20	21-30	31-40	41-50	51-60	61+	Total
Female	59	161	51	40	39	266	616
Male	57	199	27	32	13	102	430
Black	43	84	13	22	7	39	208
White	58	189	65	50	45	325	732
Other	15	87	0	0.	0	4	106
Total	116	360	78	72	52	368	1046

Table 2.5: Age range, Gender, and Race distributions of subjects in the PAL database.

The PAL database can be divided into three main races: Black subjects (208 images), White subjects (732 images) and other races subjects (106 images). See Fig 2.10 for some samples.

Adience [36] is a database for age and gender classification collected in real-world conditions, including variations in appearance, pose, lighting and image quality. The whole database consists of 26,580 face images with different yaws from frontal angle and only 13,649 almost frontal. The faces are divided in eight not balanced age categories: 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60+.

Figure 2.10: Some facial images from the PAL database.

Figure 2.11: Some facial images from the Adience database.

These characteristics make the database particularly suited for the benchmark of age group classification algorithms in challenging conditions. Some examples from Adience database are shown in Fig 2.11.

Cross-Age Celebrity Dataset (CACD): [19] is mainly used for face recognition and retrieval purposes, but it is one of the largest publicly available database of celebrities images, with age annotations. Its main problems are on the one hand the uncertainty of the annotations (collected with the same protocol proposed for IMDB-WIKI) and on the other hand the reduced age range [14-62 years]. For these reasons, the authors themselves discourages its use for age estimation purposes. Some examples from CACD database are shown in Fig 2.12.

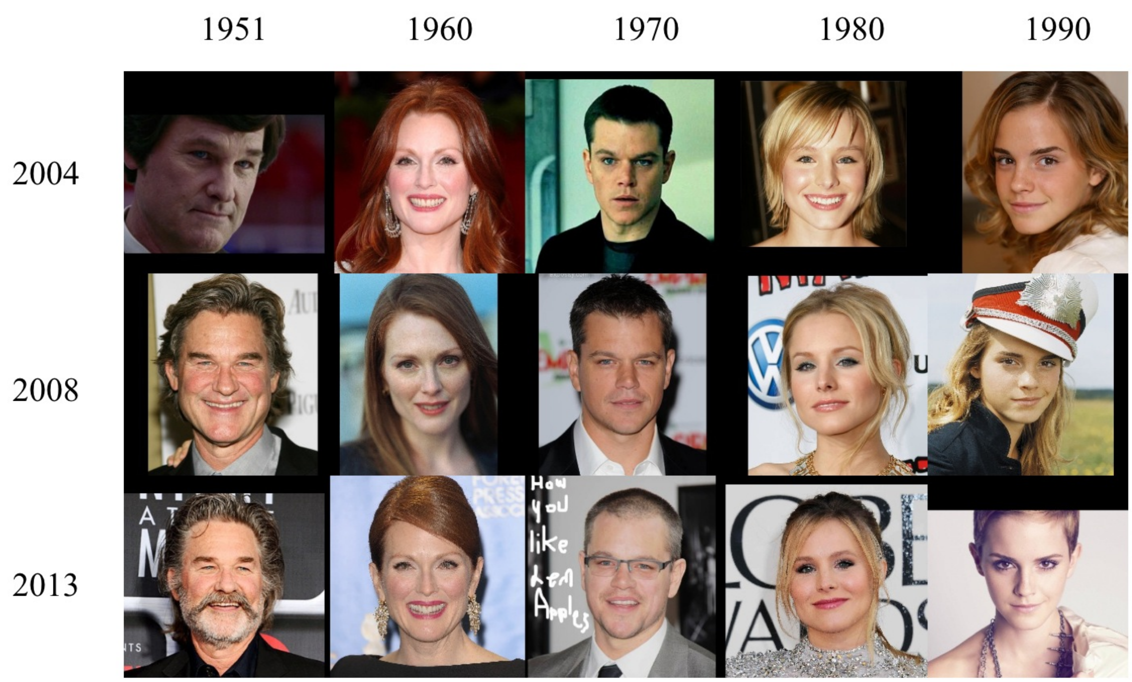


Figure 2.12: Some facial images from the CACD database [19].

Images of Groups (GROUPS): [44] is the database proposed by the BeFIT for age group classification. The images of the database are divided in seven age categories: 0-2, 3-7, 8-12, 13-19, 20-36, 37-65, and 66+. Considering that the class

20-36 contains more than 15,000 faces, this database is significantly unbalanced. Therefore, the set up proposed by the authors [44] is made of 3,500 images used for training and 1,050 for testing, being careful that training and testing images are equally distributed across the seven age groups.

Figure 2.13: Some facial images from the GROUPS database.

It is worth pointing out that the complexity of this data- base is similar to that of Adience. Images of Groups exhibits less variations in terms of pose, consists of less images in its balanced version and has a single age group 37-65 for the adulthood. Consequently, Adience is widely better and this is the reason why GROUPS is less used. Fig 2.13 shows some facial examples from the GROUPS database.

LAP databases: Since 2011 ChaLearn Looking at People (LAP) organizes several computer vision competitions aimed at recognizing people in images. One of the most successful events is the Age Estimation Challenge, organized for the first time during ICCV 2015 [37] and CVPR 2016 [38]. A database for estimating the apparent age of people, based on the opinion of web users, was in these occasions made available to the community. Although these databases do not provide large number of images, they are rightly considered the most challenging ones in terms of face variations and consequently the most commonly adopted in recent years. These reasons make them the most suited for benchmarking the performance of age estimation methods both in terms of completeness and of available performance data.

2.2.1.8 Databases analysis

Fig 2.5 points out that FG-NET and MORPH-II are the most used databases for evaluating the performance of facial soft biometrics estimation algorithms, since they were the first database available (2002 and 2006 respectively). The availability of several results on those databases in the last five years is surely a stimulus to their use. On the contrary, the recent ChaLearn LAP databases, collected for estimating the apparent age, make them particularly suited for benchmarking apparent age methods. Adience and Images of Groups are databases as challenging as the ones proposed for the ChaLearn LAP competitions, since the available images exhibit strong variations in terms of image resolution and quality and facial poses. They may be considered as good alternatives, even because the age group annotation is available, but not the real age. Recently, many researchers have experienced the benefits of using facial soft biometrics, a good database for analysing the impact of the facial expression on the perceived age is FACES. Indeed the other databases typically contain people smiling or with a neutral expression (including PAL).

To perform an effective training of a network for age estimation it is fundamental to have a very large and heterogeneous database. Currently, the database most suited to this need is the IMDB-WIKI, used by almost all the researchers in their experiments. Despite the remarkable effort done for the collection of IMDB-WIKI, the authors themselves point out that the procedure used for gathering images and labels is not optimal to guarantee the accuracy of the age annotations. Moreover, being available the identity and the gender annotations, IMDB-WIKI can be used likewise for face recognition pre-training.

2.2.2 Body Soft Biometrics

In this subsection, we focus on soft biometric attributes that have been historically used to quantify the geometry and shape of the body and skeleton (i.e., of classical anthropometric measures). The studies on anthropometric measures are not generally driven by biometric use. While at the beginning anthropometry was a technique used in physical anthropology to study the physical development of the human species. Nowadays it is employed in industrial/clothing design, ergonomics, and architecture to optimize the products to the customers needs. Other interesting works regards the study of population statistics so as to monitor changes in lifestyle, and nutrition to track body dimensions (e.g. obesity epidemic)[16].

The first biometric application of anthropometric measurement is due to the aforementioned work by Alphonse Bertillon [68]. His anthropometric measurement-based classification method was used to identify criminals, and it is indeed one of the few examples of anthropometric attrimeasurement used as biometric identifier (see Fig 2.14).

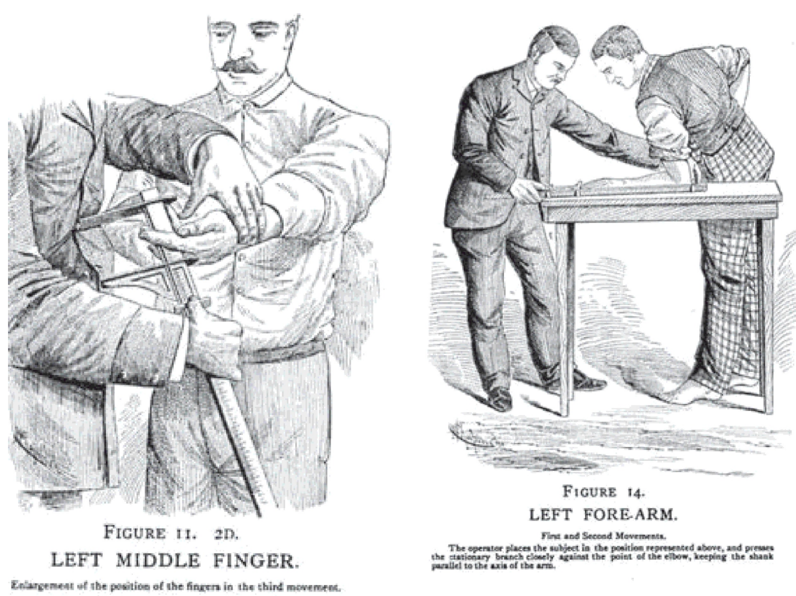


Figure 2.14: Examples of Bertillon's gathering of measurements [10]

Among the many anthropometric attributes, height, gait, body weight and color of clothes concern the body and are the main attributes that can be extracted from distance. The best distinctiveness is provided by gait detection, which is why gait occasionally is referred to as a classical biometric.

2.2.2.1 Body height

Among the many geometric attributes, body height is the most prominent. In extracting this measure, different challenges remain, including that of the human pose which can serve as a primary biasing factor. However, height estimation is an already mature topic in the literature and it has been exploited several times. One of the earliest approaches is presented in [51]. The authors used the content of the image to compute geometrical properties of objects that lie on the same plane, later they can compare objects dimensions. By knowing the height of given objects in the scene they are able to measure height of people in the Camera Field of View (FOV). Extending this last work, the authors of [95] proposed further improvements using multiple measures and a statistical approach to remove outliers. Using the proposed approach the authors reached a precision of 1cm for subjects walking in an unconstrained scenario. Precise measurement of height has been already used in combination with other features so as to track people across multiple camera systems, and to allow the identification of the same person in multiple video streams [108]. The estimation is performed via the computation of height related to the real world coordinates estimated in camera images.

Height is possibly one of the most used feature in real cases and can become under certain circumstances a crime evidence. It is indeed one of the main factors used in photogrammetry. This technique is nowadays widely used to estimate anthropometric measures from images or video surveillance footages. The Netherlands Forensic Institute has performed a comparison [35] of two methods for obtaining body height measurements from images. One is based on projective geometry and the other one on 3D modeling of the crime scene. Keeping the same camera settings setup, the authors demonstrated that the predictions of both methods are accurate, but changing camera position makes the first algorithm less reliable. Moreover, also the 3D reconstruction of the environment can be helpful as this kind of analysis greatly simplify the extraction of measurements. The possibility of using such a technique is explored in [28] where the authors used landmarks within the scene to enable the automatic collection of subjects' height measurements.

BenAbdelkader and Yacoob [12] estimated height from full body images and from images which contain only the upper part. They evaluated their methodology on synthetic data (generated randomly) and real images.

2.2.2.2 Body weight

Since the beginning, weight was introduced within the list of the soft biometric attributes [74]. However it was not fully explored as soft biometric attributes. A field where weight is considered an important feature is represented by medical studies. Here the main interest is represented by the visual extraction capability and reliability of human operators in case of emergency situations where there is no possibility of using scales like in [111], [60].

Other interests are represented by the use of weight as a foremost important feature that helps to monitor body health status [89]. Additionally, a branch of medical studies explore the forensic aspect of weight estimation [90] so as to recover information from latent traces that help to recognize victims or crime suspects. To the best of our knowledge, the only paper which involves weight directly referring to it as a soft biometrics is [3], where the authors use a scale to weight clients of a fingerprint recognition system. By exploiting weight and body fat measurements the authors reduce the total error rate of the system by 2.4%. Another work [131] considered weight as a discrete value visually defined by subjects participating to a psycho-visual experiment. However, the values used (Very Thin, Thin, Average, Fat, Very Fat) show that rather than the weight itself, the description refers to the way fat is distributed on the inspected body. That is to say users described the body build of subjects rather than their body mass.

2.2.2.3 Gait

Gait is a complex pattern that involves not only some anthropometric parameters but especially behavioral information. Among all the soft biometric attributes, it is one of the most explored, and because of its distinctiveness it is debated as being actually a hard biometric. One of the first experiments (1973) on gait analysis is presented in [80], where the author uses lights attached to the joints of the human body to record subjects' gait patterns. The author demonstrates how observers can recognize walking style people familiar to them just by the light traces they leave while walking. Since 1970's, many other authors were interested in the topic of automatic gait recognition. In [142], a spatiotemporal signature is extracted by the moving silhouette, later on a principal component analysis is employed to discard irrelevant information. Finally, supervised pattern classification techniques are performed in the lower–dimensional eigenspace. In order to provide more discriminative power, both the structural and behavioral characteristics of gait are captured. Another interesting work is proposed in [131], where gait is chosen as primary biometric attribute to be coupled with "semantic biometrics", that seems to be a very similar concept to soft biometrics. Using ANOVA they first outline the most important semantic attributes. After they merge the results of the signature generated by gait with the one generated by the semantic information so as to identify users of the biometric system. Other ways of performing human identification via gait analysis are based on the human silhouette and on model based systems like [64]. Gait analysis is not only used to identify people or extract soft biometric attributes, it is also actively used also in the medical field. The use of markers has been widely exploited, lately some studies have started to involve new techniques like computer vision, or new sensors like accelerometers [153] to analyze this attribute.

2.2.2.4 Databases

USF/NIST HumanID Gait Challenge Database: In 2005, the University of South Florida (USF) published USF/NIST HumanID Gait Challenge Database [132]. The database as formulated to facilitate objective, quantitative measurement of gait research progress on a large data set. It is presently the largest available data set. Fig. 2.15 shows some sample frames.

Figure 2.15: Samples from the HumanID gait challenge data set: subject walking on grass (a) along the frontal half of the elliptical path and (b) along the back half of the elliptical path.

The full data set consists of 1,870 sequences from 122 individuals, all these individuals walked around an ellipse. While walking, the following five covariates were changed: with or without a briefcase, left viewpoint and right viewpoint, two different shoes types, two different time instants, and grass surface or concrete surface. It is the only data set to include walking on a grass surface.

UMD Gait Database There are two UMD gait Databases: Dataset-1 consists of walking sequences of 25 subjects and Dataset-2 contains walking sequences of 55 subjects walking along a T-shape pathway. Dataset-2 is larger than Dataset-1, it taken outdoor by two surveillance cameras at a height of 4.57 meters. Fig 2.16 shows one sample frame. Each video sequence has approximately 10 gait cycles, viewed frontally and sideways. The database is diverse in terms of gender, age, and race. Moreover, the databases collected on different days differ with respect to clothing as well.

Figure 2.16: Sample frame from the UMD gait database.

NHANES Database: National Health and Nutrition Examination Survey database (NHANES)[81]. This database is unique because of its characteristics: size of the population, and time span analysis. This database was collected from a large population of individuals (more than 28000 people), over a period of 6 years (from 1999 to 2005) by the Centers for Disease Control and Prevention during the National Health and Nutrition Examination Survey. The purpose of this survey was the monitoring of American population, and the assessment of health and nutritional conditions of adults and children in the United States. The database is a significant source of data regarding a wide range of different statistics: health conditions, physical body measurements (weight, height, leg and arm length and so on).

CAESAR CAESAR dataset [125] that contains 3D scans of the full body of more than 4000 subjects (see fig 2.17 for an example). In this case the body measures might be extracted by the 3D body shape and the face recognition might be performed on the 3D face model.

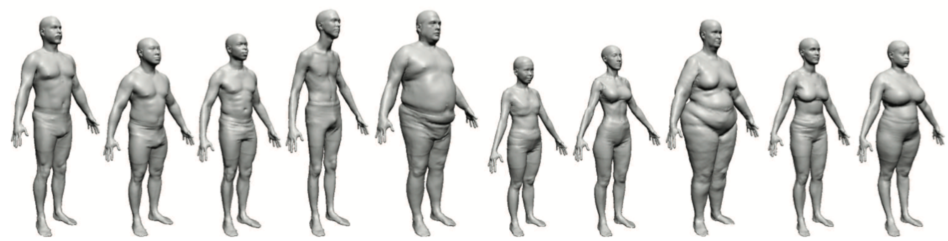


Figure 2.17: CAESAR commercial database that contains 3D full body scans of more than 4000 subjects.

2.2.3 Accessory soft biometrics

The new soft biometrics definitions allow the inclusion of accessories among the aforementioned attributes. Accessories can indeed be related to personal characteristics (as vision problems in the case of glasses), or personal choices (as adornment in case of jewelry and clothes). One of the first example is clothes color detection. According to the definition of soft biometrics these characteristics can be added to the list of already mentioned attributes.

This section focuses on the usage of accessory attributes, as well as color as soft biometrics. Also gender estimation from attributes other than face and body is discussed – examples are included for gender from iris, fingerprints or ear. Similar examples can be found in the literature for the estimation of age using alternative features, as briefly discussed in the following sections.

2.2.3.1 Clothes classification

Several works focused on the use of color information from clothes to re-identify people in video surveillance scenarios. Fig 2.18 illustrates a situation where clothing attributes were visible and can be used as a soft biometric, contributing to the identification of a suspect. One of the common ways to discriminate among different targets is represented by histogram based methods [77]. Content based retrieval systems have demonstrated how those techniques are well suited to retrieve similar images, however they are strongly affected by changes in appearance and illumination [70]. Another positive aspect of the histogram like methods [86] is that they are straightforward and fast to compute. For this reason, those works were taken as baseline techniques in [47] where the authors present a set of methods that show promising performances considering both the size of the database and the simplicity of implementation of the techniques themselves. Later, D'Angelo et al. [25] improved the performance proposing the probabilistic color histogram as new histogram like descriptor.

Figure 2.18: Image highlighting a marked suspect with covered face but wearing distinct clothing, and a face image (bottom right) of a suspect appearing to wear the same clothes [73].

An interesting yet preliminary result on clothes categorization is shown in [48] where authors are able to classify clothes in different categories as a side effect of 2D body silhouette analysis.

2.2.3.2 Eye Glasses detection

As for the beard and mustache, also for glasses the state of the art explore ways of removing this attribute so as to ameliorate the automatic face recognition results. One of the earliest works for glasses detection was performed by [79], they exploit edge detection on a preprocessed gray level image to enhance some characteristics of glasses. Certain face areas are observed and an indicator for glasses presence. The best identifying part is found to be the nose region between the eyes, where the bridge is usually located. A 3D method to detect glasses frames is presented in [144], where 3D features are obtained by a trinocular stereo vision system. Another approach for glasses extraction is employed in [145] where a face model is established based on the Delaunay triangulation. Up to now, the best results on glasses detection are achieved on thermal images from [113] where in the data fusion process, eyeglasses, which block thermal energy, are detected from thermal images and replaced with an eye template.

2.3 Benefits of soft biometric

Soft biometrics can be inexpensive to compute, discerned at a distance in a crowded environment, and require less or no cooperation of the observed subjects. To elaborate, we note the following benefits:

Human understandable interpretation: Soft biometric attributes have a semantic interpretation, in the sense that that they can provide a description that can be readily understood by humans. For example, the description “young, tall, female”. This makes them particularly useful in applications such as video surveillance, where they are directly compatible with how humans perceive their surroundings. In other words, when a human attempts to verbally describe a person, obvious characteristics regarding the person's appearance such as gender, age, height and clothes color are often used (e.g., in police reports). This allows soft biometrics to be used in applications where traditional biometrics may be insufficient.

Robustness to low data quality: Some soft biometric attributes can be deduced from low quality biometric data. In this context, such attributes can be extracted, when primary biometric data is not conclusive, due to poor acquisition quality. For example, if the input iris image is of poor quality, one could utilize the surrounding periocular information to perform recognition, rather than relying on the iris itself.

Consent free acquisition: Soft biometrics can often be captured without the consent and cooperation of the observed subject. For example, information about a person's height or gender can be deduced from a distance.

Privacy: Since soft biometrics are not distinctive, they only provide a partial description of a person (such as "male, tall, old"). This limitation has positive privacy ramifications when it comes to extracting and storing such soft biometric data.

Filtering: Soft biometrics can often be used to filtering and indexing the large database to limit the number of searched data according to the connected person characteristics. For example, we can restrict the search for female gender.

2.4 Extraction of soft biometrics in the context of Covid-19 pandemic situation

Many soft biometrics are actually extracted from faces (e.g. age, gender, race, mustache, beard, etc.). However, recently people are wearing masks due to Covid-19 pandemic. This situation makes the extraction of some soft biometrics traits very challenging. When people are wearing masks, the use of periocular regions for soft biometric is attracting more and more attention. Periocular region refers to the periphery of eyes which contains eye, eyebrow and pre-eye orbital region as illustrated in Fig. 2.19.

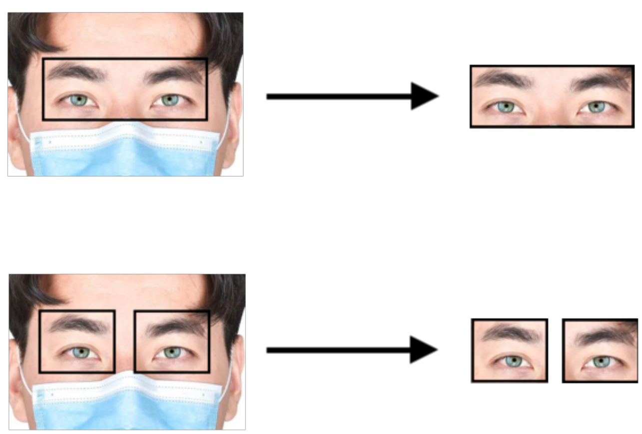


Figure 2.19: First row: Periocular region contains both eye; Second Row: Periocular region contains separate Left/Right eye.

Periocular region which can be considered as only 20 to 30% of the face area, itself is a small region of interest. Partial occlusion of such small size region may also deteriorate the performance of the system. The reason of the occlusion may be the use of eyeglasses, partial closure of eyes or pose variation.

2.5 Conclusion

This chapter was devoted to the overview of the most notable soft biometric attributes on three modalities, namely: (1) facial soft biometrics (cf. Subsection 2.2.1), (2) body soft biometrics (cf. Subsection 2.2.2) and (3) accessory soft biometrics (cf. Subsection 2.2.3). The analysis of the existing handcraft and deep learning studies are done separately in Chapters 3 and 4 respectively.
Below, we present the summary of the observations which we believe the most important in this overview chapter:

1. Among all soft biometric modalities, the face is the richest source that provides information about a person.

2. Automatic age estimation can be influenced by the gender and race differences considerably. Hence, joint treatment is not beneficial for gender and race classification. So, they can be estimated separately.

3. The database most suited to perform an effective training of a deep learning network is the IMDB-WIKI. Only MORPH II, PAL, LFW+ and PCSO databases are labeled with the gender,race and age informations.

4. Among the many anthropometric attributes height, gait and body weight concern the body and are the main attributes that can be extracted from a distance.

These observations have been utilized in the Subsection 5.3 to design our proposed approach for age estimation based on facial demographic attributes (i.e., gender, race and age group), and choose the databases that labeled with the gender,race and age to evaluate our proposed approach (cf. subsection 5.3.3).

The next chapters discuss some of existing methods (based on handcrafted features and also on deep learning) for extracting soft biometrics traits.

Chapter 3

Facial Soft Biometric Review

3.1 Introduction

Generally, facial soft biometric systems can be categorized based on handcrafted or deep learning algorithms. The main difference between these two algorithms, is the process of features extraction and selection which is accomplished manually for handcrafted models. While in case of deep learning this process is performed automatically without human interference.

In this chapter, we present a facial soft biometric review. The structure of this chapter is based on the aforementioned algorithms. We start with handcraft based methods in Section 3.2, then we present deep learning based methods in Section 3.3.

3.2 Handcraft based methods

Principally, handcrafted-based methods depend on extracting the features from images manually using set of descriptors. In fact, this process requires extra knowledge by hand. Two types of features can describe the facial components, these features are the local features and the global features [51]. The facial wrinkles and the different landmarks such as eyes, cheeks, and nose can be represented using the local features. While, global features, or holistic features is arguably the most typical technique to be used in handcrafted-based methods due to its capability to preserve configural (e.g.,the inter relations between facial regions) information, which is essential for discriminating facial soft biometric.

In this section, we will present a typical pipeline of facial soft biometric systems based on handcraft features and makes an review of the existing methods in Subsections 3.2.1 and 3.2.2 respectively. We end this section by illustrating some classification methods in Subsection 3.2.3.

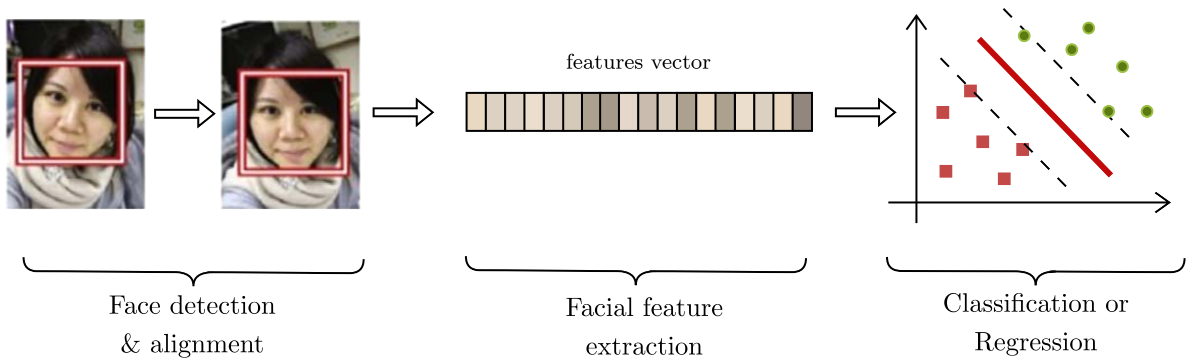


Figure 3.1: Typical pipeline of automatic facial soft biometric extraction systems. It consists of three phases:face face detection & alignment, facial feature representations and classification or regression.

3.2.1 General framework

Facial soft biometric systems based on handcrafted features, are normally follow a typical pipeline which is presented in Fig 3.1. It usually composed of three main phases, namely: (1) Face detection and alignment, (2) Facial feature representations and (3) Classification or regression.

The objective of the first phase is to extract a face or faces from an input image then scale and crop them in the same way using a set of reference points located at fixed locations in the image. while the objective of the second phase is to extract facial features based on the detected and aligned face. In the third phase the decision for the soft biometric is made. In the next subsections, we will detail each phase separately.

3.2.2 Face Detection and Alignment

Face extraction and alignment phase is normally composed of two smaller steps. The first one, face detection, is indispensable and is present in all systems of automatic face analysis. Basically, its goal is to detect a face (or faces) in an input image and to output the respective delimiting region (or regions) in the image. Face detection is a classical problem of computer vision with plenty of existing open-source solutions such as Viola and Jones [141], MTCNN [149] and Retinaface [30]. It should be noted that the form of the delimiting face region (square, rectangle, oval etc.) depends on the particular face detector. For example, in this manuscript, we employ Dlib face detector [87] which is based on Kazemi and Sullivan algorithm [85] and outputs square-sized face regions.

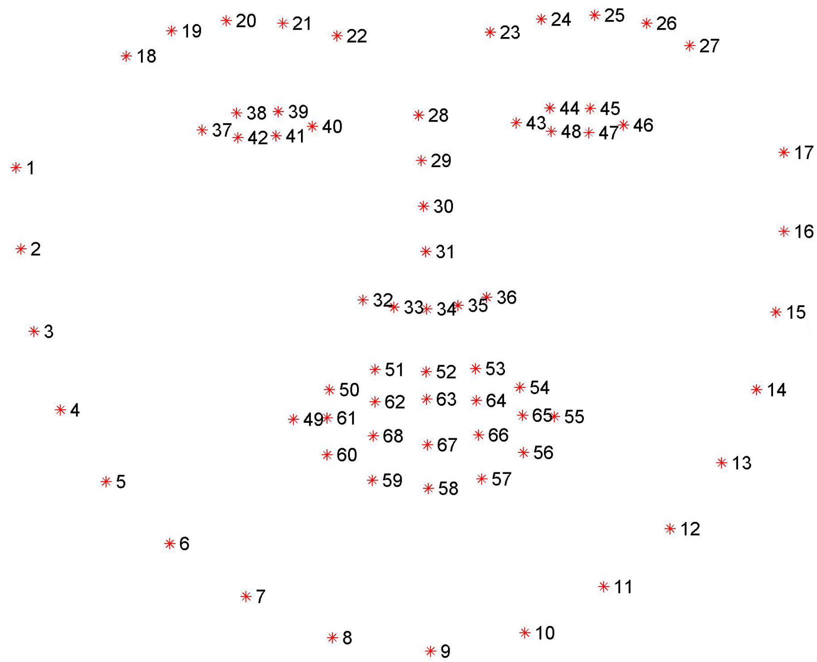


Figure 3.2: 2D facial landmarks detected by Kazemi and Sullivan algorithm [85] (figure from google image)

Sometimes, the face detected is directly given to classification or regression phase. But more often, the input faces are expected to be in a certain normalized form. For example, a face image can be aligned by rotating it so that the two eyes lie on a horizontal line, and by scaling the resulting image in order to set the distance between the eyes in pixels to some predefined value (See Fig 3.3). The described alignment process is simple but it is very sensible to precise estimation of the position of the eyes. Instead, a much more general and robust approach consists in detecting a set

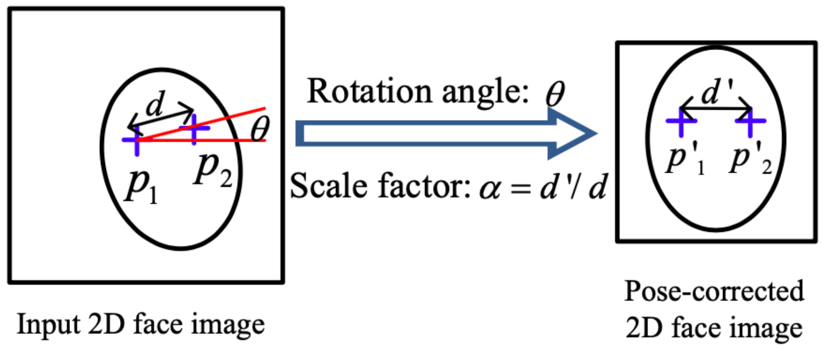


Figure 3.3: Pose correction using 2D affine transformation based on the two eye locations in [61]

of landmarks (their exact number depends on the particular implementation). In the present manuscript, we used dlib face landmark detection [87]. Fig 3.2 illustrates location of 68 facial landmarks detected by Kazemi and Sullivan algorithm [85].

3.2.3 Facial Feature extraction

Feature extraction means extracting the features from the image so that recognition is made accurate and easy. Both global and local features are crucial for face representation and recognition [9]. A number of various hand-crafted feature representations have been designed for different problems of computer vision. Often such features are well-adapted for certain problems, and are much less effective for another ones. Below, we present different approaches for description of face images, which have been utilized in previous soft biometric studies. Principally, the hand-crafted methods for feature extraction can be divided into three main groups:

3.2.3.1 Anthropometry-Based Features

Anthropometry-based features are a set of distances and ratios which are based on dimensions of the whole face/salient features (e.g., mouth, eyes, nose, etc.) and distance ratios measured from facial landmarks. The idea is to use these distances in order to describe the topological differences between male and female faces or between faces of different ages or races. Basically, the anthropometric features derive the geometric dimensions of the skull based on the provided face image. Notice that only frontal faces can be used to compute such representations, which are sensitive to head pose variations. Another disadvantage is these models consider only facial geometry while ignore texture information.

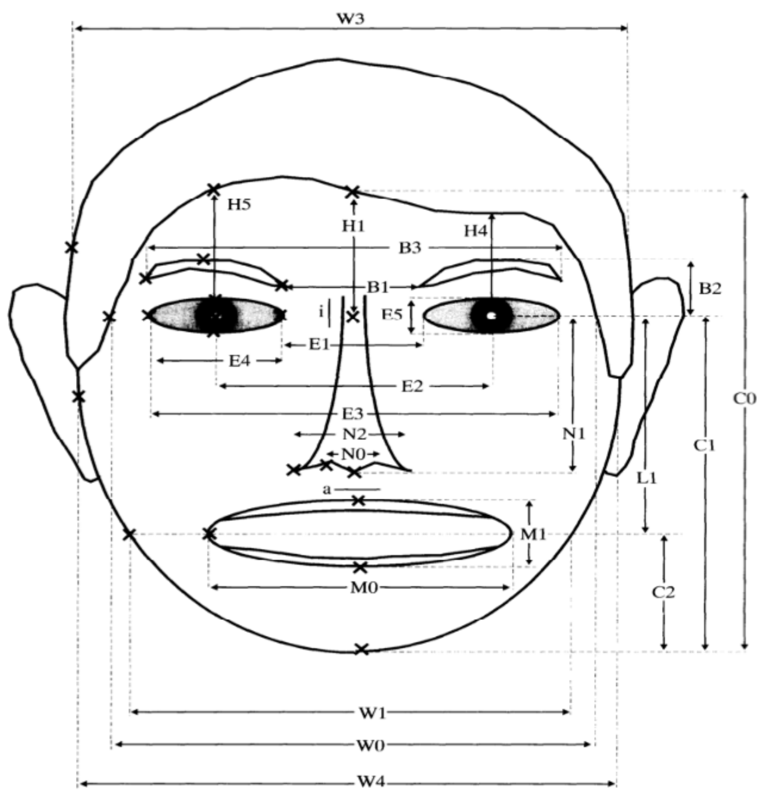


Figure 3.4: The set of 24 face fiducial distances used by *Fellous* [40].

For example, one of the first attempts to classify age from faces was performed by Kwon and Lobo [93], where the authors presented an age classification method based on facial images by computing distance ratios of different facial features (i.e., eyes, nose, mouth, chin,...etc) and detecting the presence of wrinkles. Their method can classify the input faces into one of three age groups (babies, young adults, and senior adults). Poggio et al.[120] attempted to distinguish cranial shapes of male and female faces. To this end, the authors used 15 fiducial distances: pupil to eyebrow separation and nose width appeared to be the most discriminative among them. Later, Fellous [40] extended the previous study [120] by proposing 24 fiducial distances for gender recognition (See Fig. 3.4). Indeed, in addition, to the distance between the eyes and the eyebrows and to the width of the nose, the total width of the face and the distance between the two eye pupils were also found useful for race classification in [151].

Dehshibi and Bastanfard [29] computed the geometric ratios from the facial images to extract the related features of facial components and texture and used these features for image classification into four age groups. While this approach is considering only the geometric features, it might be inappropriate for adults and old people since the appearance of the skin is the noticeable feature that represents the aging information. The set of 8 face landmarks and 6 geometric ratios are presented in Fig. 3.5.

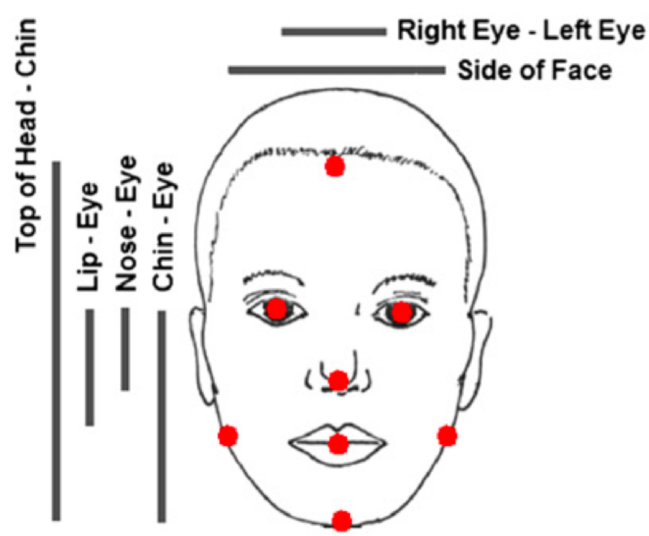


Figure 3.5: The set of 8 face landmarks and 6 geometric ratios for anthropometric recognition of age used in [29].

3.2.3.2 Texture-Based Features

Many studies on soft biometric from face images are based on extraction of both global and local texture features. These texture features can be directly extracted from images using the intensities (gray level or color) of pixels. One of the most primary and effective texture descriptor is Local Binary Patterns (LBP) [2]. This approach was employed by several facial soft biometric studies with various classification algorithms. Yang and Ai [148], introduced a demographic classification method involving age group classification, namely child, youth and old. They considered Local Binary Patterns Histogram (LBPH) feature for ordinary binary classification problems. The step descent method was applied to find an optimal reference template which is used as a measurement of confidence for classification. Their method achieved 6.82%, 7.88% and 12.5% error rates for age classification on SNAPSHOT, FERET and PIE databases respectively.

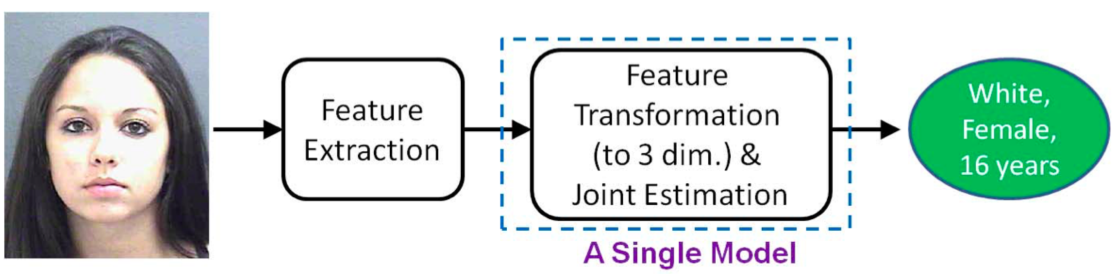


Figure 3.6: Han et al [55] hierarchical approach consisting of three binary classifiers.

A few other hand-crafted features rather than LBP were also experimented for facial soft biometric. Guo and Mu [55] proposed a complete framework that can estimate the age, gender, and race traits jointly within a single step (See Fig. 3.6). They adopted two methods of linear dimensionality, which are the Partial Least Squares (PLS) model [127] and Canonical Correlation Analysis (CCA) based methods [66]. These methods are also used to reduce the dimensionality of the original feature space. Their framework produced a Mean Absolute Error of 3.92 years on the MORPH-II database.

Another approach to determine the demographic attributes proposed by Hadid and Pietikäinen [59] considered LBP based spatio-temporal representation as a baseline system for combining spatial (i.e., facial structure information) and temporal (i.e., dynamics information) features for facial demographic classification from video sequences. Moreover, the correlation between the frames through manifold learning has been encoded. Chang and Chen [17] introduced a cost-sensitive ordinal hyperplanes ranking method to estimate human age from facial images while a novel multistage learning system which is called grouping estimation fusion (DEF) was proposed to classify human age.

However, many researchers have been attempted to develop new approaches that can estimate the facial age based on the results of race and gender classification. For instance, Han et al [62] presented a hierarchical approach consisting of three binary classifiers to build a two-level binary decision to predict gender and race attributes followed by a separate SVM regressor trained within each group to make an accurate age prediction (See fig 3.7). More recently, the same idea has been explored by Bekhouche et al. [9]. They proposed a hierarchical classifier with three layers to adopt a learning system for facial demographic attributes estimation. According to the highest accuracy of race estimation in their experiments, they chose to estimate the race of the input face as root of their hierarchical estimator. The gender is then estimated based on the race predicted using a corresponding classifier. Finally, based on the predicted race and gender, the age is estimated.

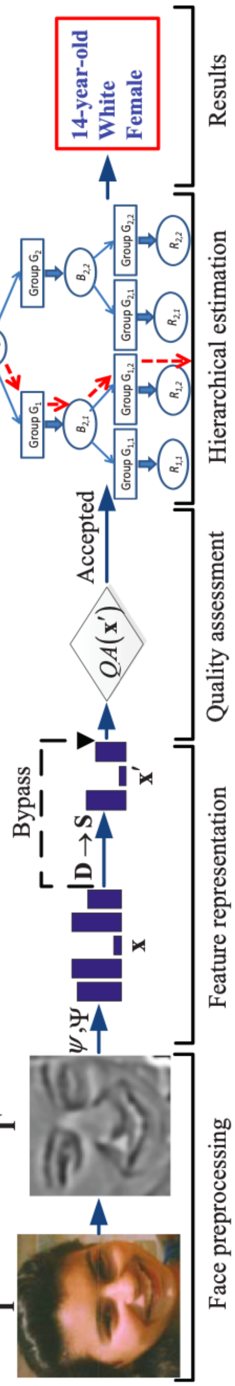


Figure 3.7: Han et al [62] hierarchical approach consisting of three binary classifiers.

Further, in [61] the authors addressed the facial appearance variations which are considered as the more challenging problem of automatic demographic estimation from face images acquired in the wild (real-life face images) by presenting a complete framework that involves face normalization, feature extraction, and age group (or exact age), gender, and race estimation. The Biologically inspired features (BIF) and Support Vector Machines (SVM) were used in extraction and classification steps respectively. The same problem was addressed in [36], by presenting a robust face alignment technique. They provided a unique more challenging benchmark of face images, acquired under real-world conditions, labeled only for age and gender, to develop their proposed method and to evaluate performances.

3.2.3.3 Appearance-Based Features

The appearance features are based both on shape and texture information. A typical method for extracting appearance features is Active Appearance Models (AAM). It initially proposed by Cootes et al. [24] and commonly used to represent the facial image. The learning step for the shape and texture models is performed through the training process on some images. AAM separately applies PCA to learn a statistical shape model and an intensity model of face images.

Later, Lanitis et al. [94] extended AAM for age modelling by proposing an aging function to explain variations in ages. Every image in the face database is described by a set of parameters b, and for each subject the best age function is drawn depending on his/her b. The greatest advantage of this approach is that different subject-based age functions allow taking into account for external factors which contribute towards the age variations. The authors tested this approach on a database of 12 people, with 80 images in the gallery and 85 in the probe. They reported an improvement of about 4–8% and 12–15% swapping the probe and gallery set.

Appearance-Based Features was applied for gender classification in [146], where

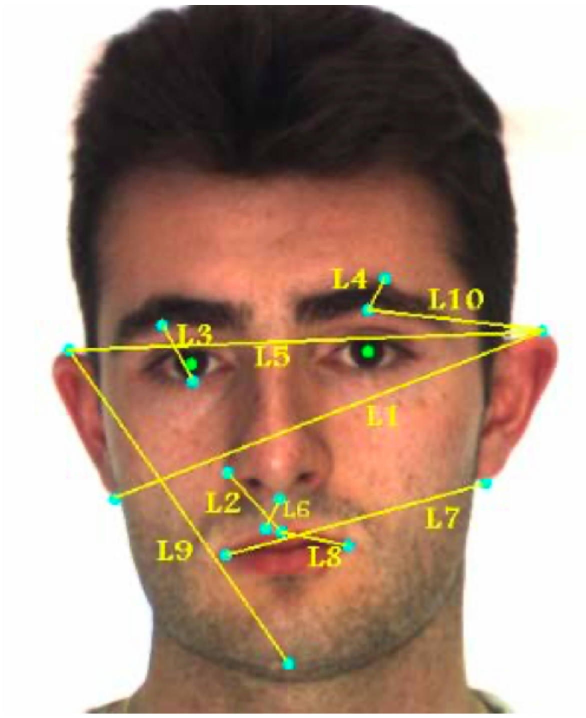


Figure 3.8: 10 features extracted with AAM that used in [146].

the authors proposed a hybrid method using local features (10 features extracted with AAM) as illustrated in Fig. 3.8 and global features (extracted using AdaBoost with Haar-like features). The authors show that a better accuracy can be obtained by fusing these features before classification. Overall accuracy using 5-fold cross validation on 1000 images is 92.38%.

Significantly, AAM technique shows a clear advantage over anthropometric model that it can deal with any age and gender consider both texture and anthropometric models. Nevertheless, a serious weakness with this model is the loss of some skin areas and wrinkles information because of using dimensionality reduction [56]. Another crucial issue of this techniques is the intensive computations and the need of large number of images to learn the features that are related to the shape and appearance. Moreover, the performance of AAM model depends on the image quality, so dealing with image intensities in the gray-level, may lead to a vulnerable model.

3.2.4 Classification/Regression

After extracting the facial feature vector, a classifier or regressor is used in training stage on the available datasets. In this subsection, we focus on the last step of facial soft biometric systems based on handcrafted features pipeline presented in Fig 3.1, i.e. on algorithms which perform classification or regression based on the image descriptors discussed in Subsection 3.2.3.

3.2.4.1 Classfication

Classification is a process of categorizing a given set of data into classes, It can be performed on both structured or unstructured data. The process starts with predicting the class of given data points. The classes are often referred to as target, label or categories. The classification predictive modeling is the task of approximating the mapping function from input variables to discrete output variables. The main goal is to identify which class/category the new data will fall into.

3.2.4.2 Regression

In machine learning, regression algorithms attempt to estimate the mapping function f from the input variables x to numerical or continuous output variables y. In this case, y is a real value, which can be an integer or a floating point value. Therefore, regression prediction problems are usually quantities or sizes. For example, Instead of dealing ages prediction as a multi-class problem, labels can be considered as numerical values. The common regression algorithms include linear regression, Support Vector Regression (SVR), and regression trees. Some algorithms, such as logistic regression, have the name ”regression” in their names but they are not regression algorithms.

3.3 Deep learning based methods

Deep learning methods enhancing the performance of many tasks in computer vision. In facial soft biometric system, deep learning have been effectively utilized to make a transformation of face images to output the label space [92]. A different approach is taken by deep learning based methods and especially Convolutional Neural Networks (CNN) comparing with hand-crafted based methods. The features are automatically extracted during the training process instead of defining a set of algorithms. The basic idea of using deep learning is based on trying to solve the problem in a hierarchical manner of concepts. This way of learning gives us the power to replace the hand-crafted methods with deep learning based methods [101]. In the next subsections, we firstly explain the basic principles and the intuitions behind CNNs in Subsection 3.3.1, and then in Subsections 3.3.2 and 3.3.3, we introduce, the two main common ways are used to training deep learning approaches. Deep Learning frameworks are introduced in details in Section 3.3.4.

3.3.1 Convolutional Neural Networks

Convolutional neural networks (CNN/ConvNet) is a type of deep learning model which is specifically applied on Image, video and some other 2D and/or 3D data. Since facial age estimation research deals with image and video data only, most of the recent research works on facial soft biometric deal with CNN models. CNNs perceive images as three- dimensional volumes. These input and output volumes after each layer of CNN model are expressed mathematically in terms of multi-dimensional matrices. As the image moves deeper into a CNN, the dimensions of these matrices get transformed after every layer. A general structure of a CNN model is depicted in Fig. 3.9.

CNN frameworks are built up of a well decided combinations of a few layers

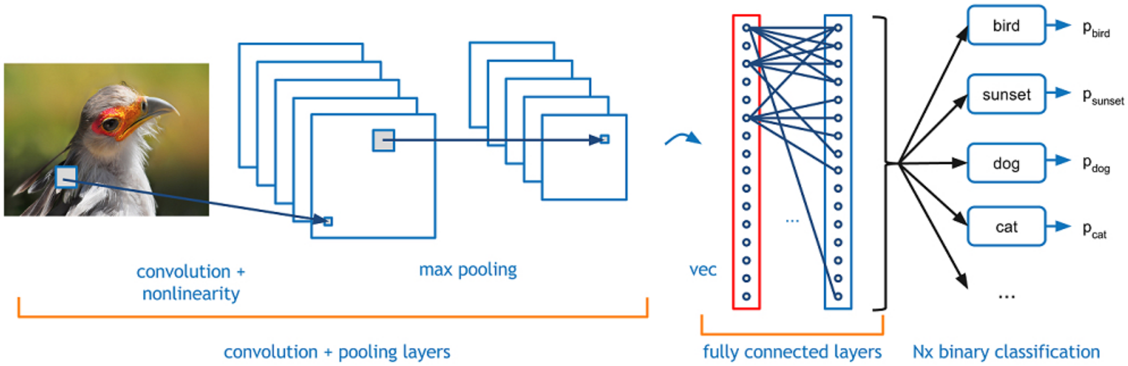


Figure 3.9: A general structure of a convolutional neural networks.

named as convolutional layers, sub-sampling or layers and Fully connected layers. Convolutional layers are used to perform the basic task of convolution. This layer acts as the main building block of a CNN. Sub-sampling layers like max pooling reduce the special size and/or the parameter size using MAX or AVERAGE operation. This layer performs a major task of controlling the problem of overfitting. Finally, neurons of a fully connected layer map all the activations in the previous layer. This layer is connected at the end of a CNN model. Moreover, there is an additional RELU layer which can be used to implement the function of non-linearity and rectification in CNNs. A SoftMax layer is often used as the final layer and performs the function of assigning decimal probabilities to each outputted neuron. Depending on this basic concept, researchers have proposed many Convolutional Neural Network architectures (Sabharwal et al. 2018) in order to implement their research ideas. Popular CNN models such as AlexNet, VGG-16, GoogLeNet and ResNet are shortly introduced in the following subsections.

3.3.1.1 AlexNet

AlexNet is named after Alex Krizhevsky, who has designed this neural network model in the year 2012 [91]. As shown in Fig. 5, AlexNet is a combination of a total of

8 layers. 5 of them are convolutional layers. Some of these convolutional layers are followed by Max-Pooling layers. Last three layers are fully connected layers. A ReLU layer is attached after every convolutional and fully connected layer. A dropout layer is also attached after every fully connected layer in order to control the problem of overfitting.

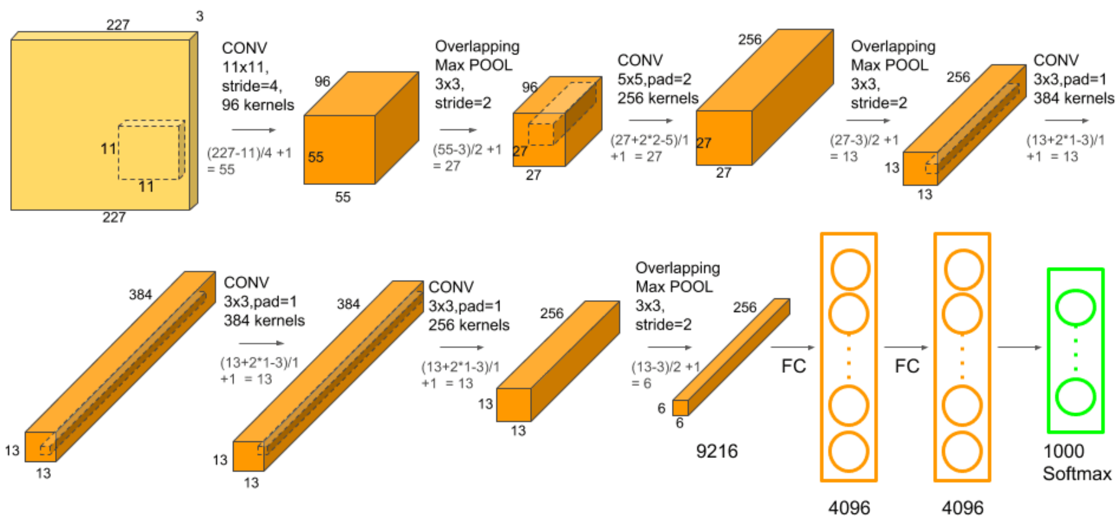


Figure 3.10: AlexNet architecture

3.3.1.2 VGG16

VGG-16 architecture is designed by Simonyan and Zisserman [137]. It is considered to be one of the excellent vision model architecture till date. VGG-16 is a very uniform architecture made up of 16 convolutional layers, 5 Max-pooling layers, 3 fully connected layers and a SoftMax layer at the output. A ReLU layer is provided after all the hidden layers. The basic architecture of VGG-16 is presented in Fig. 3.11. This network is a pretty large network and it has about 138 million (approx) parameters.

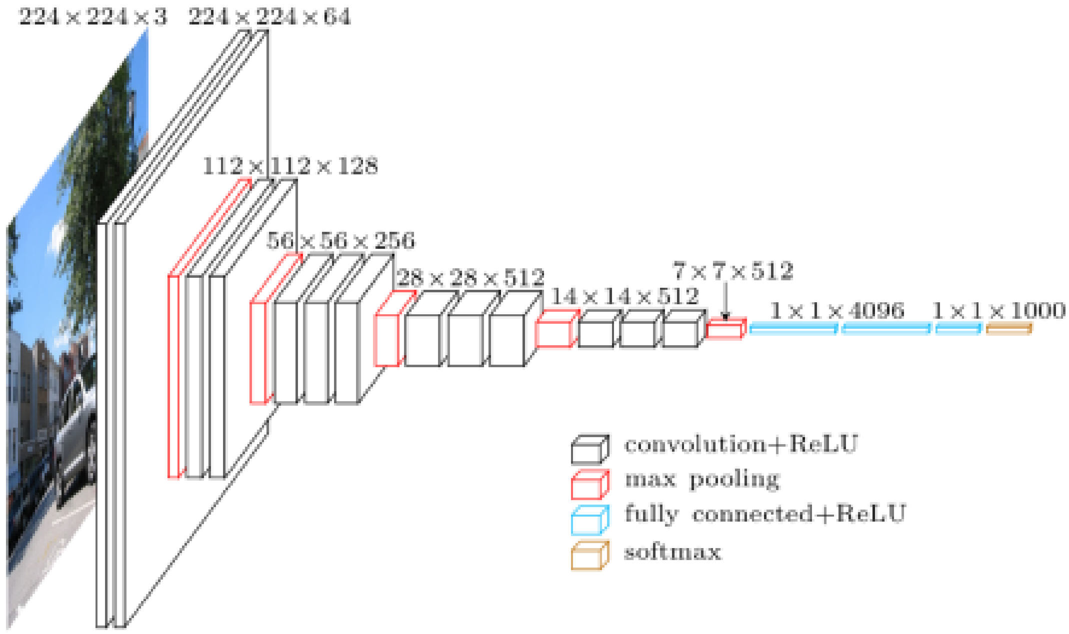


Figure 3.11: Architecture of Vgg16

3.3.1.3 GoogLeNet

GoogLeNet [140] is designed by inter-connecting nine Inception modules to recommend a deeper architecture. As a part of basic CNN architecture, Maximum pooling layers that helps in reducing the parameter size follows some inception modules for GoogLeNet architecture. Unlike Alex-Net and VGG-Net architecture, GoogLeNet does not consider a firm Convolution size. Generally, sizes of convolutional kernels are 1×1, 3×3 and 5×5 (See fig. 3.12). The 1×1 convolution layers implements feature dimensionality reduction function that is processed to broad network. GoogLeNet exceed AlexNet and VGGNet as conclusion to top-5 error that reduces to 6.67% by using 22 convolutional layers.

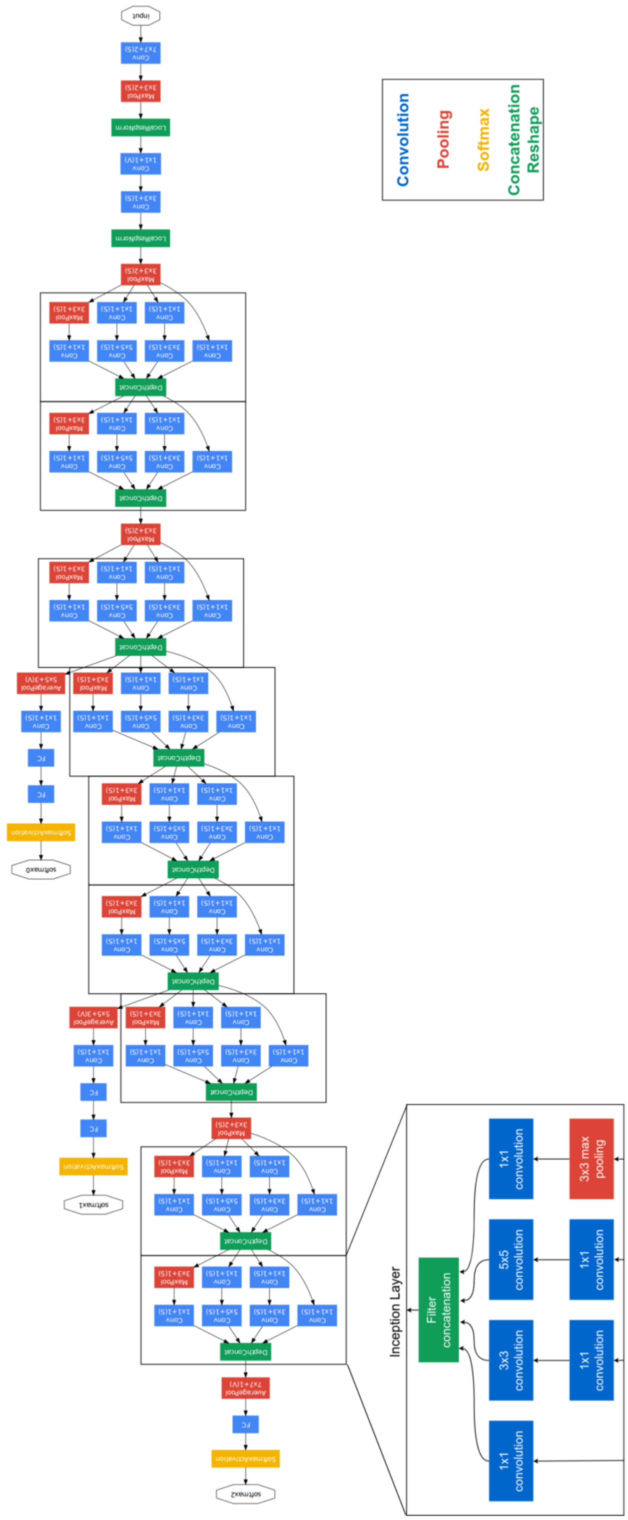


Figure 3.12: GoogLeNet architecture.

3.3.1.4 ResNet

ResNets [67] work on the principle of skip connections and allow researchers to go very deeper the neural networks without vanishing gradient and higher training error. A ResNet module can be built by using identity mapping to create a path between the input and the output layer and skipping some in between layers (See Fig. 3.13). Various types of ResNet models studied so far are ResNet-18, ResNet-34, ResNet-50, ResNet-101 and ResNet-152.

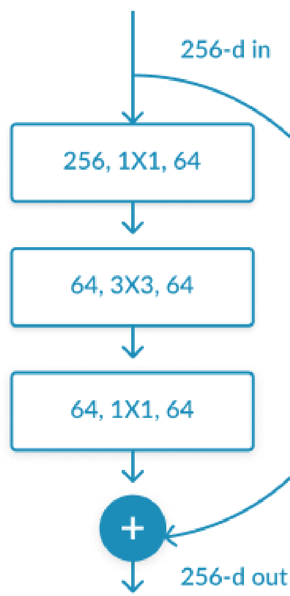


Figure 3.13: ResNet skip connection

Canziani et al. [14] provide an excellent comparison of popular CNN architectures shown in Fig. 3.14. It provides more informative view of the accuracy values of CNN architectures, because it also visualises computational cost and number of network's parameters. The first thing that is very apparent is that VGG, even though it is widely used in many applications, is by far the most expensive architecture, both in terms of computational requirements and number of parameters. The other ResNet architectures form a curve growing line.

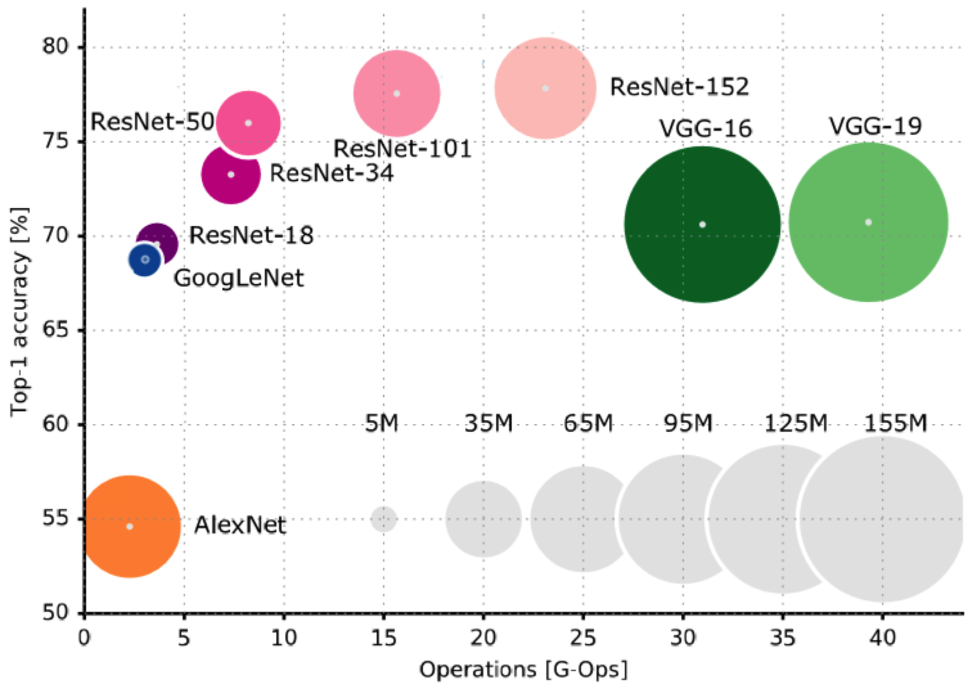


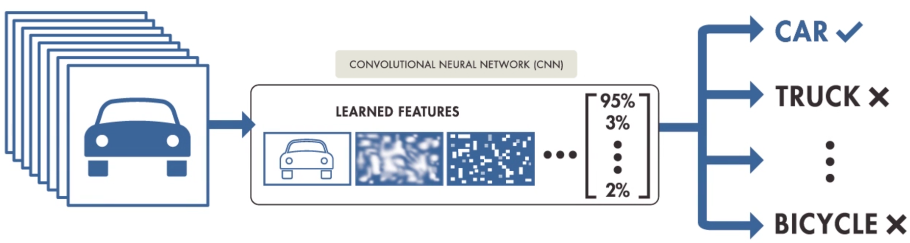
Figure 3.14: Comparison of popular CNN architectures. The vertical axis shows top 1 accuracy on ImageNet classification. The horizontal axis shows the number of operations needed to classify an image. Circle size is proportional to the number of parameters in the network [14].

3.3.2 Training deep learning approach

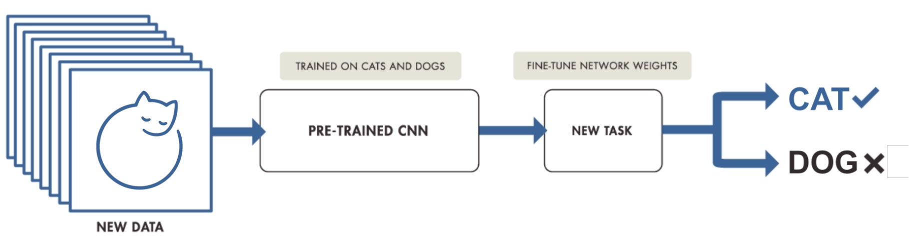
Learn a series of nonlinear features using deep learning enhances the performance of many tasks in computer vision. All CNN architectures can be employed by doing minor modifications in the architecture. To perform deep learning, two main common ways are used:

3.3.2.1 Training from scratch

In case of having new application or large number of classes as output, it can be useful to train the network from scratch on large labeled data to allow learning the features (See Fig. 3.15.a). However, training from scratch (full-training) of CNN may not be very easy as CNN requires plenty of training data for better performance [7]. This approach is less common due to the need of high-performance graphical processing units (GPUs) also needed in training of CNN for fast processing because the training with such a big collection of data is a time-consuming process [7]. Besides, the process of training in CNN is very complicated and required constant adjustments of parameters to ensure equivalent learning of all the layers due to issues of convergence and overfitting [72].



(a)



(b)

Figure 3.15: (a) Training from scratch, (b) Transfer learning.

3.3.2.2 Transfer Learning

Due to the vastness and design complexity of deep neural network architecture, most of machine learning tasks have too few training and verification sets. More image data training is needed to further improve accuracy, sensitivity, and specificity. Transfer learning is an approach is more suitable when limited training data is available for the problem under consideration. In transfer learning, one can learn a complex model using data from a source domain where large-scale annotated images are available (e.g. ImageNet, FaceNet dataset,...etc.). Then, the model is further fine-tuned with data of the target domain where only a small number of annotated images are available (e.g. race images) [5](Fig 3.15.b)

3.3.3 Deep Learning frameworks

Along with the success of deep learning techniques in diverse areas and the availability of large number of training databases, researchers have turned their attention to learning facial representations by deep neural networks in recent years. These deep learned features [38], [128] are more representative and discriminative for facial soft biometric tasks than traditional handcrafted features.
The first use of CNN was by Yang et al. [147] to estimate demographic attributes. They created a CNN model of 5 layers and trained it from scratch. In their model, a recognition system was built to recognize the face from videos and images then extract the corresponding human profile. Many outputs were obtained from the model related to gender, age and race.

Later, Wang et al. [143] developed a relatively deeper CNN (i.e., 3 convolution + 2 pooling + full connection) to address the problem of age estimation. However, the proposed CNN is only used to extract features, which is then fed to a regressor (i.e., a linear SVR regressor) for final age prediction.

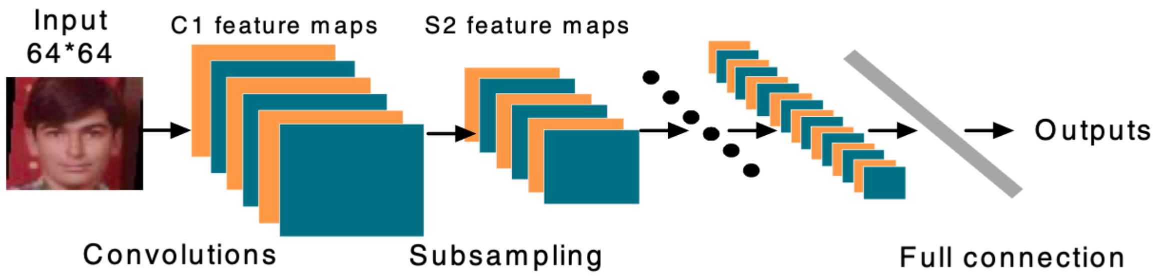


Figure 3.16: Yang et al [147] Convolutional neural network of 5 layers.

Levi and Hassner [36] used a shallow CNN architecture to classify the Adience dataset into eight age and two gender groups. They compromised between the complexity of the network with the performance to reduce the chance for overfitting. Hu et al. [69] proposed a CNNs scheme to facial age estimation without age labels by using the age difference information with three kinds of loss functions, (i.e., cross entropy loss, entropy loss, and K-L divergence distance). While Rothe et al. [128] offered a solution to real and apparent age estimation by proposing an approach called Deep EXpectation (DEX). They used a deep CNN network pre-trained on the large ImageNet images [129] with VGG-16 architecture [137] followed by a softmax function to expect value formulation for age regression. Their results are reported on FG-NET, MORPH II and CACD datasets for estimating the biological (real) age. Fig 3.17 illustrates pipeline of DEX method. . Liu et al. [101], exploited the label correlation among face samples in the transformed subspace and proposed a label-sensitive deep metric learning (LSDML) approach for facial age estimation. Then they extended it to a multi-source LSDML (M-LSDML) by using the correlation of multi-source face aging datasets to learn the label-sensitive feature similarity. Their experimental results showed the effectiveness of their approach on MORPH II, AdienceFaces, FG-NET, FACES and ChaLearn databases.

Fudong Nian et al. [116] propose to use CNN for robust gender classification in unconstrained environment. They test their method on the LFWA database and

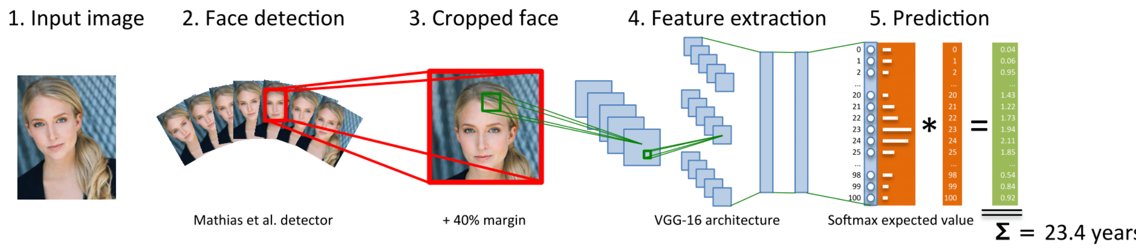


Figure 3.17: Pipeline of DEX method [128]

get the state-of-art performance of 98.8% for gender classification. In [78], Jia et al. modified VGG-16 network to create networks of depths 6, 10 and 16 and found that the increase in depth allowed larger sets having noisier training data. The results demonstrated that large depth networks perform better on gender classification tasks. Tested on the LFW dataset, the network of depth 6 achieved an accuracy of 95.80%, the network of depth 10 achieved an accuracy of 96.94% while the network of depth 16 achieved a higher accuracy of 98.25%.

3.4 Conclusion

This chapter has been devoted to a brief overview of the most prominent techniques of hand-crafted and deep learning for facial soft biometric. Therefore, we have mainly focused on the handcrafted algorithms and methods which are essential for understanding the rest of the manuscript. As explained above, the analysis of the existing facial soft biometric studies is done separately. In Section 2.2, we have introduced the most notable handcraft based methods on the main problems which are addressed in the present manuscript.

CNNs are primary deep learning models which are used in all contributions of the recent works. They have been presented in Section 2.3 as the models which have revolutionized in facial soft biometric field.

In the next chapter, we describe our contributions comparing the resulting for age prediction based on demographic classification and editing with the state-of-the-art studies presented above.

Chapter 4

Experimental Results

4.1 Introduction

In this chapter, we will describe the principle of the proposed approach in Section 4.2. Our approach consists of three main parts; 1) Automatic face detection and alignment to extract only the regions of interest (facial regions) and to correct the position and the size of faces. 2) Feature extraction from the facial regions image including both global and local texture features. 3) Tow-Stages-Estimator, where the input face is first classified into a specific demographic class using Support Vector Machines (SVM) and then a specific regressor is selected to estimate the exact age using Support Vector Regression (SVR). In Section 4.3, we will present and discuss the results of our proposed approach is evaluated on MORPH-II, PAL and a subset of LFW databases.

4.2 Proposed Approach

In this section, we present our proposed approach for facial age estimation relying on demographic attributes. In our work, we focus on the gender and the race attributes, since they are two of the most frequently facial demographic attributes reported in literature that can influence on age prediction process. Our approach consists of 3 main parts:

- Automatic face detection and alignment.
- Feature extraction.
- Demographic classification & age estimation.

The general scheme of our complete approach is illustrated in Fig 4.1. Detailed description of each part is provided in the following subsections.

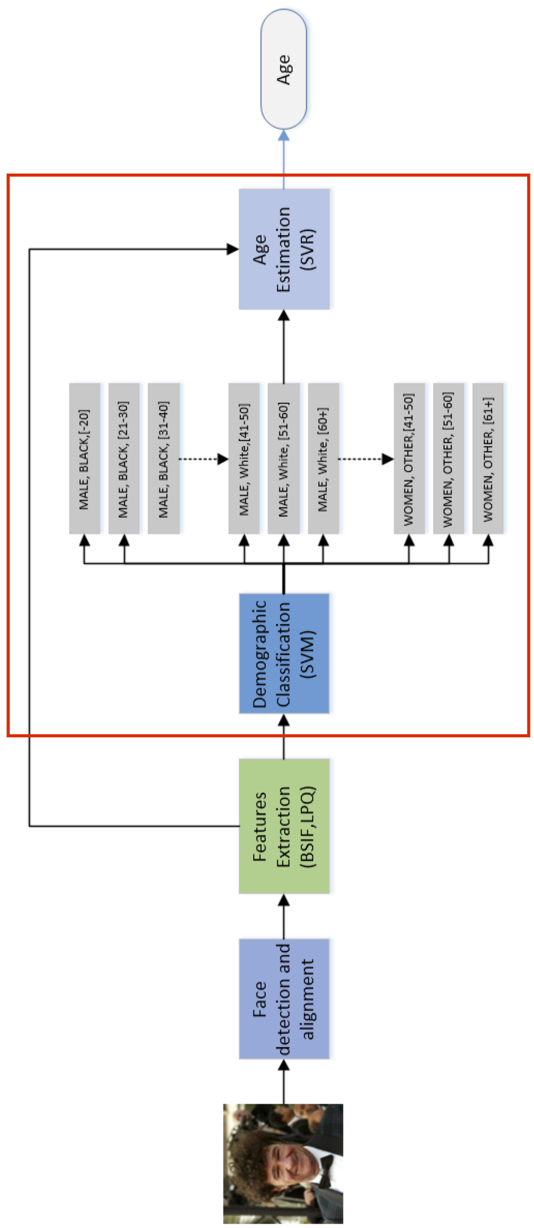


Figure 4.1: The general scheme of the complete proposed approach.

4.2.1 Automatic face detection and alignment

The original face images can be affected by the many factors such as illumination conditions, pose variation, etc. These factors led to misalignment between face images. As already discussed in the Subsection 3.2.2 of Chapter3, we adopted an alignment process based on eyes center localization, which includes:(i) face and landmarks detection, (ii) eyes localization, (iii) pose correction, and (vi) regions of interest (ROI) cropping (see Fig 4.2).

Firstly, to alleviate the influence of inconsistent colors, we transformed the input face color images to gray scale images. Next we detected the 2D facial landmarks by using Dlib open source Python library [88] which is implemented by Kazemi and Sullivan framework [85]. It detects 68 facial landmarks located on the mouth, nose, eyes, and eyebrows (See Fig 3.2). In our study, we only used 12 facial landmarks which are located around the eyes to set (X_{right}, Y_{right}) and (X_{left}, Y_{left}) coordinates of right and left eye respectively, by applying the center of pressure formula. These two-points coordinates enable us to correct the pose by rotating the face images according to the rotation with an angle β which is defined as follows:

$$\beta = \tan^{-1}(\frac{Y_{right} - Y_{left}}{X_{right} - X_{left}}) \tag{4.1}$$

Next, the rotated images are scaled to the same inter-pupillary distance d that is measured from the eyes centers. Finally, the ROIs (aligned faces) are cropped to a standard size of 200 x 200 pixels. Fig 4.2 schematizes the above automatic face detection and alignment procedure.

4.2.2 Feature extraction

The feature extraction step is performed on each aligned face. In order to extract the features, the Multi-level (ML) face representation [115] is used. The main idea behind

(a) Input face (b) Landmarks detection (c) Eyes localization (d) Pose correction (e) Regions of interest cropping

Figure 4.2: Automatic face detection and alignment.65

applying the ML face representation is to get various global and local texture features from the whole ROI. The ML representation represents the aligned face image at different levels (see Fig 4.3). In each level, the ROI is divided into n^2 non-overlapping local blocks, where n is the levels number (e.g., the first level is the aligned face without dividing). The features extracted from blocks in each level are concatenated to construct the level features vector. The order of concatenating is from left to right and from up to down. The features vectors of all levels are concatenated later to form the global face features vector. Fig 4.3 depicts an example of features vector constructing with Multi-Level face representation for 3 levels.

In many researches on age estimation, variety of feature descriptors have been proposed in order to extract aging features from images. They fall into two categories: learning-based and handcrafted-based. Learning-based methods include Discriminant Face Descriptor (DFD) [96], Cost-Sensitive Local Binary Feature Learning (CS-LBFL) [106], Deep Binary Descriptor with Multi-Quantization (DBD-MQ) [32], Discriminative Deep Metric Learning (DDML) [104] and Context-Aware Local Binary Feature Learning (CA-LBFL) [33]. In the other hand, handcrafted-based methods include Biologically Inspired Features (BIF) [57], AGing pattErn Subspace (AGES) [46], Local Binary Pattern (LBP) [2], Weber Local Descriptor (WLD) [20], and Histograms of Oriented Gradients (HOG) [26]. Some of the latter methods, such as LBP and HOG, consider the face image as a texture pattern.

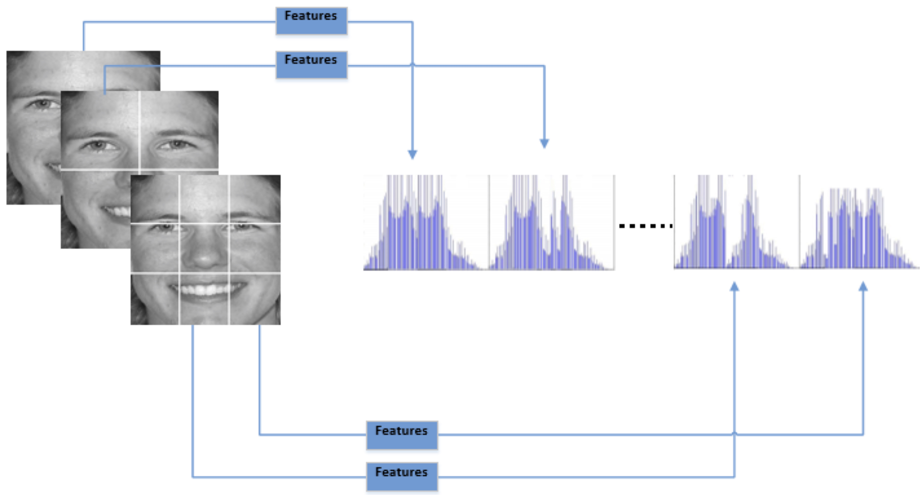


Figure 4.3: Exemple of Multi-Level face representation for 3 levels.

In this manuscript, two texture descriptors are used: *Binarized Statistical Image Feature (BSIF)* and *Local Phase Quantization (LPQ)*. The choice is based on feature extraction descriptors analyse that was done in the section 3.2.3 of Chapter 3, where demonstrated that BSIF and LPQ descriptors are powerful texture descriptors for demographic estimation [9].

4.2.2.1 Local Phase Quantization (LPQ)

Ojansivu et al in [118] proposed a spatial blurring method which used to build LPQ face description. In the frequency domain, the Fourier transforms of the blurred image $G(u)$ is the multiplication of an original image $F(u)$ with the Point Spread Function (PSF) of the blur $H(u)$.

$$G(u) = F(u).H(u) \tag{4.2}$$

Where u is a vector of coordinates $[\mathrm{u,v}]^T$. In LPQ, the phase is examined in local neighborhoods N_x at each pixel position x of the image $f(x)$. The short-term Fourier transform (STFT) is applied to compute local Fourier coefficients at four frequency points $u_1 = [a, 0]^T$,$u_2 = [0, a]^T$,$u_3 = [a, a]^T$,$u_4 = [a, -a]^T$ where a is a sufficiently small scalar to satisfy $H(u) > 0$.
The phase information in these coefficients can be counted by using a simple scalar quantization

$$q_i(x) = \begin{cases} 1 & \text{if } g_i > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

Where g_i is the j-th component of the vector $G(u)$. The resulting eight binary coefficients $q_i(x)$ are represented as integer values between 0-255 using binary coding:

$$f_{LPQ}(x) = \sum_{j=1}^{8} q_j(x) * 2^{j-1} \tag{4.4}$$

Finally, the histogram of these integer values is considered as a feature vector.

4.2.2.2 Binarized Statistical Image Feature (BSIF)

A Binarized Statistical Image Feature (BSIF) [83] is a local image texture descriptor that characterizes a pixels surrounding (patch) by a binary code. The bits in the code string are generated by binarizing the convolution results between the patch X of size $M \times M$ pixels with a linear filter H_i ($i = 1....n$) of the same size. The convolution operation is defined as:

$$S_i = \sum_{v,u} X(v, n) * H_n(v, u) \tag{4.5}$$

Each bit $b_i(i = 1...n)$ of the binary code is considered by setting a threshold at zero for the response S_i as shown in equation (3):

$$b_i = \begin{cases} 1 & \text{if } S_i > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4.6}$$

Consequently, the number of filters used determines the length of the binary code. Notice that the filters are learnt from natural image patches by maximizing the statistical independence of the filter responses. In all our experiments, the BSIF descriptor has been used with filters of size 13x13 and 8 bits.

4.2.3 Two-stages age estimation

The human age estimation using a generic classifier or regressor is an intrinsically challenging problem and many studies have been proposed to achieve small estimation error. As we mentioned aforementioned (Subsection 2.1.2 of Chapter 2), the age group classification (i.e., age grouping) is one of the famous solutions proposed for this problem. Another solution is introduced in [53] by dividing the prediction age into two phases: The first one classifies the input face images only into gender-race groups. Next, in the second phase, an age prediction model is learned on each classified gender-race group to estimate the age. This approach is conducted on a pre-selected subset of MORPH II database with the same number of face images from two races (Black and White) and two genders (Male and Female).

Unlike these studies, in this manuscript we present a novel Two-Stages Estimator (TSE), for human age estimation based on three facial demographic attributes (i.e., gender, race and age ranges). This is the main contribution of this manuscript. The TSE consists of two stages as shown in Fig 4.1:

1. Demographic classification;

2. Age estimation within demographic classes;

In the first stage, the test subject features vector is automatically classified into one of different demographic classes using support vector machine (SVM), where each class has three demographic information integrated in the same label $Y_{i,j,k}$, where i, j and k denote gender, race, and age range labels respectively. The number of demographic classes is decided based on the number of races and genders in the available training set. For example, if the training set is labeled with two genders, three races and six age ranges, we will have thirty-six demographic classes.

After the demographic classification stage is carried out, in the second stage, for each demographic class, an age estimation model is learnt using support vector regression (SVR). Noteworthy that the same features are used in both stages.
The recent CNNs based methods typically implement the extracting features and classifying in just one step with billions of arithmetic operations. They require high-cost hardware to implement high computational complexity of their algorithms [58]. In contrary to these methods, we intentionally adopted a pipeline based on feature extraction and feature classification, which are performed separately, like in [15], [8] and [1]. This pipeline allows making series of stacked classifiers and avoiding disadvantages of the CNN methods.

4.3 Experiments

In this section, we first introduce the databases and the evaluation protocols of our experiments. After that, we will give description about experimental setups. Finally, we discuss the results obtained on each databases.

4.3.1 Databases

In Subsection 2.1.7 of Chapter 2, we have introduced different databases that have been used to evaluate facial soft biometric systems. Since, our proposed method based on facial demographic attributes. We only used the databases that labeled with the gender, race and age, to evaluate the performance of the proposed approach, namely: MORPH II, PAL and LFW+. We did not use PCSO database due to its large size (1.5 million images).

The facial images of MORPH-II and PAL databases were collected under controlled and cooperative scenarios: frontal facial images, illumination variation, and facial expression. On the other hand, LFW+ database is a very challenging database, since their facial images were captured in uncontrolled environments (from the real life), non-frontal views, occlusions, and law image quality. All these databases are labeled with the gender, race and age needed for our algorithm.

In our experiments, the entire age of each database is divided into six non-overlapping ranges of 10 years (i.e., decades of life: -20, 21−30, 31−40, 41−50, 51−60 and 61+). We avoided dividing the whole database into more than 6 ranges since the number of training samples in each group becomes too small.

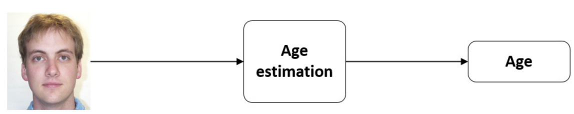
4.3.2 Experiment Setup

The demographic classification is a classification problem. While age estimation is naturally formulated as a regression task. To this end, we used LIBLINEAR library [39], which supports logistic regression and linear support vector machines (SVM). Their optimal parameters c and ϵ were founded by a grid search on the training phase for the both stages of our proposed approach (demographic classification and age estimation).

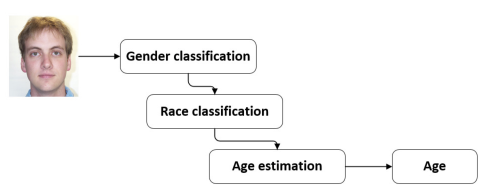
In order to evaluate our proposed approach, we considered 5-fold cross validation and reported the average performance over the 5 folds, which is frequently used for MORPH II and PAL [114], [139], [9], [102]. This protocol randomly considers 1/5 of data samples as test set, and the rest is used as training set. It assures that a sample is not in the both sets simultaneously. For this reason, the 5-fold cross validation protocol is used for all databases.

4.3.3 Experimental Results

In this study, our purpose is to investigate whether our proposed approach provides more accurate results than other age estimation approach. Towards this goal and to investigate the effectiveness of our approach, we conducted a comparative study for three automatic age estimation methods: 1) direct age regression without relying on results of other attributes (Fig 4.4.a); 2) Hierarchical age estimation method (Fig 4.4.b) using two different orders; 3) Our proposed approach TSE to estimate the age based on the result of combining gender, race and age ranges classification. In the next subsections, we present the experimental results of each method.



(a)



(b)

Figure 4.4: Two different age estimation methods, (a) direct age estimation (b) Hierarchical age estimation.

4.3.3.1 Direct age estimation

Direct age estimation is the basic pipeline which has been proposed for age estimation [152], [8], using a single level age regression. The MAE results for this method on MORPH II, PAL and a subset of LFW databases, with different image descriptors (BSIF, LPQ and BSIF+LPQ). The cumulative scores on PAL database are shown in Table 4.1 and Fig 4.6.

Table 4.1 lists the results of direct age estimation using different texture features (BSIF, LPQ features and their combination) in terms of MAE. We can observe, when using BSIF features for MORPH II database, the MAE is 3.68 years, which slightly out-performs LPQ features providing a MAE of 3.88 years. However, the combination of BSIF and LPQ features achieves much lower MAE (3.41 years). In PAL and subset of LFW databases, the combination between BSIF and LPQ features also performs better than BSIF or LPQ features used separately.

Database	BSIF	LPQ	BSIF+LPQ
MORPH II	3.68	3.88	3.58
PAL	6.01	5.88	5.55
Subset of LFW	9.74	10.18	9.07

Table 4.1: Mean Absolute Error (MAE) of age estimation on MORPH II, PAL and subset of LFW databases (in years) using direct age estimation.

Fig 4.5, 4.6 and 4.7, show the cumulative scores of age estimation on MORPH II, PAL and subset of LFW databases at the error levels from 0 to 10 years respectively. From Table 4.1, the age estimation performance is improved by the combination of BSIF and LPQ features in both MAE and cumulative score.

4.3.3.2 Hierarchical age estimation

In order to compare our proposed approach with an age estimation method based on demographic attributes, we performed the hierarchical age estimation method

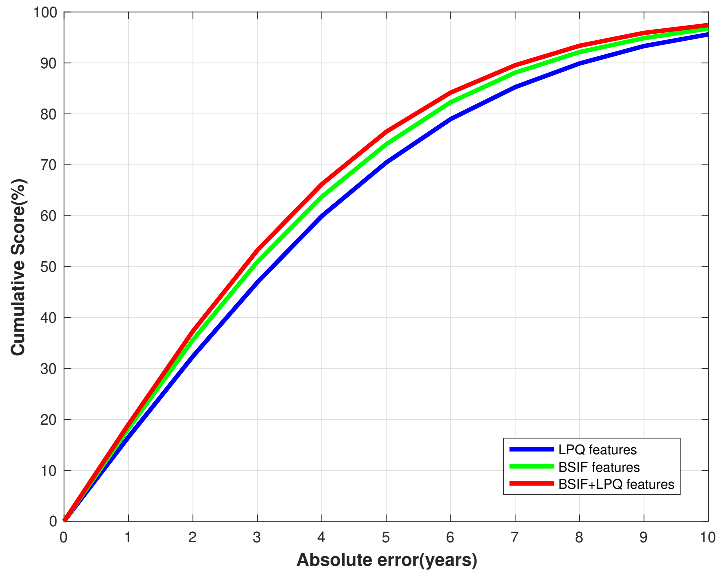


Figure 4.5: CS curves of direct age estimation at error levels from 0 to 10 years on MORPH II database.

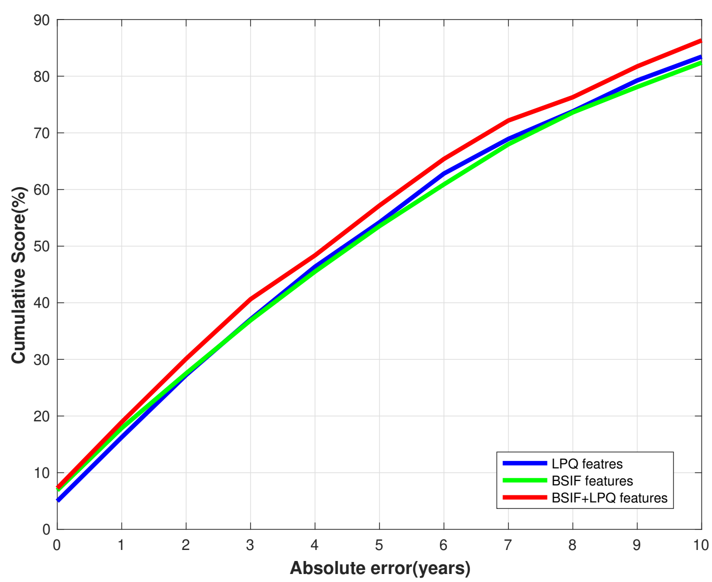


Figure 4.6: CS curves of direct age estimation at error levels from 0 to 10 years on PAL database.

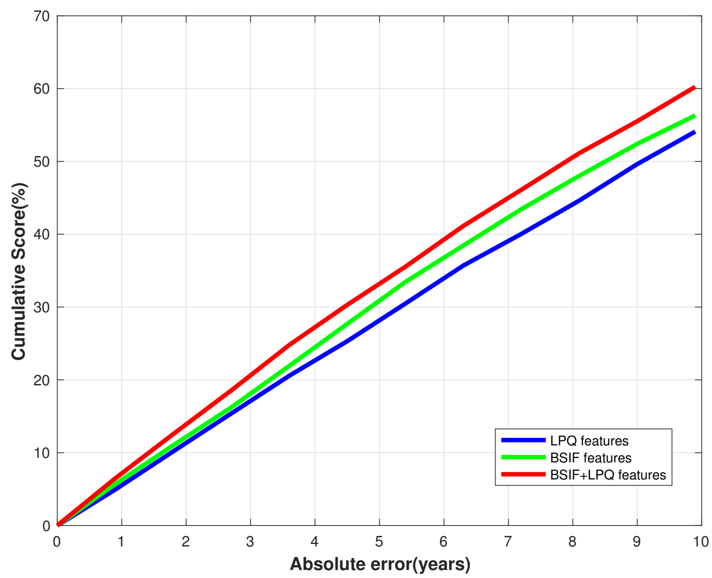


Figure 4.7: CS curves of direct age estimation at error levels from 0 to 10 years on ubset of LFW database.

for LPQ and BSIF features using two different demographic attributes orders (G→R→AR→Real age) and (R→G→AR→Real age). Where G, R and AR denote gender, race and age range attributes respectively.

The hierarchical method has been studied for age estimation task. It consists of a series of classifiers to classify each attribute separately followed by a regressor to predict age (Fig 4.4.b).

Table 4 shows the performance of hierarchical age estimation method on MORPH II, PAL, and subset of LFW databases. One can see that the best age estimation results are always obtained with (R→G→AR→real age) order and BSIF+LPQ features on MORPH II and PAL databases. These results can answer the question: Which of the two orders leads to good performances? They make sure that the order adopted in [9], where the race is chosen as the root of the hierarchical estimator, provides better MAE on MORPH II and PAL databases than the order adopted in

[63], where the gender is chosen as root. We can see also that the change in MAE value, which depends as said before on the order of attributes, is occurring regardless of their accuracy. However, we note the opposite case on subset of LFW database, where the (G→R→AR→real age) order provides better MAE than the other one. It seems that the better order must start by the attribute providing the higher accuracy (Gender as root provides 94.49% of accuracy whereas Race as root provides only 91.81%).

Database	Order of attributes		Features		
			BSIF	LPQ	BSIF+LPQ
MORPH II	G→R→AR	Gender(%)	98.53	98.36	98.91
		Race(%)	98.03	97.83	98.22
		Age ranges(%)	65.42	63.48	67.85
		Age(years)	4.02	3.70	3.70
	R→G→AR	Race(%)	98.12	97.74	98.22
		Gender(%)	98.89	98.50	99.05
		Age ranges(%)	67.25	64.03	68.58
		Age(years)	3.70	4.05	3.52
PAL	G→R→AR	Gender	96.27	94.65	96.75
		Race	97.04	96.65	97.23
		Age ranges(%)	82.79	80.11	82.41
		Age(years)	4.87	5.36	4.84
	R→G→AR	Race(%)	97.32	97.04	97.71
		Gender(%)	96.18	94.93	96.27
		Age ranges(%)	83.08	80.21	83.17
		Age(years)	4.70	5.25	4.69
Subset of LFW	G→R→AR	Gender(%)	93.90	93.56	94.49
		Race(%)	90.48	90.55	91.52
		Age range(%)	35.15	34.08	35.67
		Age(years)	8.54	8.91	8.20
	R→G→AR	Race(%)	90.93	90.81	91.81
		Gender(%)	93.26	92.54	94.04
		Age range(%)	33.84	34.48	35.64
		Age(years)	8.66	8.74	8.27

Table 4.2: Age estimation results of hierarchical method for MORPH II, PAL, and subset of LFW databases.

Fig 4.8, shows the comparison of hierarchical and direct age estimation method using BSIF+LPQ features. From the figure, it is clear that the hierarchical method provides more accurate results than direct age estimation method.

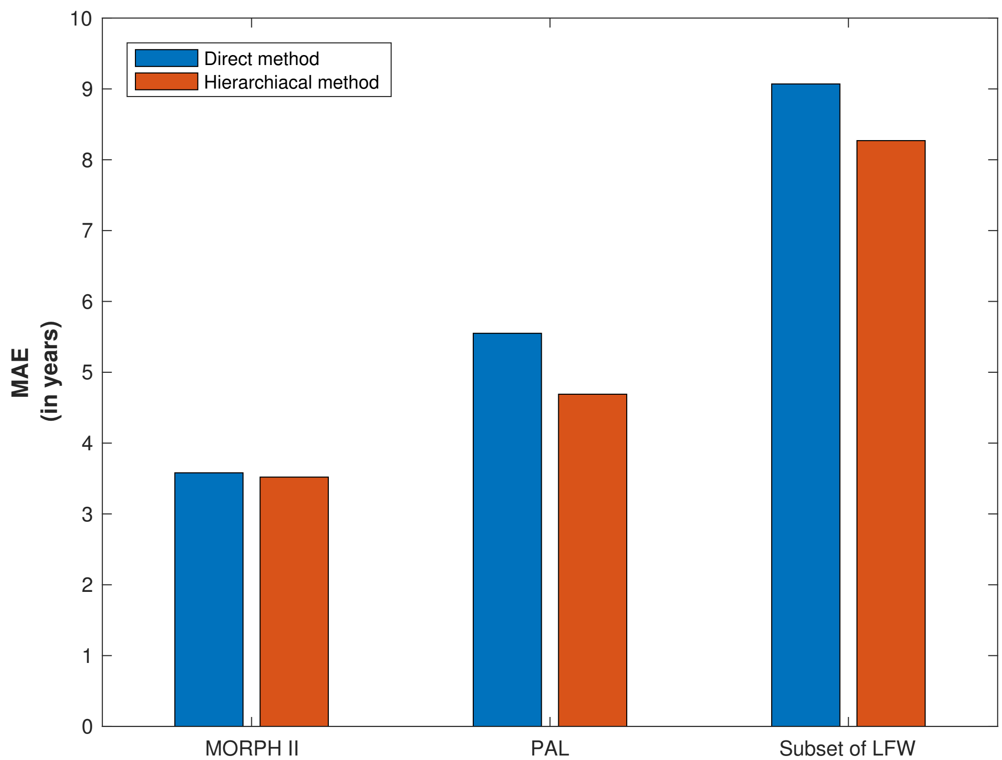


Figure 4.8: Comparison of hierarchical and direct age estimation methods.

4.3.3.3 Proposed Two-Stages age Estimator (TSE)

According to the results of the previous subsection, we found that the MAE can be decreased when the age estimation is performed across demographic attributes. In this subsection, the experimental evaluation of our TSE to age estimations is reported. The basic idea of TSE is to classify first the input test face image into one of the demographic classes. Then, based on this classification result, age estimation is performed on each classified group. We believe that estimating age within the same race, gender and age range class will provide smaller estimation errors. We conducted our experiments on MORPH, PAL and subset of LFW databases to evaluate the performance of the TSE. In Table 4.3, we organized the proposed approach results by grouping them according to the three databases. In each database, 2 results (2 sub-rows) are reported, demographic classification accuracy and MAE.

Database	Stage	Features		
		BSIF	LPQ	BSIF+LPQ
MORPH II	Demographic Classification(%)	65.49	61.75	66.78
	MAE(years)	3.39	3.68	3.21
PAL	Demographic Classification(%)	79.54	75.81	79.73
	MAE(years)	4.67	4.95	4.49
Subset of LFW	Demographic Classification(%)	29.64	26.29	30.49
	MAE(years)	8.19	8.25	7.93

Table 4.3: TSE performance results

We can see from the last column in Table 4.3 that our proposed TSE improves the age estimation accuracy for all databases. MAEs of 3.21 years for MORPH II database, 4.49 years for PAL database and 7.93 years for subset of LFW database are obtained when using the combined (BSIF+LPQ) features. These results are much smaller than direct and hierarchical age estimation results which are shown in Tables 4.1 and 4.2 (direct: 3.58/5.55/9.07 years; hierarchical: 3.52/4.69/8.27 years for MORPH II, PAL and subset of LFW databases respectively). The improvements are achieved according to the accuracy of the demographic classification. For example, the demographic classification accuracy of LPQ and BSIF texture features for PAL database are of 75.81% and 79.54% respectively. After the combination of

these features, the demographic classification accuracy is increased to 79.73%. The MAEs are reduced from 4.95 to 4.67 then to 4.49 years. The improvement of this demographic classification accuracy is not too much, but it has a great influence on age estimation result. The demographic classification accuracy increases the estimation error decreases.
Considering the performance of the proposed approach on subset of LFW database, demographic classification accuracies did not exceed the threshold of 30.49 %. They are not high compared to accuracies on MORPH II and PAL databases, even the MAE is fairly high (7.93 years). Two possible explanations for these poor accuracies:

1. The race and the age which were used as the ground truth on the training phase are not available in LFW database, they are estimated by human [63] using (MTurk) crowdsourcing service [6].

2. Most of subject images in LFW database are celebrity figures, collected from internet with large variation in pose, illumination, facial makeup and facial dynamics which make it difficult to estimate their real age.

Despite the age is estimated after classifying the integrated demographic attributes in our propose TSE, and after series of demographic attribute classifiers in hierarchical age estimation method, a question arises: why the proposed TSE approach outperforms the hierarchical method?. To answer this question, we reported in Table 4.4 the overall accuracy of the classification stage (race, gender and age range classifiers) of hierarchical age estimation method on MORPH-II, PAL and subset of LFW database. We reported also in Table 4.4 the classification stage accuracies (extracted from Table 4.3) of the proposed TSE. Comparing the two results of Table 4.4, we can see that the overall accuracies of the classification stage of hierarchical method in the two orders are higher. Whereas, we observe the opposite in terms of MAE. This can be explained by misclassification that occurs in one or more of the hierarchical classifiers, which may cause accumulation of errors. For example, on PAL database, the overall accuracies of the classification stage of the hierarchical method are 92.13% and 92.38% with 4.84 years and 4.69 years in MAE for G→R→AR and R→G→AR orders, respectively. While, in our proposed TSE, the demographic classification accuracy is 79.73 % with 4.49 years in MAE. If we look at the correct classification rates in race, gender and age range attributes taken simultaneously, the classification stage of the hierarchical method reached only 78.75% and 79.42% in G→R→AR and R→G→AR orders respectively, which are lower than the overall rates obtained by the classification stage of TSE method. This show that the errors accumulation, occurring in the classification stage of the hierarchical method, was discarded in TSE method thanks to integrating race, gender and age range attributes in a single classifier. Therein lies the outperformance of the proposed TSE than hierarchical age estimation method.

Method		Classification accuracy	Database		
			MORPH-II	PAL	Subset of LFW
Hierarchical	G→R→AR Order	Overall (%)	88.32	92.13	73.89
		Correct (%)	65.23	78.75	30.03
		MAE(years)	3.70	4.84	8.20
	R→G→AR Order	Overall (%)	88.61	92.38	73.83
		Correct (%)	65.26	79.42	3.10
		MAE(years)	3.52	4.69	8.27
TSE	Demographic classification	Overall (%)	66.78	79.73	30.49
		MAE(years)	3.21	4.49	7.93.

Table 4.4: Overall accuracy of the classification stage of hierarchical age estimation method and proposed TSE method.

Fig 4.9, 4.10 and 4.11 show the CS curves obtained by the proposed TSE approach, on MORPH, PAL and subset of LFW databases at error levels from 0 to 10 years. The ages of approximately 11.19% of the subjects in the MORPH II database, 12.05% of the subjects in the PAL database and 4.51% of the subjects in the subset of LFW database can be estimated with zero error level. Whereas when the error level increases the estimation accuracy also increases. Our proposed BSIF+LPQ based TSE is able to achieve cumulative scores of 83.13%, 74.75% and 43.64% for an absolute error of 5 years and 97%, 90.34% and 72.29% for an absolute error of 10 years for MORPH, PAL and subset of LFW databases respectively.

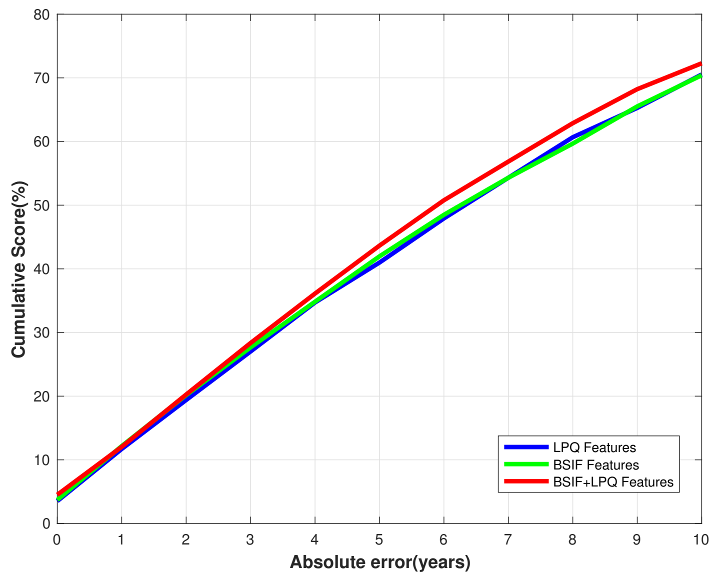


Figure 4.9: CS curves of proposed the TSE approach at error levels from 0 to 10 years on Subset of LFW database.

Next, in Tables 4.5, 4.6 and 4.7 we compare the performance of our proposed age estimation approach with recent state-of-the-art approaches in terms of MAE and CS on MORPH, PAL and subset of LFW databases. The experimental setup is the same for all compared methods. The results show that our proposed approach is most effective and can offer better performance in the age estimation task than others. Fig 4.12 illustrates some examples of correct and incorrect age estimated by the proposed approach on PAL database. Next, in Tables 4.5 and 4.6 we compare the performance of our proposed age estimation approach with known and recent state-of-the-art approaches in terms of MAE and CS on MORPH II and PAL databases

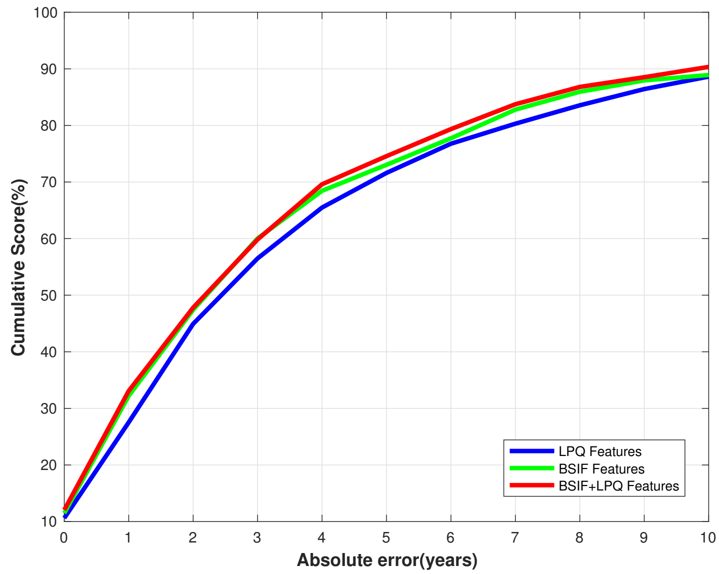


Figure 4.10: CS curves of proposed the TSE approach at error levels from 0 to 10 years on PAL database.

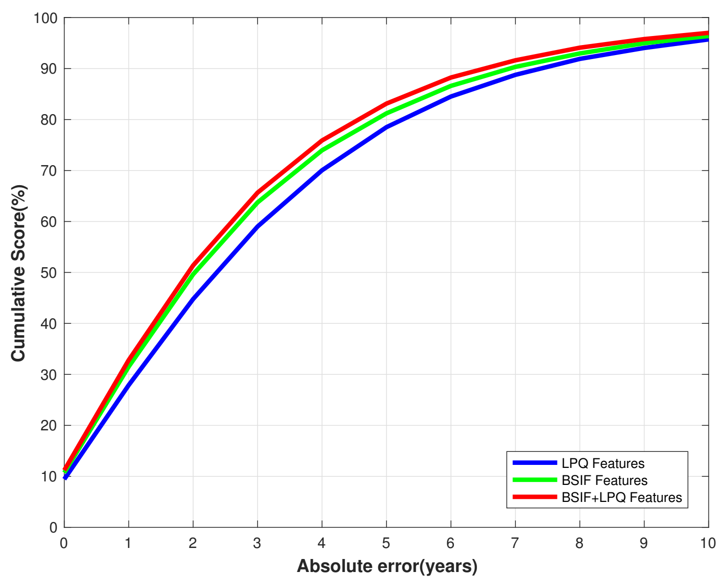


Figure 4.11: CS curves of proposed the TSE approach at error levels from 0 to 10 years on MORPH II database.

respectively. Existing methods adopted different experimental settings on MORPH-II. The 5-folds cross validation is introduced in [71], [41], [18], [9]. 10-fold cross validation is another setting which is used in [105], [100],[45]. To compare performance of our proposed TSE with different state-of-art age estimation methods on MORPH-II database, we performed experimental evaluation with both settings. The upper part of Table 4.5 shows that the proposed method out performs all the compared handcrafted state-of-the-art methods on MORPH-II database, thanks to adopting demographic classification based age estimation. Further, in the lower part of Table 4.5, we compared the performance of our proposed TSE with age estimation CNN based methods. The results show that our proposed approach can offer better performance than some CNN-based age estimation methods such as [98], [31], [143], [100] and [71]. In the literature, some works [126], [128] reported a MAE below 3 years on MORPH-II database. Based on deep learning these approaches, reached 2.56 and 2.58 years in MAE, respectively. However, they did not consider the whole database in their evaluation procedure. Indeed, they used a subset of 5,475 face images of persons, while the original database contains about 55,134 face images. Furthermore, deep learning based methods require important resources in terms of computation and memory.

Table 4.7, shows the comparison of the proposed approach with [63] which is, to the best of our knowledge, the only work on age estimation including subset of LFW database as part of their experiments. One can see from the results that the TSE method provides competitive performance. Fig 4.12 illustrates some examples of correct and incorrect age estimated by the proposed approach on PAL database.

Publication	Approach	Performance		Protocol
		MAE (years)	CS (%)	
Guo & Mu, (2011)[54]	BIF+KPLS	4.4	-	different split
Chang et al, (2011)[18]	OHRank	6.07	56.4	5-fold cross validation
Fernández et al, (2015)[41]	HOG+SVR	4.83	63.4	5-fold cross validation
Huerta et al, (2015)[71]	rCCA	4.24	71.17	5-fold cross validation
Lu et al, (2015)[105]	CS-LBMFL	4.37	74.10*	10-fold cross validation
Guo & Mu, (2013)[55]	BIF	3.98	-	different split
Bekhouche et al, (2017)[9]	Pyramid ML	3.50	75.13*	5-fold cross validation
Our	TSE	**3.21**	**83.13**	5-fold cross validation
Our	TSE	**3.17**	**83.72**	10-fold cross validation
Huerta et al, (2015) [71]	CNN	3.88	-	5-fold cross validation
Wang et al, (2015)[143]	CNN	4.77	-	5-fold cross validation
Niu et al, (2016)[117]	OR-CNN	3.27	73.42*	5-fold cross validation
Liu et al, (2017)[100]	GA-DFL (MP-CNN)	3.25	80.40*	10-fold cross validation
Duan et al, (2018)[31]	CNN+ELM	3.44	70.01	Private
Liao et al, (2018)[98]	DLF+FAM.	3.48	71.03*	5-fold cross validation
Our	TSE	**3.21**	**83.13**	5-fold cross validation
Our	TSE	**3.17**	**83.72**	10-fold cross validation
Chen et al, (2017)[21]	Ranking-CNN	2.96	85.26*	5-fold cross validation
Liu et al, (2018)[101]	M-LSDML	2.89	87.03*	5-fold cross validation
Gao et al, (2017)[45]	DLDL+VGG-Face	2.42	88.88*	10-fold cross validation

*CS result is obtained from the curves in the original paper.
The numbers in bold represent the performance of our method.

Table 4.5: Comparison of the proposed approach with known and recent state-of-the-art approaches on MORPH II database.

Publication	Approach	Performance		Protocol
		MAE (years)	CS (%)	
Choi et al, (2011)[22]	GHPF + SVR	8.44	-	5-fold cross validation
Luu et al, (2011)[107]	CAM + SVR	6.00	-	LOPO
Nguyen et al, (2014)[115]	MLBP+GABOR	6.50	-	2-fold cross-validation
Günay & Nabiyev, (2016)[50]	AAM+GABOR+LBP	5.38	-	3-fold cross validation
Günay & Nabiyev, (2017)[49]	WLD+LBP+LPQ	5.75	54.28	3-fold cross validation
Choi et al, (2017)[23]	AAM+MLBP	5.50	-	Private
Agrawal et al, (2017)[1]	HOG+GABOR	6.55	-	2-fold cross validation
Bekhouche et al, (2017)[9]	Pyramid ML	5.00	57.44*	5-fold cross validation
Our	TSE	**4.49**	**74.57**	5-fold cross validation

*CS result is obtained from the curves in the original paper.

Table 4.6: Comparison of the proposed approach with known and recent methods for age estimation on PAL database.

Publication	Approach	Performance	
		MEA (years)	CS (%)
Han et al. (2015) [63]	DIF + SVM+SVR	7.8	43.35
Our	TSE	7.9	**43.65**

Table 4.7: Comparison of the proposed approach with [63] method for age estimation on subset of LFW database.

Figure 4.12: Examples of good and poor age estimation on PAL database using BSIF+LPQ features . (a) Correct age estimates. (b) Incorrect age estimates.

4.3.3.4 Demographic attributes classification effect

In the proposed two-stages age estimator, the age is estimated after classifying the demographic attributes. The performance of age estimation can deteriorate when poor accuracies are obtained in demographic classification stage. To investigate the effect of misclassification in demographic attributes classification stage on age estimation stage, we have performed age estimation with our proposed TSE by considering the ground-truth of demographic attributes classes as the output of the first stage. In other words, we estimated the age with 100% accuracy of demographic classification stage.

	Databases		
	MORPH-II	PAL	Subset of LFW
TSE with ground-truth classes	1.17	2.50	2.74
TSE with predicted classes	3.21	4.49	7.93

Table 4.8: MAE of TSE using the ground-truth and the predicted demographic attributes classes in the first stage (in years).

Table 4.8 illustrates MAEs on MORPH-II, PAL and Subset of LFW databases, obtained by the proposed TSE method with ground-truth demographic attributes classes and with predicted demographic attributes classes. As can be seen, performing our proposed TSE with ground-truth of demographic attributes classes has improved the performance of age estimation (1.17, 2.5 and 2.74 years for MORPH II, PAL and subset of LFW databases respectively) when compared to the results obtained by using predicted demographic attributes classes (3.21, 4.49 and 7.93). The explanation is very intuitive. Estimating age within the correct (race, gender and age range) class provides smaller estimation errors and misclassified images in predicted demographic attributes classes (output of the first stage) can damage the performance of age estimation.

To study the influence of misclassified images in predicted demographic attributes classes on MAE values, we identified them in Table 4.9, and calculated their MAE for

MORPH-II, PAL and subset of LFW databases. We can clearly see, that the MAE of misclassified images in demographic classification stage are high, which adversely affects the performance of age estimation.

Database	Number of misclassified images	MAE (years)
MORPH II	18278	5.29
PAL	212	10.07
Subset of LFW	2927	9.07

Table 4.9: MAE of misclassified images in demographic classification stage of the TSE method

4.4 Conclusion

In this chapter, we proposed a novel age estimation approach based on facial demographic attributes. Face detection and alignment to extract only the regions of interest and correct the position and the size of the input face image are first performed. The features are then extracted by Multi-level face representation, using the LPQ and BSIF texture descriptors. Finally, a Two-stages estimator (TSE) for age prediction is proposed. When performing age estimation experiments over three databases using direct and hierarchical (with two different orders) age estimation methods, we observed that TSE approach outperforms significantly the two other methods. A conclusion can be drawn that age estimation can have much small error if age estimation is carried out within the same race, gender and age range. The comparison with known and recent state-of-art age estimation methods, including deep learning based methods, showed that the proposed TSE approach provides the lower MAE on PAL and MORPH II databases and competitive performance on subset of LFW database. The simplicity and efficiency of the proposed approach can provide a helpful guide to facial age estimation on mobile devices with less memory and computation resources.

Chapter 5

Conclusions

5.1 Conclusion

Human civilization has come a long way since its inception in small tribal primitive societies where every person in the community knew every other person. In today's complex, geographically mobile, increasingly electronically inter-connected information society, accurate person identification is becoming very important and increasingly difficult. It is widely believed that biometrics will become a significant component of future identification technologies and it is already of universal interest. The goal of a biometric system is to determine the identity of an individual using his/her physical/biological characteristics (known as biometric modalities). Biometric systems have many applications such as criminal identification, airport checking, computer or mobile devices log-in, building gate control, digital multimedia access, transaction authentication, voice mail, or secure teleworking. Various characteristics (or modalities) can be used: from the most conventional biometric modalities such as face, voice, fingerprint, iris, hand geometry or signature, to the so–called emerging biometric modalities such as gait, hand-grip, ear, body odour, body salinity or electro-physiological signals (EEG and ECG). Each modality has its strengths and drawbacks.

Despite the great deal of progress during the recent years, biometrics is still a major area of research. Wide range of viewpoints, occlusions, aging of subjects and complex outdoor lighting are challenges in biometric recognition. While there is a significant number of works addressing these issues, the extracting soft biometric traits has attained less attention.

Soft biometrics refers to human characteristics which are relatively easier to extract especially from distance. These characteristics include age, gender, ethnicity etc. Extracting soft biometrics is doable when considering controlled environment. However, soft biometrics in the wild is still an open research area. In this thesis, we described the problematic and proposed few solutions especially for human age prediction from face images. We obtained very interesting results, outperforming the state of the art.

5.2 Perspectives and future work

Although the extraction of soft biometrics has significantly improved during the recent years, there are still many remaining challenges including:

- **The lack of clean and well labeled data** : The question of building models from noisy and unlabeled data is still challenging. Unsupervised learning can help solving some problems. Data augmentation can also help.
- **The interdependence between different soft biometric traits remains unclear**: Although the thesis has investigated this important topic, the results showed that age can affect gender and other soft biometric modalities.
- **The potential privacy issues raised by soft biometrics are well studied**: No much has work yet investigated the possible privacy consequences of using soft biometrics especially in public places.
- **The combination of soft and hard biometrics is not well studied**: The combination of soft and hard biometrics is shown to be useful but no clear results are reported.
- **The use of soft biometrics is mainly limited to shopping and profiling**: Soft biometrics can be used to many applications but not only shopping and profiling. Among the possible applications include criminal investigations and forensics.
- **Behavioral soft biometrics have not been much explored**: Most existing work on soft biometrics are based on still images. The use of dynamic and

behavioral features is challenging and hence not well investigated.

To overcome some of the remaining challenges, efforts should be done in exploring the use of deep learning and weakly supervising learning for handling noisy and incomplete data.

References

[1] Subhash Chand Agrawal, Rajesh Kumar Tripathi, and Anand Singh Jalal. "Automatic human age estimation system using fusion of local and global features". In: *Multimedia, Signal Processing and Communication Technologies (IMPACT), 2017 International Conference on*. 2017, pp. 5–9.

[2] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. "Face description with local binary patterns: Application to face recognition". In: *IEEE transactions on pattern analysis and machine intelligence* 28.12 (2006), pp. 2037–2041.

[3] Heikki Ailisto et al. "Soft biometrics—combining body weight and fat measurements with fingerprint biometrics". In: *Pattern Recognition Letters* 27.5 (2006), pp. 325–334.

[4] A. Midori Albert, Karl Ricanek, and Eric Patterson. "A review of the literature on the aging adult skull and face: Implications for forensic science research and applications". In: *Forensic Science International* 172.1 (2007), pp. 1–9. ISSN: 0379-0738. DOI: `https://doi.org/10.1016/j.forsciint.2007.03.015`.

[5] Fouzia Altaf et al. "Going deep in medical image analysis: concepts, methods, challenges, and future directions". In: *IEEE Access* 7 (2019), pp. 99540–99572.

[6] *The Amazon Mechanical Turk (MTurk) crowdsourcing service*. `https://www.mturk.com/`. [Online; accessed 19-November-2020]. 2020.

[7] Marios Anthimopoulos et al. "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network". In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1207–1216.

[8] Salah Eddine Bekhouche et al. "Facial age estimation and gender classification using multi level local phase quantization". In: *Control, Engineering & Information Technology (CEIT), 2015 3rd International Conference on*. IEEE. 2015, pp. 1–4.

[9] Salah Eddine Bekhouche et al. "Pyramid multi-level features for facial demographic estimation". In: *Expert Systems with Applications* 80 (Sept. 2017), pp. 297–310.

[10] Alphonse Bertillon and Gallus Müller. *Instructions for taking descriptions for the identification of criminals and others by the means of anthropometric indications*. 1889.

[11] Rebecca S Bigler and Lynn S Liben. "A developmental intergroup theory of social stereotypes and prejudice". In: *Advances in child development and behavior*. Vol. 34. Elsevier, 2006, pp. 39–89.

[12] Marilynn B Brewer. *Intergroup relations.* Oxford University Press, 2010.

[13] Robert Brym and John Lie. *Sociology: your compass for a new world, brief edition: enhanced edition*. Nelson Education, 2009.

[14] Alfredo Canziani, Adam Paszke, and Eugenio Culurciello. "An analysis of deep neural network models for practical applications". In: *arXiv preprint arXiv:1605.07678* (2016).

[15] Pierluigi Carcagnı́ et al. "A study on different experimental configurations for age, race, and gender estimation problems". In: *EURASIP Journal on Image and Video Processing* 2015.1 (2015), p. 37.

[16] CFDCaP Cdc. “National health and nutrition examination survey”. In: *NCFHS (NCHS), editor (Hyattsville, MD: US Department of Health and Human Services, Centers for Disease Control and Prevention* (2005).

[17] Kuang-Yu Chang and Chu-Song Chen. “A learning framework for age rank estimation based on face images with scattering transform”. In: *IEEE Transactions on Image Processing* 24.3 (2015), pp. 785–798.

[18] Kuang-Yu Chang, Chu-Song Chen, and Yi-Ping Hung. “Ordinal hyperplanes ranker with cost sensitivities for age estimation”. In: *Computer vision and pattern recognition (cvpr), 2011 ieee conference on*. IEEE. 2011, pp. 585–592.

[19] Bor-Chun Chen, Chu-Song Chen, and Winston H Hsu. “Cross-age reference coding for age-invariant face recognition and retrieval”. In: *European conference on computer vision*. Springer. 2014, pp. 768–783.

[20] Jie Chen et al. “WLD: A robust local image descriptor”. In: *IEEE transactions on pattern analysis and machine intelligence* 32.9 (2010), pp. 1705–1720.

[21] Shixing Chen et al. “Using ranking-cnn for age estimation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 5183–5192.

[22] Sung Eun Choi et al. “Age estimation using a hierarchical classifier based on global and local facial features”. In: *Pattern Recognition* 44.6 (2011), pp. 1262–1281.

[23] Sung Eun Choi et al. “Age face simulation using aging functions on global and local features with residual images”. In: *Expert Systems with Applications* 80 (2017), pp. 107–125.

[24] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor. "Active appearance models". In: *IEEE Transactions on pattern analysis and machine intelligence* 23.6 (2001), pp. 681–685.

[25] Angela D'Angelo and Jean-Luc Dugelay. "People re-identification in camera networks based on probabilistic color histograms". In: *Visual Information Processing and Communication II*. Vol. 7882. International Society for Optics and Photonics. 2011, 78820K.

[26] Navneet Dalal and Bill Triggs. "Histograms of oriented gradients for human detection". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 1. IEEE. 2005, pp. 886–893.

[27] Antitza Dantcheva et al. "Bag of soft biometrics for person identification". In: *Multimedia Tools and Applications* 51.2 (2011), pp. 739–777.

[28] Danilo De Angelis et al. "New method for height estimation of subjects represented in photograms taken from video surveillance systems". In: *International Journal of Legal Medicine* 121.6 (2007), pp. 489–492.

[29] Mohammad Mahdi Dehshibi and Azam Bastanfard. "A new algorithm for age recognition from facial images". In: *Signal Processing* 90.8 (2010), pp. 2431–2444.

[30] Jiankang Deng et al. "Retinaface: Single-stage dense face localisation in the wild". In: *arXiv preprint arXiv:1905.00641* (2019).

[31] Mingxing Duan et al. "A hybrid deep learning CNN–ELM for age and gender classification". In: *Neurocomputing* 275 (2018), pp. 448–461.

[32] Yueqi Duan et al. "Learning deep binary descriptor with multi-quantization". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 1183–1192.

[33] Yueqi Duan et al. "Context-aware local binary feature learning for face recognition". In: *IEEE transactions on pattern analysis and machine intelligence* 40.5 (2018), pp. 1139–1153.

[34] Jennifer L Eberhardt, Nilanjana Dasgupta, and Tracy L Banaszynski. "Believing is seeing: The effects of racial labels and implicit beliefs on face perception". In: *Personality and Social Psychology Bulletin* 29.3 (2003), pp. 360–370.

[35] Gerda Edelman, Ivo Alberink, and Bart Hoogeboom. "Comparison of the performance of two methods for height estimation". In: *Journal of forensic sciences* 55.2 (2010), pp. 358–365.

[36] Eran Eidinger, Roee Enbar, and Tal Hassner. "Age and gender estimation of unfiltered faces". In: *IEEE Transactions on Information Forensics and Security* 9.12 (2014), pp. 2170–2179.

[37] Sergio Escalera et al. "Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results". In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2015, pp. 1–9.

[38] Sergio Escalera et al. "Chalearn looking at people and faces of the world: Face analysis workshop and challenge 2016". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2016, pp. 1–8.

[39] Rong-En Fan et al. "LIBLINEAR: A Library for Large Linear Classification". In: *Journal of Machine Learning Research* 9 (2008), pp. 1871–1874.

[40] Jean-Marc Fellous. "Gender discrimination and prediction on the basis of facial metric information". In: *Vision research* 37.14 (1997), pp. 1961–1973.

[41] Carles Fernández, Ivan Huerta, and Andrea Prati. "A comparative evaluation of regression learning algorithms for facial age estimation". In: *Face and Facial Expression Recognition from Real World Videos*. Springer, 2015, pp. 133–144.

[42] Jonathan B Freeman, Nalini Ambady, and Phillip J Holcomb. "The face-sensitive N170 encodes social category information". In: *Neuroreport* 21.1 (2010), p. 24.

[43] Yun Fu, Guodong Guo, and Thomas S Huang. "Age synthesis and estimation via faces: A survey". In: *IEEE transactions on pattern analysis and machine intelligence* 32.11 (2010), pp. 1955–1976.

[44] Andrew C Gallagher and Tsuhan Chen. "Understanding images of groups of people". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2009, pp. 256–263.

[45] Bin-Bin Gao et al. "Deep label distribution learning with label ambiguity". In: *IEEE Transactions on Image Processing* 26.6 (2017), pp. 2825–2838.

[46] Xin Geng, Zhi-Hua Zhou, and Kate Smith-Miles. "Automatic age estimation based on facial aging patterns". In: *IEEE Transactions on pattern analysis and machine intelligence* 29.12 (2007), pp. 2234–2240.

[47] Douglas Gray, Shane Brennan, and Hai Tao. "Evaluating appearance models for recognition, reacquisition, and tracking". In: *Proc. IEEE international workshop on performance evaluation for tracking and surveillance (PETS)*. Vol. 3. 5. Citeseer. 2007, pp. 1–7.

[48] Peng Guan, Oren Freifeld, and Michael J Black. "A 2D human body model dressed in eigen clothing". In: *European conference on computer vision*. Springer. 2010, pp. 285–298.

[49] Asuman Günay and Vasif Nabiyev. "A new facial age estimation method using centrally overlapped block based local texture features". In: *Multimedia Tools and Applications* (2017), pp. 1–27.

[50] Asuman Günay and Vasif V. Nabiyev. "Age Estimation Based on Hybrid Features of Facial Images". In: *Information Sciences and Systems 2015*. Ed. by Omer H. Abdelrahman et al. Cham: Springer International Publishing, 2016, pp. 295–304. ISBN: 978-3-319-22635-4.

[51] Semih Gunel, Helge Rhodin, and Pascal Fua. "What face and body shapes can tell us about height". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. 2019.

[52] Guodong Guo and Guowang Mu. "A study of large-scale ethnicity estimation with gender and age variations". In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE. 2010, pp. 79–86.

[53] Guodong Guo and Guowang Mu. "Human age estimation: What is the influence across race and gender?" In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE. 2010, pp. 71–78.

[54] Guodong Guo and Guowang Mu. "Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression". In: *Computer vision and pattern recognition (cvpr), 2011 ieee conference on*. IEEE. 2011, pp. 657–664.

[55] Guodong Guo and Guowang Mu. "Joint estimation of age, gender and ethnicity: CCA vs. PLS". In: *Automatic face and gesture recognition (fg), 2013 10th ieee international conference and workshops on*. IEEE. 2013, pp. 1–6.

[56] Guodong Guo and Xiaolong Wang. "A study on human age estimation under facial expression changes". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2012, pp. 2547–2553.

[57] Guodong Guo et al. "Human age estimation using bio-inspired features". In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE. 2009, pp. 112–119.

[58] Philipp Gysel, Mohammad Motamedi, and Soheil Ghiasi. "Hardware-oriented approximation of convolutional neural networks". In: *arXiv preprint arXiv:1604.03168* (2016).

[59] Abdenour Hadid and Matti Pietikäinen. "Demographic classification from face videos using manifold learning". In: *Neurocomputing* 100 (2013), pp. 197–205.

[60] William L Hall II et al. "Errors in weight estimation in the emergency department: comparing performance by providers and patients". In: *The Journal of emergency medicine* 27.3 (2004), pp. 219–224.

[61] H. Han and Anil K. Jain. "Age , Gender and Race Estimation from Unconstrained Face Images". In: 2014.

[62] Hu Han et al. "Demographic estimation from face images: Human vs. machine performance". In: *IEEE transactions on pattern analysis and machine intelligence* 37.6 (2014), pp. 1148–1161.

[63] Hu Han et al. "Demographic estimation from face images: Human vs. machine performance". In: *IEEE transactions on pattern analysis and machine intelligence* 37.6 (2015), pp. 1148–1161.

[64] Ju Han and Bir Bhanu. "Statistical feature fusion for gait-based human recognition". In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* Vol. 2. IEEE. 2004, pp. II–II.

[65] James A Hanley and Barbara J McNeil. "The meaning and use of the area under a receiver operating characteristic (ROC) curve." In: *Radiology* 143.1 (1982), pp. 29–36.

[66] David R Hardoon, Sandor Szedmak, and John Shawe-Taylor. "Canonical correlation analysis: An overview with application to learning methods". In: *Neural computation* 16.12 (2004), pp. 2639–2664.

[67] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[68] T F Rhodes Henry. *Alphonse Bertillon, Father of Scientific Detection*. Harrap, 1956.

[69] Zhenzhen Hu et al. "Facial Age Estimation With Age Difference". In: *IEEE Transactions on Image Processing* 26.7 (2017), pp. 3087–3097.

[70] Jing Huang et al. "Image indexing using color correlograms". In: *Proceedings of IEEE computer society conference on Computer Vision and Pattern Recognition*. IEEE. 1997, pp. 762–768.

[71] Ivan Huerta et al. "A deep analysis on age estimation". In: *Pattern Recognition Letters* 68 (2015), pp. 239–249.

[72] Ozan Irsoy and Ethem Alpaydın. "Continuously constructive deep neural networks". In: *IEEE transactions on neural networks and learning systems* 31.4 (2019), pp. 1124–1133.

[73] Emad Sami Jaha and Mark S Nixon. "Soft biometrics for subject identification using clothing attributes". In: *IEEE international joint conference on biometrics*. IEEE. 2014, pp. 1–6.

[74] Anil K Jain, Sarat C Dass, and Karthik Nandakumar. "Can soft biometric traits assist user recognition?" In: *Biometric technology for human identification*. Vol. 5404. International Society for Optics and Photonics. 2004, pp. 561–572.

[75] Anil K Jain, Patrick Flynn, and Arun A Ross. *Handbook of biometrics*. Springer Science & Business Media, 2007.

[76] Anil K Jain and Unsang Park. "Facial marks: Soft biometric for face recognition". In: *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE. 2009, pp. 37–40.

[77] Omar Javed et al. "Tracking across multiple cameras with disjoint views". In: *Computer Vision, IEEE International Conference on*. Vol. 3. IEEE Computer Society. 2003, pp. 952–952.

[78] Sen Jia, Thomas Lansdall-Welfare, and Nello Cristianini. "Gender classification by deep learning on millions of weakly labelled images". In: *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. IEEE. 2016, pp. 462–467.

[79] Xiaoyi Jiang et al. "Towards detection of glasses in facial images". In: *Pattern Analysis & Applications* 3.1 (2000), pp. 9–18.

[80] Gunnar Johansson. "Visual perception of biological motion and a model for its analysis". In: *Perception & psychophysics* 14.2 (1973), pp. 201–211.

[81] Clifford L Johnson et al. "National health and nutrition examination survey. Analytic guidelines, 1999-2010". In: (2013).

[82] Christopher R Jones and Russell H Fazio. "Person categorization and automatic racial stereotyping effects on weapon identification". In: *Personality and Social Psychology Bulletin* 36.8 (2010), pp. 1073–1085.

[83] Juho Kannala and Esa Rahtu. "Bsif: Binarized statistical image features". In: *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE. 2012, pp. 1363–1366.

[84] N Kaushik and A Mittal. "A novel beard removal method based on structural similarity and co-ordinate transformations". In: *2006 Annual IEEE India Conference*. IEEE. 2006, pp. 1–6.

[85] Vahid Kazemi and Sullivan Josephine. "One millisecond face alignment with an ensemble of regression trees". In: *27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, United States, 23 June 2014 through 28 June 2014*. IEEE Computer Society. 2014, pp. 1867–1874.

[86] Vera Kettnaker and Ramin Zabih. "Counting people from multiple cameras". In: *Proceedings IEEE International Conference on Multimedia Computing and Systems*. Vol. 2. IEEE. 1999, pp. 267–271.

[87] Davis E King. "Dlib-ml: A machine learning toolkit". In: *The Journal of Machine Learning Research* 10 (2009), pp. 1755–1758.

[88] Davis E. King. "Dlib-ml: A Machine Learning Toolkit". In: *Journal of Machine Learning Research* 10 (2009), pp. 1755–1758.

[89] Ilkka Korhonen, Juha Parkka, and Mark Van Gils. "Health monitoring in the home of the future". In: *IEEE Engineering in medicine and biology magazine* 22.3 (2003), pp. 66–73.

[90] Kewal Krishan. "Establishing correlation of footprints with body weight—forensic aspects". In: *Forensic science international* 179.1 (2008), pp. 63–69.

[91] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems* 25 (2012), pp. 1097–1105.

[92] Robert Kurzban, John Tooby, and Leda Cosmides. “Can race be erased? Coalitional computation and social categorization”. In: *Proceedings of the National Academy of Sciences* 98.26 (2001), pp. 15387–15392.

[93] Young H Kwon and Niels da Vitoria Lobo. “Age classification from facial images”. In: *Computer vision and image understanding* 74.1 (1999), pp. 1–21.

[94] Andreas Lanitis, Christopher J. Taylor, and Timothy F Cootes. “Toward automatic simulation of aging effects on face images”. In: *IEEE Transactions on pattern Analysis and machine Intelligence* 24.4 (2002), pp. 442–455.

[95] Seok-Han Lee and Jong-Soo Choi. “A single-view based framework for robust estimation of height and position of moving people”. In: *Pacific-Rim Symposium on Image and Video Technology*. Springer. 2007, pp. 562–574.

[96] Zhen Lei, Matti Pietikäinen, and Stan Z Li. “Learning discriminant face descriptor”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.2 (2014), pp. 289–302.

[97] Shan Li and Weihong Deng. “Deep facial expression recognition: A survey”. In: *IEEE Transactions on Affective Computing* (2020).

[98] Haibin Liao et al. “Age Estimation of Face Images Based on CNN and Divide and Rule Strategy”. In: *Mathematical Problems in Engineering* 2018 (2018), pp. 1–8. ISSN: 1024-123X. DOI: 10.1155/2018/1712686.

[99] Hui Lin, Huchuan Lu, and Lihe Zhang. “A new automatic recognition system of gender, age and ethnicity”. In: *2006 6th world congress on intelligent control and automation*. Vol. 2. IEEE. 2006, pp. 9988–9991.

[100] Hao Liu et al. “Group-aware deep feature learning for facial age estimation.” In: *Pattern Recognition* 66 (2017).

[101] Hao Liu et al. “Label-sensitive deep metric learning for facial age estimation”. In: *IEEE Transactions on Information Forensics and Security* 13.2 (2018), pp. 292–305.

[102] Kuan-Hsien Liu, Shuicheng Yan, and C-C Jay Kuo. “Age group classification via structured fusion of uncertainty-driven shape features and selected surface features”. In: *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*. IEEE. 2014, pp. 445–452.

[103] S.R. Loth and M.Y. İşcan. “ANTHROPOLOGY — Sex Determination”. In: *Encyclopedia of Forensic Sciences*. Ed. by Jay A. Siegel. Oxford: Elsevier, 2000, pp. 252–260. ISBN: 978-0-12-227215-8. DOI: `https://doi.org/10.1006/rwfs.2000.0756`.

[104] Jiwen Lu, Junlin Hu, and Yap-Peng Tan. “Discriminative deep metric learning for face and kinship verification”. In: *IEEE Transactions on Image Processing* 26.9 (2017).

[105] Jiwen Lu, Venice Erin Liong, and Jie Zhou. “Cost-sensitive local binary feature learning for facial age estimation”. In: *IEEE Transactions on Image Processing* 24.12 (2015), pp. 5356–5368.

[106] Jiwen Lu et al. “Learning compact binary face descriptor for face recognition”. In: *IEEE transactions on pattern analysis and machine intelligence* 37.10 (2015), pp. 2041–2056.

[107] Khoa Luu et al. “Contourlet appearance model for facial age estimation”. In: *Biometrics (ijcb), 2011 international joint conference on*. IEEE. 2011, pp. 1–8.

[108] CS Madden and M Piccardi. “Height measurement as a session-based biometric for people matching across disjoint camera views”. In: *Image and Vision Computing Conference*. Wickliffe Ltd. 2005.

[109] MARIANNA Madry-Pronobis. "Automatic gender recognition based on audiovisual cues". PhD thesis. Master Thesis, 2009.

[110] George Mather and Linda Murdoch. "Gender discrimination in biological motion displays based on dynamic cues". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 258.1353 (1994), pp. 273–279.

[111] Shyaman Menon and Anne-Maree Kelly. "How accurate is weight estimation in the emergency department?" In: *Emergency Medicine Australasia* 17.2 (2005), pp. 113–116.

[112] Meredith Minear and Denise C Park. "A lifespan database of adult facial stimuli". In: *Behavior Research Methods, Instruments, & Computers* 36.4 (2004), pp. 630–633.

[113] Ahmad Saeed Mohammad, Ajita Rattani, and Reza Derakhshani. "Eyebrows and eyeglasses as soft biometrics using deep learning". In: *IET Biometrics* 8.6 (2019), pp. 378–390.

[114] Azza Mokadem et al. "Age regression based on local image features". In: *Image and Video Technology (PSIVT), 2010 Fourth Pacific-Rim Symposium on*. IEEE. 2010, pp. 88–93.

[115] Dat Tien Nguyen et al. "Comparative study of human age estimation with or without preclassification of gender and facial expression". In: *The Scientific World Journal* 2014 (2014).

[116] Fudong Nian et al. "Robust gender classification on unconstrained face images". In: *Proceedings of the 7th International Conference on Internet Multimedia Computing and Service*. 2015, pp. 1–4.

[117] Zhenxing Niu et al. "Ordinal regression with multiple output cnn for age estimation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 4920–4928.

[118] Ville Ojansivu and Janne Heikkilä. "Blur insensitive texture classification using local phase quantization". In: *International conference on image and signal processing*. Springer. 2008, pp. 236–243.

[119] P Jonathon Phillips et al. "Overview of the face recognition grand challenge". In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. Vol. 1. IEEE. 2005, pp. 947–954.

[120] Brunelli Poggio, R Brunelli, and T Poggio. "HyberBF networks for gender classification". In: (1992).

[121] Minakshi R Rajput and Ganesh S Sable. "Age Group Estimation from Human Iris". In: *International Conference on Soft Computing and Signal Processing*. Springer. 2019, pp. 519–529.

[122] Gillian Rhodes and Jim Haxby. *Oxford handbook of face perception*. Oxford University Press, 2011.

[123] Karl Ricanek and Tamirat Tesafaye. "Morph: A longitudinal image database of normal adult age-progression". In: *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*. IEEE. 2006, pp. 341–345.

[124] K Ricanek Jr, E Boone, and E Patterson. "Craniofacial aging impacts on the eigenface face biometric". In: *Comput. Sci* 1.3 (2006).

[125] Kathleen M Robinette, Hans Daanen, and Eric Paquet. "The CAESAR project: a 3-D surface anthropometry survey". In: *Second International Conference on 3-D Digital Imaging and Modeling (Cat. No. PR00062)*. IEEE. 1999, pp. 380–386.

[126] Pau Rodrıéguez et al. “Age and gender recognition in the wild with deep attention”. In: *Pattern Recognition* 72 (2017), pp. 563–571.

[127] Roman Rosipal and Nicole Krämer. “Overview and recent advances in partial least squares”. In: *International Statistical and Optimization Perspectives Workshop" Subspace, Latent Structure and Feature Selection"*. Springer. 2005, pp. 34–51.

[128] Rasmus Rothe, Radu Timofte, and Luc Van Gool. “Deep expectation of real and apparent age from a single image without facial landmarks”. In: *International Journal of Computer Vision* 126.2-4 (2018), pp. 144–157.

[129] Olga Russakovsky et al. “Imagenet large scale visual recognition challenge”. In: *International Journal of Computer Vision* 115.3 (2015), pp. 211–252.

[130] Atsuya Sakata, Noriko Takemura, and Yasushi Yagi. “Gait-based age estimation using multi-stage convolutional neural network”. In: *IPSJ Transactions on Computer Vision and Applications* 11.1 (2019), p. 4.

[131] Sina Samangooei, Baofeng Guo, and Mark S Nixon. “The use of semantic human description as a soft biometric”. In: *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems*. IEEE. 2008, pp. 1–7.

[132] Sudeep Sarkar et al. “The humanid gait challenge problem: Data sets, performance, and analysis”. In: *IEEE transactions on pattern analysis and machine intelligence* 27.2 (2005), pp. 162–177.

[133] Richard T Schaefer. *Encyclopedia of race, ethnicity, and society*. Vol. 1. Sage, 2008.

[134] Daniela Schiller et al. “A neural mechanism of first impressions”. In: *Nature neuroscience* 12.4 (2009), p. 508.

[135] Lior Shamir. "Automatic age estimation by hand photos". In: *Computer Science Letters* 3.1 (2011).

[136] Xiangbo Shu et al. "Age progression: Current technologies and applications". In: *Neurocomputing* 208 (2016), pp. 249–261.

[137] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[138] K Sudo, J Yamato, and A Tomono. "Determining gender of walking people using multiple sensors". In: *1996 IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems (Cat. No. 96TH8242)*. IEEE. 1996, pp. 641–646.

[139] Jinli Suo et al. "Design sparse features for age estimation using hierarchical face model". In: *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE. 2008, pp. 1–6.

[140] Christian Szegedy et al. "Going deeper with convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

[141] Paul Viola and Michael Jones. "Rapid object detection using a boosted cascade of simple features". In: *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*. Vol. 1. IEEE. 2001, pp. I–I.

[142] Liang Wang et al. "Silhouette analysis-based gait recognition for human identification". In: *IEEE transactions on pattern analysis and machine intelligence* 25.12 (2003), pp. 1505–1518.

[143] Xiaolong Wang, Rui Guo, and Chandra Kambhamettu. "Deeply-learned feature for age estimation". In: *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*. IEEE. 2015, pp. 534–541.

[144] Haiyuan Wu et al. “Glasses frame detection with 3D Hough transform”. In: *Object recognition supported by user interaction for service robots*. Vol. 2. IEEE. 2002, pp. 346–349.

[145] Yi Xiao and Hong Yan. “Extraction of glasses in human face images”. In: *International Conference on Biometric Authentication*. Springer. 2004, pp. 214–220.

[146] Ziyi Xu, Li Lu, and Pengfei Shi. “A hybrid approach to gender classification from face images”. In: *2008 19th International Conference on Pattern Recognition*. IEEE. 2008, pp. 1–4.

[147] Ming Yang et al. “Correspondence driven adaptation for human profile recognition”. In: *CVPR 2011*. IEEE. 2011, pp. 505–512.

[148] Zhiguang Yang and Haizhou Ai. “Demographic classification with local binary patterns”. In: *International Conference on Biometrics*. Springer. 2007, pp. 464–473.

[149] Kaipeng Zhang et al. “Joint face detection and alignment using multitask cascaded convolutional networks”. In: *IEEE Signal Processing Letters* 23.10 (2016), pp. 1499–1503.

[150] Wenyi Zhao et al. “Face recognition: A literature survey”. In: *ACM computing surveys (CSUR)* 35.4 (2003), pp. 399–458.

[151] Ziqing Zhuang et al. “Facial anthropometric differences among gender, ethnicity, and age groups”. In: *Annals of occupational hygiene* 54.4 (2010), pp. 391–402.

[152] ME. Zighem et al. “Age Estimation Based on Color Facial Texture”. In: *10 ème Conférence sur le Génie Electrique*. 2017.

[153] Cong Zong, Mohamed Chetouani, and Adriana Tapus. "Automatic gait characterization for a mobility assistance system". In: *2010 11th International Conference on Control Automation Robotics & Vision*. IEEE. 2010, pp. 473–478.