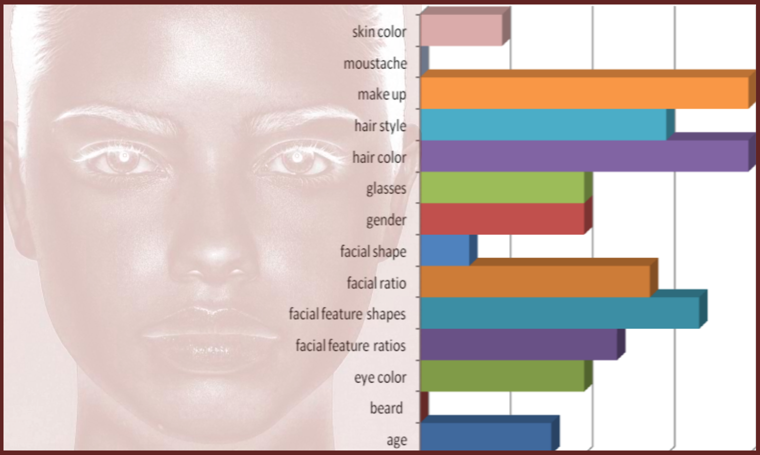
FACIAL SOFT BIOMETRICS

METHODS, APPLICATIONS AND SOLUTIONS

Doctoral Thesis



Author:

Antitza DANTCHEVA

Supervisor:

Prof. Dr. Jean-Luc DUGELAY

This thesis was defended in December, 2011 in front of the following jury.

Reviewers:

Prof. Dr. Abdenour HADID, University of Oulu, Finland

Prof. Dr. Mark NIXON, University of Southampton, United Kingdom

Examiners:

Prof. Dr. Arun ROSS, West Virginia University, USA

Prof. Dr. Bernadette DORIZZI, Telecom SudParis, France

Dr. Sami ROMDHANI, Morpho, France

Acknowledgements

This thesis would not have been possible without the help, guidance and support of many people, who I would like to acknowledge here.

I owe my deepest gratitude to my parents, Vessela Stoilova and George Pentchev, for their unwavering encouragement, devotion and love.

I am indebted to my thesis advisor Prof. Jean-Luc Dugelay for giving me the opportunity for a Ph.D. at Eurecom / Telecom ParisTech. Throughout my 3.5 Ph.D. years he provided ingenious ideas and encouraging support. He created a vastly positive and enthusiastic working atmosphere that fueled self-motivation and ambition.

I would like to express my deepest appreciation for Petros Elia, who shared his brilliance and creativity with me. It was a pleasure to work and exchange with him.

I would like to thank my committee members, the reviewers Prof. Abdenour Hadid and Prof. Mark Nixon, and furthermore the examiners Prof. Arun Ross, Prof. Bernadette Dorizzi and Dr. Sami Romdhani for their precious time, shared positive insight and guidance.

My warmest thanks to my colleagues who supported me during my Ph.D.: Angela D'Angelo, Carina Schmidt-Knorreck, Nesli Erdogmus, Christelle Yemdji, Carmello Velardo, Simon Bozonnet, Mourad Ouaret, Lionel Daniel, Neslihan Kose, Miriam Redi, Rui Min, Yingbo Li, Konstantinos Papakonstantinos, Thrasyvoulos Spyropoulos.

Lastly, special thanks to my friends for decades, Barbara Kostner, Thomas Linder, Adela Marinescu, Jutta Unterscheider and Kathrin Marghani for their unwavering friendship, moral and infinite support.

Abstract

This dissertation studies soft biometrics traits, their applicability in different security and commercial scenarios, as well as related usability aspects. We place the emphasis on human *facial soft biometric traits* which constitute the set of physical, adhered or behavioral human characteristics that can partially differentiate, classify and identify humans. Such traits, which include characteristics like age, gender, skin and eye color, the presence of glasses, moustache or beard, inherit several advantages such as ease of acquisition, as well as a natural compatibility with how humans perceive their surroundings. Specifically, soft biometric traits are compatible with the human process of classifying and recalling our environment, a process which involves constructions of hierarchical structures of different refined traits.

This thesis explores these traits, and their application in *soft biometric systems* (SBSs), and specifically focuses on how such systems can achieve different goals including database search pruning, human identification, human re–identification and, on a different note, prediction and quantification of facial aesthetics. Our motivation originates from the emerging importance of such applications in our evolving society, as well as from the practicality of such systems. SBSs generally benefit from the non-intrusive nature of acquiring soft biometric traits, and enjoy computational efficiency which in turn allows for fast, enrolment–free and pose–flexible biometric analysis, even in the absence of consent and cooperation by the involved human subjects. These benefits render soft biometrics indispensable in applications that involve processing of real life images and videos.

In terms of security, we focus on three novel functionalities of SBSs: pruning the search in large human databases, human identification, and human re–identification.

With respect to *human identification* we shed some light on the statistical properties of pertinent parameters related to SBSs, such as employed traits and trait–instances, total categories, size of authentication groups, spread of effective categories and correlation between traits. Further we introduce and elaborate on the event of interference, i.e., the event where a subject picked for identification is indistinguishable from another subject in the same authentication group.

Focusing on *search pruning*, we study the use of soft biometric traits in pre-filtering large human image databases, i.e., in pruning a search using soft biometric traits. Motivated by practical scenarios such as time–constrained human identification in biometric-based video surveillance systems, we analyze the stochastic behavior of search pruning, over large and unstructured data sets which are furthermore random and varying, and where in addition, pruning itself is not fully reliable but is instead prone to errors. In this stochastic setting we explore the natural tradeoff that appears between pruning gain and reliability, and proceed to first provide average–case analysis of the problem and then to study the atypical gain-reliability behavior, giving insight on how often pruning might fail to substantially reduce the search space. Moreover we consider actual soft biometric systems (nine of them) and the corresponding categorization algorithms, and provide a number of experiments that reveal the behavior of such systems. Together, analysis and experimental results, offer a way to quantify, differentiate and compare the presented SBSs and offer insights on design aspects for improvement of such systems.

With respect to *human re–identification* we address the problem of pose variability in surveillance videos. Despite recent advances, face-recognition algorithms are still challenged when applied to the setting of video surveillance systems which inherently introduce variations in the pose of subjects. We seek to provide a recognition algorithm that is specifically suited to a frontal-to-side re-identification setting. Deviating from classical biometric approaches, the proposed method considers color- and texture- based soft biometric traits, specifically those taken from patches of

hair, skin and clothes. The proposed method and the suitability of these patch-based traits are then validated both analytically and empirically.

Deviating from security related themes, we focus on a completely different application: employing soft biometrics in evaluation of *female facial aesthetics*. This approach is novel in that, in the context of female facial aesthetics, it combines soft biometrics with previous approaches on photo quality and beauty assessment. This study helps us to understand the role of this specific set of features in affecting the way humans perceive facial images. Based on the above objective parameters, we further construct a simple linear metric that suggests modifiable parameters for aesthetics enhancement, as well as tunes systems that would seek to predict the way humans perceive facial aesthetics. Moreover using the designed metric we evaluate beauty indices with respect to aging, facial surgery and females famous for their beauty. We simulate an automatic tool for beauty prediction with both realistic accuracy and performance.

Remaining in the realm of human perception, we also provide a comparative study of different access control systems based on fingerprint, PIN, soft biometrics and face recognition. Towards comparing these systems, we design real–life access control interfaces, each based on the above mentioned methods, and then proceeded to empirically evaluate the degree of usability for each of these interfaces. Our study is based on the recorded assessments of a set of users who rated their interaction with each interface, in terms of privacy, ease of use, user-friendliness, comfort and interaction time. The results reinforce, from a usability point of view, the employment of novel biometric authentication methods as viable alternatives to the currently predominant PIN based methods for access control.

Overall this dissertation has contributed the following:

- identification and introduction of novel applications for soft biometrics, such as human identification (bag of soft biometrics), re–identification as well as aesthetics prediction
- development of theoretical framework for SBSs in the applications: pruning the search and human identification
- application of the developed theoretical framework on existing SBSs
- construction of a novel image processing tool for classification of soft biometric traits and employing such a tool in challenging scenarios
- obtaining evidence for the high user friendliness of soft biometric based control access systems.

This work was conducted in part within the European Project ACTIBIO [ACT11] and was supported in part by the European Commission under contract FP7-215372.

Contents

Notations used in this work

$\mathbb{E}$: statistical expectation

$\alpha_{0,f}(\boldsymbol{v})\}_{f=1}^{\rho}$: instantaneous normalized distribution (histogram) of $\{|C_f|\}_{f=1}^{\rho}$ for a specific $\boldsymbol{v}$

ϵ_f : categorization or confusion error probabilities

$\widehat{\phi}$: algorithmically estimated category ϕ

λ : soft biometric trait instances

$\mathcal{G}(\boldsymbol{v})$: pruning gain $\mathcal{G}(\boldsymbol{v}) := \frac{n}{|\mathcal{S}|}$

$\mathcal{S}$: subset of n of subjects that were not pruned out

$\mathcal{U}$: goodput

$\mathcal{V}(\tau)$: set of valid $\boldsymbol{\alpha}$ for a given τ

μ : soft biometric trait

ϕ : λ–tuple of different trait–instances, one possible category: ’blue eyed, with moustache and with glasses’

Φ : $\Phi = \{\phi_i\}_{i=1}^{\rho}$ set of all ρ categories

ρ : total number of categories

τ : inverse of pruning gain, $\tau = |\mathcal{S}|/n$

$\widehat{C}(v) \in [1, \rho]$: Category that v belongs in

ACT : absolute category rating

C' : actual category of v'

$C_f \subset \boldsymbol{v},\ f = 1, \cdots, \rho$

F : number of effective or non-empty categories spanned by $\boldsymbol{v}$

MOS : mean opinion score

N : computational complexity

n : size of authentication group

P_{err} : error probability $P(\text{err}|\boldsymbol{v})$: probability of erroneously identifying a subject

P_ϕ : probability of incorrectly identifying a subject from S_ϕ

p_f : p_f : population statistics

r : relative throughput of a SBS, $r := \lim_{\rho\to\infty} \frac{n}{\rho}$

S : subset of $\boldsymbol{v}$ that remains after pruning

$S_\phi \subset \boldsymbol{v}$: set of subjects in $\boldsymbol{v}$ that belong in a specific category ϕ

S_{id} : set of subjects in $\boldsymbol{v}$ that can potentially be identified by a SBS ’endowed’ with Φ, $S_id := \cup_{\phi=1}^{F} S_\phi$

v : elements in $\boldsymbol{v}$

v' : subject of interest in the context of search pruning

$\boldsymbol{v}$: authentication group containing n subjects

$\boldsymbol{v}(i)$, $i = 1, ..., n$: i-th candidate belonging to the specific group $\boldsymbol{v}$

SB : soft biometrics

SBS : soft biometric system

Chapter 1

Introduction

Traditional biometrics offer a natural and reliable solution for establishing the identity of an individual, and for this reason, the use of human physical and behavioral characteristics has been increasingly adopted in security applications. With this approach maintaining various advantages such as universality, robustness, permanence and accessibility, it is not surprising that current intrusion detection and security mechanisms and systems include by default at least one biometric trait.

Building on this progress, the latest addition of soft biometrics builds and adds on the main advantages of classical biometrics.

The beginnings of soft biometric science were laid by Alphonse Bertillon in the 19th century, who firstly introduced the idea of a person identification system based on biometric, morphological and anthropometric determinations, see [Rho56]. In his effort, Bertillon considered traits like colors of eye, hair, beard and skin; shape and size of the head, as well as general discriminators like height or weight and also indelible marks such as birth marks, scars or tattoos. These descriptors mainly comprise what is now referred to as the family of *soft biometrics*, a term first introduced by Jain et al. [JDN04b] to describe the set of characteristics that provide (some) information about an individual, but that are not generally sufficient for fully describing and identifying a person, mainly due to the lack of distinctiveness and permanence of such traits. As stated later [JDN04a], such soft biometrics traits can be inexpensive to compute, can be sensed at a distance, do not require the cooperation of the surveillance subjects, and can be efficiently used to narrow down a search for an individual from a large set of people. Along the lines of *semantic annotation* ([SGN08] and [RN10]) we here note the human compliance of soft biometrics as a main difference between soft biometrics and classical biometrics - a difference that renders soft biometrics suitable for many applications. The terms *light biometrics* see in [ALMV04], *similes* see in [KBBN09] and *attributes* see in [VFT$^+$09] have been describing traits we associate to soft biometrics. The following definition clarifies what is considered here as soft-biometric traits.

Definition: Soft biometric traits are physical, behavioral or adhered human characteristics, classifiable in pre–defined human compliant categories. These categories are, unlike in the classical biometric case, established and time–proven by human experience with the aim of differentiating individuals. In other words soft biometric traits are created in a natural way, used by people to characterize other people.

Our interest in this thesis is in understanding the role that soft biometrics can play in security and commercial systems of the future. In brief we begin by specifying soft biometric traits that adhere to the above definition. After an overview of related work, we proceed to explore different applications that benefit from soft biometric systems (SBSs), focusing on surveillance related person identification, and on pruning of large surveillance related searches. We also consider the specific scenario of applying soft biometrics for human frontal-to-side re-identification. We then change gear and deviate from security related applications to the more commercially oriented application of employing soft biometrics in quantifying and predicting female facial aesthetics. The above approaches are then complemented by a more practical automatic soft biometric classification tool that we present. Finally, motivated by human acceptance issues, we proceed to provide a usability study relating to soft biometrics.

1.1 Achievements and structure of the dissertation

We proceed with an explicit description of the structure of the thesis, and the introduction of the scenarios / applications of interest in each chapter.

Chapter 2 - Soft biometrics: characteristics, advantages and related work

In Chapter 2 we offer general considerations related to soft biometrics. Firstly in Section 2.1 we introduce a candidate list of traits and furthermore proceed to portray pertinent advantages and limitations in Section 2.2. We then identify in Section 2.3 previous work on soft biometric traits.

Chapter 3 - Bag of facial soft biometrics for human identification

Chapter 3 considers the case where a SBS can distinguish between a set of traits (categories), which set is large enough to allow for the classification that achieves human identification. The concept of person identification based on soft biometrics originates in the way humans perform face recognition. Specifically human minds decompose and hierarchically structure complex problems into fractions and those fractions into further sub-fractions, see [Ley96], [Sim96]. Consequently face recognition performed by humans is the division of the face in parts, and subsequent classification of those parts into categories. Those categories can be naturally of physical, adhered or behavioral nature and their palette includes colors, shapes or measurements, what we refer to here as soft biometrics. The key is that each individual can be categorized in terms of such characteristics, by both humans or by image processing algorithms. Although features such as hair, eye and skin color, facial hair and shape, or body height and weight, gait, cloth color and human metrology are generally non distinctive, a cumulative combination of such features provides an increasingly refined and explicit description of a human. SBSs for person identification have several advantages over classical biometric systems, as of non intrusiveness, computational and time efficiency, human compliance, flexibility in pose- and expression-variance and furthermore an enrolment free acquirement in the absence of consent and cooperation of the observed person. Soft biometrics allow for a reduced complexity determination of an identity. At the same time though, the named reduced computational complexity comes with restrictions on the size of an authentication group. It becomes apparent that a measure of performance must go beyond the classical biometric equal error rate of the employed detectors and include a different and new parametrization. Our general interest here is to provide insightful mathematical analysis of reliability of general soft biometric systems, as well as to concisely describe the asymptotic behavior of pertinent statistical parameters that are identified to directly affect performance. Albeit its asymptotic and mathematical nature, the approach aims to provide simple expressions that can yield insight into handling real life surveillance systems.

In Chapter 3, Section 3.1 introduces the operational setting of a SBS. In this setting we elaborate on pertinent factors, such as those of the authentication group, traits, traits instances, overall categories and their interrelations. We then proceed in Section 3.5.1 to introduce and explain the event of *collision*, which is of significant character when employing SBSs for person identification. Furthermore we introduce the *number of effective categories* F which is later identified as an important parameter related to collision, and is shown to directly affect the overall performance of an SBS.

Section 3.5.2 analyzes the statistical distribution and mean of F and furthermore Section 3.5.3 offers an insight regarding the bounds of the statistical behavior of F over large populations. These bounds address the following practical question: if more funds are spent towards increasing the quality of an SBS, then what reliability gains do we expect to see? The answer and further intuition on the above bounds are provided in Section 3.5.3.1, the proofs can be found in the Appendix 3.

In Section 3.5.4 we examine the influence of algorithmic estimation errors and give an example on the overall performance of a realistic SBS. We improve the performance by a study of the distribution between population in the overall categories, see Section 3.5.4.1. We then proceed in Section 3.5.4.2 to elaborate on the human compliant aspect of soft biometrics in re–identification, hereby specifically on the quantification of traits and on the human interaction view of an SBS.

Chapter 4 - Search pruning in video surveillance systems

In Chapter 4 we explore the application using soft-biometric related categorization-based pruning to narrow down a large search.

In recent years we have experienced an increasing need to structure and organize an exponentially expanding volume of images and videos. Crucial to this effort is the often computationally expensive task of algorithmic search for specific elements placed at unknown locations inside large data sets. To limit computational cost, soft biometrics pruning can be used, to quickly eliminate a portion of the initial data, an action which is then followed by a more precise and complex search within the smaller subset of the remaining data. Such pruning methods can substantially speed up the search, at the risk though of missing the target due to classification errors, thus reducing the overall reliability. We are interested in analyzing this speed vs. reliability tradeoff, and we focus on the realistic setting where the search is time-constrained and where, as we will see later on, the environment in which the search takes place is stochastic, dynamically changing, and can cause search errors. In our setting a time constrained search seeks to identify a subject from a large set of individuals. In this scenario, a set of subjects can be pruned by means of categorization that is based on different combinations of soft biometric traits such as facial color, shapes or measurements. We clarify that we limit our use of “pruning the search” to refer to the categorization and subsequent elimination of soft biometric-based categories, within the context of a search within large databases (Figure 4.1). In the context of this work, the elimination or filtering our of the employed categories is based on the soft biometric characteristics of the subjects. The pruned database can be subsequently processed by humans or by a biometric such as face recognition.

Towards analyzing the pruning behavior of such SBSs, Chapter 4 introduces the concept of *pruning gain* which describes, as a function of pruning reliability, the multiplicative reduction of the set size after pruning. For example a pruning gain of 2 implies that pruning managed to halve the size of the original set. Section 4.5.1 provides average case analysis of the pruning gain, as a function of reliability, whereas Section 4.5 provides atypical-case analysis, offering insight on how often pruning fails to be sufficiently helpful. In the process we provide some intuition through examples on topics such as, how the system gain-reliability performance suffers with increasing confusability of categories, or on whether searching for a rare looking subject renders the search performance more sensitive to increases in confusability, than searching for common looking subjects.

In Section 4.6.1 we take a more practical approach and present nine different soft biometric systems, and describe how the employed categorization algorithms (eye color detector, glasses and moustache detector) are applied on a characteristic database of 646 people. In the same Section we furthermore provide simulations that reveal the variability and range of the pruning benefits offered by different SBSs. In Section 4.7 we derive concise closed form expressions on the measures of pruning gain and goodput, provide simulations, as well as derive and simulate aspects relating to the complexity costs of different soft biometric systems of interest.

Chapter 5 - Frontal-to-side person re-identification

Typically biometric face-recognition algorithms are developed, trained, tested and improved under the simplifying assumption of frontal-to-frontal person recognition. Such algorithms though are challenged when facing scenarios that deviate from the training setting, such as for example in the presence of non-constant viewpoints, including the frontal-to-side scenario. Most person recognition algorithms, whether holistic or based on facial features, only manage to optimally handle pose differences that are less than about 15 degrees. As a result, a variation in the pose is often a more dominant factor than a variation of subjects. This aspect of pose variation comes to the fore in video surveillance, where a suspect may be pictured firstly frontal, whereas the corresponding test images could be captured from the side, thus introducing a *frontal-to-side recognition problem*.

Towards handling this problem, we employ multiple soft biometrics related traits. One of our tasks here is to get some insight into the significance of these traits, specifically the significance of using hair, skin and clothes patches for frontal-to-side re-identification. We are working on the color FERET dataset [Fer11] with frontal gallery images for training, and side (profile) probe images for testing. Towards achieving re-identification, the proposed algorithm first analyzes the color in Section 5.2.4.1 and furthermore texture in Section 5.2.4.2 of the three patches. Then we study the intensity correlations between patches in Section 5.2.4.3. This analysis is then followed by the construction of a single, stronger classifier that combines the above measures in Section 5.2.5, to re-identify the person from his or her profile.

Deviating from the above security related applications, we consider then an application closer to entertainment, and specifically consider the application of soft biometrics in analyzing and quantifying facial aesthetics.

Chapter 6 - Soft biometrics for quantifying and predicting facial aesthetics

With millions of images appearing daily on Facebook, Picasa, Flickr, or on different social and dating sites, photographs are often seen as the carrier of the first and deciding impression of a person. At the same time though, human perception of facial aesthetics in images is a priori highly subjective.

We related among others soft biometric traits with this subjective human perception. In the provided study we quantify insight on how basic measures can be used to improve photographs for CVs or for different social and dating websites. This helps create an objective view on subjective efforts by experts / journalists when retouching images. We use the gained objective view to examine facial aesthetics in terms of aging, facial surgery and a comparison of average females relatively to selected females known for their beauty. Specifically in Section 6.3 we introduce the employed database, as well as describe the basic features and methods used in this study. In Section 6.4 we proceed with numerical results, and provide intuition on the role of features, image quality and facial features, in human perception. In Section 6.5, we use these accumulated conclusions to construct a basic linear model that predicts attractiveness in facial photographs using different facial traits as well as image properties. We then examine and validate the designed metric. In Section 6.6 we employ the developed metric to conduct experiments and answer questions regarding the beauty index in three cases: for famous attractive females, for aging females and in case of facial surgery. Finally we proceed to simulate in Section 6.7 based on both, the presented metric, as well as state of the art algorithmic accuracies an automatic tool for beauty prediction.

Chapter 7 - Practical implementation of soft biometrics classification algorithms

Towards practical implementation of the related concepts and ideas, in Chapter 7 we develop a tool (concatenation of classification algorithms) for classification of facial soft biometric traits, where we specifically emphasize on the most obvious facial identifiers, primarily mentioned by humans, when portraying an unknown individual. The constructed tool is streamlined to achieve reliability of identification at reduced complexity, and hence focuses on simple yet robust soft-biometric traits, including hair color, eye color and skin color, as well as the existence of beard, moustache and glasses. We then specifically focus on extraction and categorization of eye color, and present an additional study where we illustrate the influence of surrounding factors like illumination, eye glasses and sensors on the appearance of eye color.

In Section 7.1 a bag of six facial soft biometrics is elaborated, for which estimation algorithms are featured, along with the related experimental results, see Section 7.1.2. We then proceed to focus on eye color as a soft biometric trait in Section 7.2 and examine an automatic eye color classifier in challenging conditions, such as changing illumination, presence of glasses and camera sensors, see Section 7.4.

Chapter 8 - User acceptance study relating to soft biometrics

Finally we conclude with a usability study that verifies the user acceptance of SBSs, specifically when compared to existing PIN or fingerprint access control systems.

The pervasiveness of biometric systems, and the corresponding growth of the biometric market see [usa11a], has successfully capitalized on the strength of biometric-based methods in accurately and effectively identifying individuals. As a result, modern state-of-the-art intrusion detection and security systems include by default at least one biometric trait. It is the case though that little emphasis has been given to better understanding user-acceptance and user-preference regarding such systems. Existing usability related works, such as in [CAJ03] and [LBCK03], focus on establishing functional issues in existing ATM machines, or on studying the influence of user interaction on the performance of fingerprint based systems (see [KED11]) and interfaces (see [RJMAS09]). Other interesting works (see [usa11b], [CG05], [CJMR09]), analyze possible methods that improve interface design. Our emphasis here is on providing insight on the attitudes and experiences of users towards novel and emerging biometric verification methods, and to explore whether such novel biometric technologies can be, in terms of user acceptance, valid alternatives to existing prevalent PIN based systems. Our focus, in addition to considering the traditional PIN-based method, is to explore the usability aspects of systems based on classical biometrics such as fingerprint and face recognition, and to then proceed to study the usability of systems based on the emerging class of soft-biometric methods. Our evaluation is based on having the users rate and rank their experiences with different access methods.

In Section 8.2 we briefly describe the user test setting, as well as the conditions and the performed test procedures. We then proceed to elaborate on the chosen verification methods and on the designed interfaces. In Section 8.2 we present the results obtained from the user study, in terms of evaluation and quantification of the different usability measurement characteristics. In the same section we provide the user test outcomes of direct comparisons between the four presented methods. Finally in Section 8.3 we draw connections to other significant traits such as cost efficiency, accuracy and processing speed.

We finally note that this dissertation is supported by different journal and conference publications, which are not cited throughout the thesis, but which are listed in full in Appendix D.

Chapter 2

Soft biometrics: characteristics, advantages and related work

Soft biometrics have gained an increasing interest in the biometrics community for various reasons, as of the non–intrusiveness, computational efficiency and mostly the need for higher reliability in biometric systems. In this chapter we provide an overview of soft biometric traits, their classification, the related advantages and limitations. Furthermore we summarize work, already performed on soft biometrics traits or systems integrating soft biometrics.

2.1 Soft biometric traits

We illustrate in Table 2.1) a range of facial characteristics which accept the definition stated in chapter 1 for soft biometrics. In a first attempt to differentiate between soft biometric traits we firstly identify the affiliation to *face* or *accessory* categories. We abuse slightly annotation and include hair color in the group of facial soft biometrics. The presented traits list is not exhaustive and will naturally increase with technological progress.We here note that even though classically *accessories* do not belong to biometrics, the new stated definition clearly incorporates such traits in the class of soft biometrics. The motivation for including accessories to soft biometrics lays in the associated highly descriptiveness and discrimination of attributes such as clothes color, e.g. “the person in the red shirt”. Further significant factors for classifying soft biometric traits are *distinctiveness* and *permanence*. *Distinctiveness* is the strength with which a trait is able to distinguish between individuals. As an example ’beard’ has a low distinctiveness, since it can only be applied to the male part of the population and furthermore possesses only two sub–categories (present or not). This example points out a certain correlation between *distinctiveness* and *nature of value*. Traits with continuous sub-categories are in general more distinctive than traits with discrete and moreover binary sub-categories. In this context the difference between *nature of value* and human labeling of traits is the following: while hair color has principally different nuances and is thus of continuous character, humans tend to discrete labeling. We adopt this human approach for developed soft biometric estimation algorithms, detecting for example hair color in categories such as black, blond, brown, rather than RGB values.

The *permanence* of a trait plays a major role for the application for which a SBS is employed. As an example an application, where identification within a day is required, will accept low permanence traits like age, weight or clothing color (inter vs. intra session observation).

The final subdivision *subjective perception* refers to the degree of ambiguity associated in identifying or labelling specific soft biometric traits sub-categories. We note the relation of subjective

Table 2.1: Table of soft biometric traits

Soft Biometric trait	**Face / Accessory**	**Nature of value**	**Permanence**	**Distinctiveness**	**Subjective perception**
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
Moustache	Face	Binary	Low/Medium	Low	Medium
Facial measurements	Face	Continuous	High	Medium	Medium/High
Facial shapes	Face	Discrete	High	High	High
Facial feature measurements	Face	Continuous	High	High	Medium/High
Facial feature shapes	Face	Discrete	High	High	High
Make–up	Face	Discrete	Low	Low	Medium
Ethnicity	Face	Discrete	High	Medium	Medium
Marks	Face	Discrete	High	Medium/High	Low
Gender	Face	Binary	High	Low	Low
Age	Face	Continuous	Low/Medium	Medium	Medium
Glasses	Accessory	Binary	Low/Medium	Low	Low
Hat	Accessory	Binary	Low	Medium	Low
Scarf	Accessory	Binary	Low	Medium	Low

perception to the nature of value, where an increased amount of subcategories leads to a more difficult classification. In fact subjectivity lays even in the decision of the nature of value. In other words, colors for example can be argued to be continuous, due to the huge variance in nuances blending into each other, or to be discrete due to the fact that colors can be described by discrete RGB values.

We note that soft biometrics can be classified by additional aspects such as accuracy and importance, which are deducible from the named classification classes, depending on the cause for specification (e.g. suitability for a specific application).

2.2 Characteristics, advantages and limitations

Soft biometrics has carried in some extent the attributes of classical biometrics over, as the general idea of identification management based on *who you are* is still being pursuit. The traits provide weak biometrical information about the individual and correspondingly have inherited the predicates to be *universal*, *measurable* and *acceptable*; furthermore the trait's classification algorithm(s) *performance* should be able to meet the application's requirements. To a certain degree also the aspects *uniqueness*, *permanence* and *circumvention* play a role for soft biometrics, but are treated to a greater extent flexible.

Initially, soft biometric traits have been employed to narrow down the search of a database, in order to decrease the computational time for the classical biometric trait. An additional application is the fusion of soft biometrics and classical biometric traits to increase overall system performance. Soft biometrics impart systems substantial advantages: they can be partly derived from main detected classical biometric identifier, their acquisition is non intrusive and does not require enrolment; training can be performed in advance on individuals out of the specific identification group. Summarizing soft biometric traits typically are:

- Human compliant: Traits conform with natural human description labels.
- Computationally efficient: Sensor and computational requirements are marginal.
- Enrolment free: Training of the system is performed off–line and without prior knowledge of the inspected individuals.
- Deducible from classical biometrics: Traits can be partly derived from images captured for primary (classical) biometric identifier (e.g. eye color from eye images).
- Non intrusive: Data acquisition is user friendly or can be fully imperceptible.
- Classifiable from a distance: Data acquisition is achievable at long range.
- Classifiable pose flexible: Data acquisition is feasible from a number of poses.
- Not requiring the individual's cooperation: Consent and contribution from the subject are generally not needed.
- Preserving human privacy: The stored signatures are visually available to everyone and serve in this sense privacy.

The plethora or utilities has motivated an increasing number of research activities related to soft biometrics. In the next section we give an overview of scientific work gaining from the benefits related to soft biometrics.

2.3 Related work

In this section we outline work, pertinent to soft biometrics. This overview does not claim to be an exhaustive state of the art, but rather a highlight selection on performed scientific studies.

Soft biometrics is a relatively novel topic and related work enfolds over several research fields. Recent work can be mainly classified in three research fields:

1. The first and largest field includes the study and identification of traits and associated image processing algorithms for classification and detection of such.
2. The second fast growing field identifies operational scenarios for the aforementioned algorithms and provides experimental results for such scenarios.
3. The third and smallest field comprises of the global and theoretical investigation of the employment of soft biometrics applications and related studies.

Scientific works belonging to the first field cover algorithms for traits such as iris pattern, see in [SBS10], or facial marks, see in [JP09]. A broader overview of work from the first group is referenced in the following sections 2.3.1 and 2.3.2.

The second field can be sub-classified in subgroups which differentiate the way soft biometrics are employed, as stand–alone systems, as pre-filtering mechanisms of bigger systems, or as fused parallel systems. Related scenarios include continuous authentication [NPJ10], video surveillance see [DFBS09], [FDL$^+$10], [MKS10], person verification [ZESH04] and moreover person identification [PJ10]. An interesting recent associated scenario for SBS based person identification is the recognition of faces in triage images of mass disaster situations [CO11]. Further examples are given in section 2.3.3.

Finally the third field involves studies on the placement of soft biometrics in applications such as forensics [JKP11] and human metrology [ACPR10].

2.3.1 Facial soft biometric algorithms

Former work on soft biometrics has been performed predominantly with the aim of preprocessing. In face recognition for person identification, for instance, beard detection and removal serves an improvement of recognition results, disregarding the information of the presence of beard.

Color based facial soft biometrics: The color based facial soft biometric traits (eye, skin, and hair color) are the most obvious facial identifiers, mentioned primarily by humans, when portraying unknown individuals. Challenges for skin classification are on the one hand the low spread of different skin colors in color space, and as a consequence, on the other hand the high illumination dependance of classification. Latter is described in various skin locus papers, for example in [HPM02].

Hair color is detected by similar techniques like skin color and often researched along, but has more broadly scattered color categories. In [SBYL02] a method for human head detection based on hair–color is proposed through the use of Gaussian mixture density models describing the distribution of hair color. In [GHJW00] the fuzzy theory is used to detect faces in color images, where two fuzzy models describe the skin color and hair color, respectively.

Eye color classification, unlike the other color based facial soft biometrics is a relatively new research topic. Few publications offer insight, see [BRM$^+$06] and section 7.2, probably due to the fact that 90% of humans possess brown eyes. An advantage of eye color categorization is the availability of all necessary information in images used for iris pattern analysis, in other words iris color is a free side effect. Work on fusion between iris texture and color can be found in [ZESH04], where the authors fuse iris and iris color with fingerprint and provide performance improvement in respect with the unimodal systems. In [PS08] iris color is used to successfully support an iris indexing method.

Beard and Moustache detection: Presence of beard and moustache do not appear in the literature as an identification trait, but rather as an obstacle for face recognition, which is why their removal is performed as a preprocessing step. As an example, in [KM06] a beard removal algorithm is shown using the concept of structural similarity and coordinate transformations.

Age: Age plays an important role for long time employable systems based on face or body and is a challenging and relatively new field. An interesting study on face changes over time can be found in [PSA$^+$07], which spans a biometric, forensic, and anthropologic review, and further discusses work on synthesizing images of aged faces. In [WM07] the authors distinguish children from adults based on the face/iris size ratio. Viola–Jones face detection technique [VJ01b] is used, followed by an iterative Canny edge detection and a modified circular Hough transform for iris measuring, with good results. In [NBS$^+$08] the authors observe facial skin regions of Caucasian women and build partial least square regression models to predict the chronological and the perceived age. They find out that the eye area and the skin color uniformity are the main attributes related to perceived age.

Gender: Gender perception and recognition has been much researched already in social and cognitive psychology work in the context of face recognition. From image processing point of view, the topic offers as well many of approaches. A basic approach of understanding simple metrology in connection to gender is offered in [CCP$^+$11]. The latest efforts employ a selection of fused biometric traits to deduce gender information. For example in [CSM07] gait energy images and facial features are fused and classified by support vector machines. In [YHP11] contrast and local binary patterns are fused. Another approach in [ST06] proposes a combined gender and expression recognition system by modeling the face using an Active Appearance Model, feature extraction and finally linear, polynomial and radial based function based support vector machines for classification. The work in [BR06] proposes using adaboost on several weak classifiers, applied on low resolution grey scale images with good results. Matta et al. in [MSMD08] present a multimodal gender recognition system, based on facial appearance, head and mouth motion, employing the means of a unified probabilistic framework. Another approach based on motion and appearance is the work in [HP09]. On a different note the authors of [CR11a] employ thermal and near infra-red images for gender classification.

Ethnicity: Ethnicity recognition is an ethically and sociological hot debated trait, once again relevant for face recognition. In the context of ethnicity a uniquely defined classification is a difficult and important task. For recognition of Asian and non–Asian faces in [LJ04] machine learning framework applies a linear discriminant analysis (LDA) and multi scale analysis. A further framework, integrating the LDA analysis for input face images at different scales, further improves the classification performance. In the paper [HTK04] an ethnicity recognition approach is based on Gabor Wavelets Transformation, combined with retina sampling for key facial features extraction. Finally support vector machines are used for ethnicity classification providing very good results, even in the presence of various lighting conditions.

Facial measurements: Facial measurements were early on found as very distinctive and helpful in the context of facial recognition [Nix85]. Later studies continue employing facial measurements, and apply on 3D [CC06].

Recent work on facial soft biometrics is performed on scars, marks and tattoos by the authors in [LJJ08]. Moreover patterns in the sclera have been employed as a soft biometric trait as well, see [CR11b].

2.3.2 Accessory soft biometrics

The new soft biometrics definition allows the inclusion of accessories among these traits. Accessories can indeed be related to personal characteristics (as sight problems in case of glases), or personal choices (as adornment in case of jewelry).

Eye Glasses detection: The forerunner for glasses detection are Jiang et al. in [JBAB00], performing classical edge detection on a preprocessed gray level image. Certain face areas are observed and an indicator for glasses is searched for. The most successful identifier region for glasses is found to be the nose part of the glasses, between the eyes. A different approach for glasses extraction is employed in [XY04], where a face model is established based on the Delaunay triangulation. A 3D method to detect glasses frames is presented in [WYS$^+$02], where 3D features are obtained by a trinocular stereo vision system. The best results on glasses detection up to now are achieved on thermal images [HKAA04].

Scarf and glasses detection: The work [MHD11] handles facial recognition in the presence of occlusions caused by glasses and scarfs. The occlusion is hereby detected by Gabor wavelets, PCA and support vector machines, followed by facial recognition of the non-occluded part based on block–based local binary patterns.

Cap detection: In [MD11] the authors work on an additional occlusion, which impedes face recognition, namely cap detection.

2.3.3 Combined soft biometrics

Since soft biometric traits are individually not distinctive and permanent, a combination of those could overcome those limits. In this context, many recent papers deal with fusion of classical biometry and soft biometry or exclusively with fusion of soft biometric traits. An example for latter is the work in [OPD94]. The authors propose algorithms for gender, body size, height, cadence, and stride using a gait analysis tool. In [DFBS09] height, and appearance are extracted from videos and exploited in a multiple camera video surveillance scenario in order to track the subjects that cross the surveillance network. In [LLZ06] an approach for recognizing the gender, ethnicity and age with facial images is proposed. The approach incorporates Gabor filter, Adaboost learning as well as support vector machine classifiers. A further hybrid classification based on gender and ethnicity is considered in [GPW98] and [GW99]. The hybrid approach consists of an ensemble of radial basis function networks and inductive decision trees. The authors show robustness and good performance. A different approach for analysis in hybrid soft biometric systems is provided in [SGN08] and [RN10], where semantic information (which corresponds to soft biometric classifiers) is manually extracted from a series of videos. Using the analysis of variance the authors select a pool of traits which are considered the most representative. Those traits are then used together with gait information. The authors demonstrate that the additional information provided by the semantic traits increases the performance of the people recognition system based on gait. Those results are extended in [RNS11] and in [RN11]. The authors in [ACPR10] go one step further and study the relation of human body measures, which allows for certain applications the prediction of missing body measures. In [VFT$^+$09] the authors propose an approach for people search in surveillance data, characterized by three main elements: sensors, body parts, and their attributes. The body parts and attributes are here closely related to soft biometrics.

2.4 Domains of application

Soft biometrics are either employed as uni modal systems, classifying a single trait classifiers, or in a combination with other systems. We differentiate following main domains of application.

Fusion with classical biometric traits : SBSs are incorporated in multi modal biometrical systems with the goal of increasing the overall reliability. Such an approach has been followed, in [JDN04b], where the benefits of soft biometrics in addition to fingerprint lead to an improvement of approximately 5% over the primary biometric system.

Pruning the search : SBS were employed in previous works to pre filter large biometric databases with the aim of higher search efficiency. Scientific work on using soft biometrics for pruning the search can be found in [KBN08, KBBN09], where a multitude of attributes, like age, gender, hair and skin color were used for classification of a face database, as well as in [GBDB97, New95] where the impact of pruning traits like age, gender and race was identified in enhancing the performance of regular biometric systems.

A third application is the employment of a multi modal SBS with the goal of human identification or human re-identification.

Human (re-)identification : For human (re-)identification the soft biometric trait related limitations of distinctiveness and permanence are overcome by combining multiple traits. The concept of Bag of Soft Biometrics(BoSB) is directly inspired from the idea of Bag of Words [Joa98, WPS06] and Bag of Features [LSP06] developed under the context of text mining and content based image retrieval. For the BoSB the “items” of the bag are soft biometric signatures extracted from the visual appearance of the subject.

Other possible applications relate to the ability to match people based on their biometric-trait preferences, acquiring statistical properties of biometric identifiers of groups, avatar modelling based on the instantaneous facial characteristics (glasses, beard or different hair color), statistical sampling of audiences, and many others.

Chapter 3

Bag of facial soft biometrics for human identification

The concept of person identification based on soft biometrics originates in the way humans perform face recognition. Specifically human minds decompose and hierarchically structure complex problems into fractions and those fractions into further sub-fractions, cf. [Ley96], [Sim96]. Consequently face recognition performed by humans is the division of the face into parts, and subsequent classification of those parts into sub-categories. Those sub-categories are associated with what refer to as soft biometrics and the key is that each individual can be categorized in terms of such characteristics, by humans or by image processing algorithms. Although features such as hair, eye and skin color, facial hair and shape, or body height and weight, gait, clothing color and human metrology are generally non distinctive, a cumulative combination of such features provides an increasingly refined and explicit description of a human.

3.1 Main parameters: authentication group, traits, trait-instances, and categories

The setting of interest corresponds to the general scenario where, out of a large population, an authentication group is randomly extracted as a random set of n people, out of which one person is picked for identification (and is different from all the other members of the authentication group). We note that this general scenario is consistent with both, the case of person verification as well as of identification. A general soft-biometric system employs detection that relates to λ soft biometric traits (hair color, skin color, etc), where each trait i, $i = 1, 2, \ldots, \lambda$, is subdivided into μ_i *trait instances*, i.e., each trait i can take one of μ_i values. We henceforth denote as category to be any λ-tuple of different trait-instances, and we let $\Phi = \{\phi_i\}_{i=1}^{\rho}$ define a set of all ρ categories, i.e., the set of all ρ combinations of soft-biometric trait-instances. The number of ρ, that the system is endowed with, is given by

$$\rho = \Pi_{i=1}^{\lambda} \mu_i \qquad (3.1)$$

We slightly abuse notation and henceforth say that *a subject belongs to category* ϕ if his or her trait-instances are the λ-tuple corresponding to category ϕ. We here note that to have conclusive identification of a subject, and subsequent differentiation from the other subjects of the authentication group, it must be the case that the subject does not belong in the same category as other members of the authentication group. Given a specific authentication group, the maximum-likelihood optimizing rule for detecting the most probable category in which a chosen subject belongs, is given by:

$$\hat{\phi} = argmax_{\phi \in \Phi} P(\phi) \cdot P(y/\phi), \tag{3.2}$$

where y is the observation vector, $P(\phi)$ is the pdf of the set of categories over the given population (note $\sum_{\phi=1}^{\rho} P(\phi) = 1$), and $P(y/\phi)$ the probability that y is observed, given that the subject belongs in category ϕ.

3.2 Design aspects in soft-biometric systems

In designing a soft-biometric system, the overall choice of the traits and trait-instances must take into consideration aspects as traditional limitations on estimation reliability, which is commonly a function of the sensor resolution, and of the capabilities of the image-processing part of detection. In addition to this traditional aspect, new concerns come into the picture when designing a soft-biometric system as of the size and statistics of the authentication group (such as the possible similarities that might exist between different subjects), as well as the statistical relationship between the authentication group and Φ. The interrelated nature of the above aspects brings to the fore different tradeoffs. Such tradeoffs include for example the fact that an increasing μ_i, and thus also an increasing ρ, generally introduce a reduction in the reliability of detection, but can potentially result in a welcomed increase in the maximum authentication group size (n) that the system can accommodate for. It then becomes apparent that designing and analyzing soft-biometric systems requires a deviation from traditional design and analysis of classical multi-biometric systems, towards considering the role of the above parameters, and their effect on the tradeoffs and the overall system performance. This approach motivates the proposed soft-biometric system design described in chapter 7, as well as the subsequent system analysis of Section 3.5.2 which also includes simulation evaluation of the proposed system in the interference limited setting of very high sensor resolution.

3.3 The proposed Soft-Biometric System

In accordance with the above design aspects, and in an effort to find a good balance between identification-reliability and complexity, we here propose a soft-biometric system that focuses on simple and robust detection from a bounded set of traits and their trait-instances. In what follows, we will describe these basic elements, as well as the employed detection algorithms.

3.3.1 Chosen features of the proposed soft-biometric system

In the presented bag of facial soft biometric traits for human identification, we allocate $\lambda = 6$ traits, which we choose and label as:

1. skin color
2. hair color
3. eye color
4. presence of beard
5. presence of moustache
6. presence of glasses.

In this setting we clearly assign $\mu_4 = \mu_5 = \mu_6 = 2$, corresponding to the binary nature of traits $i = 4, 5, 6$. On the other hand, the first three traits are of a discrete character (see Table I) and had to be categorized in consideration to the tradeoff between reliability of detection and trait importance. Towards this we chose to subdivide trait 1 (skin color) into $\mu_1 = 3$ instances and label them (following a recommendation provided by the ethical partner of a former EU project, ACTIBIO [ACT11] to avoid any assumptions about race or ethnicity based on skin color) as:

– {skin color type 1, skin color type 2, skin color 3} using numbers that increase from light to dark,

to subdivide trait 2 (hair color) into $\mu_2 = 8$ instances

– {light-blond, dark-blond, brown-, black-, red-, grey-, white-haired, and bald}

and to subdivide trait 3 (eye color) into $\mu_3 = 6$ instances:

– {blue-, green-, brown-, grey-, green-, black-eyed}

As a result, the proposed system is endowed with the ability to detect

$$\rho = \Pi_{i=1}^{6} \mu_i = 1152 \tag{3.3}$$

distinct categories. For the sake of clarification, we note two simple examples of such categories in Φ:

– {skin type 1, brown hair, blue eyes, no beard, no moustache, no glasses} $\epsilon \Phi$

– {skin type 3, black hair, black eyes, beard present, moustache present, glasses present} $\epsilon \Phi$

3.4 Statistical aspects of the constructed scheme

Relevant parameters, in addition to λ, μ_i, and ρ, also include the size and statistics of the authentication group (revealing possible similarities between different subjects), as well as the statistical relationship between the authentication group and Φ. In what follows we aim to gain insight on the behavior of the above, in the specific setting of the proposed soft-biometric design. The following analysis, which is by no means conclusive, focuses on providing insight on parameters such as: The spread of the effective categories for a given authentication group, where this spread is used as a measure of the suitability of Φ in authenticating subjects from a certain authentication group. The relationship between n, and the corresponding probability of interference as a function of Φ (the probability that two users share the same category and will thus be indistinguishable). The probability of interference-induced identification error, again to be considered as a measure of the system's reliability).

3.4.1 Spread of the category set Φ

We here consider the case where a soft biometric system is designed to distinguish among ρ distinct categories, but where the randomly introduced authentication group only occupies a smaller fraction of such categories, and where these categories are themselves substantially correlated. Leaving correlation issues aside for now, we first define the set of *effective categories* Φ_e to be the set of categories that are present (are non empty) in the specific authentication group. A pertinent measure of system diversity and performance then becomes the cardinality $\rho_e = |\Phi_e|$. We note that clearly both Φ_e and ρ_e are random variables, whose realizations may change with each realization of the authentication group. To gain insight on the above randomness, we consider the case where the authentication groups are each time drawn from general population that is a fixed set of $K = 646$ subjects taken from the FERET database [Fer11], with $\rho = 1152$ categories, corresponding to a pdf $P(\phi)$ as shown in Figure 3.1, where this pdf itself corresponds to the traits and trait-instances of the proposed system.

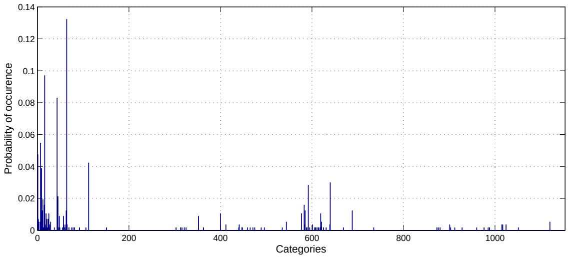


Figure 3.1: $P(\phi)$ corresponding to FERET distribution and the proposed system.

Given the above, Figure 3.2 describes the average number of empty categories, $\rho - \mathbb{E}[\rho_e](n)$, as a function of n, where the expectation is taken over the different realizations of authentication groups.

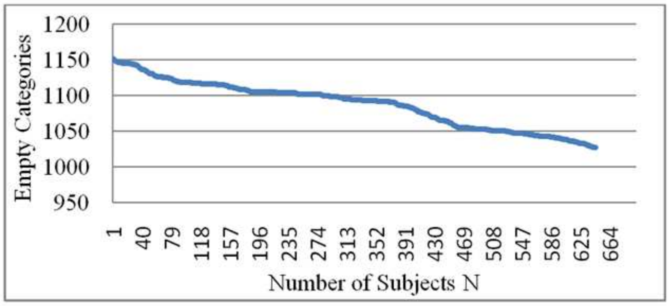


Figure 3.2: Expected number of empty categories as a function of n (FERET).

It becomes apparent that a natural remedy for increasing $\mathbb{E}[\rho_e]$ is to increase the overall ρ, which brings to the fore the natural question as to whether this increase in ρ should be more a result of an increase in the number of traits, or rather more a result of the number of trait-instances (given that the trait-instances do not already span the entire possible range per trait). We address this resource allocation problem, under the simplifying assumption of symmetry, where $\mu_i = \mu$, for all $i = 1, , \lambda$. In this symmetric setting, where clearly

$$\rho = \mu^{\lambda} \tag{3.4}$$

and where ρ increases polynomially with μ and exponentially with λ, a simple comparison of the two derivatives $\frac{d\rho}{d\mu}$, $\frac{d\rho}{d\lambda}$, identifies the *trait-limited region* of a soft biometric system to be the region

$$\lambda < \mu \cdot ln\mu \tag{3.5}$$

in which ρ increases faster with λ than with μ, and where emphasis should be placed on increasing λ rather than μ. This means that for $\rho > 16$, ρ increases faster with λ than with μ, see Table 3.1.

Example 1 *Practical system augmentation for increasing ρ: We propose the bag structure of an augmented system, where an increase in resources (such as an improved resolution of the sensors, or an increased computational capability), can be allocated to include the increased set of traits, and trait-instances, as described in Table 3.2, yielding an impressive ρ in the order of eighty million, which may be suitable for several applications.*

Table 3.1: SBSs with symmetric traits instances

$\mu \backslash \lambda$	2	3	4	5	6
2	4	8	16	32	64
3	9	27	81	243	729
4	16	64	256	1024	4096
5	25	125	625	3125	15625
6	36	216	1296	7776	46656

Table 3.2: Augmented set of facial soft biometric traits and corresponding number of instances

Skin Color	**Hair Color**	**Eye Color**	**Glasses Presence**	**Beard Presence**	**Moustache Presence**	**Age**	**Gender**
3	8	6	2	2	2	3	2
Make up	**Facial Shapes**	**Facial feature Shapes**	**Facial measurements**	**Facial feature measurements**	**Facial moles and marks**	**Hair style**	
4	3	3	3	6	6	3	

This approach in turn, brings to the fore the issue that increasing ρ, may indeed result in an increased $\mathbb{E}[\rho_e]$, but might affect the correlation between the different categories. This would subsequently result in a reduced spread of Φ, which would imply a reduced distinctiveness in identification. In regards to this, we give some intuition on the distinctiveness of some non-empty categories of the proposed system, by computing the correlation between these categories using Pearson's product-moment coefficient

$$r_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{3.6}$$

The resulting correlation parameters shown below

$$r_{EyeColor,HairColor} = -0.1964 \tag{3.7}$$
$$r_{HairColor,SkinColor} = -0.1375 \tag{3.8}$$
$$r_{EyeColor,SkinColor} = 0.370 \tag{3.9}$$
$$r_{Moustache,Beard} = 0.6359 \tag{3.10}$$

revealed as expected the highest correlation to be that between moustache and beard.

Further intuition on related correlations is given in Figure 3.3, which shows the joint pdf with respect to the eye, skin and hair color.

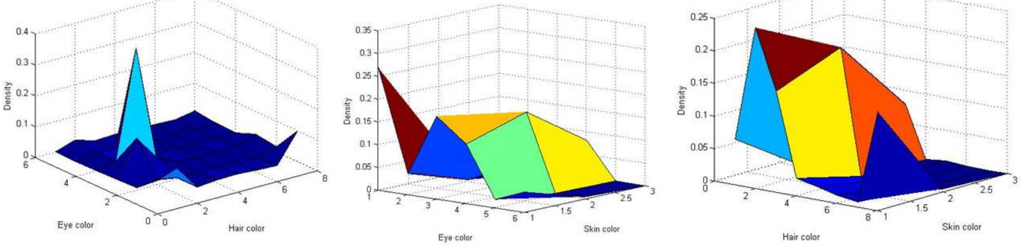


Figure 3.3: Color based soft biometric traits (eye-hair, eye-skin and hair-skin color) distributions for FERET database.

3.4.2 Bounding n for a given interference probability

We are here interested in describing the relationship between n, and the corresponding probability of interference, as a function of Φ. We proceed to properly define the event of collision or interference. *Definition*: The event of *collision*, or equivalently of *interference*, describes the event where any two or more subjects belong in the same category ϕ. Focusing on a specific subject, we say that this subject experiences interference if he/she belongs in a category which also includes other subjects from the authentication group. In regards to this, we are interested in gaining insight on two probability measures. The first measure is the probability $p(n;\rho)$ that the authentication group of size n, chosen randomly from a large population of subjects, is such that there exist two subjects within the group that collide. We briefly note the relationship of $p(n;\rho)$ to the famous *birthday paradox*. For the other measure of system reliability, we consider the case where an authentication group of size n is chosen randomly from a large population of subjects, and where a randomly chosen subject from within this authentication group, collides with another member of the same group. We denote this probability as $q(n)$, and note that clearly $q(n) < p(n)$. To clarify, $p(n)$ describes the probability that interference exists, even though it might not cause error, whereas $q(n)$ describes the probability of an interference induced error. *Example:* In a group of N subjects $p(n)$ would describe the probability that any two subjects will belong to the same category ϕ_x. On the other hand $q(n)$ reflects the probability that a specific subject will interfere with one or more of the $N-1$ remaining subjects. We first focus on calculating and plotting $p(n)$, under the simplifying assumption of statistical uniformity of the categories. The closed form expression for this probability is derived (see [Das05]) to be

$$p(n) = 1 - \bar{p}(N) \tag{3.11}$$

$$p(n) = 1 - 1 \cdot \left(1 - \frac{1}{\rho}\right) \cdot \left(1 - \frac{2}{\rho}\right) \cdots \left(1 - \frac{N-1}{\rho}\right) \tag{3.12}$$

$$p(n) = 1 - \frac{\rho!}{\rho^n\,(\rho - n)!}. \tag{3.13}$$

We note that under the uniformity assumption, the above described $p(n;\rho)$ forms a lower bound on this same probability (in the absence of the same assumption). Equivalently, from the above, we can also compute the maximum n that will allow for a certain probability of collision. In terms of a closed form expression, this is accommodated by using the approximation from [AM00]:

$$p(n;\rho) \approx 1 - e^{-\frac{n(n-1)}{2\rho}} = 1 - \left(\frac{\rho - 1}{\rho}\right)^{\frac{n(n-1)}{2}} \tag{3.14}$$

$$n(p;\rho) \approx \sqrt{2\rho \cdot ln\left(\frac{1}{1-p}\right)}, \tag{3.15}$$

corresponding to the value of n for which the system will introduce interference probability equal to p. As an example, we note that for $\rho = 1152$, and $p = 0.5$, then n = 39. In regards to $q(n)$, the closed form expression is readily seen to be

$$q(n) = 1 - (\frac{\rho - 1^n}{\rho}). \tag{3.16}$$

As an example we note that under the uniformity assumption, and given $\rho = 1152$, and $q = 0.5$, then $n > 700$, which, as expected, is much higher than the pessimistic equivalent corresponding to $p(n,\rho)$. Towards generalizing, we deviate from the uniformity assumption, to rather consider a more realistic setting where the category distribution originates from an online survey (see [hai10]), of 5142 subjects from Central Germany. For computational simplicity we choose to consider a simpler, reduced version of our proposed system, where the traits are limited to hair color and eye color. In this setting, the hair color trait has 7 trait-instances, and the eye color trait has 5 trait instances, resulting in a total of $\rho = 35$ categories, with probabilities $P(\phi_i), i = 1, \dots, 35$.

In this case the probability that all n subjects are in different categories is the sum of the products of all non-colliding events [JDP92]:

$$p_{non\_collision}(n) = \sum_{i \neq j \neq \cdots \neq z} P(\phi_i)P(\phi_j)\dots P(\phi_z) \tag{3.17}$$

where the summation indexing corresponds to the non-empty categories with respect to the authentication group.

3.4.3 Simulation evaluation of the system in the interference limited setting

In the following we provide a simulation of the probability of identification error, in the setting of interest, under the assumption that the errors are due to interference, i.e., under the assumptions that errors only happen if and only if the chosen subject shares the same category with another person from the randomly chosen authentication group. This corresponds to the setting where the soft-biometric approach cannot provide conclusive identification. In the simulation, the larger population consisted of 646 people from the FERET database, and the simulation was run for different sizes n of the authentication group. The probability of identification error is described in the following figure.

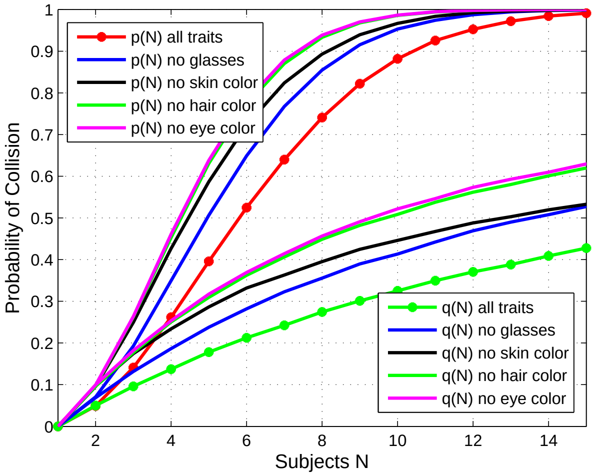


Figure 3.4: Collision probability in an n sized authentication group.

As a measure of the importance of each trait, Figure 3.4 describes the collision probability when different traits are removed. The presence of moustache and beard seem to have the least influence on the detection results, whereas hair and eye color have the highest impact on distinctiveness.

3.5 Reliability and scaling laws

In this section we seek to provide insightful mathematical analysis of reliability of general soft biometric systems (SBSs), as well as to study error events and underlying factors. Furthermore we will proceed to articulate related closed form expressions for the single and averaged SBS behavior by concisely describing the asymptotic behavior of pertinent statistical parameters that are identified to directly affect performance. Albeit its asymptotic and mathematical nature, the approach aims to provide simple expressions that can yield insight into handling real life surveillance systems.

Along with the general setting from section 3.1, in which we referred to the randomly extracted set of n people, out of which one person is picked for identification, we here introduce the notation of $\boldsymbol{v}$ for such a n-tuple of people. Furthermore we denote by $\boldsymbol{v}(i)$, $i = 1, ..., n$ the i-th candidate belonging to the specific group $\boldsymbol{v}$.

3.5.1 Error events, interference, and effective categories

Let the randomly chosen subject for identification, belong in category $\phi \in \Phi$. The SBS first produces an estimates $\widehat{\phi}$ of ϕ, and based on this estimate, tries to identify the chosen subject, i.e., tries to establish which candidate in $\boldsymbol{v}$ corresponds to the chosen subject. An error occurs when the SBS fails to correctly identify the chosen subject, confusing him or her with another candidate from the current n-tuple $\boldsymbol{v}$. An error can hence certainly occur when the category is incorrectly estimated [1], i.e., when $\widehat{\phi} \neq \phi$, or can possibly occur when the chosen subject $\boldsymbol{v}(i)$ interferes with another subject $\boldsymbol{v}(j)$ from the authentication group $\boldsymbol{v}$, i.e., when the chosen subject is essentially indistinguishable to the SBS from some other candidates in $\boldsymbol{v}$.. We recall that interference occurs, whenever two or more subjects belong in the same category.

For a given $\boldsymbol{v}$, let $S_\phi \subset \boldsymbol{v}$ be the set of subjects in $\boldsymbol{v}$ that belong in a specific category ϕ. Furthermore let S_0 denote the set of people in $\boldsymbol{v}$ who do not belong in any of the categories in Φ. We here note that no subject can simultaneously belong to two or more categories, but also note that it is entirely possible that $|S_\phi| = 0$, for some $\phi \in \Phi$. Hence an error is caused due to estimation noise (resulting in $\widehat{\phi} \neq \phi$), due to interference (subject indistinguishable from other subjects in $\boldsymbol{v}$), or when the chosen candidate belongs in S_0 (system is not designed to recognize the subject of interest).

For a given $\boldsymbol{v}$, let

$$F(\boldsymbol{v}) := |\{\phi \in \Phi : |S_\phi| > 0\}|$$

denote the number of effective categories, i.e., the number of (non-empty) categories that fully characterize the subjects in $\boldsymbol{v}$. For notational simplicity we henceforth write F to denote $F(\boldsymbol{v})$, and we let the dependence on $\boldsymbol{v}$ be implied.

3.5.1.1 The role of interference on the reliability of SBSs

Towards evaluating the overall probability of identification error, we first establish the probability of error for a given set (authentication group) $\boldsymbol{v}$. We note the two characteristic extreme instances of $F(\boldsymbol{v}) = n$ and $F(\boldsymbol{v}) = 1$. In the first case, the random n-tuple $\boldsymbol{v}$ over which identification will take place, happens to be such that each subject in $\boldsymbol{v}$ belongs to a different category, in which case none of the subjects interferes with another subject's identification. On the other hand, the second case corresponds to the (unfortunate) realizations of $\boldsymbol{v}$ where all subjects in $\boldsymbol{v}$ fall under

[1] this possibility will be addressed later on

the same category (all subjects in $\boldsymbol{v}$ happen to share the same features), and where identification is highly unreliable.

Before proceeding with the analysis, we briefly define some notation. First we let $P_\phi,\ \phi \in \Phi$, denote the probability of incorrectly identifying a subject from S_ϕ, and we adopt for now the simplifying assumption that this probability be independent of the specific subject in S_ϕ. Without loss of generality, we also let $S_1, \cdots, S_F$ correspond to the $F(\boldsymbol{v}) = F$ non-empty categories, and note that $F \leq n$ since one subject can belong to just one category. Furthermore we let

$$S_{id} := \cup_{\phi=1}^{F} S_\phi$$

denote the set of subjects in $\boldsymbol{v}$ that can potentially be identified by the SBS 'endowed' with Φ, and we note that $S_{id} = \cup_{\phi=1}^{\rho} S_\phi$. Also note that $|S_0| = n - |S_{id}|$, that $S_\phi \cap S_{\phi'} = \emptyset$ for $\phi' \neq \phi$, and that

$$|S_{id}| = \sum_{\phi=1}^{F} |S_\phi|.$$

We proceed to derive the error probability for any given $\boldsymbol{v}$.

Lemma 1 *Let a subject be drawn uniformly at random from a randomly drawn n-tuple $\boldsymbol{v}$. Then the probability $P(err|\boldsymbol{v})$ of erroneously identifying that subject, is given by*

$$P(err|\boldsymbol{v}) = 1 - \frac{F - \sum_{\phi=1}^{F} P_\phi}{n}, \tag{3.18}$$

where $F(\boldsymbol{v}) = F$ is the number of effective categories spanned by $\boldsymbol{v}$.

The following corollary holds for the interference limited case where errors due to feature estimation are ignored [2], i.e., where $P_\phi = 0$.

Corollary 0.1 *For the same setting and measure as in Lemma 1, under the interference limited assumption, the probability of error $P(err|\boldsymbol{v})$ is given by*

$$P(err|\boldsymbol{v}) = 1 - \frac{F}{n}, \tag{3.19}$$

for any $\boldsymbol{v}$ such that $F(\boldsymbol{v}) = F$.

The above reveals the somewhat surprising fact that, given n, the reliability of an SBS for identification of subjects in $\boldsymbol{v}$, is independent of the subjects' distribution $\boldsymbol{v}$ in the different categories, and instead only depends on F. As a result this reliability remains identical when employed over different n-tuples that fix F.

Proof of Lemma 1: See Appendix A.

We proceed with a clarifying example.

Example 2 *Consider an SBS equipped with three features ($\rho = 3$), limited to (correctly) identifying dark hair, gray hair, and blond hair, i.e., $\Phi = \{$'dark hair' $= \phi_1$, 'gray hair' $= \phi_2$, 'blond hair' $= \phi_3\}$. Consider drawing at random, from a population corresponding to the residents of Nice, three n-tuples, with $n = 12$, each with a different subject categorization, as shown in Table 3.3. Despite their different category distribution, the first two sets $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ introduce the same number of effective categories $F = 3$, and hence the same probability of erroneous detection $P(err|\boldsymbol{v}_1) = P(err|\boldsymbol{v}_2) = 3/4$ (averaged over the subjects in each set). On the other hand for $\boldsymbol{v}_3$ with $F = 2$, the probability of error increases to $P(err|\boldsymbol{v}_3) = 5/6$.*

[2] we assume that estimation causes substantially fewer errors than interference does and ignore them for now

Table 3.3: Illustration of Example 2

	ϕ_1	ϕ_2	ϕ_3	F	$P(\text{err}\|\boldsymbol{v})$
$\boldsymbol{v}_1$	10	1	1	3	$3/4$
$\boldsymbol{v}_2$	4	4	4	3	$3/4$
$\boldsymbol{v}_3$	10	2	0	2	$5/6$

Up to now the result corresponded to the case of specific realizations of $\boldsymbol{v}$, where we saw that the probability of error for each realization of length n, was a function only of the realization of $F(\boldsymbol{v})$ which was a random variable describing the number of categories spanned by the specific group $\boldsymbol{v}$. We now proceed to average over all such realizations $\boldsymbol{v}$, and describe the overall probability of error. This analysis is better suited to evaluate an ensemble of distributed SBSs deployed over a large population. *We henceforth focus on the interference limited setting* [3] *i.e., we make the simplifying assumption that* $P_\phi = 0,\ \phi > 0$.

Lemma 2 *The probability of error averaged over all n-tuples $\boldsymbol{v}$ randomly drawn from a sufficiently large population, is given by*

$$\mathbb{E}_{\boldsymbol{v}}[P(err|\boldsymbol{v})] = 1 - \frac{\mathbb{E}_{\boldsymbol{v}}[F(\boldsymbol{v})]}{n}, \tag{3.20}$$

and is dependent only on the first order statistics of F.

Proof: The proof follows directly from Lemma 1. □

An example follows, related to the above.

Example 3 *Consider the case where the city of Nice installs throughout the city a number of independent SBSs* [4] *and is interested to know the average reliability that these systems will jointly provide, over a period of two months* [5]. *The result in Lemma 2 gives the general expression of the average reliability that is jointly provided by the distributed SBSs, indexed by n, for all n. Indexing by n simply means that the average is taken over all cases where identification is related to a random set $\boldsymbol{v}$ of size n.*

We now proceed to establish the statistical behavior of F, including the mean $\mathbb{E}[F]$.

Despite the fact that the probability of error in (3.24) is a function only of the first moment of F, our interest in the entire probability density function stems from our desire to be able to understand rare behaviors of F. More on this will be seen in the asymptotic analysis that will follow.

[3] It is noted though that with increasing ρ, the probability of erroneous identification is, in real systems, expected to increase. This will be considered in future work. Toward motivating the interference limited setting, we note that such setting generally corresponds to cases where a very refined SBS allows for ρ to be substantially larger than n, thus resulting in a probability of interfence that is small but non negligible and which has to be accounted for.

[4] Independence follows from the assumption that the different SBSs are placed sufficiently far apart.

[5] In this example it is assumed that the number of independent SBSs and the time period are sufficiently large to jointly allow for ergodicity.

3.5.2 Analysis of interference patterns in SBSs

Given ρ and $n < \rho$, we are interested in establishing the probability $P(F)$ that a randomly drawn n-tuple of people will have F active categories out of a total of $\min(\rho, n)$ possible active categories [6]. We here accept the simplifying assumption of *uniform distribution* of the observed subjects over the categories ρ, i.e., that

$$P(\boldsymbol{v}(i) \in S_\phi) = \frac{1}{\rho}, \ \forall \phi \in \Phi, \ i \leq n. \tag{3.21}$$

We also accept that $n < \rho$. The following then holds.

Lemma 3 *Given ρ and n, and under the uniformity assumption, the distribution of F is described by*

$$P(F) = \frac{F^{n-F}}{(\rho - F)!(n-F)! \sum_{i=1}^{n} \frac{i^{n-i}}{(n-i)!(\rho-i)!}}, \tag{3.22}$$

where F can take values between 1 and n.

Proof of Lemma 3: See Appendix A.

Example 4 *Consider the case where $\rho = 9, n = 5, F = 3$. Then the cardinality of the set of all possible n-tuples that span $F = 3$ effective categories, is given by the product of the following three terms.*

- *The first term is $(\rho \cdot (\rho - 1) \cdots (\rho - F + 1)) = \frac{\rho!}{(\rho - F)!} = 9 \cdot 8 \cdot 7 = 504$ which describes the number of ways one can pick which $F = 3$ categories will be filled.*
- *Having picked these $F = 3$ categories, the second term is $(n \cdot (n-1) \cdots (n - F + 1)) = \frac{n!}{(n-F)!} = 5 \cdot 4 \cdot 3 = 60$, which describes the number of ways one can place exactly one subject in each of these picked categories.*
- *We are now left with $n - F = 2$ subjects, that can be associated freely to any of the $F = 3$ specific picked categories. Hence the third term is $F^{n-F} = 3^2 = 9$ corresponding to the cardinality of $\{1, 2, \cdots, F\}^{n-F}$.*

Motivated by Lemma 2, we now proceed to describe the first order statistics of F. The proof is direct.

Lemma 4 *Under the uniformity assumption, the mean of F is given by*

$$\mathbb{E}_{\boldsymbol{v}}[F(\boldsymbol{v})] = \sum_{F=1}^{n} F P(F) = \sum_{F=1}^{n} \frac{\frac{F^{n-F+1}}{(\rho-F)!(n-F)!}}{\sum_{i=1}^{n} \frac{i^{n-i}}{(n-i)!(\rho-i)!}}. \tag{3.23}$$

Remark 1 *The event of no interference corresponds to the case where $F = n$. Decreasing values of $\frac{F}{n}$ imply higher degrees of interference. An increasing ρ also results in reduced interference.*

Related cases are plotted in Figure 3.5.

Finally, directly from the above, we have the following.

[6] Clarifying example: What is the statistical behavior of F that is encountered by a distributed set of SBSs in the city of Nice?

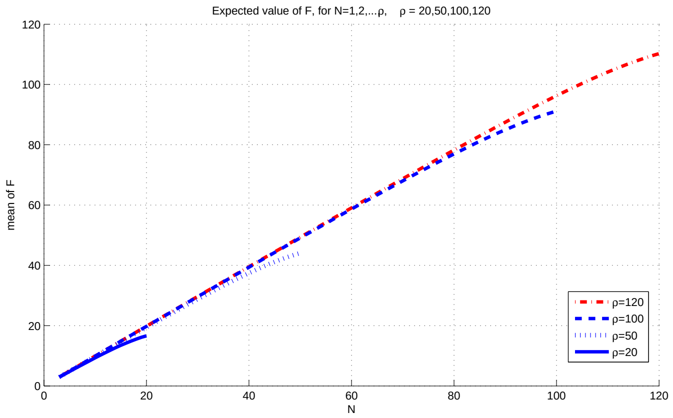


Figure 3.5: $\mathbb{E}_{\boldsymbol{v}}[F]$ for $\rho = 20, 50, 100, 120$, $n \in [3, 4, ..., \rho]$. We note that for ρ sufficiently larger than n, then $\mathbb{E}_{\boldsymbol{v}}[F] \approx n$.

Theorem 1 *In the described operational setting of interest, under the interference limited and uniformity assumptions, the probability of error averaged over all possible n-tuples $\boldsymbol{v}$, that is provided by an SBS endowed with ρ categories, is given by*

$$\mathbb{E}_{\boldsymbol{v}}[P(err)] = 1 - \frac{F^{n-F+1}}{(\rho - F)!(n - F)!n \sum_{i=1}^{n} \frac{i^{n-i}}{(n-i)!(\rho-i)!}}. \tag{3.24}$$

Proof of Theorem 1: The proof is direct from Lemma 2 and from (3.23). □
Related examples are plotted in Figure 3.6.

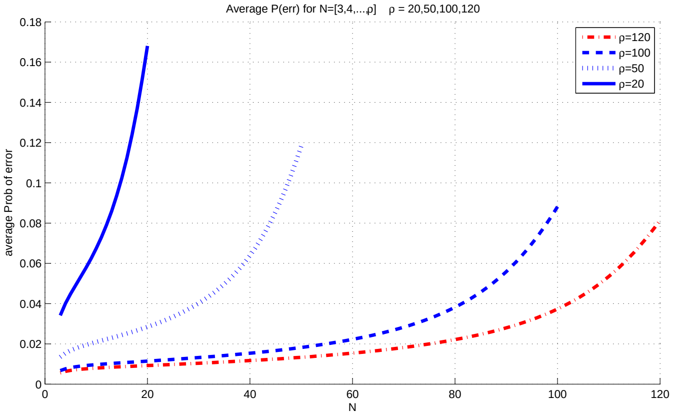


Figure 3.6: $\mathbb{E}_{\boldsymbol{v}}[P(\text{err})]$ for $\rho = 20, 50, 100, 120$, $n \in [3, 4, ..., \rho]$.

We proceed to explore scaling laws of SBS employed for human identification.

3.5.3 Asymptotic bounds on subject interference

In this section we seek to gain insight on the role of increasing resources (increasing ρ) in reducing the subject interference experienced by an SBS. Specifically we seek to gain insight on the following practical question: if more funds are spent towards increasing the quality of an SBS by increasing ρ, then what reliability gains do we expect to see? This question is only partially answered here, but some insight is provided in the form of bounds on the different subject-interference patterns seen by an SBS. The asymptotic bounds simplify the hard to manipulate results of Lemma 3 and Theorem 1, and provide insightful interpretations. A motivating example is presented before the result.

Example 5 *Consider an SBS operating in the city of Berlin, where for a specific n, this system allows for a certain average reliability. Now the city of Berlin is ready to allocate further funds, which can be applied towards doubling the number of categories ρ that the system can identify. Such an increase can come about, for example, by increasing the number and quality of sensors, which can now better identify more soft-biometric traits. The natural question to ask is how this extra funding will help to improve the system? The bounds, when tight, suggest that doubling ρ, will result in a doubly exponential reduction in the probability that a specific degree of interference will occur.*

Further clarifying examples that motivate this approach are given in Section 3.5.3.1.

The following describes the result.

Lemma 5 *Let*

$$h := \lim_{\rho \to \infty} \frac{n}{\rho}, \tag{3.25}$$

define the relative throughput *of a soft biometrics system, and let $F := fn,\ 0 \leq f \leq 1$. Then the asymptotic behavior of $P(F)$ is bounded as*

$$-\lim_{\rho \to \infty} \frac{1}{\rho \log \rho} \log P(f) \geq 2 - h(1 + f). \tag{3.26}$$

Proof of Lemma 5: See Appendix A.

3.5.3.1 Interpretation of bounds

Lemma 5 bounds the statistical behavior of $P(F)$ in the high ρ regime (for large values of ρ). To gain intuition we compare two cases corresponding to two different relative-throughput regimes. In the first case we ask that n is close to ρ, corresponding to the highest relative-throughput of $r = 1$, and directly get from (3.26) that $d(h, f) := 2 - h(1 + f) = d(1, f) = 1 - f,\ 0 < f < 1$. In the second case we reduce the relative-throughput to correspond to the case where n is approximately half of ρ ($h = 1/2$), which in turn gives $d(h, f) = d(\frac{1}{2}, f) = \frac{3}{2} - \frac{f}{2},\ 0 < f < 1/2$. As expected $d(\frac{1}{2}, f) > d(1, f),\ \forall f \leq \frac{1}{2}$.

Towards gaining further insight, let us use this same example to shed some light on how Lemma 5 succinctly quantifies the increase in the probability that a certain amount of interference will occur, for a given increase in the relative-throughput of the soft biometrics system. To see this, consider the case where there is a deviation away from the typical $f = h$ by some small *fixed* ϵ, to a new $f = h - \epsilon$, and note that the value of ϵ defines the extend of the interference [7], because a larger ϵ implies a smaller f, and thus a reduced F for the same n. In the high relative-throughput case of our example, we have that $f = h - \epsilon = 1 - \epsilon$, and thus that $d(1, 1 - \epsilon) = \epsilon$, which implies that the probability of such deviation (and of the corresponding interference) is in the order of $\rho^{-\rho d(1,1-\epsilon)} = \rho^{-\rho\epsilon}$. On the other hand, in the lower relative-throughput case where $f = h - \epsilon = \frac{1}{2} - \epsilon$, we have that $d(\frac{1}{2}, \frac{1}{2} - \epsilon) = \frac{5}{4} + \frac{\epsilon}{2}$, which implies that the probability of the same deviation in the lower throughput setting is in the order of $\rho^{-\rho d(\frac{1}{2},\frac{1}{2}-\epsilon)} = \rho^{-\rho(\frac{5}{4}+\frac{\epsilon}{2})} << \rho^{-\rho\epsilon}$.

[7] Note that interference may occur only if $\epsilon > 0$.

In other words the bound in Lemma 5 implies that, a reduction of the relative-throughput from its maximal value of $n/\rho \approx 1$ to a sufficiently smaller $n/\rho \approx \frac{1}{2}$, for high enough ρ, results in a substantial and exponential reduction in the probability of interference, from $P(h = 1) \approx \rho^{-\rho\epsilon}$ to $P(h = \frac{1}{2}) \approx \rho^{-\rho(\frac{5}{4}+\frac{\epsilon}{2})}$.

We have up to now focused on the interference limited scenario, where errors occur only due to more than one subject belonging to one category. In the next section 3.5.4 we consider estimation error and a more pragmatic way to improve the overall reliability of a SBS.

3.5.4 Estimation reliability

In the aforementioned operational setting of interest, the *reliability* of an SBS captures the probability of false identification of a randomly chosen person out of a random set of n subjects. In such a setting, the reliability of an SBS is generally related to:

- the number of categories that the system can identify.
- the degree with which these features/categories represent the chosen set (of subjects) over which identification will take place
- n, where a higher n corresponds to identifying a person among an increasingly large set of possibly similar-looking people
- robustness with which these categories can be detected

We here proceed to study the general SBS error probability, containing inevitably all above mentioned factors including the algorithmic categorization error–probabilities. With other words we examine, the re–identification error probability, regardless of the underlying source, which can be both due to misclassifcation or due to interference.

Given the knowledge of the population statistics and moreover the exact algorithmic reliabilities (true detection rates and additionally the confusion probabilities), we can use a maximum–likelihood (ML) optimizing rule to compute the maximal posterior probability for each category. We note here that the ML optimizing rule for the most probable category in which a chosen subject belongs, is given by:

$$\hat{\phi} = \text{argmax}_{\phi \in \Phi} P(\phi) \cdot P(y/\phi), \tag{3.27}$$

where y is the observation vector, $P(\phi)$ is the pdf of the set of categories over the given population (note $\sum_{ß=1}^{\rho} P(\phi_i) = 1$), and $P(y/\phi)$ the probability that y is observed, given that the subject belongs in category ϕ.

3.5.4.1 Improving SBS reliability

In the most common case of a training set, which provides insufficient information on all confusion factors as of all $P(y/\phi)$, we can find heuristic rules to combat the overall error probability P_{err}. Given the large amount of empty categories, see the distribution of over-all-categories in the FERET population in Figure 3.1, and furthermore the above presented correlations between traits, certain misclassifications can be identified and reconciled. An example for a heuristic error conciliation attempt can be the following.

Example 6 *We take into account the given large FERET population and the SBS presented in section 3.3.1. We simulate again the same identification scenario, where here we simulate an estimation error of* 10% *for the color soft biometric traits (hair, skin and eye color). When fusing the traits on decision level (hard fusion) those errors naturally add up. That is why a soft decision, taking into account the confidence levels of the extracted features and also the reliability of the*

Table 3.4: Example for a heuristic rule. SBS endowed with $\rho = 4$ categories and a given known population (distribution in the 4 categories). If our SBS estimates category ϕ_2 for a subject to belong into, due to the 0 probability of occurrence, the system decides for the next probable category, namely ϕ_1.

Category	μ_1	μ_2	**Probability for occurrence of ϕ_i given n**
ϕ_1	0	0	0.5
ϕ_2	0	1	0
ϕ_3	1	0	0.3
ϕ_4	1	1	0.2

underlying algorithm is used, since it will discriminate some error cases. In the classification step we can easily identify subjects classified into "empty" categories. Since those categories are of probability 0 to occur, due to the known population, an intelligent classification system recognizes these cases and reclassifies those subjects in "similar" categories with higher probabilities of occurrence(=non–empty categories). A "similar" category hereby is a category with highest probability for misclassification, given the wrongly detected category. We here assume that an error caused by misclassification of one trait is more probable than the misclassification of two or more traits. For visualization see Table 3.4: if a subject is classified into the category ϕ_2, the system recognizes that ϕ_2 is an empty category and searches for the next probable category, which in this case would be ϕ_1. This simple heuristic rule leads already to a significant reconciliation of the added up estimation error of the SBS, see Figure 3.7.

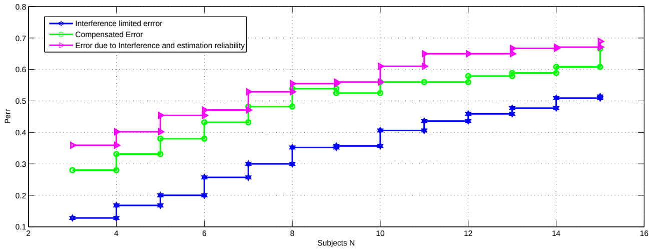


Figure 3.7: Errors in a SBS system: interference limited error, estimation reliability error, compensated error.

In the following we outline an additional error, which can be associated to SBSs when used in a scenario, where the traits are re–identified based on a human description.

3.5.4.2 Human machine interaction

The immense benefit of having human understandable and compliant soft biometric traits over classical biometrics, enables a computer aided biometric search to have as an input a human description of the target subject. The importance of related applications becomes evident in cases, such as a loss of a child in a mall, where the mother can just provide a description of the child and computer based search can be performed on available security video material. Along with the benefit of human compliance come though additional quantification and human-machine interaction errors. Such errors can have different causes.

- Quantification error: the discrete values of soft biometric traits are mapped onto a limited amount of bins and cause such an error. A lower amount of bins corresponds to less misclassifications at the cost though of decreasing distinctiveness of this trait, as elaborated above.
- The nomenclature of traits varies and is ambiguous. For example a hair color that might be denoted with "red" can be labeled by different subjects as a variety of synonyms: auburn, orange, copper, reddish; but also as a completely different trait e.g. brown. A related study establishing labels for soft biometric traits with the Mechanical Turk was recently conducted by the authors in [CO11].
- Different people perceive different traits (e.g. colors) differently. Specifically if the witness of a crime has a different color understanding than the SBS performing the search it can lead to an erroneous search. This aspect though can be minimized if the witness is asked to point at reference colors than just human labeling.
- The awareness of people can be bad or wrong in how they remember traits.
- Often occurring mixed categories like red-brown for hair color can be challenging for all, human perception, the SBS - training and - classification step.

To visualize just the quantification error introduced by a human understandable SBS we have the following simulation. We display in Figure 3.8 on the one hand purely the collision probability of subjects with 8 quantification bins(=traits instances) of hair color (light blond, dark blond, red, brown, black, grey, white and bald). On the other hand we have the re–identification of non–quantified and discrete computer–to–computer search. It is of interest, that even in the presence of an estimation error the over all error probability is decreased in the absence of quantification error. A full computer–to–computer is presented in Chapter 5 and specifically in Figure 5.4, where the performance of a SBS employing AdaBoost [FHT98] boosted algorithms for hair, skin and cloths color, their textures and patch histograms is illustrated. The system is used for frontal–to–side re–identification.

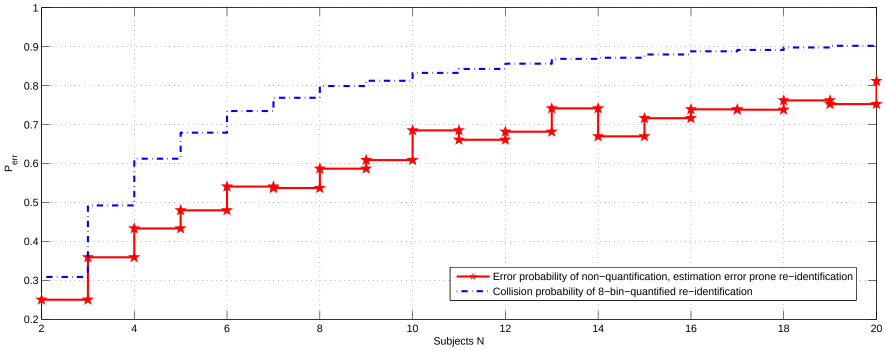


Figure 3.8: Re–identification error for hair color in an n-sized authentication group.

3.6 Summary

In this chapter we explored the use of multi-trait SBSs for human identification, studying analytically the relationship between an authentication group $\boldsymbol{v}$, its size n, the featured categories ρ, and the effective categories F. Then we proceeded to show that in the interference limited setting, for a given randomly chosen authentication group $\boldsymbol{v}$, of a given size n, the reliability of identification (averaged over the subjects in $\boldsymbol{v}$) is a function only of the number of non-empty categories $F(\boldsymbol{v})$. Furthermore we provided statistical analysis of this reliability, over large populations. The latter part provided bounds that, in the interference limited setting suggest an *exponential* reduction in the probability of interference patterns, as a result of a *linear* increase in ρ. Finally we made some observations regarding algorithmic estimation and gave an example of how to counteract, given known population statistics.

Having analyzed in this chapter pertinent measures and dynamics in the process of human identification based on SBSs, we proceed in the next chapter 4 to study the process of employing SBSs for pruning a large database search. The goal will be then not to identify a subject, but rather to pre-filter such a large database for a consecutive processing with a more reliable algorithm, e.g. classical face recognition.

Chapter 4

Search pruning in video surveillance systems

In recent years we have experienced an increasing need to structure and organize an exponentially expanding volume of data that may take the form of, among other things, images and videos. Crucial to this effort is the often computationally expensive task of algorithmic search for specific elements placed at unknown locations inside large data sets. To limit computational cost, pre-filtering such as pruning can be used, to quickly eliminate a portion of the initial data, an action which is then followed by a more precise and complex search within the smaller subset of the remaining data. Such pruning methods can substantially speed up the search, at the risk though of missing the target, thus reducing the overall reliability. Common pre-filtering methods include video indexing and image classification with respect to color [SB91], patterns, objects [AAR04], or feature vectors [NJT06].

4.1 Categorization-based pruning

Our interest in analyzing this speed vs. reliability tradeoff, focuses on the realistic setting where the search is time-constrained and where, as we will see later on, the environment in which the search takes place is stochastic, dynamically changing, and can cause search errors. We note here that there is a fundamental difference between search in unstructured versus structured data, where the latter can be handled with very efficient algorithms, such as the sphere decoding algorithm. One widely known practical scenario that adheres to the above stochastic setting, is the scenario of biometric-based video surveillance. In this setting a time constrained search seeks to identify a subject from within a large set of individuals that may consist of, for example, the people surrounding the subject in a specific instance at a specific location. In the language of biometrics we provide analysis on the general speed-reliability behavior in search pruning. In this scenario, a set of subjects can be pruned by means of categorization that is based on different combinations of soft biometric traits such as facial color, shapes or measurements. The need for such biometrically-based search pruning often comes to the fore, such as in the case of the 2005 London bombing and the 2011 London riots where a sizeable fraction of the police force worked for days to screen a fraction of the available surveillance videos relating to the event.

We stay focused on search pruning based on soft biometrics but remind the reader that this analysis can be generally applied to several domains of computer vision or other disciplines that adhere to the setting of categorization-based pruning in time-constrained searches over error-prone stochastic environments. We clarify that we refer to *pruning the search* as the categorization and further elimination of categories, which limits large databases of subjects to a fraction of the initial database, see Figure 4.1. In the context of this chapter the elimination or filtering of the employed categories is based on the soft biometric characteristics of the subjects. The pruned database can be subsequently processed by humans or by a biometric such as face recognition.

The approach of pruning the search using SBSs, can apply to several re-identification scenarios, including the following:

- A theft in a crowded mall is observed by different people who give partial information about the thief's appearance. Based on this information, a first-pass search applies SBS methods to cut down on the long surveillance video recordings from several cameras.
- A mother has lost her child and can describe traits like clothes color and height of the child. Video surveillance material can be pruned and resulting suggestions can be displayed to the mother.

The above cases support the applicability of SBSs, but also reveal that together with the benefits of such systems, come considerable risks such as that of erroneously pruning out the target of the search. This brings to the fore the need to jointly analyze the gains and risks of such systems.

In the setting of human identification, we consider the scenario where we search for a specific *subject of interest*, denoted as v', belonging to a large and randomly drawn *authentication group* $\boldsymbol{v}$ of n subjects, where each subject belongs to one of ρ *categories*. The elements of the set (authentication group) $\boldsymbol{v}$ are derived randomly from a larger population, which adheres to a set of population statistics. A category corresponds to subjects who adhere to a specific combination of soft biometric characteristics, so for example one may consider a category consisting of blond, tall, females. We note the analogy to the scenario from chapter 3, but proceed to elaborate on the different goal of the current chapter.

With n being potentially large, we seek to simplify the search for subject v' within $\boldsymbol{v}$ by *algorithmic pruning* based on categorization, i.e., by first identifying the subjects that potentially belong to the same category as v', and by then pruning out all other subjects that have not been estimated to share the same traits as v'. Pruning is then expected to be followed by careful search of the remaining unpruned set. Such categorization-based pruning allows for a search speedup through a reduction in the search space, from $\boldsymbol{v}$ to some smaller and easier to handle set $\mathcal{S}$ which is the subset of $\boldsymbol{v}$ that remains after pruning, see Figure 4.1 and Figure 4.4. This reduction though happens in the presence of a set of categorization error probabilities $\{\epsilon_f\}$, called confusion probabilities, that essentially describe how easy it is for categories to be confused, hence also describing the probability that the estimation algorithm erroneously prunes out the subject of interest, by falsely categorizing it. This confusion set, together with the set of population statistics $\{p_f\}_{f=1}^{\rho}$ which describes how common a certain category is inside the large population, jointly define the statistical performance of the search pruning, which we will explore. The above aspects will be precisely described later on.

Example 7 *An example of a sufficiently large population includes the inhabitants of a certain city, and an example of a randomly chosen authentication group (n-tuple) $\boldsymbol{v}$ includes the set of people captured by a video surveillance system in the aforementioned city between 11:00 and 11:05 yesterday. An example SBS could be able to classify 5 instances of hair color, 6 instances of height and 2 of gender, thus being able to differentiate between $\rho = 5{\cdot}6{\cdot}2 = 60$ distinct categories. An example search could seek for a subject that was described to belong to the first category of, say, blond and tall females. The subject and the rest of the authentication group of $n = 1000$ people, were captured by a video-surveillance system at approximately the same time and place somewhere in the city. In this city, each SBS-based category appears with probability $p_1, \cdots, p_{60}$, and each such category can be confused for the first category with probability $\epsilon_2, \cdots, \epsilon_{60}$. The*

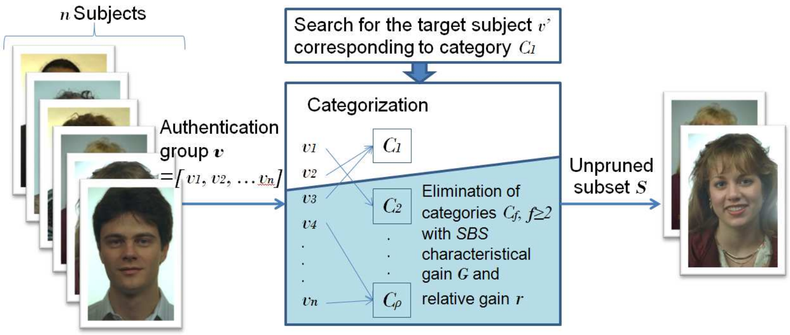

n Subjects
Search for the target subject v'
corresponding to category C_1
Categorization
Authentication group $\boldsymbol{v}$ $=[v_1, v_2, \ldots v_n]$
v_1
v_2
v_3
v_4
v_n
C_1
C_2
C_ρ
Elimination of categories C_f, $f \geq 2$ with *SBS* characteristical gain G and relative gain r
Unpruned subset S

Figure 4.1: System overview.

SBS makes an error whenever v' is pruned out, thus it allows for reliability of ϵ_1. To clarify, having $p_1 = 0.1$ implies that approximately one in ten city inhabitants are blond-tall-females, and having $\epsilon_2 = 0.05$ means that the system (its feature estimation algorithms) tends to confuse the second category for the first category with probability equal to 0.05.

What becomes apparent though is that a more aggressive pruning of subjects in $\boldsymbol{v}$ results in a smaller $\mathcal{S}$ and a higher pruning gain, but as categorization entails estimation errors, such a gain could come at the risk of erroneously pruning out the subject v' that we are searching for, thus reducing the system reliability.

Reliability and pruning gain are naturally affected by, among other things, the distinctiveness and differentiability of the subject v' from the rest of the people in the specific authentication group $\boldsymbol{v}$ over which pruning will take place that particular instance. In several scenarios though, this distinctiveness changes randomly because $\boldsymbol{v}$ itself changes randomly. This introduces a stochastic environment. In this case, depending on the instance in which v' and its surroundings $\boldsymbol{v} - v'$ were captured by the system, some instances would have $\boldsymbol{v}$ consist of bystanders that look similar to the subject of interest v', and other instances would have $\boldsymbol{v}$ consist of people who look sufficiently different from the subject. Naturally the first case is generally expected to allow for a lower pruning gain than the second case.

The pruning gain and reliability behavior can also be affected by the system design. At one extreme we find a very conservative system that prunes out a member of $\boldsymbol{v}$ only if it is highly confident about its estimation and categorization, in which case the system yields maximal reliability (near-zero error probability) but with a much reduced pruning gain. At the other extreme, we find an effective but unreliable system which aggressively prunes out subjects in $\boldsymbol{v}$, resulting in a potentially much reduced search space ($|\mathcal{S}| << n$), at a high risk though of an error. In the above, $|\mathcal{S}|$ denotes the cardinality of set $\mathcal{S}$.

4.2 Contributions

In the next section we elaborate on the concept of *pruning gain* which describes, as a function of pruning reliability, the multiplicative reduction of the set size after pruning: for example a pruning gain of 2 implies that pruning managed to halve the size of the original set. Section 4.5.1 provides average case analysis of the pruning gain, as a function of reliability, whereas Section 4.5 provides atypical-case analysis, offering insight on how often pruning fails to be sufficiently helpful. In the process we try to provide some intuition through examples on topics such as, how the system gain-reliability performance suffers with increasing confusability of categories, or on

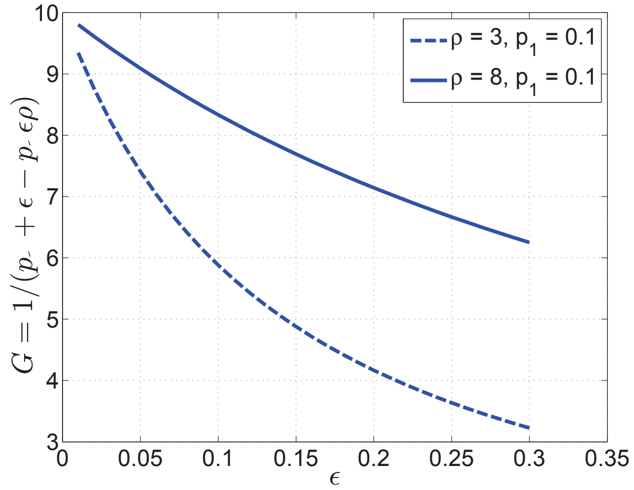


Figure 4.2: Pruning gain, as a function of the confusability probability ϵ, for the uniform error setting, and for $p_1 = 0.1$. Plotted for $\rho = 3$ and $\rho = 8$.

whether searching for a rare looking subject renders the search performance more sensitive to increases in confusability, than searching for common looking subjects. We then present nine different soft biometric systems, and describe how the employed categorization algorithms (eye color detector, glasses and moustache detector) are applied on a characteristic database of 646 people. In Section 4.6.1 we provide simulations that reveal the variability and range of the pruning benefits offered by different SBSs. In Section 4.7 we provide concise closed form expressions on the measures of pruning gain and goodput, provide simulations, as well as derive and simulate aspects relating to the complexity costs of different soft biometric systems of interest.

Before proving the aforementioned results we hasten to give some insight, as to what is to come. In the setting of large n, Section 4.5.1 easily tells us that the average pruning gain takes the form of the inverse of $\sum_{f=1}^{\rho} p_f \epsilon_f$, which is illustrated in an example in Figure 4.2 for different (uniform) confusability probabilities, for the case where the search is for an individual that belongs to a category that occurs once every ten people, and for the case of two different systems that can respectively distinguish 3 or 8 categories. The atypical analysis in Section 4.5 is more involved and is better illustrated with an example, which asks what is the probability that a system that can identify $\rho = 3$ categories, that searches for a subject of the first category, that has 80 percent reliability, that introduces confusability parameters $\epsilon_2 = 0.2, \epsilon_3 = 0.3$ and operates over a population with statistics $p_1 = 0.4, p_2 = 0.25, p_3 = 0.35$, will prune the search to only a fraction of $\tau = |\mathcal{S}|/n$. We note that here τ is the inverse of the pruning gain. We plot in Figure 4.3 the asymptotic rate of decay for this probability,

$$J(\tau) := -\lim_{N \to \infty} \frac{\log}{n/\rho} P(|\mathcal{S}| > \tau n) \tag{4.1}$$

for different values of τ. From the $J(\tau)$ in Figure 4.3 we can draw different conclusions, such as:

– Focusing on $\tau = 0.475$ where $J(0.475) = 0$, we see that the size of the (after pruning) set $\mathcal{S}$ is typically (most commonly - with probability that does not vanish with n) 47.5% of the original size n. In the absence of errors, this would have been equal to $p_1 = 40\%$, but the errors cause a reduction of the average gain by about 15%.
– Focusing on $\tau = 0.72$, we note that the probability that pruning removes less than $1-0.72 = 28\%$ of the original set is approximately given by e^{-n}, whereas focusing on $\tau = 0.62$, we

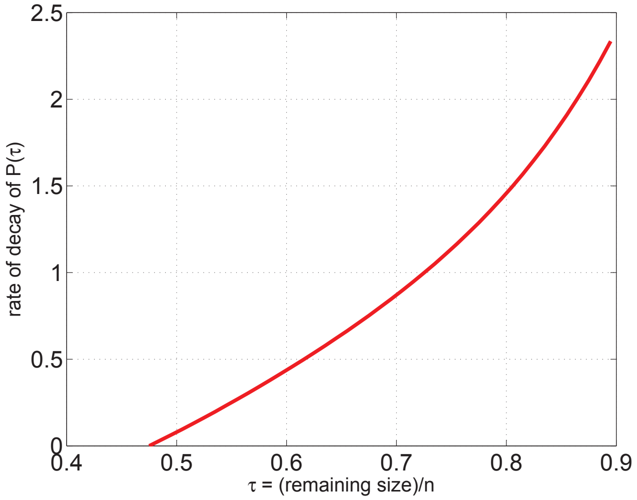


Figure 4.3: Asymptotic rate of decay of $P(|\mathcal{S}| > \tau n)$, for $\rho = 3$, reliability 0.8, population statistics $p_1 = 0.4, p_2 = 0.25, p_3 = 0.35$ and confusability parameters $\epsilon_2 = 0.2, \epsilon_3 = 0.3$.

note that the probability that pruning removes less than $1 - 0.62 = 38\%$ of the original set, is approximately given by $e^{-n/2}$. The probability that pruning removes less than half the elements is approximately $P(\tau > 0.5) \approx e^{-n/10}$.

The expressions from the above graphs will be derived in detail later.

4.3 Gain vs. reliability in soft biometric systems

As an intermediate measure of efficiency we consider the (instantaneous) *pruning gain*, defined here as

$$\mathcal{G}(\boldsymbol{v}) := \frac{n}{|\mathcal{S}|}, \tag{4.2}$$

which simply describes [1] the size reduction, from $\boldsymbol{v}$ to $\mathcal{S}$, and which can vary from 1 (no pruning gain) to n. In terms of system design, one could also consider the *relative gain*,

$$r(\boldsymbol{v}) := 1 - \frac{|\mathcal{S}|}{n} \in [0, 1], \tag{4.3}$$

describing the fraction of people in $\boldsymbol{v}$ that was pruned out.

It is noted here that $\mathcal{G}(\boldsymbol{v})$, and by extension $r(\boldsymbol{v})$, vary randomly with, among other things, the relationship between $\boldsymbol{v}$ and v', the current estimation conditions as well as the error capabilities of the system. For example, we note that if $\boldsymbol{v}$ and v' are such that v' belongs in a category in which very few other members of $\boldsymbol{v}$ belong to, then the SBS-based pruning is expected to produce a very small $\mathcal{S}$ and a high gain. If though, at the same time, the estimation capabilities (algorithms and hardware) of the system result in the characteristics of v' being easily confusable with the characteristics of another populous category in $\boldsymbol{v}$, then $\mathcal{S}$ will be generally larger, and the gain smaller.

As a result, any reasonable analysis of the gain-reliability behavior must be of a statistical nature and must naturally reflect the categorization refinement, the corresponding estimation error capabilities of the system, as well as the statistics of the larger population.

[1] We here assume that the SBS is asked to leave at least one subject in $\mathcal{S}$.

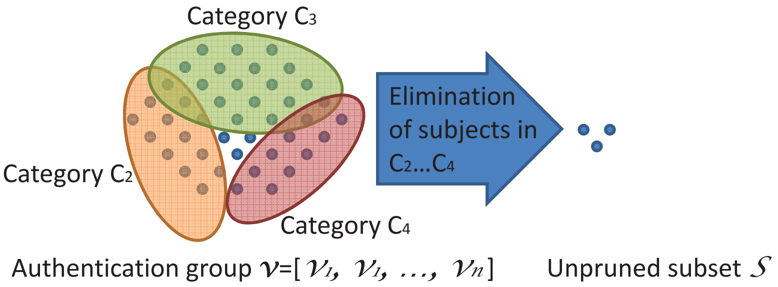

Category C3
Elimination of subjects in C2...C4
Category C2
Category C4
Authentication group $\mathcal{V}$=[$\mathcal{V}_1$, $\mathcal{V}_1$, ..., $\mathcal{V}_n$]
Unpruned subset $\mathcal{S}$

Figure 4.4: Pruning process: categorization and elimination of categories.

4.4 General setting

For this chapter, as mentioned above, we consider the setting where there is a search for a *subject of interest* v', from within a larger *authentication group* of n subjects, $\boldsymbol{v}$. The subject of interest v' is randomly placed inside $\boldsymbol{v}$, and in turn $\boldsymbol{v}$ is randomly drawn from a larger population. Each member of $\boldsymbol{v}$ belongs to one of ρ categories $C_f \subset \boldsymbol{v},\ f = 1, \cdots, \rho$, with probability equal to

$$p_f := \mathbb{E}_{\boldsymbol{v}} \frac{|C_f|}{n}, \quad f = 1, \cdots, \rho, \tag{4.4}$$

where $\mathbb{E}$ is used to denote the statistical expectation. Such category can be for example (labeled as) 'blue eyed, with moustache and with glasses'. The soft biometric system goes through the elements $v \in \boldsymbol{v}$, and provides an estimate $\widehat{C}(v) \in [1, \rho]$ of the category that v belongs in. For C' denoting the actual category of v', where this category is considered to be known to the system, then each element v is pruned out if and only if $\widehat{C}(v) \neq C'$. Specifically the SBS produces a set

$$\mathcal{S} = \{v \in \boldsymbol{v} \ : \ \widehat{C}(v) = C'\} \subset \boldsymbol{v}$$

of subjects that were not pruned out. The pruning gain comes from the fact that $\mathcal{S}$ is generally smaller than $\boldsymbol{v}$.

It is the case that pruning which results in generally smaller $\mathcal{S}$, is associated to a higher gain, but also a higher risk of erroneously pruning out the target subject v', thus reducing the *reliability* of the SBS. Both reliability and pruning gain are naturally affected by different parameters such as

- the category distribution of the authentication group $\boldsymbol{v}$,
- the distinctiveness of the category to which v' belongs
- the system design: a conservatively tuned system will prune only with low risk to prune out v', allowing for a high false acceptance rate FAR, on the other hand an aggressive system will prune stronger with the cost of a higher false rejection rate FRR.

Furthermore, the gain is clearly a function of $\boldsymbol{v}$. Consequently any meaningful analysis of an SBS will have to be statistical in nature. We here consider the average behavior of such systems. In such a case we will see that two aspects prove to be crucial in defining the average case behavior of the system. The first aspect is the population statistics and the second is the error behavior of the different categorization algorithms. Specifically we here consider the vector

$$\boldsymbol{p} := [p_1, p_2, \cdots, p_\rho]^T \tag{4.5}$$

which defines the entire population statistics. In terms of error behavior, we define

$$\epsilon_{ij} := P(\widehat{C}(v) = C_j \ : v \in C_i) \tag{4.6}$$

to be the probability that the algorithms will categorize into the jth category C_j, an element which actually belongs to the ith category C_i (see Figure 4.5 for a graphical illustration). Simply ϵ_{ij} $i, j \in [1, \rho]$ is the element of the ith row and jth column of what is known as the $\rho \times \rho$ *confusion matrix*, which we denote here as $\mathbf{E}$:

$$\mathbf{E} := \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{1\rho} \\ \epsilon_{21} & \epsilon_{22} & \cdots & \epsilon_{2\rho} \\ & \ddots & & \\ \epsilon_{\rho 1} & \epsilon_{\rho 2} & \cdots & \epsilon_{\rho\rho} \end{bmatrix}. \tag{4.7}$$

Related to these parameters we also define

$$\epsilon_f := \sum_{i=1, i \neq f}^{\rho} \epsilon_{fi} \tag{4.8}$$

to denote the probability that a member of category C_f is wrongly categorized. Finally we use the notation

$$\mathbf{e} := [\epsilon_1, \epsilon_2, \cdots, \epsilon_\rho]. \tag{4.9}$$

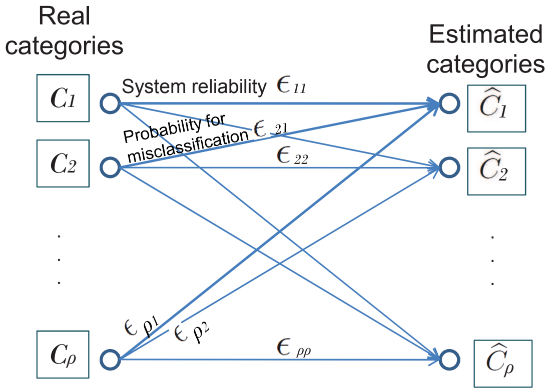


Figure 4.5: Confusion parameters $\{\epsilon_f\}$.

4.5 Statistical analysis using the method of types and information divergence

Let us consider a scenario where a search for a subject v' turned out to be extremely ineffective, and fell below the expectations, due to a very unfortunate matching of the subject with its surroundings $\boldsymbol{v}$. This unfortunate scenario motivates the natural question of how often will a system that was designed to achieve a certain average gain-reliability behavior, fall short of the expectations, providing an atypically small pruning gain and leaving its users with an atypically large and unmanageable $\mathcal{S}$. It consequently brings to the previously related questions such as for example, how will this probability be altered if we change the hardware and algorithmic resources of the system (change the ϵ_f and ρ), or change the setting in which the system operates (change the p_i).

We proceed to analyze these issues and first recall that for a given authentication group $\boldsymbol{v}$, the categorization algorithm identifies set $\mathcal{S}$ of all unpruned subjects, defined as $\mathcal{S} = \{v \in \boldsymbol{v} : \widehat{C}(v) = 1\}$. We are here interested in the size of the search after pruning, specifically in the parameter

$$\tau := \frac{|\mathcal{S}|}{n/\rho},\ 0 \leq \tau \leq \rho, \tag{4.10}$$

which represents [2] a relative deviation of $|\mathcal{S}|$ from a baseline n/ρ. It can be seen that the typical, i.e., common, value of τ is (see also Section 4.5.1)

$$\tau_0 := \mathbb{E}_{\boldsymbol{v}} \frac{|\mathcal{S}|}{n/\rho} = \rho \sum_{f=1}^{\rho} p_f \epsilon_f. \tag{4.11}$$

We are now interested in the entire tail behavior (not just the typical part of it), i.e., we are interested in understanding the probability of having an authentication group $\boldsymbol{v}$ that results in atypically unhelpful pruning ($\tau > \tau_0$), or atypically helpful pruning ($\tau < \tau_0$).

Towards this let

$$\alpha_{0,f}(\boldsymbol{v}) := \frac{|C_f|}{n/\rho}, \tag{4.12}$$

let $\boldsymbol{a}_0(\boldsymbol{v}) = \{\alpha_{0,f}(\boldsymbol{v})\}_{f=1}^{\rho}$ describe the *instantaneous* normalized distribution (histogram) of $\{|C_f|\}_{f=1}^{\rho}$ for the specific, randomly chosen and fixed authentication group $\boldsymbol{v}$, and let

$$\boldsymbol{p} := \{p_f\}_{f=1}^{\rho} = \{\mathbb{E}_{\boldsymbol{v}} \frac{|C_f|}{n}\}_{f=1}^{\rho}, \tag{4.13}$$

denote the *normalized statistical* population distribution of $\{|C_f|\}_{f=1}^{\rho}$.

Furthermore, for a given $\boldsymbol{v}$, let

$$\alpha_{1,f}(\boldsymbol{v}) := \frac{|C_f \cap \mathcal{S}|}{n/\rho},\ 0 \leq \alpha_{1,f} \leq \rho, \tag{4.14}$$

let $\boldsymbol{\alpha}_1(\boldsymbol{v}) := \{a_{1,f}(\boldsymbol{v})\}_{f=1}^{\rho}$, and $\boldsymbol{\alpha}(\boldsymbol{v}) := \{\boldsymbol{\alpha}_0(\boldsymbol{v}), \boldsymbol{\alpha}_1(\boldsymbol{v})\}$, and let [3]

$$\mathcal{V}(\tau) := \big\{0 \leq \alpha_{1,f} \leq \min(\tau, \alpha_{0,f}), \sum_{f=1}^{\rho} \alpha_{1,f} = \tau\big\}, \tag{4.15}$$

denote the set of valid $\boldsymbol{\alpha}$ for a given τ, i.e., describe the set of all possible authentication groups and categorization errors that can result in $|\mathcal{S}| = \tau \frac{n}{\rho}$.

Given the information that $\boldsymbol{\alpha}_1$ has on $\boldsymbol{\alpha}_0$, given that τ is implied by $\boldsymbol{\alpha}_1$, and given that the algorithms here categorize a subject independently of other subjects, it can be seen that for any $\boldsymbol{\alpha} \in \mathcal{V}(\tau)$, it is the case that

$$\begin{aligned} P(\boldsymbol{\alpha}, \tau) &= P(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1) = P(\boldsymbol{\alpha}_0) P(\boldsymbol{\alpha}_1 | \boldsymbol{\alpha}_0) && (4.16)\\ &= \prod_{f=1}^{\rho} P(\alpha_{0,f}) \prod_{f=1}^{\rho} P(\alpha_{1,f} | \alpha_{0,f}). && (4.17) \end{aligned}$$

[2] Note the small change in notation compared to Section 4.2. This change is meant to make the derivations more concise.

[3] For simplicity of notation we will henceforth use $\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}, \alpha_{0,f}, \alpha_{1,f}$ and let the association to $\boldsymbol{v}$ be implied

The following lemma describes the asymptotic behavior of $P(\boldsymbol{\alpha}, \tau)$, for any $\boldsymbol{\alpha} \in \mathcal{V}(\tau)$. To clarify, the lemma describes the asymptotic rate of decay of the joint probability of an authentication group with histogram $\boldsymbol{\alpha}_0$ and an estimation/categorization process corresponding to $\boldsymbol{\alpha}_1$, given that the group and categorization process result in an unpruned set of size

$$|\mathcal{S}| = \tau \frac{n}{\rho} \tag{4.18}$$

for some $0 \leq \tau \leq \rho$. This behavior will be described below as a concise function of the binomial rate-function (see [CT06])

$$I_f(x) \ = \left\{ \begin{array}{ll} x\log(\frac{x}{\epsilon_f}) + (1-x)\log(\frac{1-x}{1-\epsilon_f}) & f \geq 2 \\ x\log(\frac{x}{1-\epsilon_1}) + (1-x)\log(\frac{1-x}{\epsilon_1}) & f = 1. \end{array} \right. \tag{4.19}$$

The lemma follows.

Lemma 6

$$-\lim_{N\to\infty} \frac{\log}{n/\rho} P(\boldsymbol{\alpha}, \tau) = \rho D(\boldsymbol{\alpha}_0 || \boldsymbol{p}) + \sum_{f=1}^{\rho} \alpha_{0,f} I_f\big(\frac{\alpha_{1,f}}{\alpha_{0,f}}\big),$$

where

$$D(\boldsymbol{\alpha}_0||\boldsymbol{p}) = \sum_f \alpha_{0,f} \log \frac{\alpha_{0,f}}{p_f}$$

is the informational divergence between $\boldsymbol{\alpha}_0$ and $\boldsymbol{p}$ (see [CT06]).

The proof follows soon after. We now proceed with the main result, which averages the outcome in Lemma 6, over all possible authentication groups.

Theorem 2 *In SBS-based pruning, the size of the remaining set $|\mathcal{S}|$, satisfies the following:*

$$J(\tau) \ := \ -\lim_{N\to\infty} \frac{\log}{n/\rho} P(|\mathcal{S}| \ \approx \ \tau\frac{n}{\rho}) \ = \ \inf_{\boldsymbol{\alpha}\in\mathcal{V}} \rho \sum_{f=1}^{\rho} \alpha_{0,f} \log \frac{\alpha_{0,f}}{p_f} + \sum_{f=1}^{\rho} \alpha_{0,f} I_f\big(\frac{\alpha_{1,f}}{\alpha_{0,f}}\big). \tag{4.20}$$

Furthermore we have the following.

Theorem 3 *The probability that after pruning, the search space is bigger (resp. smaller) than $\tau\frac{n}{\rho}$, is given for $\tau \geq \tau_0$ by*

$$-\lim_{N\to\infty} \frac{\log}{n/\rho} P(|\mathcal{S}| > \tau\frac{n}{\rho}) = J(\tau) \tag{4.21}$$

and for $\tau < \tau_0$

$$-\lim_{N\to\infty} \frac{\log}{n/\rho} P(|\mathcal{S}| < \tau\frac{n}{\rho}) = J(\tau). \tag{4.22}$$

The above describe how often we encounter authentication groups $\boldsymbol{v}$ and feature estimation behavior that jointly cause the gain to deviate, by a specific degree, from the common behavior described in (4.11), i.e., how often the pruning is atypically ineffective or atypically effective. We offer the intuition that the atypical behavior of the pruning gain is dominated by a small set of authentication groups, that minimize the expression in Theorem 2. Such minimization was presented in Fig. 4.3, and in examples that will follow after the proofs.

Please see the Annex B for the proofs.

The following examples are meant to provide insight on the statistical behavior of pruning.

Example 8 (How often will the gain be very small?) *We recall the discussion in section 4.2 where a pruning (surveillance) system can identify $\rho = 3$ categories, operates with reliability $\epsilon_1 = 0.8$ over a population with statistics $p_1 = 0.4, p_2 = 0.25, p_3 = 0.35$ and has confusability parameters $\epsilon_2 = 0.2, \epsilon_3 = 0.3$. In the context of the above theorems we note that the result already shown in Figure 4.3 applies by substituting τ with $\tau/\rho = \tau/3$. Consequently from Figure 4.3 we recall the following. The size of the (after pruning) set $\mathcal{S}$ is typically 47.5% of the original size n. The probability that pruning removes less than $1 - 0.72 = 28\%$ of the original set, is approximately given by $e^{-\rho n} = e^{-3n}$ because, as Figure 4.3 shows, $J(0.72) \approx 1$ (recall $\rho = 3$). Similarly the same Figure tells us that the probability that pruning removes less than $1 - 0.62 = 38\%$ of the original set, is approximately given by $e^{-\rho n/2} = e^{-3n/2}$ because $J(0.62) \approx 1/2$.*

In the following example we are interested in understanding the behavior of the search pruning in the case of rare authentication groups.

Example 9 (Which groups cause specific problems?) *Consider the case where a (soft biometrics based) pruning system has $\rho = 2$ identifiable categories, population probabilities $\boldsymbol{p} = [p,\ 1-p]$, and confusion probabilities $\boldsymbol{\epsilon} = [1-\epsilon,\ \epsilon]$ (this means that the probability that the first category is confused for the second, is equal to making the reverse error). We want to understand what types of authentication groups will cause our pruning system to prune out only, for example, a fourth of the population ($|\mathcal{S}| \approx 3n/4$). The answer will turn out to be that the typical groups that cause such reduced pruning, have 43% of the subjects in the first category, and the rest in the other category.*

To see this we recall that (see Theorem 2) $|\mathcal{S}| \approx \tau\frac{n}{2}|\mathcal{S}| = 3n/4$ which implies that $\tau = 3/2$. For α denoting the fraction of the subjects (in $\boldsymbol{v}$) that belong in the first category, and after some algebra that we ignore here, it can be shown that $\alpha = \frac{3-\tau}{5-\tau}$ which yields $\alpha = 3/7 \approx 43\%$.

A further clarifying example focuses on the case of a statistically symmetric, i.e., maximally diverse population.

Example 10 (Male or female?) *Consider a city with 50% male and 50% female population ($\rho = 2, p_1 = p_2 = 0.5$). Let the confusion probabilities as before to be equal, in the sense that ($\boldsymbol{\epsilon} = [1-\epsilon,\ \epsilon]$). We are interested in the following questions. For a system that searches for a male (first category), how often will the system prune out only a third of the population (as opposed to the expected one half)? How often will we run across an authentication group with $\alpha \triangleq a_{0,1} = 20\%$ males, and then have the system prune out only 45% of the overall size (as opposed to the expected 80%)? As it turns out, the first answer reveals a probability of about $e^{-n/4}$, and the second answer reveals a probability of about $e^{-n/5}$. For $n \approx 50$, the two probabilities are about five in a million and forty-five in a million respectively.*

To see this, first note that $\alpha_{1,1} = \alpha\tau$. Then we have that

$$
\begin{aligned}
I(\alpha, \alpha_{1,1}, \tau) := \lim_{N\to\infty} \frac{\log}{n/2} P(|\mathcal{S}| = \frac{n}{2}\tau, \alpha, \alpha_{1,1}),& \\
\inf_{\delta} I(\alpha, \alpha_{1,1}, \tau) = I(\alpha, \tau, \alpha_{1,1} = \alpha\tau) =& \\
2\alpha \log 2\alpha + 2(1-\alpha)\log 2(1-\alpha) + \tau\log\tau + (2-\tau)\log(2-\tau).&
\end{aligned}
$$

To see the above, just calculate the derivative of I with respect to $\alpha_{1,1}$. For the behavior of τ we see that $I(\tau) = \inf_\alpha \inf_{\alpha_{1,1}} I(\alpha, \alpha_{1,1}, \tau) = I(\alpha = p_1, \alpha_{1,1} = p_1\tau, \tau)$ $= \tau\log\tau + (2-\tau)\log(2-\tau)$, which can be seen by calculating the derivative of $\inf_\delta I(\alpha, \alpha_{1,1}, \tau)$ with respect to α.

4.5.1 Typical behavior: average gain and goodput

We here provide expressions for the average pruning gain as well as the goodput which jointly considers gain and reliability. This is followed by several clarifying examples.

In terms of the average pruning gain, it is straightforward that this takes the form $\mathcal{G} := \mathbb{E}_{\boldsymbol{v},\boldsymbol{w}}\mathcal{G}(\boldsymbol{v}) = \left(\sum_{f=1}^{\rho} p_f \epsilon_f\right)^{-1}$, and similarly the average relative gain takes the form $\mathbb{E}_{\boldsymbol{v},\boldsymbol{w}} r(\boldsymbol{v}) = \sum_{f=1}^{\rho} p_f(1-\epsilon_f)$. We recall that reliability is given by ϵ_1.

In combining the above average gain measure with reliability, we consider the (average) *goodput*, denoted as $\mathcal{U}$, which for the sake of simplicity is here offered a concise form of a weighted product between reliability and gain,

$$\mathcal{U} := \epsilon_1^{\gamma_1}\mathcal{G}^{\gamma_2} \tag{4.23}$$

for some chosen positive γ_1, γ_2 that describe the importance paid to reliability and to pruning gain respectively.

We proceed with clarifying examples.

Example 11 (Average gain with error uniformity) *In the* uniform error setting *where the probability of erroneous categorization of subjects is assumed to be equal to ϵ for all categories , i.e., where $\epsilon_f = \epsilon = \frac{1-\epsilon_1}{\rho-1}$, $\forall f = 2, \cdots, \rho$, it is the case that*

$$\mathcal{G} = \left(p_1 + \epsilon - p_1\epsilon\rho\right)^{-1}. \tag{4.24}$$

This was already illustrated in Fig. 4.2. We quickly note that for $p_1 = 1/\rho$, the gain remains equal to $1/p_1$ irrespective of ϵ and irrespective of the rest of the population statistics $p_f,\ f \geq 2$.

Example 12 (Average gain with uniform error scaling) *Now consider the case where the uniform error increases with ρ as $\epsilon = \frac{\max(\rho-\beta,0)}{\rho}\lambda$, $\beta \geq 1$. Then for any set of population statistics, it is the case that*

$$\mathcal{G}(\lambda) = \left(p_1[1+(\rho-\beta)\lambda] + \frac{\rho-\beta}{\rho}\lambda\right)^{-1}, \tag{4.25}$$

which approaches $\mathcal{G}(\lambda) = (p_1[1+(\rho-\beta)\lambda]+\lambda)^{-1}$ as ρ increases. We briefly note that, as expected, in the regime of very high reliability ($\lambda \to 0$), and irrespective of $\{p_f\}_{f=2}^{\rho}$, the pruning gain approaches $\frac{1}{p_1}$. In the other extreme of low reliability ($\lambda \to 1$), the gain approaches ϵ_1^{-1}.

We proceed with an example on the average goodput.

Example 13 (Average goodput with error uniformity) *Under error uniformity where erroneous categorization happens with probability ϵ, and for $\gamma_1 = \gamma_2 = 1$, the goodput takes the form*

$$\mathcal{U}(\epsilon) = \frac{\epsilon + (1-\epsilon\rho)}{\epsilon + p_1(1-\epsilon\rho)}. \tag{4.26}$$

To offer insight we note that the goodput starts at a maximum of $\mathcal{U} = \frac{1}{p_1}$ for a near zero value of ϵ, and then decreases with a slope of

$$\frac{\delta\mathcal{U}}{\delta\epsilon} = \frac{p_1 - 1}{[\epsilon + p_1(1-\rho\epsilon)]^2}, \tag{4.27}$$

which as expected[4] *is negative for all $p_1 < 1$. We here see that $\frac{\delta}{\delta p_1}\delta\frac{\mathcal{U}}{\delta\epsilon}|_{\epsilon\to 0} \to \frac{2-p_1}{p_1^3}$ which is positive and decreasing in p_1. Within the context of the example, the intuition that we can draw is that, for the same increase in ϵ*[5]*, a search for a rare looking subject (p_1 small) can be much more sensitive, in terms of goodput, to outside perturbations (fluctuations in ϵ) than searches for more common looking individuals (p_1 large).*

[4] We note that asking for $|\mathcal{S}| \geq 1$, implies that $\epsilon + p_1(1-\rho\epsilon) > \frac{1}{n}$ (see (4.24)) which guarantees that $\frac{\delta\mathcal{U}}{\delta\epsilon}$ is finite.

[5] An example of such a deterioration that causes an increase in ϵ, can be a reduction in the luminosity around the subjects.

We now proceed with the application of the derived measures on real life SBSs and furthermore quantify these SBSs.

4.6 Practical application of above analysis on proposed SBSs

We adopt the results of 3 real soft biometric trait categorization algorithms from chapter 7 by specifically taking over the related confusion matrices. These algorithms are error prone systems for

- categorization of 4 eye colors: based on Gaussian Mixture Models with expectation maximization classification of hue and saturation values in the iris,
- moustache detection: based on skin color and hair color comparison in the region below the nose,
- glasses detection: based on edge and line detection between the eyes.

Using the three traits, we construct nine different SBSs which we list below in Table 4.1. All related confusion matrices and related population statistics are listed in Appendix B. The large population in which we employ those SBSs is based on the statistics of the FERET database [Fer11]. For this purpose we annotated the 646 subjects in the FERET database in terms of glasses, moustache and eye color.

SBS	Description	ρ
'2e'	Categorization of 2 eye colors	2
'm'	Moustache detection	2
'g'	Glasses detection	2
'4e'	Categorization of 4 eye colors	4
'mg'	Moustache and Glasses detection	4
'2em'	2 eye color categories and moustache detection	4
'2eg'	2 eye color categories and glasses detection	4
'2emg'	2 eye color classes, moustache and glasses detection	8
'4emg'	4 eye color classes, moustache and glasses detection	16

Table 4.1: SBSs labeling and description of the ρ associated categories

In the following we analyze the pruning gain related to the presented systems.

4.6.1 Simulations: the instantaneous pruning gain

A pertinent characteristic of an SBS is the amount by which the initial database is reduced. As a measure of this we adopt the pruning gain from above to be:

$$r(\boldsymbol{v}) := 1 - \frac{|\mathcal{S}|}{n} \in [0, 1], \tag{4.28}$$

describing the fraction of subjects from $\boldsymbol{v}$ which was pruned out. This ranges from 0 (no pruning gain) to 1.

We proceed to illustrate the variability of $r(\boldsymbol{v})$, as a function of $\boldsymbol{v}$ but also of v'. To understand this variability we can note that if v' belongs in a rare category (e.g. green eyes), then we generally expect a higher gain, than if v' belonged in a more common category (e.g. black eyes). Similarly, if $\boldsymbol{v}$ happens to comprise of people who look similar to v' then the gain will be smaller than the case where another $\boldsymbol{v}$ comprised of people who looked sufficiently different from v'.

To illustrate these relationships we proceed with some clarifying simulations involving the presented SBSs. In these simulations we randomly pick 100 realizations of $\boldsymbol{v}$, each consisting of

$n = 50$ subjects, out of which we randomly pick v', and we perform error-prone pruning according to the confusion matrices presented in the Appendix.

In Figure 4.6 we demonstrate this pruning gain in an experiment involving the SBS labeled as ‘4e’ (see Table 4.1), which employs $\rho = 4$ categories, and which acts on the population that shares the distribution of the FERET database.

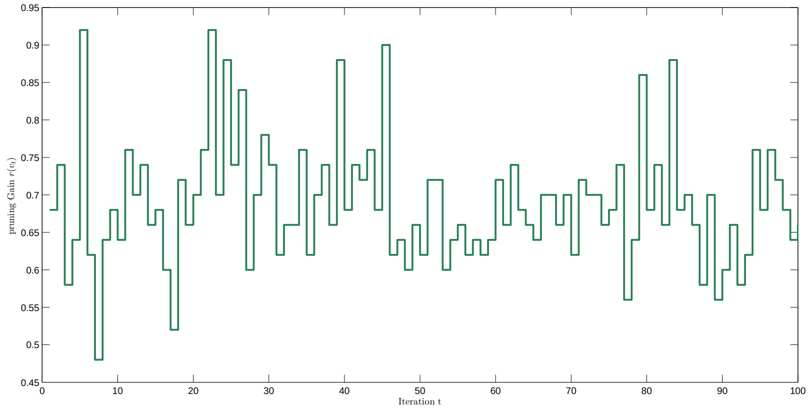


Figure 4.6: SBS ‘4e’, n=50, target subject v' belongs to a random category C'.

For further understanding we proceed with some variations of this simulation.

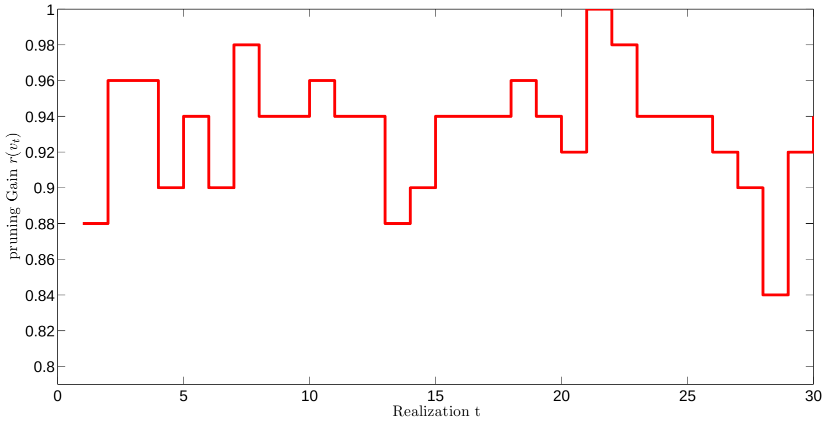


Figure 4.7: SBS ‘mg’, n=50, target subject v' belongs to category C'=“moustache - glasses".

One point we observed is that generally, even in the presence of the confusion matrices, having more categories generally (but not always) translates to a higher gain.

We note that for $\rho = 16$, $n = 20$, and a target person v' who is blue eyed and has no glasses and no moustache, Figure 4.10 reports pruning gain values up to 0.95. This corresponds to the pruning out of 95% of $\boldsymbol{v}$, which in this specific case is equivalent to saying that the SBS has entirely identified v'.

Furthermore we note that, as expected, an increasing authentication group size n generally introduces a smaller variability in the gain; see for example Figure 4.9 with $n = 200$.

The Figures 4.7 and 4.8 allow for a gain comparison, where specifically the first considers the case where v' belongs in the rare category of people having glasses and moustache, and in

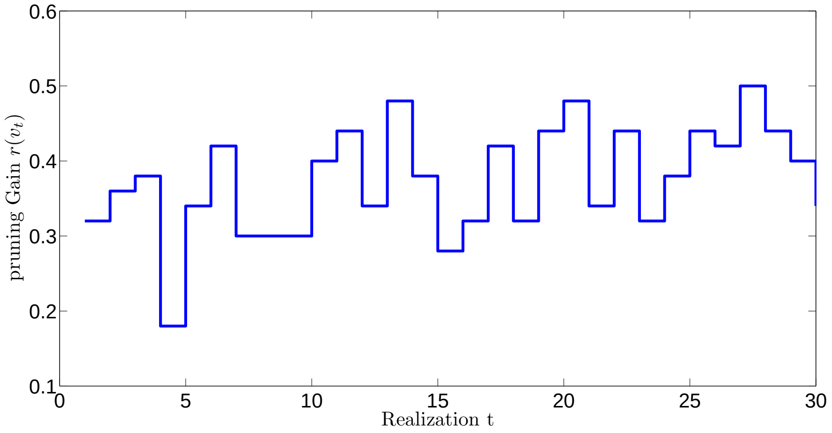


Figure 4.8: SBS ‘mg’, n=50, target subject v' belongs to category C'=“no moustache - no glasses".

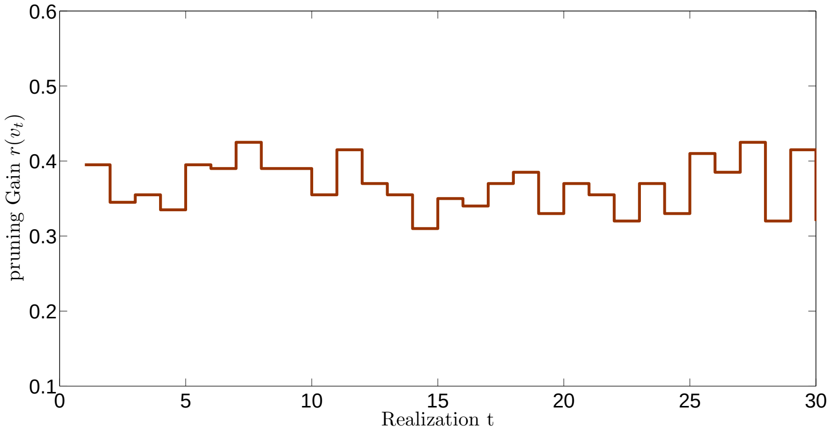


Figure 4.9: SBS ‘mg’, n=200, target subject v' belongs to category C'=“no moustache - no glasses”.

the other more common case where v' belongs in the category of people without glasses and moustache. The operating ranges of the pruning gain reflects directly the distinctiveness of the target subject, given in both cases the same population and system characteristics.

4.7 Average case analysis of gain, reliability and computational-cost effects of pruning

4.7.1 Average pruning gain and related error

We proceed with presenting a concise description of the average gain of an SBS, where the gain $r(\boldsymbol{v})$ is averaged over all possible authentication groups $\boldsymbol{v}$, and over the randomness of the categorization errors $\boldsymbol{w}$, as elaborated in 4.3.

The following describes the average gain and reliability of an SBS.

Proposition 1 *An SBS system endowed with a categorization confusion matrix* $\mathbf{E}$ *and error vector* $\mathbf{e}$*, and operating over a general population with statistics given by* $\mathbf{p}$*, allows for a probability of*

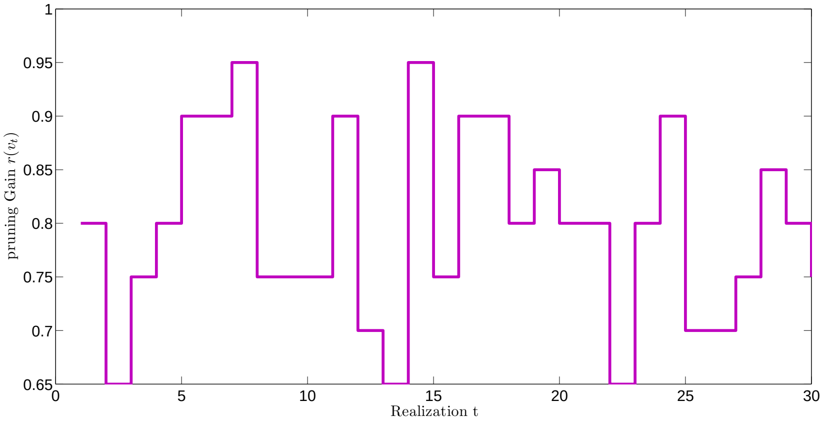


Figure 4.10: SBS ‘emg’, n=20, target subject v' belongs to category C'=“Blue eyes - no moustache - no glasses".

error given by

$$P_{err} = \boldsymbol{p}^T \mathbf{e}$$

and an average gain of

$$r := \mathbb{E}_{\boldsymbol{v},\boldsymbol{w}} r(\boldsymbol{v}) = 1 - \boldsymbol{p}^T \mathbf{E} \boldsymbol{p}. \tag{4.29}$$

A relevant question is whether it is better, in terms of increasing the average gain, to invest in soft biometric traits like tattoos, scars and birth marks, which are rare, but distinctive, or if it is of more value to invest in facial measures and facial colors, in which subjects are distributed more uniformly. The above proposition addresses this question and can show that investing towards a uniform category distribution for a given population is most valuable in terms of gain.

We illustrate the average gain and pruning error for the proposed SBSs in Figure 4.11 and provide the exact values in Table 2.

At this point we can establish also the measure of *goodput*, which was introduced in 4.5.1 as a measure that jointly considers both the gain and the reliability capabilities of an SBS.

Average goodput of search pruning The measure of goodput, combines as introduced in 4.5.1 the pruning gain with reliability. For the sake of simplicity the measure, denoted here as $\mathcal{U}$, takes the form of a weighted product between reliability and gain

$$\mathcal{U} := (1 - P_{\text{err}})^{\gamma_1} r^{\gamma_2} \tag{4.30}$$

for some chosen positive γ_1, γ_2 that respectively describe the importance paid to reliability and to pruning gain. We note the change of the expression from section 4.5.1, which forms though are both equivalent. We proceed to evaluate and rank the given SBSs in terms of the introduced characteristics gain, error and goodput and set hereby the tuning variables $\gamma_1 = \gamma_2 = 1$.

Table 4.7.1 provides the results on the proposed nine SBSs. We observe that the highest goodput is attributed to system ‘4e’ endowed with 4 eye color categories. The enhanced systems ‘2emg’ and ‘4emg’ introduce a gain increase, but at the cost of an increased error probability. On the other hand the systems ‘2e’, ‘m’, ‘g’, and ‘2eg’ introduce lower error probabilities but at a cost of low average pruning gain. The intertwined relationship between error, gain and goodput is illustrated in Figure 4.11. Given the measure of goodput we can compare SBSs, by prioritizing

SBS	P_{err}	r	$\mathcal{U}$
‘2e’	0.0750	0.4743	0.4388
‘m’	0.1420	0.2538	0.2177
‘g’	0.0690	0.3275	0.3049
‘4e’	0.1522	0.7039	0.5968
‘mg’	0.2012	0.4982	0.3979
‘2em’	0.2063	0.6077	0.4823
‘2eg’	0.1388	0.6465	0.5568
‘2emg’	0.2611	0.7362	0.5440
‘4emg’	0.3227	0.8514	0.5766

Table 4.2: Pruning error, gain and goodput of the proposed SBSs

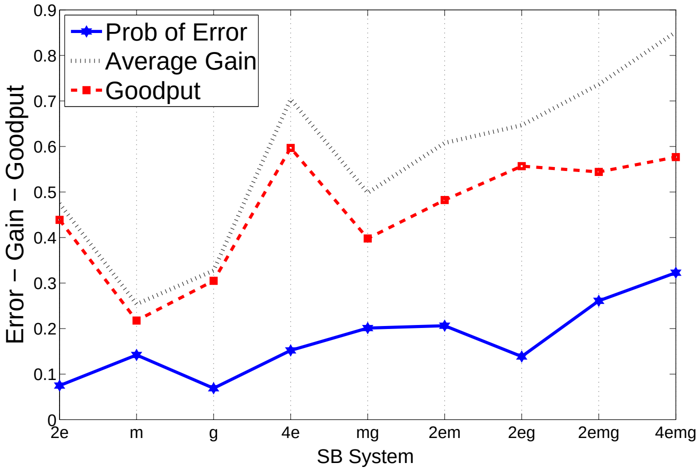


Figure 4.11: Pruning error, gain and goodput of the proposed SBSs.

the gain or the error contribution depending on the application scenario. If we are interested in pruning aggressively a big database, we would emphasize on the gain and choose a system like ‘4emg’, at the cost though of an increased FRR. If we want to make a conservative search prune with a low risk of pruning out the target subject, then system ‘2eg’ is more suitable to be our system of choice. The question arises here of how to bound the case when an enhancement of a system is profitable in terms of goodput. With other words is an addition of a soft biometric trait a positive contribution to the SBS or does the related average error outweighs the related pruning gain? This is addressed in Equation (4.30) and can be answered depending on a given application, population statistics and error probability.

4.7.2 Computational cost reduction

We briefly discuss the computational savings that result by preceding a computationally expensive algorithm (e.g. full person recognition) with pruning based on generally simpler categorization algorithms. In terms of analysis, let T be the total number of soft biometric traits (e.g. system ‘4emg’ corresponds to $T = 3$ traits: $t = 1$ for eyes, $t = 2$ for moustache, $t = 3$ for glasses). Furthermore let N_t be the average computational complexity required to implement categorization, for one person, for trait t, $t = 1, 2, \cdots, T$. At the same time Let N_{fr} be the average complexity associated with the face recognition algorithm, per person. Given a naive way of pruning, we conclude that the overall computational complexity N is given as

$$N = n \sum_{i=1}^{T} N_t + |S| N_{fr}, \tag{4.31}$$

which reflects the fact that the categorization algorithms for the T traits would be employed on n people, and the computationally expensive face recognition algorithm would be employed only on $|\mathcal{S}|$ people. We relate this computational complexity N to the computational complexity required if only the face recognition system was employed on the entire $\boldsymbol{v}$. Towards this we have the following.

Proposition 2 *Pruning results in a computational cost reduction to a fraction of* nN_{fr} *that is equal to*

$$\frac{\mathbb{E}N}{nN_{fr}} = \frac{\sum_{i=1}^{T} N_t}{N_{fr}} + \mathbf{p}^{\mathbf{T}}\mathbf{E}\mathbf{p}. \tag{4.32}$$

Example: We proceed with the evaluation of our presented SBSs with three different estimated relationships between the computational complexity of the face recognition system N_{fr}, and that of implementing a trait, i.e., to N_t.

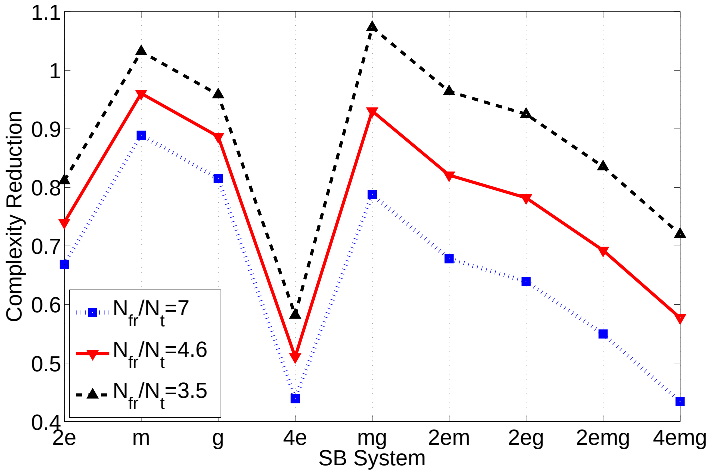


Figure 4.12: Complexity cost reduction provided by the presented SBSs. In the above, the ratio N_{fr}/N_t varies.

Figure 4.12 reflects on the computational cost reduction of the proposed SBSs. In this context the best systems are ‘4e’ and ‘4emg’, which reduce the computational cost by more than half. The system ‘4e’ specifically has very low complexity, since the computation involves just one trait, but where the four categories are sufficient to increase the gain, and hence reduce the number of people on which the face recognition system is applied. Similarly system ‘4emg’ achieves good complexity, mainly because of its high pruning gain, which substantially reduces the cost of applying the face recognition system. Also interesting to see is the cost increase for systems ‘m’ and ‘mg’. This means that those systems reduce the initial database insufficiently and cannot justify their computational costs.

4.8 Summary

The current chapter provided statistical analysis of the gain and reliability in pruning the search over large data sets, where these sets are random and where there is a possibility that the pruning may entail errors. In this setting, pruning plays the role of pre-filtering, similar to techniques such as video indexing. The analysis may offer insight on better designing pre-filtering algorithms for different search settings. We further studied nine different, actual, soft biometric systems, as well as analyzed and experimented with factors like average error, pruning gain and goodput. Using these factors, we provided a quantifiable comparison of these systems. Furthermore we identified relations between SBS enhancement, error probabilityP_{err}, pruning gain r and goodput $\mathcal{U}$. These findings bring to the fore some SBS design aspects. Finally we gave insight on the computational cost reduction related to person recognition systems with a pruning mechanism. This insight revealed some of the benefits of applying SBS for pre filtering.

We here studied and analyzed the pertinent characteristics related to search pruning performed by SBSs. In the next chapter we examine a third scenario (after human identification and pruning the search), namely human re-identification. In such a scenario in what follows, we explore the capability and limitations of existing SB algorithms. We hereby introduce an additional challenge of frontal-to-side pose variation.

Chapter 5

Frontal-to-side person re–identification

Typically biometric face-recognition algorithms are developed, trained, tested and improved under the simplifying assumption of frontal-to-frontal person recognition. Such algorithms though are challenged when facing scenarios that deviate from the training setting, such as for example in the presence of non-constant viewpoints, including the frontal-to-side scenario. Most person recognition algorithms, whether holistic or based on facial features, only manage to optimally handle pose differences that are less than about 15 degrees. As a result, a variation in the pose is often a more dominant factor than a variation of subjects. This aspect of pose variation comes to the fore in video surveillance, where a suspect may be pictured firstly frontal, whereas the corresponding test images could be captured from the side, thus introducing a *frontal-to-side recognition problem*.

Towards handling this problem, we draw as already in the chapters 3 and 4 from the way humans perform frontal-to-side recognition, that is by using simple and evident traits like hair, skin and clothes color. One of our tasks here is to get some insight into the significance of these traits, specifically the significance of using hair, skin and clothes patches for frontal-to-side re-identification. We mention that we work on the color FERET dataset [Fer11] with frontal gallery images for training, and side (profile) probe images for testing. Towards achieving re-identification, the proposed algorithm first analyzes the color and texture of the three patches, as well as their intensity correlations. This analysis is then followed by the construction of a single, stronger classifier that combines the above measures, to re-identify the person from his or her profile.

5.1 Related work

Pose invariant face recognition has been addressed in different approaches which, as described in [PEWF08], can be classified in following three categories:

– mapping methods: construction of a 3D model based on more than one 2D images (see [IHS05])
– geometric methods: construction of a 3D model based on a single 2D image (see [SVRN07])
– statistical methods: statistical learning methods that relate frontal to non-frontal poses (see [PEWF08].

An overview of these frontal-to-side face recognition methods was given in [ZG09], which work also addressed some of the methods' limitations in handling different pose variations. Such methods can be originally found in [WMR01] and [WAS$^+$05], which recorded a true recognition rate of 50-60% over an authentication group of 100 subjects. Better results on pose-variant face recognition were recorded in [PEWF08] which employed statistical methods to achieve reliability of 92% over an authentication group with the same size.

We here take a rather different and more direct approach in the sense that the proposed method does not require mapping or a-priori learning. In applying this approach, we provide a preliminary study on the role of a specific set of simple features in frontal-to-side face recognition. These selected features are based on soft biometrics, as such features are highly applicable in video surveillance scenarios. To clarify, we refer to *re-identification* as the correctly relating of data to an already explicitly identified subject (see [MSN03]). Our used traits of hair, skin and clothes color and texture also belong in this category of soft biometric traits, and are thus of special interest for video surveillance applications.

5.2 System model, simulations and analysis

In what follows we empirically study the error probability of a system for extraction and classification of properties related to the above described color soft biometric patches. We first define the operational setting, and also clarify what we consider to be system reliability, addressing different reliability related factors, whose impact we examine. The clarifying experiments are then followed by a preliminary analytical exposition of the error probability of a combined classifier consisting of the above described traits.

5.2.1 Operational scenario

The setting of interest corresponds to the re-identification of a randomly chosen subject (*target subject*) out of a random authentication group, where subjects in gallery and probe images have approximately 90 degrees pose difference.

Patches retrieval: We retrieve automatically patches corresponding to the regions of hair, skin and clothes, see Figure 5.1, based on the coordinates of face and eyes. The size of each patch is determined empirically by one half and one third distance of center-to-center eye distance. The frontal placement of the patches is following, horizontally / vertically respectively:

- hair patch: from nose towards left side / top of the head,
- skin patch: centered around left eye / centered between mouth and eyes,
- clothes patch: from left side of the head towards left / mouth-(mouth-top of head distance).

The horizontal side face placements of the patches are: nose-(nose-chin distance), eye towards left, head length-chin. Those coordinates were provided along with the images from the FERET database, but are also easily obtainable by state of the art face detectors, like the OpenCV implementation of the Viola and Jones algorithm [VJ01a].

Following the extraction of the patches we retrieve a set of feature vectors including the color and texture information. In this operational setting of interest, the *reliability* of our system captures the probability of false identification of a randomly chosen person out of a random set of n subjects. This reliability relates to the following parameters:

- Factor 1. The number of categories that the system can identify, as in the previous chapters 3 and 4 we refer to as ρ.
- Factor 2. The degree with which these features / categories represent the chosen set of subjects over which identification will take place.
- Factor 3. The robustness with which these categories can be detected.

Finally reliability is related (empirically and analytically) to n, where a higher n corresponds to identifying a person among an increasingly large set of possibly similar-looking people.

We will proceed with the discussion and illustration of the above three named pertinent factors.

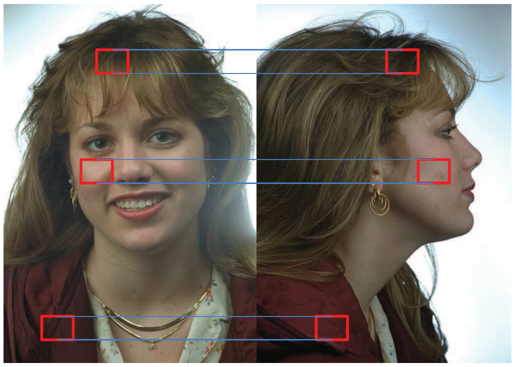


Figure 5.1: Frontal / gallery and profile / probe image of a subject. Corresponding ROIs for hair, skin and clothes color.

5.2.2 Factor 1. Number of categories that the system can identify

Our system employs soft biometric traits, where each trait is subdivided into *trait-instances*, that directly have an impact on the number of overall categories ρ. We here note that in the scientific work [SGN08] the terminology trait and semantic term was used instead. For this specific scenario, our system is endowed with the traits hair, skin and shirt color as well as texture and intensity correlation. It is to be noted that, as already shown in the chapters 3 and 4 the more categories a system can classify subjects into, the more reliable the system is in terms of distinctiveness.

We proceed with the description of population of categories.

5.2.3 Factor 2. Category representation of the chosen set of subjects

The distribution of subjects over the set of ρ categories naturally has an impact on the specific importance of the different traits and thus also affects the behavior of the system for a given population. Table 5.1 and Table 5.2 give example distributions, with respect to skin and hair color traits, based on 265 subjects from the color FERET database. For clarifying experiments we attributed this subset into three subcategories of skin color and eight subcategories of hair color. It is evident that the trait hair color has a higher distinctiveness and thus greater importance than skin color, since it has more instances in which the subjects are subdivided into. This observation is illustrated in Figure 5.2 and we follow with the explanation.

Categories	1	2	3
Skin Color	62.64%	28.66%	8.7%

Table 5.1: Distribution of FERET subjects in the three skin color categories.

Categories	1	2	3	4	5	6	7	8
Hair Color	4.2%	26.8%	46.4%	3%	3%	3.4%	1.5%	11.7%

Table 5.2: Distribution of FERET subjects in eight hair color categories.

In the case of (re-)identification the only way to unambiguously recognize a person is an exclusive category membership, with other words a subject can only be (re-)identified if he or she is the only one assigned to a category. Hereby it is of interest, that the target subject does not collide with any other subject inside the authentication group in terms of category, see 3.4.2.

In the example, of subjects-categories distribution probabilities illustrated in Table 5.1 and Table 5.2 the probabilities for collision for skin and hair color as functions of n are portrayed in Figure 5.2. For this experiment we randomly pick n subjects out of the color FERET subset, pick randomly one of the n subjects as target subject and examine if the target subject collides with somebody else from the given authentication group. We repeatedly perform this experiment and compute finally an averaged collision error probability as a function of n.

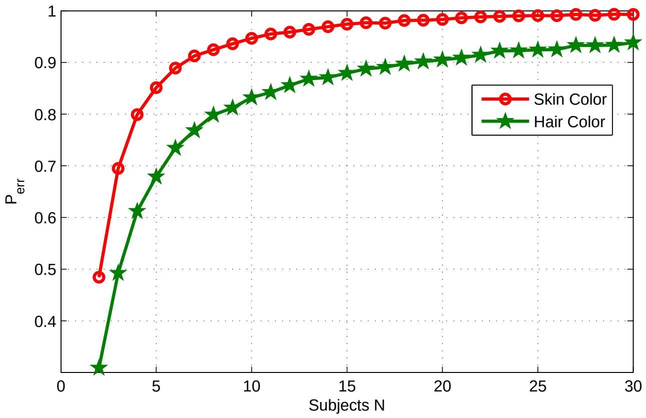


Figure 5.2: Probability of collision in the FERET database for skin and hair color.

This collision error can be decreased by considering more distinctive or multiple traits. We note here that the collision error probability is the minimum error bound a system can achieve given the specific traits in a certain population.

In what follows we will include automatic category classification and analyze the influence of estimation errors caused by algorithm limitations.

5.2.4 Factor 3. Robustness of categories estimation

In this section we take into consideration the over all error probability, containing inevitably of the above discussed collision error- and furthermore of the algorithmic categorization error-probabilities. With other words we examine, how often the target subject is wrongly re-identified, regardless of the underlying source, which can be both estimation- or collision error character. Thereby we again randomly pick an authentication group of n subjects and randomly declare one of the n subjects as the target subject for re–identification. Then we proceed to train our algorithms with the feature vectors extracted from the patches of the frontal gallery images. As feature vectors we here firstly consider the color information provided by the selected patches. For this purpose we work in the HSV (hue, saturation, value) color space and use the hue and saturation information as feature vectors. We train an AdaBoost classifier [FHT98] with these feature vectors and subsequently re-identify the randomly picked target subject by matching their feature vectors of the probe patches (retrieved from the profile image) with the trained ones (frontal). We repeat the procedure and average the error probability over all iterations for all values of n. We note here that we do not consider anymore the manual annotation of categories used for the experiment in

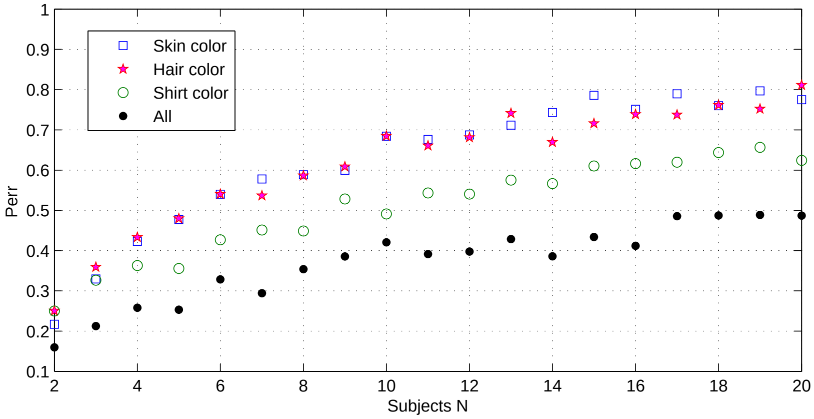


Figure 5.3: Boosting of soft biometric patches.

Section 5.2.3, instead we use the discrete HS distribution of colors. By doing so we do not have anymore human compliant categories, but a more refined and efficient classification.

5.2.4.1 Boosting of patch color

Boosting is referred to *additive logistic regression* and is the combination of weak classifiers, which classifiers alone should perform slightly better than random into one higher accuracy classifier. This combined classifier is a sum of the weighed weak classifiers, where the weight of each classifier is a function of the accuracy of the corresponding classifier. This concept is consistent with the concept of soft biometrics, where we have a multitude of weak biometric information that we want to combine to a stronger classification hypothesis. We selected a discrete AdaBoost.MH algorithm, see [SS98] for multi class classification, for its good performance and robustness against overfitting.

We train hue and saturation (HS) per gallery patch using AdaBoost and evaluate posterior probabilities of the HS vectors of the probe patch related to the target subject and each patch of the n trained subjects. An error occurs, if the HS vectors of gallery and probe patch of the target subject do not match. The results on error probability as a function of the authentication group size n for each patch color and the combined color patches are illustrated in 5.3. As expected, and similar to the collision experiment in Section 5.2.3, the re-identification error probability is increasing with an increasing authentication group. We further observe that clothing color has the strongest distinctiveness. The explanation therefore is that clothing color is distributed in a high range of colors and is thus easier to distinguish. Furthermore we surprisingly notice a much higher performance of skin or hair color than in the pure collision analysis, see Figure 5.2. For example, in an authentication group of 5 subjects, AdaBoost re-identification provides an error probability for skin color of about 0.48, where in the collision analysis we achieve only 0.85. This interesting error decrease is due to the limited human capability for distinguishing and classification of skin color in only three categories, which classification was only used in the collision analysis. In the current AdaBoost estimation setting the classification is hidden, non human compliant but of a higher efficiency.

As expected the combined classifier has a stronger classification accuracy than the single classifiers. The performance of the combined classifier is though enticed by several factors. Firstly the traits used are partly correlated, see 3.4.1. Furthermore the estimation errors of traits are correlated as well, $r_{Estimationerror(HairColor,SkinColor)} = 0.22$, which shows a tendency of jointly occurrence of classification errors, for both hair and skin color classification. On a different note we point out that each further trait and its classification error contributes negatively to the over all categorization error and thus the over all error probability is increasing with an increasing number of traits and categories ρ. On the other hand with each further trait the collision probability decreases.

We proceed with the description and inclusion of two further properties of the employed patches, namely texture and intensity difference.

5.2.4.2 Patch texture

We formalize a descriptor for texture $\overrightarrow{x}$ including following four characteristics and compute them on the graylevel images for each patch.

Contrast: measure of the intensity contrast between a pixel and its neighbor over the whole image. The contrast in an image is related to its variance and inertia and is:

$$x_1 = \sum_{i,j} |i-j|^2 p(i,j), \tag{5.1}$$

where i and j denote the gray scale intensities of two pixels, p refers to the gray level co-occurrence matrix, which describes the co-occurrence of gray scale intensities between two image areas. Each element (i, j) in a gray level co-occurrence matrix specifies the number of times that the pixel with value i occurred horizontally adjacent to a pixel with value j.

Correlation: measure for correlation of neighboring pixels and is denoted as:

$$x_2 = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)\vec{p}(i,j)}{\sigma_i \sigma_j}, \tag{5.2}$$

where μ_i and μ_j stand for the mean values of the two areas around i and j, σ_i and σ_j represent the related standard deviations.

Energy: sum of squared elements or angular second moment. Energy equal to one corresponds to a uniform color image.

$$x_3 = \sum_{i,j} p(i,j)^2 \tag{5.3}$$

Homogeneity: measure of the closeness of distribution of elements.

$$x_4 = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \tag{5.4}$$

5.2.4.3 Patch histogram distance

Along with the color information we integrate into our classifier a simple relation measure for the divergence between the intensity probability density functions (pdf) of patches concerning one subject. With other words we express the three relationships between intensities within a subject: hair–skin, skin–clothes and hair–clothes. Speaking in an example we expect to have a higher distance measure for a person with brown hair and light skin than for a person with blond hair and light skin. For the computation we convert the patches to gray level intensities and assess the

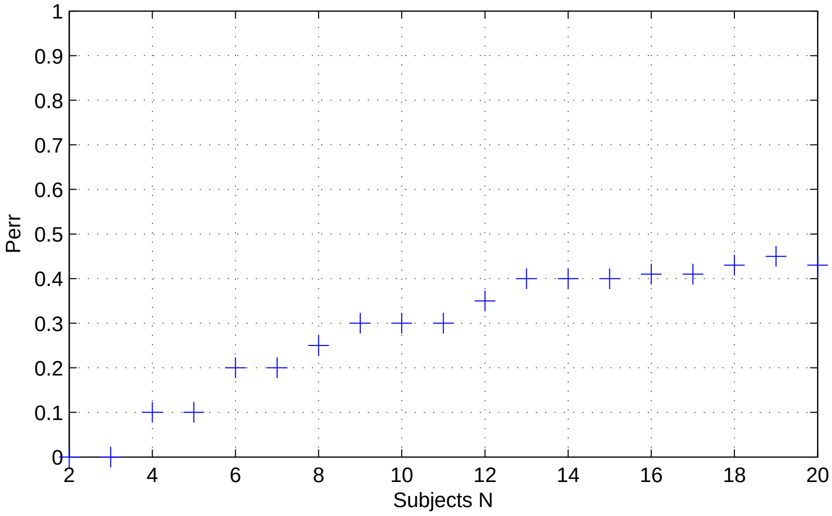


Figure 5.4: Overall-classifier obtained by boosting color, texture and intensity differences.

L1–distance three times per person for all relations between the patches. For two distributions r and s of discrete random character the measure is given as:

$$D = \|r - s\|_1 = \sum_{k=1}^{255} |r(k) - s(k)|, \tag{5.5}$$

where k represents a bin of the 255 intensity bins in a gray scale image.

5.2.5 Combined overall-classifier

The combined over–all–classifier, which boosts all described traits, color, texture and intensity differences performs with a decreased error probability and thus outperforms expectedly the color classifier shown in Figure 5.3. Still the achieved error probability of 0.1 in an authentication group of 4 subjects is not sufficient enough for a robust re-identification system. This limited enhanced performance is due to the strong illumination dependence of color and furthermore due to correlations between traits, e.g. hair color–skin color or skin color–skin texture, see 3.4.1. We here note that the FERET database is a database captured with controlled lighting conditions, so with a different testing database we expect the performance to decrease additionally. Towards increasing the performance the amount of sub-classifiers can be extended, whereas emphasis should be placed on classifiers not based on color information. The system in its current constellation can be used as a pruning system for more robust systems or as an additional system for multi-trait biometric systems.

5.3 Summary

Motivated by realistic surveillance scenarios, we addressed in this chapter the problem of frontal-to-side facial recognition, providing re–identification algorithms/classifiers that are specifically suited for this setting. Emphasis was placed on classifiers that belong in the class of soft biometric traits, specifically color–, texture– and intensity– based traits taken from patches of hair, skin and clothes. Towards providing insight, the work presented different identification experiments that adhere to the frontal–to–side setting, as well as presented a preliminary analytical study that seeks to impart intuition on the role of the above traits in improving algorithmic reliability. Our analysis described the overall error probability, both as a function of collisions and of erroneous categorizations for given sizes of authentication groups. In the presence of a moderate reliability of the patches-based method, the analysis suggests promising applications of this method in settings such as pruning of searches.

After the analysis of the three security related applications of human identification, pruning the search and human re–identification in the chapters 3, 4 and 5, in the following chapter deviate from security and introduce a commercial applications of female facial aesthetics. We note that in the employment of a SBSs, the system remains the same, solely the last analytic step changes when moving from security to entertainment.

Chapter 6

Soft biometrics for quantifying and predicting facial aesthetics

With millions of images appearing daily on Facebook, Picasa, Flickr, or on different social and dating sites, photographs are often seen as the carrier of the first and deciding impression of a person. At the same time though, human perception of facial aesthetics in images is a priori highly subjective. The nature of this perception has long been explored *separately* from psychological and photographical points of view, respectively focusing on the properties of the subject and of the image. The photographical point of view, corresponding to photo-quality assessment and enhancement, has recently attracted further attention, partly due to the vast amount of digital images that are now available, as well as due to the ease with which digital image manipulation can now be achieved.

6.1 Related work

The present work draws from former work in three areas, namely classical facial aesthetics, photo–quality and aesthetics and image processing based face recognition.

6.1.1 Facial aesthetics

There are substantial amounts of works, both from psychological and sociological points of view, studying human perception of facial attractiveness and beauty. Such perception is highly subjective and is influenced by sociological and cultural factors and furthermore by individual preferences. Although appreciation of beauty is subjective or in other words "beauty is in the eye of the beholder", there are some characteristics that scientists have identified to evoke superior pleasure when looking at. One such characteristic generally associated to beauty and perfection is the *golden ratio* $\varphi \approx 1.6180339887$. When this divine proportion appears in both nature or art, they are perceived harmonic and aesthetic, see [Doc05]. An attractive human face contains φ in several proportions, e.g face height / width and face height / location of eyes, see Figure 6.1.

A further main characteristic is *symmetry*, which was evolutionary beneficial in its direct analogy to health [McK06]. Another sign for health and fertility is *averageness* of facial characteristics, not to be confused with faces of average persons. In [LR90] the authors present a study showing that mathematically average faces are considered beautiful. This study though contradicts with other theorems stating that attractiveness implies distinctive facial features. In terms of such features in literature following specifications are associated with beauty: a narrow face,

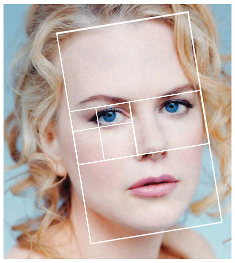


Figure 6.1: Golden ratio applied in a human face.

fuller lips, long and dark eyelashes, high cheek bones and a small nose [bea11]. An overview of psychological studies based on face proportions and symmetries is presented in [BL10].

Finally the *babyfaceness* or cute faces elicit sympathy and protective urges. Traits in babyfaces include a big round forehead, low located eyes and mouth, big round eyes and small chin, see Figure 6.2.

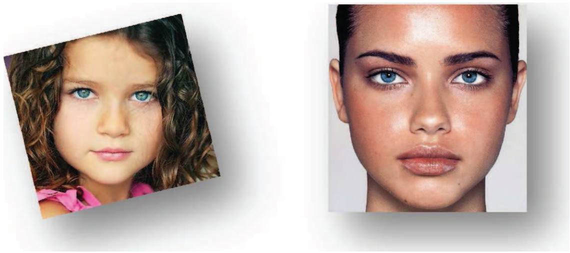


Figure 6.2: Babyfacesness theorem: females with child like traits are perceived sympathetic.

6.1.2 Beauty and image processing

From an image processing point of view, few attempts seek to exploit and validate some of the aforementioned psychological results and even introduce early methods for beauty prediction. In [GKYG10] for example, the authors present a multi-layer neuronal network for beauty learning and prediction regarding faces without landmarks. Such approaches often accept interesting applications, as the automatic forecast of beauty after a plastic surgery in [GPJ04]. The same work deals with beauty classification, considering facial measurements and ratios, such as ratios of distances from pupil to chin and from nose to lips (see also the work in [AHME01] and [MJD09]).

6.1.3 Photo–quality and aesthetics

Broad background work on image quality assessment (IQA) is needed in applications such as image transmission, lossy compression, restoration and enhancement. The subjective criteria intertwined with image quality are assessed in numerous metrics for mobile phones, electronic tabs or cameras. A number of automatic IQA algorithms has been built on those metrics. For an overview of related works, see [WBSS04] and [SSB06].

From a photographic point of view, the presence of people, their facial expressions, image sharpness, contrast, colorfulness and composition are factors which play a pivotal role in subjective perception and accordingly evaluation of an image, see [SEL00]. Recent works on photography considerations include [BSS10] and [CBNT10]. Hereby the authors reveal that appealing photographs draw from single appealing image regions as well as their location and the authors use this proposition to automatically enhance photo-quality. Photo-quality can be also influenced by image composition, see [OSHO10]. Finally there are current studies, which model aesthetic perception of videos [MOO10]. Such methods have become increasingly relevant due to the prevalence of low price consumer electronic products.

6.2 Contribution

In this chapter we study the role of objective measures in modeling the way humans perceive facial images. In establishing the results, we incorporate a new broad spectrum of known aesthetical facial characteristics, as well as consider the role of basic image properties and photograph aesthetics. This allows us to draw different conclusions on the intertwined roles of facial features in defining the aesthetics in female head-and-shoulder-images, as well as allows for further insight on how aesthetics can be influenced by careful modifications.

Towards further quantifying such insights, we construct a basic linear metric that models the role of selected traits in affecting the way humans perceive such images. This model applies as a step towards an automatic and holistic prediction of facial aesthetics in images.

The study provides quantitative insight on how basic measures can be used to improve photographs for CVs or for different social and dating websites. This helps create an objective view on subjective efforts by experts / journalists when retouching images. We use the gained objective view to examine facial aesthetics in terms of aging, facial surgery and a comparison of average females relatively to selected females known for their beauty.

The novelty in here lies mainly in two aspects. The first one is that we expand the pool of facial features to include non permanent features such as make-up, presence of glasses, or hair-style. The second novelty comes from the fact that we seek to combine the results of both research areas, thus to jointly study and understand the role of facial features and of image processing states.

6.3 Study of aesthetics in facial photographs

In our study we consider 37 different characteristics that include facial proportions and traits, facial expressions, as well as image properties. All these characteristics are, manually or automatically extracted from a database of 325 facial images. The greater part of the database, 260 images, is used for training purposes and further 65 images are tested for the related validation. Each image is associated with human ratings for attractiveness, as explained in Section 6.3.1. The database forms the empirical base for the further study on how different features and properties relate to attractiveness.

We proceed with the details of the database and related characteristics.

6.3.1 Database

The database consists of 325 randomly downloaded head-and-shoulders images from the web site HOTorNOT [Hot11]. HOTorNOT has been previously used in image processing studies (see [GKYG10] [SBHJ10]), due to the sufficiently large library of images, and the related ratings and demographic information.

Each image depicts a young female subject (see for example Fig. 6.3 and Fig. 6.4.) and was rated by a multitude of users of the web site. The rating, on a scale of one to ten, corresponds to the notion of attractiveness. The relevance and robustness of the provided ratings was confirmed in an experiment in [LLA$^+$08], where subjects re-rated a collection of images. For increasing robustness, we consider only images that have received a sufficiently high number of ratings, specifically more than 70 ratings. We will henceforth refer to these ratings as the *Mean Opinion Score* (MOS). Among the chosen images of the database the mean MOS was 7.9, the standard deviation was 1.4, whereas the minimum value was 4.3 and the maximum value was 9.9.

The JPEG images in our database are of different resolutions and of different qualities.

We now proceed with the description of the two groups of considered features: the photograph-aesthetics (image properties), and the facial aesthetics. All characteristics, from both groups, are stated in Table 6.1.

6.3.2 Photograph aesthetics

The considered photograph aesthetic features are here specifically chosen to be simple and objective. Firstly we include characteristics such as image resolution, image format (portrait or landscape) and illumination. Furthermore, we consider the relative angle of the face in the photograph (this angle is denoted as α in Fig. 6.3). We also incorporate the zoom-factor, specifically how large the face appears in comparison to the image height. Finally we also connect three novel image quality traits with facial aesthetics, which in previous work have been associated to photograph-aesthetics: the *relative foreground position*, the *BIQI* and the *JPEG quality measure*.

Regarding the *relative foreground position*, we essentially compute if the distance of the foreground's center of mass, (left eye, right eye or nose tip, respectively, see [cul]) to one of the stress points (see [GPJ04]) is shorter than to the center of the image. For clarity Figure 6.3 illustrates the stress points of the image, where each of the four stress points is in a distance of $1/3^{rd}$ the image width and $1/3^{rd}$ the image height from the boundary of the image, an aspect derived from the "Rule of thirds". In case that the foreground's center of mass is equidistant to all stress points, which is the case in the image center, it has been shown that subjects lose their attention and interest.

The *BIQI measure* is based on the distorted image statistics and it employs support vector machines for classification (see [MB09b] and [MB09a]). It is a blind quality measure; specifically it is a no-reference assessment measure on image quality.

The *JPEG quality measure* on the other hand considers artifacts caused by JPEG compression, such as blockiness and blurriness, evaluating again a no-reference score per image [WSB02].

6.3.3 Facial characteristics

Literature related to facial beauty (see [bea11]) identifies pertinent traits including the size of the face, the location and size of facial features like eyes, nose, and mouth, brows, lashes and lids, facial proportions, as well as the condition of the skin. Such literature confirms the role of these facial features in affecting human perception of beauty (see [bea11] and [BL10]). Drawing from this previous work, we also consider ratios of facial features and/or their locations by relating a multitude of measures, specifically including known facial beauty ratios adhering to the golden ratio, e.g. x_{16} (see Table 6.1 for notations).

Moreover we proceed a step further and consider soft biometric characteristics, such as eye-, hair- and skin-color, face- and brows-shape, as well as presence of glasses, make-up style and hair style.

The full set of facial features is listed in Table 6.1 and can be categorized in the following five groups:

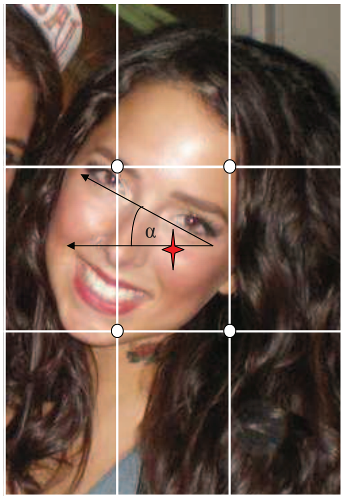


Figure 6.3: Example image of the web site HOTorNOT, $MOS = 9.8$. The white disks represent the stress points, the red cross the image center.

- Ratios of facial features and their locations,
- Facial color traits,
- Shapes of face and facial features,
- Non permanent traits, and
- Expression.

Features related to the mouth and nose width were not exploited, due to the variety of expressions within the database. This expression variety causes significant diversity in the measurements of both, mouth and nose. All selected traits are listed in Table 6.1 (the photograph aesthetics are highlighted for a better overview). Table C.1 and Table C.2 in the Appendix exhibit the traits, trait instances and furthermore the range of magnitude for all photograph aesthetics and facial

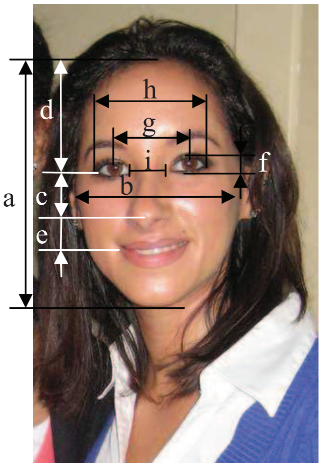
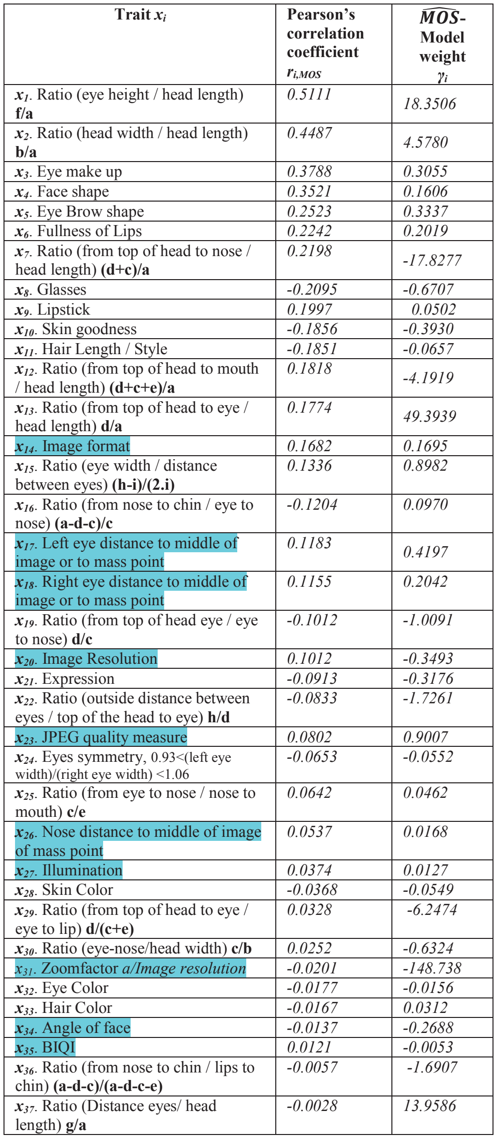


Figure 6.4: Example image of the web site HOTorNOT, $MOS = 9.2$ with employed facial measures.

aesthetics respectively.

Table 6.1: Characteristics listed in decreasing order with respect to the absolute Pearson's correlation coefficient, related Pearson's correlation coefficients and related $\widehat{MOS}$-model weights, see Figure 6.3 and Figure 6.4 for notations of facial measures



Trait x_i	**Pearson's correlation coefficient** $r_{i,MOS}$	$\widehat{MOS}$**-Model weight** γ_i
x_1. Ratio (eye height / head length) **f/a**	0.5111	18.3506
x_2. Ratio (head width / head length) **b/a**	0.4487	4.5780
x_3. Eye make up	0.3788	0.3055
x_4. Face shape	0.3521	0.1606
x_5. Eye Brow shape	0.2523	0.3337
x_6. Fullness of Lips	0.2242	0.2019
x_7. Ratio (from top of head to nose / head length) **(d+c)/a**	0.2198	-17.8277
x_8. Glasses	-0.2095	-0.6707
x_9. Lipstick	0.1997	0.0502
x_{10}. Skin goodness	-0.1856	-0.3930
x_{11}. Hair Length / Style	-0.1851	-0.0657
x_{12}. Ratio (from top of head to mouth / head length) **(d+c+e)/a**	0.1818	-4.1919
x_{13}. Ratio (from top of head to eye / head length) **d/a**	0.1774	49.3939
x_{14}. Image format	0.1682	0.1695
x_{15}. Ratio (eye width / distance between eyes) **(h-i)/(2.i)**	0.1336	0.8982
x_{16}. Ratio (from nose to chin / eye to nose) **(a-d-c)/c**	-0.1204	0.0970
x_{17}. Left eye distance to middle of image or to mass point	0.1183	0.4197
x_{18}. Right eye distance to middle of image or to mass point	0.1155	0.2042
x_{19}. Ratio (from top of head eye / eye to nose) **d/c**	-0.1012	-1.0091
x_{20}. Image Resolution	0.1012	-0.3493
x_{21}. Expression	-0.0913	-0.3176
x_{22}. Ratio (outside distance between eyes / top of the head to eye) **h/d**	-0.0833	-1.7261
x_{23}. JPEG quality measure	0.0802	0.9007
x_{24}. Eyes symmetry, 0.93<(left eye width)/(right eye width) <1.06	-0.0653	-0.0552
x_{25}. Ratio (from eye to nose / nose to mouth) **c/e**	0.0642	0.0462
x_{26}. Nose distance to middle of image of mass point	0.0537	0.0168
x_{27}. Illumination	0.0374	0.0127
x_{28}. Skin Color	-0.0368	-0.0549
x_{29}. Ratio (from top of head to eye / eye to lip) **d/(c+e)**	0.0328	-6.2474
x_{30}. Ratio (eye-nose/head width) **c/b**	0.0252	-0.6324
x_{31}. Zoomfactor *a/Image resolution*	-0.0201	-148.738
x_{32}. Eye Color	-0.0177	-0.0156
x_{33}. Hair Color	-0.0167	0.0312
x_{34}. Angle of face	-0.0137	-0.2688
x_{35}. BIQI	0.0121	-0.0053
x_{36}. Ratio (from nose to chin / lips to chin) **(a-d-c)/(a-d-c-e)**	-0.0057	-1.6907
x_{37}. Ratio (Distance eyes/ head length) **g/a**	-0.0028	13.9586

6.4 Results

6.4.1 Effect of traits on the MOS rating

Our first goal is to find correlation measures for each of the 37 extracted traits and the MOS in order to observe the importance of each characteristic for human perception. The preprocessing step for the MOS related study includes the removal of about 5% of the images, due to their outlier character (i.e. $> 2\sigma_X$, given that x_i is each function of the described traits).

A direct way to find a relationship between the MOS and each of the 37 traits is using Pearson's correlation coefficient. We remind the reader that for two vectors, $X = x_1, x_2, \ldots, x_n$ and $Y = y_1, y_2, \ldots, y_n$, the Pearson's correlation coefficient is given by

$$r_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \tag{6.1}$$

where σ_X and σ_Y are being the standard deviations for X and Y, respectively. The coefficient ranges between -1 and 1, with the two extreme points being obtained when the variables are maximally linearly related.

Pearson's correlation coefficients are calculated for all 37 vectors, each vector corresponding to a feature. Per feature, a 260-values X vector describes each feature for each one of the 260 training images [1]. The 260–values vector Y describes each image related MOS rating. Table 6.1 itemizes these coefficients in decreasing order of importance with respect to the absolute Pearson's correlation coefficient.

6.4.2 Insight provided from empirical data

The first notable result reveals the strong correlation between the best ranked traits and the MOS, which even exceeds a Pearson's correlation coefficient of 0.5 for the trait 'ratio eye-height/face-height'. Particularly in regard to an automatic MOS prediction image processing tool these results are very encouraging. Further we observe that photo-quality features play a less significant role than facial aesthetics, as expected, but they are not to be neglected, since they achieve an $r_{14,MOS} = 0.168$. Moreover we note that the high ranked traits x_1, x_2 and x_4, which represent the ratios (eye-height/face-height) and (head-width/head-height), and furthermore face shape, see Table 6.1 are features corresponding strongly to person's weight. This outcome brings to the fore the strong importance of low human weight for aesthetics. Furthermore it is worth noting that Table 6.1 reveals the surprising fact among others, that non permanent traits place a pivotal role in raising the MOS rating. Eye make-up, lipstick, glasses and hair-style are all among the top 11 of the obtained ranking. These results hint the high modifiability of facial aesthetics perception by simple means like make-up or hair styling. The relevance of eye make-up had been previously observed in [GKYG10]. Together with the different conclusions that one may draw from Table 6.1, it also becomes apparent that different questions are raised, on the interconnectedness of the different traits. This is addressed in Section 6.4.3. Finally we note that traits, such as x_1, x_7, x_{12} and x_{13} directly comply with the well known babyfaceness hypothesis (see [bea11]), which describes that childlike facial features in females increase attractiveness, such features include big eyes, e.g. x_1 and a relative low location of facial elements, e.g. x_7, x_{12} and x_{13}. One measure known for increasing attractiveness, if equal to the golden ratio $\phi = 1.618$, is x_{16}.

6.4.3 Interconnectedness of different traits

To get a better understanding of the role of the different traits in raising the MOS, it is helpful to study the inter-relationship between these traits. This is addressed in Table C.3 in the Appendix, which describes the correlation between selected traits. Due to lack of space we limit the correlation matrix to just a group of the first six traits. Table C.3 can answer different questions such as for example the validity of the conclusion in Table 6.1 on the importance of the make-up feature. In this case, the question arises whether it is truly the make-up that affects the MOS or whether

[1] For information on denotation of features and according X–values, please refer to the Appendix, Table C.2

already attractive subjects use make-up more heavily. Table C.3 suggests a low correlation between the facial proportions (representing beauty) and eye make-up, which validates the strong role of makeup in raising the MOS.

6.5 Model for facial aesthetics

We choose a linear metric due to its simplicity and the linear character of the traits with increasing MOS. We perform multiple regression with the multivariate data and obtain a MOS estimation metric with the following form:

$$\widehat{MOS} = \sum_{i=1}^{37} \gamma_i x_i. \tag{6.2}$$

The resulting weights γ_i corresponding to each trait are denoted in Table 6.1.

We here note that the weights of the model are not normalized and do not give information about the importance of each characteristic. With other words, we did not normalize for the sake of reproducibility - $\widehat{MOS}$ can be computed with features labeled as in Table C.1 and Table C.2 in Appendix C and related weights from Table 6.1. The importance of the characteristics is conveyed by the Pearson's correlation coefficients $r_{X_i,MOS}$.

6.5.1 Validation of the obtained metric

To validate our model we compute the following three parameters.

- Pearson's correlation coefficient. As described above, and it is computed to be

$$r_{\widehat{MOS},MOS} = 0.7690. \tag{6.3}$$

- Spearman's rank correlation coefficient, which is a measure of how well the relation between two variables can be described by a monotonic function. The coefficient ranges between -1 and 1, with the two extreme points being obtained when the variables are purely monotonic functions of each other. This coefficient takes the form

$$r_S = 1 - \frac{6\sum_i d_i}{n(n^2-1)}, \tag{6.4}$$

  where $d_i = rank(x_i) - rank(y_i)$ is the difference between the ranks of the i^{th} observation of the two variables. The variable n denotes the number of observations. The coefficient, which is often used due to its robustness to outliers, was calculated here to be

$$r_{S\widehat{MOS},MOS} = 0.7645. \tag{6.5}$$

- Mean standard error of the difference between the estimated objective $\widehat{MOS}$ and the actual subjective MOS.

$$MSE = 0.7398 \tag{6.6}$$

These results clearly outperform the outcomes from Eigenfaces of $r_{\widehat{MOS},MOS)} = 0.18$, as well as neural networks $r_{\widehat{MOS},MOS} = 0.458$ (see [GKYG10]), but the comparison is not very adequate as we would compare manual extraction with automatic extraction of facial aesthetics. Nevertheless the potential of our approach is evident and we proceed with a robust validation of the facial aesthetics metric. For this purpose we annotated the 37 traits beyond the training set, in an extra testing set of 65 images. Once more we excluded outliers (3 images) and we computed the metric verification measures for the estimated $\widehat{MOS}$ and the according actual MOS

– Pearson's correlation coefficient:

$$r_{\widehat{MOS},MOS} = 0.7794. \tag{6.7}$$

– Spearman's rank correlation coefficient:

$$r_{S\widehat{MOS},MOS} = 0.7860. \tag{6.8}$$

– Mean standard error:

$$MSE = 1.158. \tag{6.9}$$

The high Pearson's coefficient implies a robust prediction accuracy of the facial aesthetics metric. The Spearman's coefficient gives an indication about the correlation between estimated and real MOS, but without the restriction of linear dependence. It considers each monotonic function connecting the two vectors. In our case this coefficient is relatively high as well. The MSE on the other hand gives an idea about the absolute error between the predicted and actual values. It is interesting to observe that the testing set provides even higher correlation coefficients than the calibration set, but the MSE reveals that the absolute error increases for the testing set, and thus that the actual performance decreases.

We proceed with three experiments using the validated MOS–prediction metric.

6.6 Experiments with $\widehat{MOS}$

The above designed MOS prediction metric is in this section employed towards (partial) quantification of the general concept of beauty. We are specifically interested in addressing questions such as:

- Are famous females known for their beauty more beautiful than average females?
- What is the influence of age on beauty?
- How much does facial surgery change the beauty score?

Towards addressing the above, we proceed to apply our metric on images drawn from the internet and from official databases such as the FG–NET and the Plastic surgery database.

6.6.1 Metric verification on highly ranked females

Towards verification of its usefulness, we applied the above designed MOS–prediction metric on images of females who have been highly ranked by the popular media. Specifically we considered images of females leading the lists of *People's magazine* as the 'most beautiful people' from 1991 to 2011, as well as the top 10 entries from the same list for the year 2010. The considered images included, among others, those of Jennifer Lopez (winner 2011), Julia Roberts and Angelina Jolie. After annotation and calculation of the related beauty indices, we contrasted the results from the above lists, to those we obtained when we considered images from the HOTorNOT database (see Figure 6.5). The test validated our choice of metrics, with the entries from the above 'beautiful people' lists, consistently scoring significantly higher scores, as well as exhibiting a lower variance. We displayed, for both image sets, the average MOS values, as well as those within a confidence interval of 95%.

6.6.2 Dependence between beauty and age: FG–NET aging database

Towards investigating the dependence between beauty and age, we considered images from the FG–NET database [fgn11], as this database provides us with multiple images of subjects as

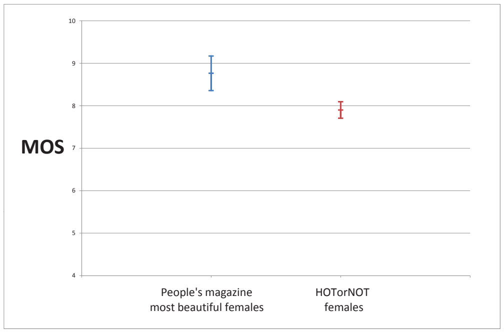


Figure 6.5: Comparison of average MOS for subjects of the HOTorNOT database and of average $\widehat{MOS}$ for subjects of the People's magazine most beautiful people list. Average MOS and $\widehat{MOS}$ values with related a confidence interval for 95%.

they age. We specifically selected females with images available from a broad time spectrum, e.g. images available from an age about 18 years old to 60 years old. We annotated labeled these images with the facial and photographic traits from Section 6.3.2 and Section 6.3.3 and computed the corresponding $\widehat{MOS}$ values. We obtained per subject several beauty scores spanned over time. Since the range of these beauty functions differed on the MOS scale between different females, we normalized the functions to 1, with 1 being the maximum $\widehat{MOS}$ per female. We then averaged the normalized beauty over time functions and estimated based on the result a polynomial function of the 5th degree. Figure 6.6 displays the merged functions and the related estimation function. The resulting beauty function over time bares a maximum between the ages 23 to 33. The outcome can be explained on the one hand by traits changes like wrinkles and presence of glasses with advancing age, as well as on the other hand by a reduced interest in regards to make up or hair style.

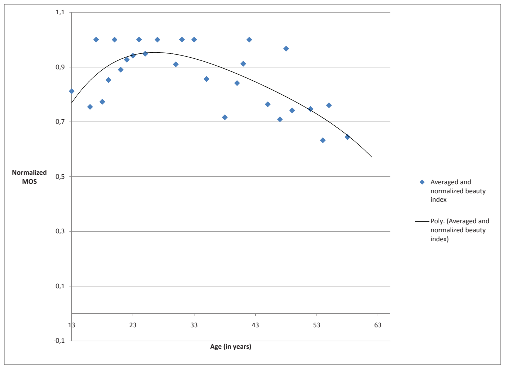


Figure 6.6: $\widehat{MOS}$ for females of different ages normalized to 1 per female, with 1 being the maximum $\widehat{MOS}$ per female, and furthermore averaged over all considered females.

6.6.3 Facial surgery

We also examined the effect of *blepharoplasty* (eyelid lifting surgery) on the beauty index. Our choice of this specific parameter and surgery was motivated by the fact that eye size has been shown to have a high impact on our chosen beauty metric. We randomly selected 20 image pairs (before and after the surgery [2], see Figure 6.7) from the *plastic surgery database* [SVB$^+$10], and after annotation, we computed the related beauty indices. Interestingly our analysis suggested a relatively small surgery gain in the $\widehat{MOS}$ increase. Specifically the increase revealed a modest surgery impact on the beauty index, with variations ranging in average between 1% and 4%.

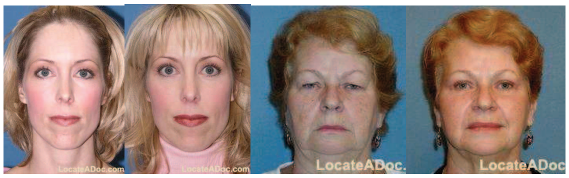


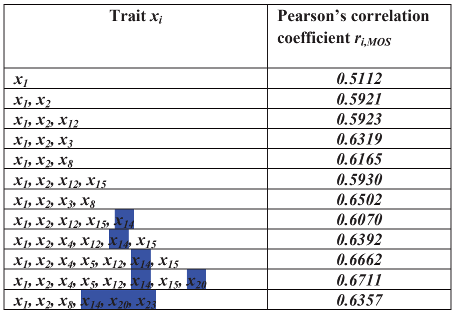
Figure 6.7: Examples of the Plastic Surgery Database. The left images depict the subjects before surgery, the right images after surgery.

We proceed with the analysis and simulation of an automatic tool for facial beauty prediction.

6.7 Towards an automatic tool for female beauty prediction

An automatic tool including classification regarding all 37 above presented traits will have the benefit of a maximal achievable prediction score, at the same time though each automatically detected trait will bring an additional classification error into the prediction performance. Thus in designing such an automatic tool a tradeoff between possible prediction score and categorization error has to be considered. We illustrate an analysis of prediction scores evoked by different combinations of traits in Table 6.2.

Table 6.2: Sets of combined traits and related Pearson's correlation coefficients



Trait x_i	**Pearson's correlation coefficient $r_{i,MOS}$**
x_1	*0.5112*
x_1, x_2	*0.5921*
x_1, x_2, x_{12}	*0.5923*
x_1, x_2, x_3	*0.6319*
x_1, x_2, x_8	*0.6165*
x_1, x_2, x_{12}, x_{15}	*0.5930*
x_1, x_2, x_3, x_8	*0.6502*
$x_1, x_2, x_{12}, x_{15}, x_{14}$	*0.6070*
$x_1, x_2, x_4, x_{12}, x_{14}, x_{15}$	*0.6392*
$x_1, x_2, x_4, x_5, x_{12}, x_{14}, x_{15}$	*0.6662*
$x_1, x_2, x_4, x_5, x_{12}, x_{14}, x_{15}, x_{20}$	*0.6711*
$x_1, x_2, x_8, x_{14}, x_{20}, x_{23}$	*0.6357*

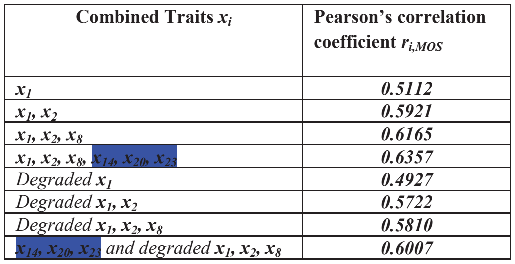
Motivated by this Table 6.2 and towards simulating a realistic automatic tool for beauty prediction, we select a limited set of significant traits, x_1, x_2, x_8, with other words factors describing how big the eyes of a person are, the ratio head width/head height and the presence of glasses. Moreover we add acquisition traits with no extra error impact, such as x_{14}, x_{20}, x_{23}, namely image format, JPEG quality measure and image resolution. We then proceed to appropriate reliability scores related to x_1, x_2, x_8 based on state of the art categorization algorithms:

[2] For this experiment all values attached to non permanent traits were artificially kept constant for "before surgery" and "after surgery" images.

– facial landmark recognition with accuracy of of 6.23 pixels and 2.1%, reported in the work [DM08],
– face localization with accuracy between 90% and 98% depending on the database presented in [GLW$^+$11] and
– glasses detection with accuracy of 94% shown in [WAL04].

We deteriorate the manually annotated data with the above realistic algorithmic estimation accuracies and compute the Pearson's correlation coefficient between user MOS rating and the predicted $\widehat{MOS}$ based on simulated error prone algorithms. We obtain a realistic simulated beauty prediction performance presented in Table 6.3. Such an automatic tool, based only on three traits provides related results that would outperform outcomes from Eigenfaces of $r_{\widehat{MOS},MOS)} = 0.18$ (see [GKYG10]) and neural networks $r_{\widehat{MOS},MOS} = 0.458$ (see [GKYG10]).

Table 6.3: Simulated automatic tool for beauty prediction



Combined Traits x_i	Pearson's correlation coefficient $r_{i,MOS}$
x_1	*0.5112*
x_1, x_2	*0.5921*
x_1, x_2, x_8	*0.6165*
$x_1, x_2, x_8, x_{14}, x_{20}, x_{23}$	*0.6357*
Degraded x_1	*0.4927*
Degraded x_1, x_2	*0.5722*
Degraded x_1, x_2, x_8	*0.5810*
x_{14}, x_{20}, x_{23} *and degraded* x_1, x_2, x_8	*0.6007*

6.8 Summary

In this chapter, we presented a study on facial aesthetics in photographs, where we compared objective measures (namely photograph quality measures, facial beauty characteristics and soft biometrics), with human subjective perception. Our analysis revealed a substantial correlation between different selected traits, and the corresponding MOS-related beauty indices. Specifically we presented that non permanent features can influence highly the MOS, and based on our analysis we conclude that facial aesthetics in images can indeed be substantially modifiable. With other words parameters such as the presence of makeup and glasses, the image quality as well as different image post–processing methods can significantly affect the resulting MOS. Furthermore we constructed a linear MOS–based metric which was successfully employed to quantify beauty-index variations due to aging and surgery. Our work applies towards building a basis for designing new image-processing tools that further automate prediction of aesthetics in facial images. Towards this we provided a simulation of an automatic prediction tool based on state of the art categorization algorithms and the designed MOS–prediction metric.

By now we ensured the user of the practicality of SBS for security as well as entertainment applications. In a next step we provide a chapter 7 featuring classification algorithms of a SBS, as employed and analyzed in the chapters 3 and 4.

Chapter 7

Practical implementation of soft biometrics classification algorithms

7.1 Set of facial soft biometrics

As elaborated in Chapter 2 higher and more satisfactory distinctiveness can be achieved by using more than one trait, rather than a single trait. Thus we here propose a set of facial soft biometrics that can be exploited for human identification, as shown in Chapter 3. In an effort to find a good balance between identification–reliability and complexity, we here propose a soft–biometric system that focuses on simple and robust classification from a bounded set of traits and their trait–instances. In what follows, we will describe these basic elements, as well as the employed classification algorithms.

In the presented set of facial soft biometric traits, we allocate 6 traits, which we choose and label as shown in Table 7.1.

Table 7.1: Table of Facial soft biometric traits

SB trait	**Algorithm**	**Database**
Skin color	Deduced from [KMB]	FERET
Hair color	Deduced from [ZSH08]	FERET
Eye color	Own developed	UBIRIS2
Beard	Own developed	FERET
Moustache	Own developed	FERET
Glasses	Deduced from [JBAB00]	FERET

We proceed now to specify basic aspects of the classification algorithms that were used for trait–instance identification.

7.1.1 Classification algorithms

The basic classification tool consisted of an automatic frontal face and facial features detector, which was partially drawn and modified from the algorithms in [vio]. Implementation of the different classification algorithms (see Table 7.1 for an overview) was performed using OpenCV[1].

Before describing some basic aspects of the implemented trait classification algorithms, we note few pertinent issues that accompany categorization. Regarding coordinate determination, we note that typical eye, skin and hair color classifiers require knowledge of the eye coordinates, and similarly hair color categorization requires knowledge of the coordinates for the upper head region. The precise computation and extraction of the characteristic regions of interest (ROI) (see Figure 7.1) for the eyes, mouth, nose and upper face coordinates, are essential for the subsequent classification. For higher accuracy, only in the training step, all coordinates were manually annotated. The considered ROIs for the selected soft biometric traits are illustrated in Figure 7.1. Identification of the ROI was generally followed by acquisition of the Hue, Saturation and Value (HSV) values. We note that the HSV color–space was chosen for being robust to illumination changes, as well as for the fact that it allows for a high degree of independence between the H, S, and V parameters, which renders the system capable to better handle light changes or shadows. Regarding outlier filtering, we used a simple threshold on the HSV values, based on the color standard–deviation σ. This was followed by HSV normalization. Regarding the statistical modelling, the probability density functions of skin, eye and hair color were computed using 3–component Gaussian mixture models whose parameters were estimated using the EM algorithm. Posterior probabilities over the observed HSV vectors for all trained trait instances were computed, followed by a majority vote decision on the categorized trait instance.

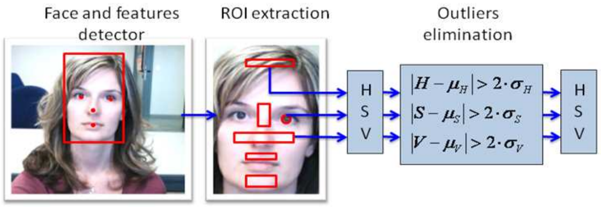

Face and features detector
ROI extraction
Outliers elimination
H
S
V
$|H-\mu_H|>2\cdot\sigma_H$
$|S-\mu_S|>2\cdot\sigma_S$
$|V-\mu_V|>2\cdot\sigma_V$
H
S
V

Figure 7.1: ROI for the set of facial soft biometrics. Outlier filtering was a function of the standard deviation σ and the mean μ for each of the H,S and V parameters.

1) Eye Color classification: In this setting, careful and precise consideration of the ROI was particularly important, due to the region's inherently small size. The specific ROIs were retrieved using the circular Hough transform, followed by pupil and reflection extraction, and then by acquisition of the HSV vectors. Regarding the training step, each eye color group was trained using images from the UBIRIS2[2] database. A more elaborate study on eye detection and eye color classification follows in Section 7.2.

2) Hair color classification: The hair color ROI was chosen as a thin bar in the upper head region, as indicated in Figure 7.1. Training utilized 30 FERET[3] images for each of the hair colors, where the annotation was done manually.

3) Skin color classification: Classification of skin color was done in accordance to the eye coordinates which defined the ROI for the skin color classification to be the area underneath the ocular region. Training utilized 33 FERET images per skin color group, which were again annotated manually.

[1] OpenCV webpage on Source forge http://souceforge.net/projects/opencvlibrary/

[2] available for download at http://iris.di.ubi.pt/ubiris2.html

[3] available for download at http://www.itl.nist.gov/iad/humanid/feret

4) Eye glasses detection: Towards glasses detection, we considered that the areas around the eyes can be searched both for hints of glasses as well as for glass reflections. Challenges related to the fact that glasses frames are either occasionally absent, or that they often resemble wrinkles, brows, shades and hair. A further challenge came from the fact that illumination variations hindered the appearance of reflections. These challenges were handled by placing emphasis on a ROI corresponding to the nose part of the glasses. The specific algorithm consisted of eye position estimation, grey–level conversion, histogram equalization, extraction of region between the eyes, Laplacian edge detection and finally line detection.

5) Beard and Moustache Detection: In this case, face detection and feature localization were followed by identification of the ROIs. These ROIs include the chin for the beard, and the area between the mouth and nose for the moustache. The color estimation was followed by outlier extraction and HSV normalization. The presence of beard and / or moustache was based on the Euclidean distance between the processed observation and skin and hair–color information respectively. The presence of moustache was determined independently.

Algorithmic dependencies: As is the case with general optimization problems, identification of algorithmic dependencies endows the system with increased reliability and computational efficiency. Towards this we refer to notable examples of such dependencies, such as that between skin color and glasses where, due to ROI overlap, the presence of glasses has an impact on the perceived skin color. This information can be utilized and employed by modifying the ROI for skin color classification. Additionally we recall that skin color is employed in the classification of hair, detection of beard and moustache, where furthermore the latter two traits are also contingent upon hair color. Figure 7.2 sketches further dependencies of the mentioned facial soft biometric traits. Some of these dependencies were partly exploited in the process of classification.

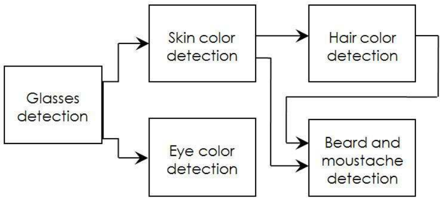


Figure 7.2: Facial Soft Biometric traits algorithmic dependencies.

7.1.2 Experimental results

The above introduced algorithms for categorization of the chosen facial soft biometric traits are here examined and evaluated. It is to be noted that the traits glasses, beard and moustache are of a binary character, whereas the color based facial traits possess discrete traits instances.

Glasses: Tests for eye glasses detection were performed on a testing set of images of FERET[4] database. The threshold based algorithm provided a correct classification rate (containing the true positive and true negative rate) of 87.17% (see Table 7.2) comparable to the results in [JBAB00].

Color based Facial Soft biometric traits: Eye, Skin and Hair Color: In the context of the color based facial soft biometrics it is to be noted, that the number of the established classification groups was adjusted to both, the performance and limitations of human perception and estimation capabilities. Results are presented in true positive rates and confusion matrices in Figure 7.3.

[4] available for download at http://www.itl.nist.gov/iad/humanid/feret

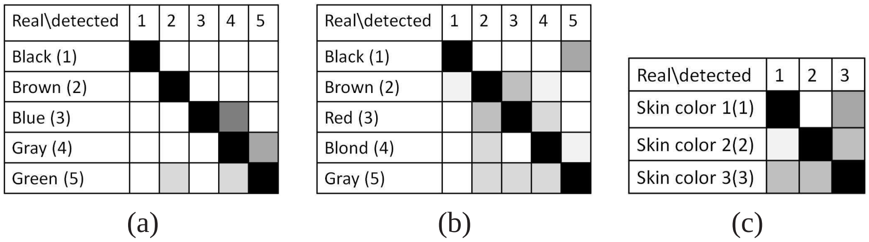
Table 7.2: Glasses, beard, and moustache detection results. The experiments are conducted on the well known FERET database.

SB trait	**Detection rate**	**FPR**	**FNR**
Glasses	87.17%	7.17%	5.66%
Beard	80.7%	8.1%	11.2%
Moustache	72.8%	12.7%	14.5%

Table 7.3: Eye, Skin and Hair Color True Positive Rates

	Eye Color	**Skin Color**	**Hair Color**
True Positive Rate	72.6%	79.2%	70.08%

For the latter the values range from white (no confusion) to black (maximum confusion). The diagonal fields correspond to the true positive rates. Eye color results were performed on a testing set containing 5 eye color groups, namely black, brown, blue, gray and green. The images were retrieved from the UBIRIS2 database and results are presented in Table 7.3 and in Figure 7.3.(a). We here briefly note the peak confusion rate between blue and gray eye color, mostly responsible for the overall break–in in the true positive rate. Hair color is classified in 5 groups, black, brown, red, blond and grey. A testing set of FERET images provided the in Table 7.3 and Figure 7.3.(b) presented results. Skin color exhibits low variation in color spaces and thus slight illumination changes result in wrong classifications. Due to this challenge the limitation of 3 skin color groups was adopted with related results presented in Table 7.3 and Figure 7.3.(c). The confusions were mostly due to illumination variances and detected shadows, which result in a shift on the skin color shades.



Real\detected	1	2	3	4	5
Black (1)	■				
Brown (2)		■			
Blue (3)			■	▓	
Gray (4)				■	▒
Green (5)		░		░	■

(a)

Real\detected	1	2	3	4	5
Black (1)	■				▒
Brown (2)	░	■	▒	░	
Red (3)		▒	■	░	
Blond (4)		░		■	░
Gray (5)		░	░	░	■

(b)

Real\detected	1	2	3
Skin color 1(1)	■		▒
Skin color 2(2)	░	■	▒
Skin color 3(3)	▒	▒	■

(c)

Figure 7.3: Confusion matrices: (a) Eye Color (b) Hair Color and (c) Skin Color.

Beard and Moustache detection: Once more a set of FERET images was employed for the validation of beard an moustache. The binary character of the traits (present or not present) is in real images ambiguous, due to various lengths and shapes of beard and moustache. This factor made a unique annotation and then in turn estimation difficult and led to the results shown in Table 7.2. A small fraction of the wrong detections is due to the not correspondence between hair color and beard/moustache color, which we assumed in the detection algorithm.

To understand the presented experimental results, we perform a detailed study on one of the traits namely eye color.

7.2 Eye color as a soft biometric trait

In this section we focus on eye color as a soft biometric trait, where we spot eye color information, previously mostly disregarded by classical iris pattern and texture recognition methods. We then specifically examine extraction and categorization of eye color and conduct an additional study where we illustrate the influence of surrounding factors like illumination, eye glasses and sensors on the appearance of eye color.

7.2.1 The human iris color

The human ocular region bears a plethora of biometric traits with different importance, such as iris, sclera and retina patterns, eye shape, eye brow shape and size and eye color. Eye color has been mainly neglected, probably for the reason that 90% of humans have brown eyes. The distribution of eye colors characterizing Caucasian and specifically European subjects is of a bigger deviation than in the rest of the world, as an example of the German speaking region proves in [hai10]. Eye color is determined by the amount and type of pigments in the eye's iris and is genetically inherited. It is of more permanent character than other soft biometric traits. We note though that 'hazel' eye color is a special case, where it is known to change its colors. A study [BMC$^+$97] shows that eye color changes slightly over the span of 40 years. The extraction of eye color is of very sensitive character, since the area is small (around 11mm) and color itself is difficult to deduce because of its illumination sensitivity. A further difficulty is the variable size of the pupil (mainly due to illumination), from $3-4$mm up to $5-9$mm. Then again, positive factors for the feasibility of eye color classification are the smaller and lower-priced surveillance sensors, which are increasingly available, and furthermore provide higher resolutions.

7.2.2 Related work

Few preliminary scientific works on eye color exist, as on subjectivity of human eye color grading [SSG$^+$90], [FCB08], on human analysis of eye color photographs [TSC$^+$01] and first preliminary classification attempts [MRGL00] and [FDH03]. We clearly differentiate our work, by presenting a full automatic eye color categorization system and furthermore by providing insight on related pertinent factors.

We present here a preliminary study towards a robust eye color classifier. In Section 7.3 we describe an automatic eye color classification algorithm, which contains automatic iris extraction and Gaussian mixture models for classification. Simultaneously related results on the reliability are presented. Section 7.4 offers a preliminary study on factors with impact on eye color classification, such as illumination, camera sensor, presence of glasses and consideration of the left or right eyes. Such an eye color classifier can serve as a preprocessing step of an iris pattern classification system, where we do not expect to increase the reliability of the overall system (iris patterns are considered as highly reliable biometrics), but our system rather can pre prune the database to save computational complexity and time, see Chapter 4 and [HJP99].

7.3 Iris color classification

In designing an automatic eye color classification system, the choice of the method for iris extraction as well as the color classification have to meet the criteria of reliability, of time and of

Table 7.4: GMM eye color results for manually segmented irides

Black	Brown	Green	Blue
100%	100%	87.5%	81.8%

computational efficiency. In accordance with these aspects in this section we present an iris extraction technique and a classification method and jointly examine them on a large color eyes database captured in visual light, the UBIRIS2 [PFS+09]. It contains 261 subjects featuring diverse illuminations and iris positions. We manually annotated a subset of the database and obtained the following four eye color sets for both, training and testing (about $3/4$ and $1/4$, respectively):

- Black: 100 images
- Brown: 85 images
- Blue: 81 images
- Green: 60 images

Those colors were specified, as on the one hand they are straightforward and human distinguishable and on the other hand, enough images are available for the following modeling.

7.3.1 Gaussian Mixture Models and color spaces

Manual region of interest (ROI) extraction: Towards the statistical modeling, the selected and annotated subset of UBIRIS2 images are manually cropped to avoid effects of reflections or traces of the pupil. An illustration of the manual performed color extraction is shown in Figure 7.4. The four probability density functions (pdf), one for each color, are then computed considering all pixels of the extracted region of interest (ROI), using 3-component Gaussian mixture models (GMM). The GMM parameters are estimated using the expectation maximization (EM) algorithm. We refer to this step as training and perform it for the four color spaces: RGB, HSV, CieLAB and CieLuv in order to assess the color space best suited for eye color description.

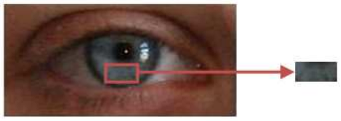


Figure 7.4: Manual iris color region extraction.

For the testing step, the set of images is again manually cropped and posterior probabilities over all observed pixels are computed, followed by a majority vote decision on the categorized eye color. The analysis is performed on manually extracted images to prove the suitability of GMM for color distinguishing. The best results are acquired surprisingly on the RGB color space; see Table 7.4, for which reason, the rest of the study considers solely the RGB color space.

In the first case of wrong classifications, blue is confused for green and in the second case, green for brown.

Automatic region of interest (ROI) extraction: Furthermore, the test was performed on automatically extracted color irides of UBIRIS2 images. We hereby segment the color part of the irides by the automatic extraction method presented in [ED10] based on circular Hough transfor-

Table 7.5: GMM eye color results for automatically segmented irides

Det \ Real	Black	Brown	Green	Blue
Black	90.9%	5.26%		6.25%
Brown		89.47%	14.28%	
Green	4.54%		78.57%	18.75%
Blue	4.54%	5.26%	7.14%	75%

mation. We then proceed to again compute, pixel by pixel, 4 pdf-s by the means of GMM and EM for the training. For the testing again the posterior probabilities for a new, again automatically segmented testing set are computed. This time the majority vote rule is extended and contains following considerations. An eye appears as brown, green or blue, if it contains pigments representing those colors, but the brown, green or blue fraction does not necessarily have to possess the highest percentage. Mainly the pupil and the dark boundary contribute to a higher occurrence of black color, and often brighter irises enclose a multitude of further darker pigments, which constitute in patterns. In black eyes on the other hand, the black percentage is at least $2/3$ of the iris pixels. Following rules were deduced from the above considerations and adhere with priority to the majority vote rule:

1. If the iris contains more than 70% of black pixels, the categorized color is black.
2. If black is the majority, but accounting less than 50%, then the second strongest color is the categorized color.
3. If black is the majority, but accounting less than 50% and brown and green are in the same range, the categorized color is green.

The related results following those rules can be found in Table 7.5. The values in the diagonal, highlighted in gray, represent the true classification rates. All other fields illustrate the confusions between the real and estimated eye colors.

Two examples of confusions are provided in Figure 7.5. In the first case a green eye is confused with black. It is to be noted that the pupil, iris boundary, lashes and an unfavorable illumination establish a high percentage of the image and thus of the black pixel fraction. In the second case a blue eye is categorized as green. On the one hand the eye contains greenish pigmentation, and on the other hand the presence of the lid and pupil account for the wrong estimation.

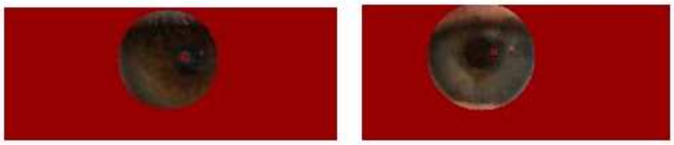


Figure 7.5: Examples of wrong classified eye colors.

7.4 Influential factors

Eye color classification is a challenging task, especially under real life conditions, mainly due to the small size of the ROI and the glass-like surface of the apple of the eye. Smallest

Table 7.6: Eye color database for influential factors study

Subjects	8
Eye colors	8: black, brown, hazel, green-brown, light green, green-blue, blue, light blue
Camera sensors	2: white balanced Cannon 400D, webcam Logitech 1.3 Megapixel
Illuminations	4: (in office) daylight, daylight+room lights, flashlight, fluorescent table light
Pair of glasses	2

external changes may have impact on the perception and categorization of eye color. To understand the magnitude of the impact we here study following pertinent and frequently occurring factors: illumination variation, presence of glasses, difference in perception of observation of left and right eye, and finally the influence of two camera sensors. For this study we captured a small database of eight subjects (see Table 7.6), each with a different color of eyes. For each subject we produced 7 different images: four of the images under real life illuminations, one image with a second camera sensor, and finally two images, one for each pair of glasses. We note that for the further analysis the ROIs were extracted manually (see Figure 7.4) to eliminate any traces of the pupil and light reflections.

In the following study performed on the presented database, we consider red and green chromaticities, following similar studies, regarding skin locus and we note that: $r = R/(R+G+B)$ and $g = G/(R+G+B)$.

7.4.1 Illumination variation

The spectrum of incoming light plays a major role in many biometrics applications and especially in color based ones. Intuitively, it is expected that illumination has also a strong impact on eye color. That is why we here study the subjects of the own recorded database in 4 real illumination conditions.

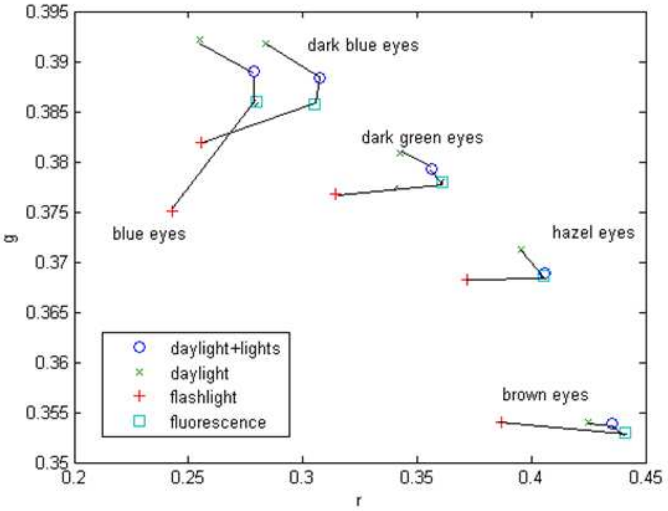


Figure 7.6: Illumination variation of eye colors.

We here captured the subjects of our database in following 4 illuminations conditions: daylight and office lights, daylight, flashlight and a fluorescent table light. Clear shifts in the color space can be observed. For conciseness and to avoid overlap we portray only 5 subjects. It is clear that a robust eye color categorization technique must consider and cope with illumination variations. A possible solution is to estimate the illumination condition. A self suggesting illumination estimation method is proposed by [DYC06], where the color of the sclera serves as estimator of the ROI.

7.4.2 Glasses influence

The presence of glasses is another interfering factor, primarily examined in the context of face recognition and naturally of importance in the current eye color categorization study.

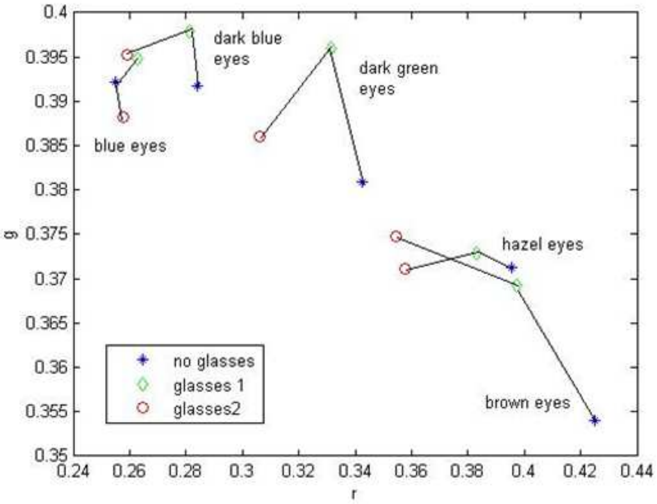


Figure 7.7: Eye colors behavior with and without glasses (constant illumination).

The subjects were asked for this test to wear 2 different pairs of glasses. It is interesting to compare this graph with the illumination variation one (Figure 7.6) in order to comprehend the immense drift eye glasses cause in eye color. It is evident that a stable eye color categorization system should include a priori glasses detection. To detect the presence of glasses in an efficient and robust manner, we can perform histogram normalization, followed by Canny edge detection on the area between the eyes. A further line detection indicates the presence of the frame part of the glasses. This algorithm was deduced from [JBAB00]. For eye color classification in the case of presence and absence of glasses the eye color classifier should be able to estimate and compensate the color shift of the estimated values.

7.4.3 Consideration of left and right eye color

We here show that the strong illumination influence has not only impact on images captured under different illumination conditions, but also on the color perception of left and right eye. Although none of our subjects has the seldom condition of heterochromia (different iris colors of left and right eye), a drift between the colors of left and right eye can be observed. The illumination for this graph was constant daylight falling sidely on the face of the subjects, to achieve in order a maximum illumination difference between left and right eye.

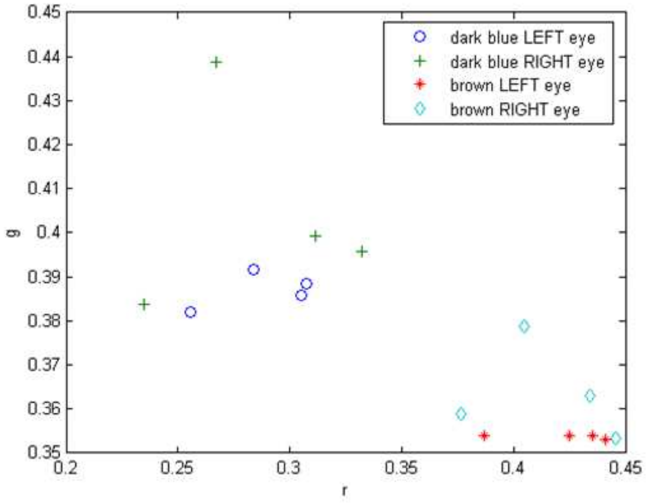


Figure 7.8: Eye colors of left and right eyes for 2 subjects (for 4 different illuminations).

7.4.4 Camera sensors

For the sake of completeness we proceed to provide a graph on the shift between two camera sensors (Logitech Webcam and Cannon 400D). The measured color data is clearly influenced by the characteristics of the cameras.

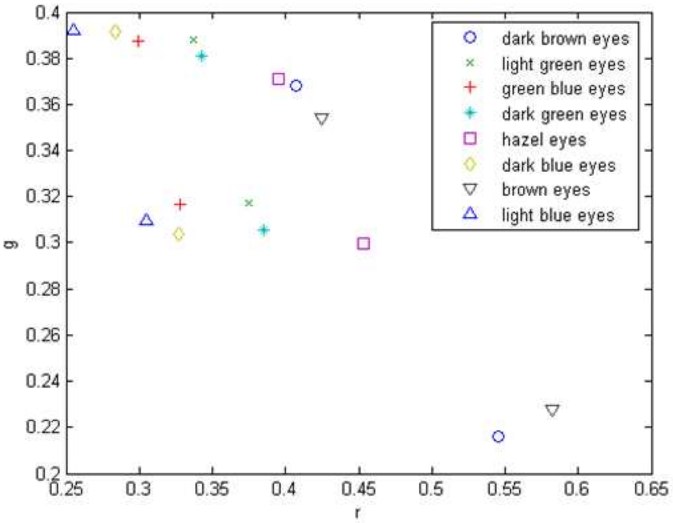


Figure 7.9: Eye colors captured with two camera sensors (constant illumination).

We note that the presented study identifies each one of the examined influential factors as disturbing for eye color categorization. The measure of importance for each one of them is ascertained by the embedding application.

7.5 Summary

This chapter presented classification algorithms related to six facial soft biometric traits, namely the color of eye, skin and hair and moreover beard, moustache and glasses and provide according results. We then specifically focused on and examined eye color, developed an automatic eye classification system and studied the impact of external factors on the appearance of eye color. We have identified and illustrated color shifts due to variation of illumination, presence of glasses, the difference of perception of left and right eye, as well as due to having two different camera sensors.

In the last chapter 8 we deviate from the analysis and development point of view towards SBS and examine instead the user friendliness of such a system by providing a study on user acceptance towards such systems. In this study we compare SBSs to other biometric systems, as well as to the classical PIN system toward access control.

Chapter 8

User acceptance study relating to soft biometrics

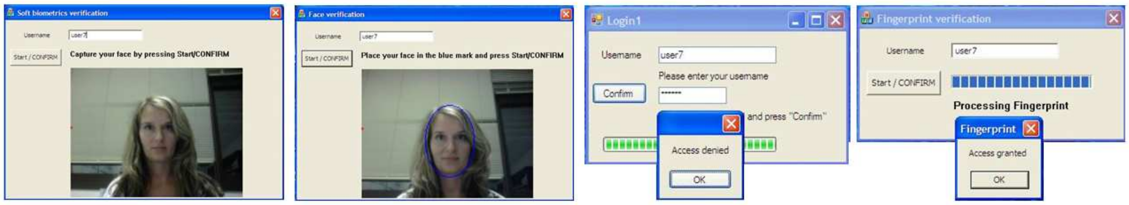
The pervasiveness of biometric systems, and the corresponding growth of the biometric market, see [usa11a], has successfully capitalized on the strength of biometric-based methods in accurately and effectively identifying individuals. As a result, modern state-of-the-art intrusion detection and security systems include by default at least one biometric trait. It is the case though that little emphasis has been given to better understanding user-acceptance and user-preference regarding such systems. Existing usability related works, such as in [CAJ03] and [LBCK03], focus on establishing functional issues in existing ATM machines, or on studying the influence of user interaction on the performance of fingerprint based systems (see [KED11]) and interfaces (see [RJMAS09]). Other interesting works (see [usa11b], [CG05], [CJMR09]), analyze possible methods that improve interface design. Our emphasis here is on providing insight on the attitudes and experiences of users towards novel and emerging biometric verification methods, and to explore whether such novel biometric technologies can be, in terms of user acceptance, valid alternatives to existing prevalent PIN based systems. Our focus, in addition to considering the traditional PIN based method, is to explore the usability aspects of systems based on classical biometrics such as fingerprint and face recognition, and to then proceed to study the usability of systems based on the emerging class of soft-biometric methods. Our evaluation is based on having the users rate and rank their experiences with different access methods.

8.1 Usability of access control methods

A successful access control system must (see [BH90] and [Nie93]) incorporate common user understanding and knowledge in order to enable a natural and unconstrained interaction with the user. Generally this interaction involves a number of different subtasks, which must be planned and structured to allow for simplicity and comfort. The subsequent mapping of such tasks onto the interface should be intuitive, obvious and user friendly. Furthermore such designs should impart feedback in the form of descriptions or state indications (progress bar, message window). Finally the selected design should be robust against human errors. The above considerations will be the guiding principles of our tests and of the corresponding usability study.

8.1.1 Testing Methodology

Test Setup : A set of 15 users (5 female, 10 male) of different nationality and ethnicity between 26 and 37 years old was randomly selected from an office complex with the condition to not work on biometrics. The participants were not paid. The test was consistent with the ITU-T recommendation [usa00], that is to say we followed methods for interactive user tests of setting, equipment and environment, as well as subjects training and solicitation of opinions. We used rating methods according to the absolute category rating (ACR). Specifically for the ratings of the four access systems we presented each one at a time and let the user rate them independently. We denote this rating as MOS (mean opinion score), which spans on a five grade scale, five being "excellent" and one being "very poor". The user study took place in a computer laboratory, with similar conditions to an office. The duration per test was about half an hour. We performed a Wizard of Oz study, see [Nie93], specifically the employed interfaces were functional, however the acquired data was not processed. Hence processing time evaluation is not part of this study, neither the verification accuracy of the presented methods. The employed laptop was a DELL E4310. The documentation of the study contains the notes of questionnaires and related observer notes.



user7

Capture your face by pressing Start/CONFIRM

user7

Place your face in the blue mark and press Start/CONFIRM

user7

Please enter your username

****

and press "Confirm"

Access denied

user7

Processing Fingerprint

Access granted

Figure 8.1: Interfaces of the soft biometrics, face, PIN and fingerprint based access methods.

Procedure : The four access methods, see Figure 8.1, were presented and demonstrated by the observer of the study. Additional information on the methods, as of the differences between the methods, was provided. Subsequently participants of the user study freely explored the available systems. In the next step users were asked to log in with each of the systems. Subsequently an interview about the user experience was conducted. Here users were asked to absolutely rate (from 1 to 5, 5 being excellent) and comment on different aspects (easiness, clarity, comfort and speed of the methods). Then users were confronted with two scenarios, where in the first scenario the user accesses his/ her personal computer; whereas in the second scenario the user acquires the right of entry for a lab in a crowded corridor. The suitability of the methods was enquired for both scenarios. Additionally to the absolute ratings the users were asked to rank the methods in terms of speed, easiness, privacy preservation and overall satisfaction. Finally the users were invited to select freely one method to log in with a task to read a file. This preference was noted as spontaneous and practical preference. In the following we give details on the different access methods.

8.1.2 Access methods

We selected four substantially different accessing methods, in terms of both, interface and technology, namely soft biometrics, face, PIN and fingerprint based access. In consent of recommendations for user friendly human computer interfaces (see [BH90], [CG05] and [Nie93]) we developed four access systems, see Figure 8.1. All four interfaces were designed to be similar in terms of structure and processing time in order to place emphasis on the four interaction elements:

Table 8.1: User experience on access methods.

Soft Biometrics	Face	PIN	Fingerprint
0/15	2/15	15/15	12/15

camera capturing, fingerprint scanning and PIN entry. All four interfaces initially asked for the entry of the username. In the following the methods differed as they acquire the following diverse information:

Soft biometrics, denoted as SB : For the soft biometric based authentication method the user is captured by the laptop integrated webcam and weak biometric facial classifiers including age, gender, hair and skin color are extracted. He / she can unconstrained capture his / her face by pressing a button. An indicator in the related window describes the one-by-one processing of each trait.

Face, denoted as Face : Similar to soft biometrics, in the face based system the face of the user is captured, but he / she is asked to place their face in a pre-defined blue elliptical mark. This constraint is originally designed in order to differentiate the first two access methods. Furthermore it indicates the common pose-constrain of face recognition.

PIN, denoted as PIN : Additionally to the username for the PIN based verification a five-digit password was requested, along with a button confirmation for a successful verification.

Fingerprint, denoted as FP : In this access method a scanner acquired the user fingerprint data of the index finger of the left hand. To initiate the scanning process the user had to press the confirm button. The processing time of all systems, indicated by progressing bars, was designed to be very similar in order not to affect the user preference.

8.2 User study related results

Firstly we inquired previous experience of users on the presented access methods, see Table 8.1. All users employ PIN based verification system daily, on personal computers, ATMs, mobile phones or web pages. This fact can be biasing on the one hand towards PIN, on the other hand it accelerates the awareness of related drawbacks. Surprisingly the majority of user had previous experience with fingerprint based verification systems incorporated in personal laptops or for border control. Soft biometrics was a novel technique to all subjects, whereas face based verification was used before by two users.

Then we proceeded with the questions about usability related measures in context of the different methods.

8.2.1 Ease of access, user friendliness, time to acquire and clarity

The first graph in Figure 8.2 reflects on how intuitive users found the system. All methods provide according to the users an intuitive access, but soft biometrics was significantly best re-

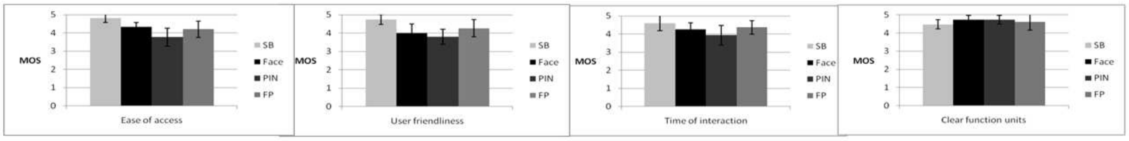


Figure 8.2: User rating of the provided access methods in terms of ease of access, user friendliness, speed and clarity: MOS and standard deviation.

ceived in terms of both, user friendliness and ease of access. We explain this result with the seamless soft biometrics based verification without cumbersome additional interaction (as in all other cases, placing finger on scanner, face in blue mark or entry of a PIN). The majority of female users expressed though concerns regarding the non-permanence of captured soft biometric traits such as hair color. Few users questioned the distinctiveness of a soft biometrics related system. Some users stated that they would prefer an access control system without any contact. The majority stated to be biased by the prevalent use of PIN based systems, but was still convinced such a system bears disadvantages as of forgetting a password or having to keep too many. Few test participants had hygienic concerns related to the fingerprint access system. As of time of acquisition, users were asked to evaluate the time used for operating the access system (different from system processing time). All systems exhibit similar acquisition time ratings. The clarity of the systems was a measure for the feedback a user gets from each system. Here users appreciated the blue mark of the face recognition based system as it helped control the capturing, whereas in regards to PIN they valued the feedback to each step (e.g. the "*" symbols for how many digits they had already have entered).

8.2.2 Access of private computer vs. crowded environment access

In the next step the access methods were associated in two scenarios, personal computer access and crowded environment access, and rated. Confronted with the thought of employing those methods for personal verification, the majority of users expressed immediate concern about the accuracy of the systems. They were asked to disregard for the study this factor. Figure 8.3 illustrates the access method preference in terms of personal computer access. There was no strong user preference; all methods were basically comparably rated. Here one user noted the illumination dependence of the camera based traits and stated not be willing to adjust to that in this scenario.

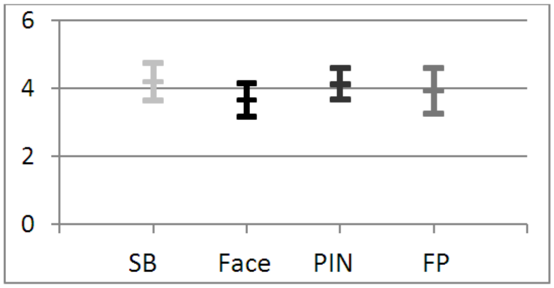


Figure 8.3: User preference of access methods for personal computer use: MOS and confidence intervals for a confidence of 0.95.

In the second scenario users showed a significant preference to use fingerprint in crowded environments. The PIN based verification is last with the reason that user had experienced or were scared of their PIN being spying. Regarding face and soft biometrics, two users reported not to be fully comfortable to be captured by a camera in front of colleagues. It is interesting to compare the two scenarios (see Figure 8.3 with Figure 8.4), where the task was basically the same, given a different environment. The ratings though are partly significantly different (see FP and PIN).

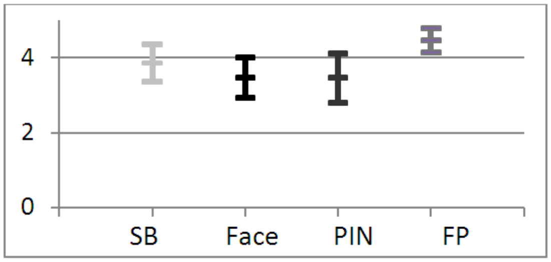


Figure 8.4: User preference of access methods in crowded environment: MOS and confidence intervals for a confidence of 0.95.

8.2.3 Time to Acquire, Ease of Use and Data Security

To evaluate statistical significance of the differences between the access methods, we performed additionally Wilcoxon tests [Sie56], due to the relatively small sample size. Hence we asked users to rank the four access methods (1 to 4, 1 being the number one) in terms of speed, ease to access and data security. For the following graphs, the smaller the bars are the better the access method was ranked.

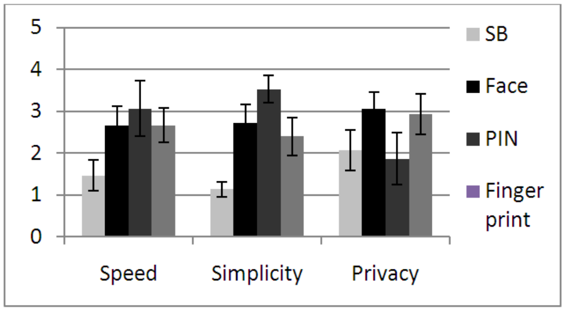


Figure 8.5: User comparison of presented access methods in terms of speed, simplicity and data security: MOS and standard deviation. Small bars represent high rankings.

In terms of acquisition time and simplicity soft biometrics were ranked significantly better than the other methods. In respect to privacy, users were made aware that their data, in particular face image, list of soft features, PIN and fingerprint, would be stored on a database. This data can be on the one hand misused by operators, and on the other hand hacked. Users felt most comfortable with providing both, PIN or a soft biometrics based list for storage on a database. Users were ambiguous about which they found is the riskiest trait to give away, fingerprint or a face image.

8.2.4 Ranking of favorite access methods

The last question regarded the overall satisfaction of the methods. Users named in this context different priorities. On the one hand privacy preservation was named by three users as justification for their ranking, on the other hand the easiness of use. The results on this ranking are displayed in

Figure 8.6. The majority of users, although not strongly familiar with biometrics, emphasized that the ranking as it is, holds only in the case of equal access verification accuracies for all methods.

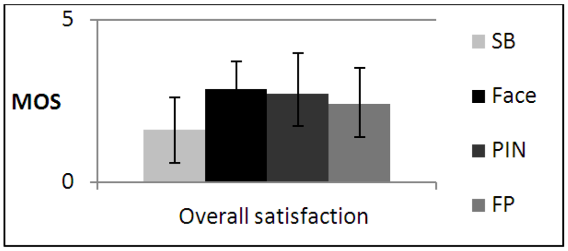


Figure 8.6: Overall user preference of access methods. Small bars represent high rankings.

However, to verify the ranked preference users were also given the final task to freely choose one of the four systems with the goal of reading a file. Figure 8.7 illustrates the ranking ordered by number of selections. The Soft biometrics module was most often chosen and thus the favorite, followed by PIN and fingerprint and finally face based verification (due to similarity with soft biometrics, but longer acquisition due to the blue mark).

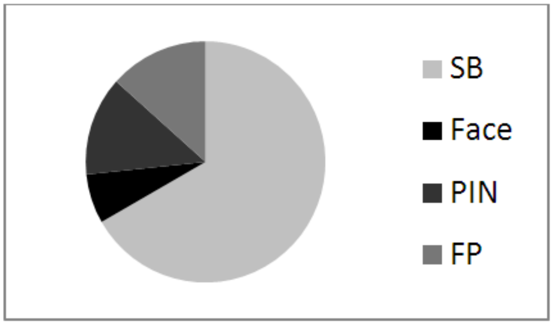


Figure 8.7: Freely selected access methods for performing a task.

To recapitulate the testing, generally users were eager to explore both the known and new access methods. The primary concern of all users was related to reliability and accuracy of the presented methods. Furthermore about half of the users asked about spoofing methods (e.g. holding a photo in front of the camera) and related countermeasures. These reactions are evidence that users are aware of novel techniques and have a need to be illuminated on system characteristic in order to gain the trust of the users for the access methods.

8.3 Comparison of access control systems

Using the above usability study we proceed to perform a broader comparison by including characteristics pertinent to access control systems, such as cost efficiency, accuracy, processing speed and maximum amount of enrolled users. We identified existing commercially available access control systems based on fingerprint, face and PIN based methods, see Table 8.2. We selected cost efficient appliances representing each method and set price and related specifications in comparison. We note that the PIN based system is the most widely available access control system, followed by fingerprint based systems, and only few face recognition based systems. There are no commercially available soft biometrics access control systems yet.

Table 8.2: Comparison of existing access control systems, source:

Trait	Cost	Accuracy	Users	Speed
FP	$\$140$	FRR = 1%, FAR=0.0001%	500	$2s$
Face	$\$320$	>99.9%	100	$< 1s$
PIN	$\$15.99$	100%	500	$2s$

We illustrate usability and presented system performance in terms of cost, accuracy, users and processing speed in Figure 8.8.

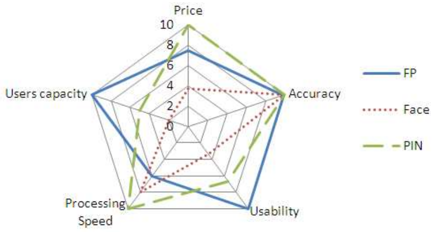


Figure 8.8: Comparison of fingerprint (FP), face recognition and PIN based access control systems.

8.4 Summary

We presented a user study investigating the preference of a set of test participants on access methods, namely soft biometrics, face, PIN and fingerprint based access methods. This preference was evaluated generally in terms of usability measures, such as ease of use, intuitiveness and log-in-speed. Furthermore two scenarios were hereby assessed, specifically personal computer access and entrance of a security lab in a crowded environment. The surprising outcome is that although all users were strongly biased towards the PIN based verification method, by daily using it, the biometric based options were overall equally or even significantly better rated than the PIN based system. Users appreciated the comfort, easiness and speed of modern technology. Specifically they favored the soft biometrics system, due to the provided privacy preservation and ease of use. We demonstrated furthermore a broader comparison of existing access control systems, taking into account the evaluated usability and moreover cost efficiency, processing speed and accuracy.

Conclusions

This dissertation explored the role of soft biometrics in security related applications, as well as in quantifying and predicting female facial aesthetics. Our analysis was accompanied by constructions of practical trait classification algorithms that were tested on existing image databases. We also performed a usability study of systems that employ such soft biometric identifiers.

In terms of security, we focused on three related applications, namely: a) applying SBSs to achieve complete person identification, b) applying SBSs to prune a large database in order to reduce the search space, and c) applying soft biometrics for person re-identification, with a focus on the frontal vs. side scenario.

Applying SBSs to achieve complete person identification

We explored in this context the use of multi-trait SBSs for human identification, studying analytically the identification capabilities of the system, as a function of the authentication group $\boldsymbol{v}$, its size n, the featured categories ρ, and the effective categories F. We showed that in the interference limited setting, for a given randomly chosen authentication group $\boldsymbol{v}$, of a given size n, then the reliability of identification (averaged over the subjects in $\boldsymbol{v}$) is a function only of the number of non-empty categories $F(\boldsymbol{v})$. Then we provided statistical analysis of this reliability, over large populations. The latter part provided bounds that, in the interference limited setting suggest an *exponential* reduction in the probability of interference patterns, as a result of a *linear* increase in ρ.

Applying SBSs to prune a large database in order to reduce the search space

We provided statistical analysis of the gain and reliability in pruning the search over large data sets, where these sets are random and where there is a possibility that the pruning may entail errors. In this setting, pruning plays the role of pre-filtering, similar to techniques such as video indexing. The average-case analysis presented here, described the typical assistance that pruning provides in reducing the search space, whereas large-deviations based analysis provided insight as to how often pruning can behave in an atypically unhelpful, or atypically helpful manner. The analysis may offer insight on better designing pre-filtering algorithms for different search settings. We further studied nine different, actual, soft biometric systems, as well as analyzed and experimented with factors like average error, pruning gain and goodput. Using these factors, we provided a quantifiable comparison of these systems. Furthermore we identified relations between SBS enhancement, error probabilityP_{err}, pruning gain r and goodput $\mathcal{U}$. These findings bring to the fore some SBS design aspects. We also gave insight on the computational cost reduction that can be introduced by SBS-based pruning in the setting of person recognition.

Applying soft biometrics for person re-identification, with a focus on the frontal vs. side scenario

Motivated by realistic surveillance scenarios, the work addressed the problem of frontal-to-side facial recognition, providing re–identification algorithms/classifiers that are specifically suited for this setting. Emphasis was placed on classifiers that belong in the class of soft biometric traits, specifically color–, texture– and intensity– based traits taken from patches of hair, skin and clothes. Towards providing insight, the work presented different identification experiments that adhere to the frontal–to–side setting, as well as presented a preliminary analytical study that seeks to impart intuition on the role of the above traits in improving algorithmic reliability. Our analysis described the overall error probability, both as a function of collisions and of erroneous categorizations for given sizes of authentication groups. In the presence of a moderate reliability of the patches-based method, the analysis suggests promising applications of this method in settings such as pruning of searches.

Applying soft biometrics quantification and prediction of female facial aesthetics

In terms of *female facial aesthetics*, we presented a study on facial aesthetics in photographs, where we compared objective measures (namely photograph quality measures, facial beauty characteristics and non permanent facial features), with human subjective perception. Our analysis revealed a substantial correlation between different selected traits, and the corresponding MOS-related beauty indices. Specifically we presented that non permanent features can influence highly the MOS, and based on our analysis we conclude that facial aesthetics in images can indeed be substantially modifiable. With other words parameters such as the presence of makeup and glasses, the image quality as well as different image post–processing methods can significantly affect the resulting MOS. Furthermore we constructed a linear MOS–based metric which was successfully employed to quantify beauty-index variations due to aging and surgery. Our work applies towards building a basis for designing new image-processing tools that further automate prediction of aesthetics in facial images. Towards this we provided a simulation of an automatic prediction tool based on state-of-art classification algorithms and the designed MOS–prediction metric.

The above approaches were accompanied by a more practically oriented part where we designed an *automatic soft biometrics classification tool*. Specifically we focused on eye, skin and hair color, as well as on the presence of beard, moustache and glasses.

In terms of *usability analysis*, we presented a user study investigating the preference of a set of test participants on access methods, namely soft biometrics, face, PIN and fingerprint based access methods. This preference was evaluated generally in terms of usability measures, such as ease of use, intuitiveness and log-in-speed. Furthermore two scenarios were hereby assessed, specifically personal computer access and entrance of a security lab in a crowded environment. The surprising outcome is that although all users were strongly biased towards the PIN based verification method, by daily use, the biometric based options were overall equally or even significantly better rated than the PIN based system. Users appreciated the comfort, easiness and speed of modern technology. Specifically they favored the soft biometrics system, due to the provided privacy preservation and ease of use.

Future Work

It is becoming apparent that surveillance will increasingly affect our quality of life and security. Research in this area has been embraced by both academia and industry. For this reason, security related biometric systems will become larger and more dynamic. We see the area of soft biometrics having from now on a solid position in such systems. Towards this we will need better understanding of the component parts of such SBSs, and a corresponding better understanding of novel trait classification algorithms, as well as novel ways of combining and analyzing such algorithms. Our aim will be to allow for more efficient SBSs, but also develop a rigorous understanding of the capabilities and limits of such systems.

Our aim in the future will also be, in addition to developing novel algorithms for SBSs, to also identify and develop new commercial applications that can benefit by the power of soft biometrics.

Appendix A

Appendix for Section 3

A.1 Proofs

Proof of Lemma 1 Let $\widehat{\phi}$ denote the estimated category and let $P(S_\phi)$ denote the probability that the chosen subject belongs to category indexed by ϕ, $\phi = 0, 1, \cdots, F$. Then we have

$$\begin{aligned}
P(\text{err}|F) = &\sum_{\phi=0}^{F} P(S_\phi, \widehat{\phi} = \phi) P(\text{err}|S_\phi, \widehat{\phi} = \phi) \\
&+ \sum_{\phi=0}^{F} P(S_\phi, \widehat{\phi} \neq \phi) P(\text{err}|S_\phi, \widehat{\phi} \neq \phi) \\
\overset{(a)}{=}& \sum_{\phi=0}^{F} P(S_\phi) P(\widehat{\phi} = \phi|S_\phi) P(\text{err}|S_\phi, \widehat{\phi} = \phi) \\
&+ \sum_{\phi=0}^{F} P(S_\phi) P(\widehat{\phi} \neq \phi|S_\phi) P(\text{err}|S_\phi, \widehat{\phi} \neq \phi) \\
\overset{(b)}{=}& \frac{N - |S|}{N} + \sum_{\phi=1}^{F} P(S_\phi) P(\widehat{\phi} = \phi|S_\phi) P(\text{err}|S_\phi, \widehat{\phi} = \phi) \\
&+ \sum_{\phi=1}^{F} P(S_\phi) P(\widehat{\phi} \neq \phi|S_\phi) P(\text{err}|S_\phi, \widehat{\phi} \neq \phi). \quad \text{(A.1)}
\end{aligned}$$

Hence

$$\begin{aligned}
&P(\text{err}|F) \\
&\overset{(c)}{=} \frac{N - |S|}{N} + \sum_{\phi=1}^{F} \big(\frac{|S_\phi|}{N} (1 - P_\phi) \frac{|S_\phi| - 1}{|S_\phi|} + \frac{|S_\phi|}{N} P_\phi \big) \\
&= \frac{N - |S|}{N} + \sum_{\phi=1}^{F} \big(|S_\phi| - 1 - P_\phi|S_\phi| + P_\phi + P_\phi|S_\phi| \big) \quad \text{(A.2)}
\end{aligned}$$

which gives

$$\begin{aligned} P(\text{err}|F) &= 1 - \frac{|S|}{N} + \frac{1}{N}\sum_{\phi=1}^{F}\big(|S_\phi| - 1 + P_\phi\big) && \text{(A.3)} \\ &\overset{(d)}{=} 1 - \frac{F - \sum_{\phi=1}^{F} P_\phi}{N}. && \text{(A.4)} \end{aligned}$$

In the above (a) is due to Bayes rule, (b) considers that

$$P(\text{err}|S_0, \widehat{\phi} = 0) = P(\text{err}|S_0, \widehat{\phi} \neq 0) = 1$$

and that

$$\begin{aligned} P(S_0, \widehat{\phi} = 0)P(\text{err}|S_0, \widehat{\phi} = 0) & \\ &+ P(S_0, \widehat{\phi} \neq 0)P(\text{err}|S_0, \widehat{\phi} \neq 0) \\ &= P(S_0, \widehat{\phi} = 0) \cdot 1 + P(S_0, \widehat{\phi} \neq 0) \cdot 1 = P(S_0) = \frac{N - |S|}{N}, \end{aligned}$$

(c) considers that $P(S_\phi) = \frac{|S|}{N}$, that $P(\widehat{\phi} = \phi|S_\phi) = 1 - P_\phi$, that $P(\text{err}|S_\phi, \widehat{\phi} \neq \phi) = 1$, and that

$$P(\text{err}|S_\phi, \widehat{\phi} = \phi) = \frac{|S_\phi| - 1}{|S_\phi|},$$

and finally (d) considers that $\sum_{\phi=1}^{F} |S_\phi| = |S|$.

□

Proof of Lemma 3 Let C_F be the total number of N-tuples $\boldsymbol{v}$ that introduce F effective feature categories. Then

$$C_F = \frac{\rho!}{(\rho - F)!}\frac{N!}{(N - F)!}F^{N-F} \qquad \text{(A.5)}$$

where the first term $\frac{\rho!}{(\rho-F)!}$ describes the total number of ways F categories can be chosen to host subjects, the second term $\frac{N!}{(N-F)!}$ describes the total number of ways F initial people, out of N people, can be chosen to fill these F categories, and where the third term F^{N-F} describes the total number of ways the F effective categories can be freely associated to the rest $N - F$ subjects. Finally we note that

$$P(F) = \frac{C_F}{\sum_{i=1}^{N} C_i},$$

which completes the proof. □

Example 14 *Consider the case where $\rho = 9, N = 5, F = 3$. Then the cardinality of the set of all possible N-tuples that span $F = 3$ effective categories, is given by the product of the following three terms.*

- *The first term is $(\rho \cdot (\rho - 1) \cdots (\rho - F + 1)) = \frac{\rho!}{(\rho-F)!} = 9 \cdot 8 \cdot 7 = 504$ which describes the number of ways one can pick which $F = 3$ categories will be filled.*
- *Having picked these $F = 3$ categories, the second term is $(N \cdot (N - 1) \cdots (N - F + 1)) = \frac{N!}{(N-F)!} = 5 \cdot 4 \cdot 3 = 60$, which describes the number of ways one can place exactly one subject in each of these picked categories.*

- *We are now left with $N-F=2$ subjects, that can be associated freely to any of the $F=3$ specific picked categories. Hence the third term is $F^{N-F}=3^2=9$ corresponding to the cardinality of $\{1,2,\cdots,F\}^{N-F}$.*

Proof of Lemma 5 Recall from (3.22) that

$$P(F)=\frac{F^{N-F}}{(\rho-F)!(N-F)!\sum_{i=1}^{N} i^{N-i}\big((N-i)!(\rho-i)!\big)^{-1}}, \tag{A.6}$$

and note that

$$\sum\nolimits_{i=1}^{N} i^{N-i}\big((N-i)!(\rho-i)!\big)^{-1}\geq(\rho-N)!$$

corresponding to the Nth summand $(i=N)$, and corresponding to the fact that all summands are non-negative. As a result

$$P(F)\leq\frac{F^{N-F}}{(\rho-F)!(N-F)!(\rho-N)!}.$$

Using Stirling's approximation [AS02] that holds in the asymptotically high ρ setting of interest, we have

$$P(F)\dot{\leq}\frac{F^{N-F}}{(\rho-F)^{\rho-F}(N-F)^{N-F}(\rho-N)^{\rho-N}e^{-(2\rho-2F)}}, \quad \text{(A.7)}$$

and as a result

$$\begin{aligned}
&P(f)\\
&\quad\dot{\leq}\frac{(fr\rho)^{r\rho(1-f)}}{(\rho-fr\rho)^{\rho-fr\rho}(r\rho-fr\rho)^{r\rho-fr\rho}(\rho-r\rho)^{\rho-r\rho}e^{2\rho(1+fr)}}\\
&\quad=\frac{\rho^{r\rho(1-f)}\rho^{-\rho(1-fr)}}{(fr)^{-r\rho(1-f)}(1-fr)^{\rho(1-fr)}}\\
&\qquad\cdot\frac{\rho^{-\rho r(1-f)}\rho^{-\rho(1-r)}}{(r-fr)^{\rho r(1-f)}(1-r)^{\rho(1-r)}e^{2\rho(1+fr)}}.
\end{aligned} \tag{A.8}$$

In the above we use $\doteq$ to denote *exponential equality*, where

$$f\doteq\rho^{-\rho B}\iff-\lim_{\rho\to\infty}\frac{\log f}{\rho\log\rho}=B, \tag{A.9}$$

with $\dot{\leq},\dot{\geq}$ being similarly defined. The result immediately follows.

Appendix B

Appendix to Section 4

B.1 Proofs

Proof of Lemma 6: We first note that

$$P(\boldsymbol{\alpha}_0) \doteq e^{-nD(\boldsymbol{\alpha}_0/\rho||\boldsymbol{p})} = e^{-\frac{n}{\rho}D(\boldsymbol{\alpha}_0||\rho\boldsymbol{p})} \tag{B.1}$$

where as previously stated $D(\boldsymbol{\alpha}_0||\boldsymbol{p}) = \sum_f \alpha_{0,f} \log \frac{\alpha_{0,f}}{p_f}$ is the information divergence (also called the Kullback-Leibler distance) between $\boldsymbol{\alpha}_0$ and $\boldsymbol{p}$. We use $\doteq$ to denote exponential equality, i.e., we write $f(n) \doteq e^{-nd}$ to denote $\lim\limits_{n\to\infty} \dfrac{\log f(n)}{n} = d$ and $\dot{\leq}$, $\dot{\geq}$ are similarly defined. In establishing $P(\boldsymbol{\alpha}_1|\boldsymbol{\alpha}_0)$, we focus on a specific category f, and look to calculate

$$P\left(|\mathcal{S} \cap C_f| = \frac{n}{\rho}\alpha_{1,f} \quad | \quad |C_f| = \frac{n}{\rho}\alpha_{0,f}\right), \tag{B.2}$$

i.e., to calculate the probability that pruning introduces $\frac{n}{\rho}\alpha_{1,f}$ elements, from C_f to $\mathcal{S}$, given that there are $\frac{n}{\rho}\alpha_{0,f}$ elements of C_f. Towards this we note that there is a total of

$$|C_f| = \frac{n}{\rho}\alpha_{0,f} \tag{B.3}$$

possible elements in C_f which may be categorized, each with probability ϵ_f, to belong to C_1 by the categorization algorithm. The fraction of such elements that are asked to be categorized to belong to C_1, is defined by $\boldsymbol{\alpha}$ to be

$$x_f := \frac{|\mathcal{S} \cap C_f|}{|C_f|} = \frac{\frac{n}{\rho}\alpha_{1,f}}{|C_f|} = \frac{\alpha_{1,f}}{\alpha_{0,f}}, \tag{B.4}$$

an event which happens with probability

$$\begin{aligned} P(x_f) = P\left(|\mathcal{S} \cap C_f| = \frac{n}{\rho}\alpha_{1,f} \quad | \quad |C_f| = \frac{n}{\rho}\alpha_{0,f}\right) \\ \doteq e^{-|C_f| I_f(x_f)}, \end{aligned} \tag{B.5}$$

where in the above, $I_f(x_f) = x_f \log(\frac{x_f}{\epsilon_f}) + (1-x_f)\log(\frac{1-x_f}{1-\epsilon_f})$ is the rate function of the binomial distribution with parameter ϵ_f (cf. [CT06]). Now given that

$$P(\boldsymbol{\alpha}_1|\boldsymbol{\alpha}_0) = \prod_{f=1}^{\rho} P\left(|\mathcal{S} \cap C_f| = \frac{n}{\rho}\alpha_{1,f} \quad | \quad |C_f| = \frac{n}{\rho}\alpha_{0,f}\right) \tag{B.6}$$

we conclude that

$$-\lim_{N\to\infty}\frac{\log}{n/\rho}\log P(\boldsymbol{\alpha}_1|\boldsymbol{\alpha}_0)=\sum_f \alpha_{0,f} I_f(\frac{\alpha_{1,f}}{\alpha_{0,f}}). \tag{B.7}$$

Finally given that $P(\boldsymbol{\alpha},\tau)=P(\boldsymbol{\alpha}_0)P(\boldsymbol{\alpha}_1|\boldsymbol{\alpha}_0)$, we conclude that $-\lim_{N\to\infty}\frac{\log}{n/\rho}\log P(\boldsymbol{\alpha},\tau)=D(\boldsymbol{\alpha}_0||\rho\boldsymbol{p})+\sum_f \alpha_{0,f} I_f(\frac{\alpha_{1,f}}{\alpha_{0,f}})$.

□

Proof of Theorem 2: The proof is direct from the *method of types* (cf. [CT06]), which applies after noting that $|\mathcal{V}(\tau)| \leq n^{2\rho} \dot{\leq} e^{n\delta}\ \forall\delta > 0$, and that $\sup_{\boldsymbol{\alpha}\in\mathcal{V}(\tau)} P(\boldsymbol{\alpha}) \leq P(\tau) \leq |\mathcal{V}(\tau)| \sup_{\boldsymbol{\alpha}\in\mathcal{V}(\tau)} P(\boldsymbol{\alpha})$.

□

Proof of Theorem 3: The proof is direct by noting that for any $\delta > 0$, then for $\tau \geq \tau_0$ we have

$$-\lim_{N\to\infty}\frac{\log}{n/\rho}P(|\mathcal{S}| > (\tau+\delta)\frac{n}{\rho}) > -\lim_{N\to\infty}\frac{\log}{n/\rho}P(|\mathcal{S}| > \tau\frac{n}{\rho}), \tag{B.8}$$

and similarly for $\tau < \tau_0$ we have

$$-\lim_{N\to\infty}\frac{\log}{n/\rho}P(|\mathcal{S}| < (\tau-\delta)\frac{n}{\rho}) > -\lim_{N\to\infty}\frac{\log}{n/\rho}P(|\mathcal{S}| < \tau\frac{n}{\rho}). \tag{B.9}$$

□

B.2 Confusion matrices and population characteristics for proposed systems

	Light eyes	Dark eyes
Light eyes	0.9266	0.0734
Dark eyes	0.0759	0.9238
p_C	0.3762	0.6238

Table B.1: Confusion matrix related to ‘2e’: 2 eye color categories

	No moustache	Moustache
No moustache	0.8730	0.1270
Moustache	0.2720	0.7280
p_C	0.7340	0.1623

Table B.2: Confusion matrix related to ‘m’: moustache detection

	No glasses	Glasses
No glasses	0.9283	0.0717
Glasses	0.0566	0.9434
p_C	0.0849	0.0188

Table B.3: Confusion matrix related to ‘g’: glasses detection

	Blue	Green	Brown	Black
Blue	0.75	0.1875	0	0.0625
Green	0.0714	0.7858	0.1428	0
Brown	0.0526	0	0.8948	0.0526
Black	0.0455	0.0455	0	0.909
p_C	0.3251	0.0511	0.2399	0.3839

Table B.4: Confusion matrix and population statistics related to ‘4e’: 4 eye color categories

	No moustache	No moustache	Moustache	No moustache
	No glasses	Glasses	No glasses	Glasses
	00	01	10	11
00	0.8104	0.0626	0.1179	0.0091
01	0.0494	0.8236	0.0072	0.1198
10	0.2525	0.0195	0.6758	0.0522
11	0.0154	0.2566	0.0412	0.6868
p_C	0.7340	0.1623	0.0849	0.0188

Table B.5: Confusion matrix and population statistics related to ‘mg’: moustache - glasses system

	light eyes	light eyes	dark eyes	dark eyes
	no moustache	moustache	no moustache	moustache
	00	01	10	11
00	0.8089	0.1177	0.0641	0.0093
01	0.2520	0.6746	0.0200	0.0534
10	0.0663	0.0096	0.8065	0.1173
11	0.0207	0.0553	0.2513	0.6725
p_C	0.3371	0.0390	0.5591	0.0647

Table B.6: Confusion matrix and population statistics related to ‘2em’: 2 eye color and moustache classification

	light eyes	light eyes	dark eyes	dark eyes
	no glasses	glasses	no glasses	glasses
	00	01	10	11
00	0.8602	0.0664	0.0681	0.0053
01	0.0524	0.8741	0.0042	0.0693
10	0.0705	0.0054	0.8575	0.0662
11	0.0043	0.0716	0.0523	0.8715
p_C	0.3080	0.0681	0.5109	0.1130

Table B.7: Confusion matrix and population statistics related to ‘2eg’: 2 eye color and glasses classification

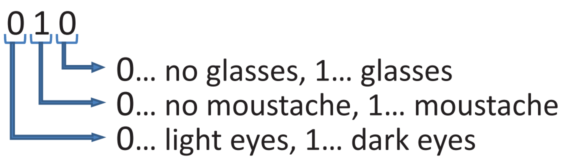

0 1 0
0… no glasses, 1… glasses
0… no moustache, 1… moustache
0… light eyes, 1… dark eyes

Figure B.1: Category nomenclature for ‘2emg’

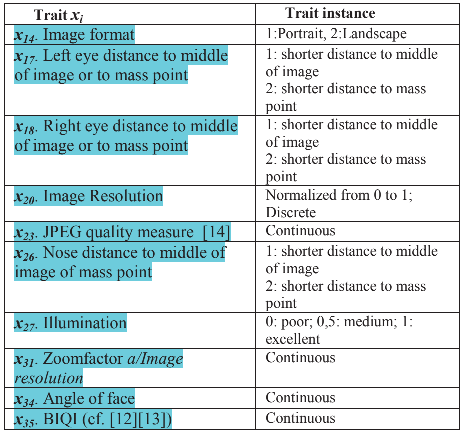
	000	001	010	011	100	101	110	111
000	0.751	0.058	0.109	0.008	0.059	0.005	0.009	0.001
001	0.046	0.763	0.007	0.111	0.004	0.060	0.001	0.009
010	0.234	0.018	0.626	0.048	0.018	0.001	0.050	0.004
011	0.014	0.238	0.038	0.636	0.001	0.019	0.003	0.050
100	0.061	0.005	0.009	0.001	0.749	0.058	0.109	0.008
101	0.004	0.062	0.001	0.009	0.046	0.761	0.007	0.111
110	0.019	0.001	0.051	0.004	0.233	0.018	0.624	0.048
111	0.001	0.019	0.003	0.052	0.014	0.237	0.038	0.634
p_C	0.276	0.061	0.032	0.007	0.458	0.101	0.053	0.012

Table B.8: Confusion matrix and population statistics related to '2emg': 2 eye color, moustache and glasses classification; see Figure B.1 for the nomenclature

Appendix C

Appendix for Section 6

Table C.1: Photograph aesthetic traits, according trait instances and annotations



Trait x_i	**Trait instance**
x_{14}. Image format	1:Portrait, 2:Landscape
x_{17}. Left eye distance to middle of image or to mass point	1: shorter distance to middle of image 2: shorter distance to mass point
x_{18}. Right eye distance to middle of image or to mass point	1: shorter distance to middle of image 2: shorter distance to mass point
x_{20}. Image Resolution	Normalized from 0 to 1; Discrete
x_{23}. JPEG quality measure [14]	Continuous
x_{26}. Nose distance to middle of image of mass point	1: shorter distance to middle of image 2: shorter distance to mass point
x_{27}. Illumination	0: poor; 0,5: medium; 1: excellent
x_{31}. Zoomfactor *a/Image resolution*	Continuous
x_{34}. Angle of face	Continuous
x_{35}. BIQI (cf. [12][13])	Continuous

Table C.2: Facial aesthetic traits, according trait instances and annotations

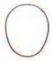
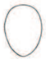
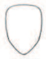
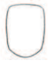
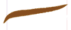
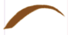
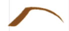
Trait x_i	**Trait instance**
x_1. Ratio (Eye height / head length) **f/a**	Continuous
x_2. Ratio (Head width / Head length) **b/a**	Continuous
x_3. Eye make up	0:No make up, 0.5: light make-up, 1:strong make-up
x_4. Face shape	1:  2:  3:  4:  5:2&3, 6:1&3, 7:1&2
x_5. Eye Brow shape	1:  2:  3: 
x_6. Fullness of Lips	0:Thin lips, 0.5:medium, 1:full lips
x_7. Ratio (from top of head to nose)/head length **(d+c)/a**	Continuous
x_8. Glasses	0:No glasses, 1:glasses
x_9. Lipstick	0:No lipstick, 1:bright lipstick, 2:flashy lipstick
x_{10}. Skin goodness	1:Clear skin, 2:not clear skin (pimples)
x_{11}. Hair Length / Style	1:Short, 2:shoulder, 3:long, 4:half tied back, 5:tied back
x_{12}. Ratio (from top of head to mouth)/head length **(d+c+e)/a**	Continuous
x_{13}. Ratio (from top of head to eye/head length) **d/a**	Continuous
x_{15}. Ratio (eye width / distance between eyes) **(h-i)/(2.i)**	Continuous
x_{16}. Ratio (from nose to chin / eye to nose) **(a-d-c)/c**	Continuous
x_{19}. Ratio (from top of head eye / eye to nose) **d/c**	Continuous
x_{21}. Expression	1:Smile + teeth, 2:smile, 3:neutral, 4:corner of the mouth facing down, 5:non of all
x_{22}. Ratio (outside distance between eyes/ top of the head to eye) **h/d**	Continuous
x_{24}. Eyes symmetry	0.93<(left eye width)/(right eye width) <1.06
x_{25}. Ratio (from eye to nose / nose to mouth) **c/e**	Continuous
x_{28}. Skin Color	1, 2, 3 (from light to dark)
x_{29}. Ratio (from top of head to eye / eye to lip) **d/(c+e)**	Continuous
x_{30}. Ratio (eye-nose/head width) **c/b**	Continuous
x_{32}. Eye Color	1:blue, 2:green, 3:brown, 4:black, 5:mix
x_{33}. Hair Color	1:blond, 2:brown, 3:black, 4:red, 5:dark blond
x_{36}. Ratio (from nose to chin / lips to chin) **(a-d-c)/(a-d-c-e)**	Continuous
x_{37}. Ratio (Distance eyes/ head length) **g/a**	Continuous

Table C.3: Correlation matrix of selected non permanent and permanent traits, see Table 1 for notations of x_i

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1	0.317	0.308	0.153	0.151	0.161
x_2	0.317	1	0.132	0.268	0.034	0.092
x_3	0.308	0.132	1	0.140	0.158	0.108
x_4	0.153	0.268	0.140	1	-0.0036	0.122
x_5	0.151	0.034	0.158	-0.0036	1	0.155
x_6	0.092	0.092	0.108	0.122	0.155	1

Appendix D

Publications

The featured list spans over all published and to be published documents of the author. None of these publications appear in the Bibliography.

Journals

A. Dantcheva and C. Velardo and A. D'Angelo and J.-L. Dugelay, "Bag of soft biometrics for person identification. New trends and challenges," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 739 - 777, 2011.

A. Dantcheva and J.-L. Dugelay, "Perception of Female Facial Beauty based on Anthropometric, Non Permanent and Acquisition Characteristics," to be submitted.

A. Dantcheva, P. Elia and J. L. Dugelay, "Human-like person re–identification using soft biometrics," to be submitted.

Conference Papers

A. Dantcheva, J.-L. Dugelay, and P. Elia, "Person recognition using a bag of facial soft biometrics (BoFSB),"*in Proc. of IEEE MMSP*, 2010.

A. Dantcheva and J.-L. Dugelay and P. Elia, "Soft biometric systems: reliability and asymptotic bounds," *in Proc. of BTAS*, 2010.

A. Dantcheva and N. Erdogmus and J.-L. Dugelay, "On the reliability of eye color as a soft biometric trait," *in Proc. of WACV*, 2011.

A. Dantcheva and J.-L. Dugelay, "Female facial aesthetics based on soft biometrics and photo-quality," *in Proc. of ICME*, 2011.

A. Dantcheva and J.-L. Dugelay, "Frontal-to-side face re–identification based on hair, skin and cloths patches," *in Proc. of AVSS*, 2011.

A. Dantcheva, A. Singh, P. Elia, J. L. Dugelay, "Search pruning video surveillance systems: Efficiency-reliability tradeoff," *in Proc. of ICCV Workshop IWITINCVPR, 1st IEEE Workshop on Information Theory in Computer Vision and Pattern Recognition in the International Conference on Computer Vision*, 2011.

A. Dantcheva, P. Elia and J. -L. Dugelay, "Gain, reliability and complexity measures in biometric search pruning based on soft biometric categorization," submitted to ICME 2011.

A. Dantcheva, J. -L. Dugelay, "User Acceptance of Access Control based on Fingerprint, PIN, Soft Biometrics and Face Recognition," submitted to ICB 2011.

M. Ouaret, A. Dantcheva, R. Min, L. Daniel, J. -L. Dugelay, “BIOFACE, a biometric face demonstrator,” ACMMM 2010, ACM Multimedia 2010, October 25-29, 2010, Firenze, Italy , pp 1613-1616.

Book Chapter

C. Velardo, J. -L. Dugelay, L. Daniel, A. Dantcheva, N. Erdogmus, N. Kose, R. Min, X. Zhao, “Introduction to biometry Book chapter of "Multimedia Image and Video Processing"” (2nd edition); CRC Press; 2011

Bibliography

[AAR04] S. Agarwal, A. Awan, and D. Roth. Learning to detect objects in images via a sparse, part-based representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(11):1475–1490, 2004.

[ACPR10] D. Adjeroh, D. Cao, M. Piccirilli, and A. Ross. Predictability and correlation in human metrology. In *Proceedings of the WIFS*, 2010.

[ACT11] European project actibio, 2011.

[AHME01] P. Aarabi, D. Hughes, K. Mohajer, and M. Emami. The automatic measurement of facial beauty. In *Proc. of IEEE SMC*, 2001.

[ALMV04] H. Ailisto, M. Lindholm, S.-M. Mäkelä, and E. Vildjiounaite. Unobtrusive user identification with light biometrics. In *Proceedings of NordiCHI*, 2004.

[AM00] S. E. Ahmed and R. J. Mcintosh. An asymptotic approximation for the birthday problem. *Crux Mathematicorum*, 26:151–155, 2000.

[AS02] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover Publications, New York, USA, 2002.

[bea11] Website on beautycheck, 2011.

[BH90] H. Beyer and K. Holtzblatt. *Contextual Design: Defining Customer-Centered Systems, Morgan Kaufmann Publishers*. Academic Press, San Diego, USA, 1990.

[BL10] A. Bottino and A. Laurentini. The analysis of facial beauty: an ermerging area of research in pattern analysis. *Lecture Notes in Computer Science, Image Analysis and Recognition*, 6111, 2010.

[BMC$^+$97] L. Z. Bito, A. Matheny, K. J. Cruickshanks, D. M. Nondahl, and O. B. Carino. Eye color changes past early childhood. The Louisville Twin Study. *Arch Ophthalmol.*, 115:659–663, 1997.

[BR06] S. Baluja and H.A. Rowley. Boosting sex identification performance. *International Journal of Computer Vision*, 71:111–119, 2006.

[BRM$^+$06] C. Boyce, A. Ross, M. Monaco, L. Hornak, and X. Li. Multispectral iris analysis: A preliminary study. In *Proceedings of CVPRW*, 2006.

[BSS10] S. Bhattacharya, R. Sukthankar, and M. Shah. A framework for photo-quality assessment and enhancement based on visual aesthetics. In *Procedings of ACM MM*, 2010.

[CAJ03] L. Coventry, A. De Angeli, and G. Johnson. Usability and biometric verification at the atm interface. In *Proceedings of ACM CHI*, 2003.

[CBNT10] B. Cheng, S. Yan B. Ni, and Q. Tian. Learning to photograph. In *Proceedings of ACM MM*, 2010.

[CC06] J. Carnicky and J.D. Chorvat. Three-dimensional measurement of human face with structured-light illumination. *Measurement science review*, 6:1–4, 2006.

[CCP$^+$11] D. Cao, C. Chen, M. Piccirilli, D. Adjeroh, T. Bourlai, and A. Ross. Can facial metrology predict gender? In *Proceedings of IJCB*, 2011.

[CG05] L. F. Cranor and S. Garfinkel. *Security and usability*. O'Reilly Media, Inc., 2005.

[CJMR09] L. Coventry, G. Johnson, T. McEwan, and C. Riley. Biometrics in practice: What does hci have to say? In *Proceedings of INTERACT*, pages 920–921, 2009.

[CO11] N. Chhaya and T. Oates. Integrating soft biometrics to extract text descriptors from triage images in mass disaster situations. In *Proceedings of HST*, 2011.

[CR11a] C. Chen and A. Ross. Evaluation of gender classification methods on thermal and near-infrared face images. In *Proceedings of IJCB*, 2011.

[CR11b] S. Crihalmeanu and A. Ross. On the use of multispectral conjunctival vasculature as a soft biometric. In *Proceedings of WACV*, 2011.

[CSM07] S. Caifeng, G. Shaogang, and P.W. McOwand. Learning gender from human gaits and faces. In *Proceedings of AVSS*, pages 505–510, 2007.

[CT06] T. M. Cover and J. A. Thomas. Elements of Information Theory. *2nd edition, ISBN: 0-471-24195-4, Wiley*, 2006.

[cul] Website wired science.

[Das05] A. DasGupta. The matching, birthday and the strong birthday problem: A contemporary review. *Journal of Statistical Planning and Inference*, 130(1–2):377–389, 2005.

[DFBS09] S. Denman, C. Fookes, A. Bialkowski, and S. Sridharan. Soft-biometrics: Unconstrained authentication in a surveillance environment. In *Proceedings of DICTA*, pages 196–203, 2009.

[DM08] L. Ding and A. M. Martinez. Precise detailed detection of faces and facial features. In *Proceedings of CVPR*, 2008.

[Doc05] G. Doczi. *The Power of Limits: Proportional harmonies in nature, art, and architecture*. Boston: Shambhala Publications, 2005.

[DYC06] H.-C. Do, J.-Y. You, and S.-I. Chien. Skin color detection through estimation and conversion of illuminant color using sclera region of eye under varying illumination. In *Proceedings of ICPR*, 2006.

[ED10] N. Erdogmus and J.-L. Dugelay. An efficient iris and eye corners extraction method. In *Proceedings of SSPR and SPR*, 2010.

[FCB08] L. Franssen, J. E. Coppenhs, and T. J. T. P. Van Den Berg. Grading of iris color with an extended photographic reference set. *Journal Optom*, 1:36–40, 2008.

[FDH03] S. Fan, C. R. Dyer, and L. Hubbard. Quantification and correction of iris color. *Technical Report 1495*, 2003.

[FDL$^+$10] C. Fookes, S. Denman, R. Lakemond, D. Ryan, S. Sridharan, and M. Piccardi. Semi-supervised intelligend surveillance system for secure environments. In *Proceedings of IEEE ISIE*, 2010.

[Fer11] NIST database FERET, 2011.

[fgn11] Web site of the database fgnet, 2011.

[FHT98] J. H. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. In *Technical Report*, pages 76670P–76670P–12, 1998.

[GBDB97] G. Givens, J. R. Beveridge, B.A. Draper, and D. Bolme. A statistical assessment of subject factors, 1997.

[GHJW00] S. Gutta, J.R.J. Huang, P. Jonathon, and H. Wechsler. Mixture of experts for classification of gender, ethnic origin, and pose of human faces. *Transactions of Neural Networks*, 11:948–960, 2000.

[GKYG10] D. Gray, W. Xu K. Yu, and Y. Gong. Predicting facial beauty without landmarks. In *Proceedings of ECCV*, 2010.

[GLW$^+$11] J. M. Guo, C. C. Lin, M. F. Wu, C. H. Chang, and H. Lee. Complexity reduced face detection using probability-based face mask prefiltering and pixel-based hierarchical-feature Adaboosting. *IEEE Signal Processing Letters*, 2011.

[GPJ04] H. Gunes, M. Piccardi, and T. Jan. Comparative beauty classification for pre-surgery planning. In *Proceedings of International Conference on Systems, Man and Cybernetics*, 2004.

[GPW98] S. Gutta, J. Phillips, and H. Wechsler. Gender and ethnic classification of face images. In *Proceedings of FG*, pages 194–199, 1998.

[GW99] S. Gutta and H. Wechsler. Gender and ethnic classification of human faces using hybrid classifiers. In *Proceedings of IJCNN*, volume 6, pages 4084–4089, 1999.

[hai10] Hair and eye colors web site, 2010.

[HJP99] L. Hong, A. K. Jain, and S. Pankanti. Can multibiometrics improve performance? In *Proceedings of AutoID*, 1999.

[HKAA04] J. Heo, S.G. Kong, B.R. Abidi, and M.A. Abidi. Fusion of visual and thermal signatures with eyeglass removal for robust face recognition. In *Proceedings of CVPRW*, page 122, 2004.

[Hot11] Dating and rating website HOTorNOT, 2011.

[HP09] A. Hadid and M. Pietikäinen. Combining appearance and motion for face and gender recognition from videos. *Pattern Recognition*, 42(11):2818–2827, 2009.

[HPM02] A. Hadid, M. Pietikaeinen, and B. Martinkauppi. Color-based face detection u sing skin locus model and hierarchical filtering. In *Proceedings of ICPR*, pages 196–200, 2002.

[HTK04] S. Hosoi, E. Takikawa, and M. Kawade. Ethnicity estimations with facial images. In *Proceedings of FG*, pages 195–200, 2004.

[IHS05] R. Ishiyama, M. Hamanaka, and S. Sakamoto. An appearance model constructed on 3-D surface for robust face recognition against pose and illumination variations. *IEEE Transactions System, Man, Cybernetics*, 3(35):326–334, 2005.

[JBAB00] X. Jiang, M. Binkert, B. Achermann, and H. Bunke. Towards detection of glasses in facial images. *Pattern Analysis and Applications*, 3:9–18, 2000.

[JDN04a] A.K. Jain, S. C. Dass, and K. Nandakumar. Can soft biometric traits assist user recognition? In *Proceedings of SPIE*, volume 5404, pages 561–572, 2004.

[JDN04b] A.K. Jain, S.C. Dass, and K. Nandakumar. Soft biometric traits for personal recognition systems. In *Proceedings of ICBA*, 2004.

[JDP92] K. Joag-Dev and F. Proschan. Birthday problem with unlike probabilities. *American Mathematical Monthly*, 99(1):10–12, 1992.

[JKP11] A. K. Jain, B. Klare, and U. Park. Face recognition: Some challenges in forensics. In *Proceedings of IEEE FG*, pages 726–733, 2011.

[Joa98] T. Joachims. Text categorization with support vector machines: Learning with many relevant features. In *Proceedings of ECML*, pages 137–142. Springer, 1998.

[JP09] A. K. Jain and U. Park. Facial marks: Soft biometric for face recognition. In *Proceedings of ICIP*, volume 1, pages 37–40, 2009.

[KBBN09] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *Proceedings of IEEE ICCV*, 2009.

[KBN08] N. Kumar, P. N. Belhumeur, and S. K. Nayar. Facetracer: a search engine for large collections of images with faces. In *Proceedings of ECCV*, 2008.

[KED11] E. P. Kukula, S. J. Elliott, and V. G. Duffy. The Effects of Human Interaction on Biometric System Performance. *Lecture Notes in Computer Science*, 4561:904–914, 2011.

[KM06] N. Kaushik and A. Mittal. A novel beard removal method based on structural similarity and coordinate transformations. In *Proceedings of INDICON*, pages 270.1–270.6, 2006.

[KMB] P. Kakumanua, S. Makrogiannis, and N. Bourbakis. A survey of skin-color modeling and detection methods. In *Proceedings of ICPR*, volume 40.

[LBCK03] L. Little, P. Briggs, L. Coventry, and D.J. Knight. *Attitudes Towards Technology Use in Public Areas: The Influence of External Factors on ATM use*, volume 2. Lawrence Erlbaum Associates: NJ, 2003.

[Ley96] M. Leyton. *The architecture of complexity: Hierarchic systems, Symmetry, Causality, Mind*. Cambridge, MA: MIT Press, 1996.

[LJ04] X. Lu and A. K. Jain. Ethnicity identification from face images. In *Proceedings of SPIE*, volume 5404, pages 114–123, 2004.

[LJJ08] J.-E. Lee, A. K. Jain, and R. Jin. Scars, marks and tattoos (SMT): Soft biometric for suspect and victim identification. In *Proceedings of BCC*, 2008.

[LLA$^+$08] L. Lee, G. Loewenstein, D. Ariely, J. Hong, and J. Young. If i'm not hot, are you hot or not? *Psychological science*, 2008.

[LLZ06] H. Lin, H. Lu, and L. Zhang. A new automatic recogntion system of gender, age and ethnicity. In *Proceedings of WCICA*, volume 2, pages 9988–9991, 2006.

[LR90] J. H. Langlois and L. A. Roggman. Attractive faces are only average. *Psychological Science*, 1:115–121, 1990.

[LSP06] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proceedings of ICPR*, volume 2, 2006.

[MB09a] A. K. Moorthy and A. C. Bovik. BIQI software release. In *http://live.ece.utexas.edu/research/quality/biqi.zip*, 2009.

[MB09b] A. K. Moorthy and A. C. Bovik. A modular framework for constructing blind universal quality indices. *IEEE Signal Processing Letters*, 2009.

[McK06] S. McKeen. A beauty fix plumps up psyche and overall health. *The Edmonton Journal*, 2006.

[MD11] R. Min and J. L. Dugelay. Cap detection for moving people in entrance surveillance. In *Proceedings of ACM MM*, 11 2011.

[MHD11] R. Min, A. Hadid, and J. L. Dugelay. Improving the recognition of faces occluded by facial accessories. In *Proceedings of IEEE FG*, 2011.

[MJD09] H. Mao, L. Jin, and M. Du. Automatic classification of chinese female facial beauty using support vector machine. In *Proceedings of IEEE SMC*, 2009.

[MKS10] D. Meltem, G. Kshitiz, and G. Sadiye. Automated person categorization for video surveillance using soft biometrics. In *Proceedings of SPIE*, pages pp. 76670P–76670P–12, 2010.

[MOO10] A. Moorthy, P. Obrador, and N. Oliver. Towards computational models of the visual aesthetic appeal of consumer videos. In *Proceedings of ECCV*, pages 6315: 1–14, 2010.

[MRGL00] M. Melgosa, M. J. Rivas, L. Gomez, and E. Litag. Towards a colorimetric characterization of the human iris. *Ophthal. Physiol. Opt.*, 20:252–260, 2000.

[MSMD08] F. Matta, U. Saeed, C. Mallauran, and J.-L. Dugelay. Facial gender recognition using multiple sources of visual information. In *Proceedings of MMSP*, pages 785–790, 2008.

[MSN03] B. Malin, L. Sweeney, and E. Newton. Trail re-identification: learning who you are from where you have been. In *Data privacy laboratory technical report*, pages LIDAP–WP12, 2003.

[NBS$^+$08] A. Nkengne, C. Bertin, G.N. Stamatas, A. Giron, A. Issachar N. Rossi, and B. Fertil. Influence of facial skin attributes on the perceived age of Caucasian women. *Journal of the European Academy of Dermatology and Venerology*, 22:982–991, 2008.

[New95] E. Newham. The biometric report, 1995.

[Nie93] J. Nielsen. *Usability Engineering*. Morgan Kaufmann, San Francisco, 1993.

[Nix85] M. Nixon. Eye spacing measurement for facial recognition. In *Proceedings of SPIE*, volume 575, pages 279–285, 1985.

[NJT06] E. Nowak, F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. In *Proceedings of ECCV*, 2006.

[NPJ10] K. Niinuma, U. Park, and A. K. Jain. Soft Biometric Traits for Continuous User Authentication. *IEEE Transactions on Information Forensics and Security*, 5(4):771–780, 2010.

[OPD94] A. O'Toole, A. Peterson, and K. Deffenbacher. Structural aspects of face recognition and the other race effect. *Memory and Cognition*, 22:208–224, 1994.

[OSHO10] P. Obrador, L. Schmidt-Hackenberg, and N. Oliver. The role of image composition in image aesthetics. In *Proceedings of ICIP*, pages 3185–3188, 2010.

[PEWF08] S.J.D. Prince, J.H. Elder, J. Warrell, and F.M. Felisberti. Tied factor analysis for face recognition across large pose differences. *IEEE Transactions on pattern analysis and machine intelligence*, 30(6):970–984, 2008.

[PFS$^+$09] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre. The UBIRIS.v2: A Database of Visible Wavelength Iris Images Captured On-The-Move and At-A-Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.

[PJ10] U. Park and A. K. Jain. Face Matching and Retrieval Using Soft Biometrics. *IEEE Transactions on Information Forensics and Security*, 5(3):406–415, 2010.

[PS08] N. B. Puhan and N Sudha. A novel iris database indexing method using the iris color. In *Proceedings of ICIEA*, pages 1886–1891, 2008.

[PSA$^+$07] E. Patterson, A. Sethuram, M. Alber, K. Ricanek, and M. King. Aspects of age variation in facial morphology affecting biometrics. In *Proceedings of BTAS*, 2007.

[Rho56] H.T.F. Rhodes. Alphonse bertillon: Father of scientific detection. *Pattern Recognition Letters*, 1956.

[RJMAS09] C. Riley, G. Johnson, H. McCracken, and A. Al-Saffar. Instruction, feedback and biometrics: The user interface for fingerprint authentication systems. In *Proceedings of INTERACT*, pages 293–305, 2009.

[RN10] D. Reid and M. Nixon. Imputing human descriptions in semantic biometrics. In *Proceedings of ACM MM, Workshop on Multimedia in Forensics, Security and Intelligence*, 2010.

[RN11] D. Reid and M. Nixon. Using comparative human descriptions for soft biometrics. In *Proceedings of IJCB*, 2011.

[RNS11] D. Reid, M. Nixon, and S. Stevenage. Identifying humans using comparative descriptions. In *Proceedings of ICDP*, 2011.

[SB91] M. J. Swain and D. H. Ballard. Color Indexing. *Inter. Journal of Computer Vision*, 7(1):11–32, 1991.

[SBHJ10] D. Sutic, I. Breskovic, R. Huic, and I. Jukic. Automatic evaluation of facial attractiveness. In *Proceedings of MIPRO*, 2010.

[SBS10] L. Stark, K. W. Bowyer, and S. Siena. Human perceptual categorization of iris texture patterns. In *Proceedings of IEEE BTAS*, 2010.

[SBYL02] Z. Sun, G. Bebis, X. Yuan, and S.J. Louis. Genetic feature subset selection for gender classification: a comparison study. In *Proceedings of WACV*, page 165, 2002.

[SEL00] A. Savakis, S. Etz, and A. Loui. Evaluation of image appeal in consumer photography. In *Proceedings of SPIE*, pages 111–121, 2000.

[SGN08] S. Samangooei, B. Guo, and Mark S. Nixon. The use of semantic human description as a soft biometric. In *Proceedings of BTAS*, 2008.

[Sie56] S. Siegel. *Non-parametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

[Sim96] H. A. Simon. *The sciences of the artificial*. Cambridge, MA: MIT Press, 1996.

[SS98] R.E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. In *Proceedings of CCLT*, 1998.

[SSB06] H. R. Sheikh, M F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451, 2006.

[SSG$^+$90] J. M. Seddon, C. R. Sahagian, R. J. Glynn, R. D. Sperduto, and E. S. Gragouda. Evaluation of an iris color classification system. *Investigative Ophthalmology and Visual Science*, 31:1590–1598, 1990.

[ST06] Y. Saatci and C. Town. Cascaded classification of gender and facial expression. In *Proceedings of FG*, pages 393–400, 2006.

[SVB$^+$10] R. Singh, M. Vatsa, H. S. Bhatt, S. Bharadwaj, A. Noore, and S. S. Nooreyezdan. Plastic surgery: a new dimension to face recognition. *IEEE Transaction on Information Forensics and Security*, 5(3):441–448, 2010.

[SVRN07] R. Singh, M. Vatsa, A. Ross, and A. Noore. A mosaicing scheme for pose-invariant face recognition. *IEEE Transactions System, Man, Cybernetics*, 5(37):1212–1225, 2007.

[TSC$^+$01] T. Takamoto, B. Schwartz, L. B. Cantor, J. S. Hoop, and T. Steffens. Measurement of iris color using computerized image analysis. *Current Eye Research*, 22:412–419, 2001.

[usa00] Interactive test methods for audiovisual communications, 2000. ITU-T Recommendation P.920.

[usa11a] BCC research market forecasting, the global biometrics market, 2011.

[usa11b] NIST usability and biometrics, 2011.

[VFT+09] D. Vaquero, R. Feris, D. Tran, L. Brown, A. Hampapur, and M. Turk. Attribute-based people search in surveillance environments. In *Proceedings of WACV*, 2009.

[vio]

[VJ01a] P. Viola and M. Jones. Robust real–time face detection. In *Proceedings of IEEE ICCV*, 2001.

[VJ01b] P. Viola and M. Jones. Robust real-time face detection. In *Proceedings of IEEE ICCV*, 2001.

[WAL04] B. Wu, H. Ai, and R. Liu. Glasses detection by boosting simple wavelet features. In *Proceedings of ICPR*, 2004.

[WAS+05] F. Wallhoff, D. Arsic, B. Schuller, G. Rigoll, J. Stadermann, and A. Stoermer. Hybrid profile recognition on the mugshot database. 2005. Proceedings of IEEE EUROCON.

[WBSS04] Z. Wang, A. C. Bovik, H. R. Sheih, and E. P. Simoncelli. Image qulaity assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(P4):600–612, 2004.

[WM07] H. Weda and Barbieri M. Automatic children detection in digital images. In *Proceedings of ICME*, 2007.

[WMR01] F. Wallhof, S. Mueller, and G. Rigoll. Recognition of face profiles from the mugshot database using a hybrid connectionist / hmm approach. In *"Proceedings of IEEE ICASSP"*, 2001.

[WPS06] F. Wolf, T. Poggio, and P. Sinha. *Bags of Words*. Citeseer, 2006.

[WSB02] Z. Wang, H. R. Sheikh, and A. C. Bovik. No-reference perceptual quality assessment of jpeg compressed images. In *Proceedings of IEEE ICIP*, 2002.

[WYS+02] H. Wu, G. Yoshikawa, T. Shioyama, S. Lao, and M. Kawade. Glasses frame detection with 3D Hough transform. In *Proceedings of ICPR*, volume 16, pages 346–349, 2002.

[XY04] Y. Xiao and H. Yan. Extraction of glasses in human face images. pages 1–23. Springer, 2004.

[YHP11] J. Ylioinas, A. Hadid, and M. Pietikäinen. Combining contrast and local binary pattrens for gender classification. In *Proceedings of SCIA*, 2011.

[ZESH04] R. Zewail, A. Elsafi, M. Saeb, and N. Hamdy. Soft and hard biometrics fusion for improved identity verification. In *Proceedings of MWSCAS*, volume 1, pages I – 225–8, 2004.

[ZG09] X. Zhang and Y. Gao. Face recognition across pose: a review. *Pattern Recognition*, 42(1):2876–2896, 2009.

[ZSH08] M. Zhao, D. Sun, and H. He. Hair-color modeling and head detection. In *Proceedings of WCICA*, pages 7773–7776, 2008.

BIOMETRIES FACIALES DOUCES
METHODES, APPLICATIONS ET SOLUTIONS

Doctoral Thesis
French Summary

Auteur:
Antitza DANTCHEVA

Directeur de these:
Prof. Dr. Jean-Luc DUGELAY

Rapporteurs:

Prof. Dr. Abdenour HADID, University of Oulu, Finland
Prof. Dr. Mark NIXON, University of Southampton, United Kingdom

Examinateurs:

Prof. Dr. Arun ROSS, West Virginia University, USA
Prof. Dr. Bernadette DORIZZI, Telecom SudParis, France
Dr. Sami ROMDHANI, Morpho, France

Ce travail présente l'idée d'utiliser la biométrie, dite douce « soft » pour l'identification et la vérification d'individus. Les systèmes de biométrie douce présentent l'avantage d'être non-intrusifs et d'avoir une faible complexité de calcul. De plus, ils permettent d'effectuer des enrôlements rapides même en l'absence du consentement ou de coopération d'individus, sous-surveillance.

Motivés par le potentiel considérable offert par la biométrie douce, nous présentons dans ce chapitre, une analyse statistique de la fiabilité de tels systèmes, qui consistent à utiliser des signes caractéristiques d'individus dans un contexte d'identification. Nous aborderons tout particulièrement certains aspects de la conception de systèmes de biométrie douce, ainsi que la caractérisation statistique des paramètres pertinents, tels que les signes caractéristiques d'individus, des instances de signes, etc. De plus, nous analyserons les performances et les limites de tels systèmes. Cette analyse nous permettra d'avoir une meilleure compréhension des erreurs d'identification qui sont dues aux interférences inter-individus (i.e. signes caractéristiques communes à plusieurs individus) ou aux erreurs de classification.

1. Introduction

La biométrie traditionnelle offre une solution naturelle et fiable d'identification des individus d'où la croissante utilisation des caractéristiques physiques et comportementales dans les systèmes de sécurité. Cette approche présente l'avantage d'être universelle, robuste, permanente et accessible : c'est pour ces raisons que les mécanismes et systèmes de sécurité actuels, peu importe leur fonction (gouvernementale ou commerciale), s'appuie sur au moins une caractéristique biométrique. Fort de cette constatation, les biométries douces s'inspirent et complètent les avantages de la biométrie classique.

Le premier système d'identification des personnes basé sur la biométrie fut introduit au 19eme siècle par Alphonse Bertillon [RHO 56]. Ce système utilisait tout aussi bien les signes caractéristiques tels que la couleur des yeux, des cheveux et de la peau, la forme et la taille de la tête que les discriminateurs généraux comme la taille, le poids ou encore des marques indélébiles telles que les taches de naissance, les cicatrices et les tatouages. Ces descripteurs constituent principalement ce qui est connu aujourd'hui sous l'appellation de *biométrie douce*.

Les signes utilisés en biométrie douce peuvent être physiques, comportementaux ou externes à l'homme (e.g. port d'écharpe ou de sac). A l'inverse de la biométrie classique, ces catégories sont établies dans le temps par l'expérience humaine afin de différencier les individus. En d'autres termes, les signes caractéristiques utilisés en biométrie douce sont créés de manière naturelle et sont utilisés par les personnes afin de caractériser d'autres personnes.

Les termes *biométrie légère « light biometrics»* [AIL 04], *signes sémantiques « semantic traits»* [SAM 08], *« similes » dans* [KUM 09] et *attribut* [VAQ 09] sont des descripteurs de signes associés à la biométrie douce.

1.1. Domaines d'application

Les biométries douces sont utilisées, soit comme système uni-modal, c'est-à-dire en ne classifiant qu'un seul signe caractéristique, soit en combinaison avec d'autres systèmes. On peut donc considérer les applications suivantes :

– *Fusion biométrie douce- biométrie classique* : les systèmes de biométrie douce sont incorporés dans des systèmes biométriques multimodaux avec comme l'objectif d'accroître la qualité globale. Cette approche a été expérimentée dans [JAI 04] ou les auteurs montrent notamment que l'utilisation des biométries douces, en plus de l'empreinte digitale, permet une amélioration des performances de l'ordre de 5%.

– *Elagage « pruning* »: les systèmes de biométries douces peuvent également être employés pour pré-filtrer de larges bases de données avec comme objectif d'augmenter l'efficacité. Des travaux scientifiques sur l'usage des biométries douces à des fins d'élagage sont présentés dans [KUM 08], [KUM 09], [GIV 97], [NEW 95]. Dans [KUM 08], [KUM 09], les auteurs utilisent les attributs tels que l'âge, le genre, les cheveux et la couleur de peau, pour la classification faciale, alors que [GIV 97], [NEW 95] montrent que les attributs tels que l'âge, le genre et l'origine permettent d'améliorer les performances des systèmes de biométrie classique.

– *(Re)-Identification d'individus* : pour l'identification des humains, les limitations de la biométrie douce (e.g. non-unicité,) sont contournées en combinant plusieurs signes caractéristiques. Le concept de « Bag of Soft Biometrics (BoSB) », est directement inspiré de l'idée de « Bag of Words» [WOL 06], [JOA 98] et de « Bag of Features» [LAZ 06] développée dans le contexte de l'extraction automatique de texte et de la recherche d'images basée sur le contenu. Dans le cadre du BoSB, les éléments du « Bag » sont les signatures de la biométrie douce extraites de l'apparence visuelle du sujet.

1.2. Travaux connexes

Dans cette section, nous présentons les travaux les plus pertinents de la biométrie douce. Cet aperçu ne prétend pas être un état de l'art exhaustif mais serait plutôt une mise en évidence sélective de quelques études scientifiques de la littérature existante.

La biométrie douce est un domaine de recherche très récent et les travaux sur le sujet s'étendent sur plusieurs domaines de recherche. Les contributions les plus récentes peuvent être divisées en trois axes de recherche :

– 1er axe : il est considéré comme l'un axe, le plus exploré. Il comprend l'étude et l'identification des signes caractéristiques des individus, notamment en termes d'algorithmes de traitement d'image, de classification et de détection.

– 2ème axe : cet axe est en plein expansion. Il consiste à identifie les scénarios opérationnels pour les différents algorithmes et fournit des résultats expérimentaux pour ces scénarios. L'objectif principal étant de réduire le temps de calcul tout en augmentant l'efficacité du système.

– 3ème axe : c'est l'axe le moins exploré pour l'instant. Il comprend, notamment, l'étude globale et théorique de l'utilisation des applications relatives à la biométrie douce.

Les travaux scientifiques appartenant au premier domaine englobent les algorithmes de signes caractéristiques tels que l'iris [STA 10] ou les signes du visage. Pour une présentation plus large et détaillée de ces algorithmes, le lecteur peut consulter [DAN 10a].

Le deuxième axe peut, à son tour, être divisé en sous-domaines qui se différencient les uns des autres par leur manière d'utiliser la biométrie douce. On distingue notamment le cas où la biométrie douce est employée comme système autonome, comme mécanisme de pré-filtrage ou comme système parallèle. Les applications incluent l'identification continue [NII 10], la vidéo-surveillance (voir [DEN 09], [FOO 10], [MEL 10]), la vérification de personnes [PAR 10] et enfin l'identification de personnes [ZEW 04]. Un exemple récent de système d'identification de personnes basé sur la biométrie douce consiste en la reconnaissance faciale dans le tri d'images de catastrophes [CHH 11].

Enfin, le troisième axe inclut l'étude du placement de la biométrie douce dans des applications telles que la criminologie [JAI 11] et la métrologie de l'homme [ADJ 10].

Les autres applications possibles concernent la capacité de faire ce qu'on appelle le « *matching* » des individus en se basant, sur leurs préférences de signes biométriques, l'acquisition de propriétés statistiques d'identificateurs biométriques de groupes de personnes, la modélisation d'avatar basée sur les caractéristiques faciales instantanées (lunettes, barbe, couleur des cheveux), l'échantillonnage statistique d'audiences ainsi que plusieurs autres domaines.

2. La biométrie douce pour l'identification humaine

Dans cette partie, nous analysons un scénario dans lequel un ensemble de biométries douces est utilisé pour l'identification des personnes. Nous essaierons de donner un aperçu des facteurs pertinents à la conception et des limitations

Le dispositif d'étude correspond à un scénario général selon lequel un groupe d'authentificatio*n* «*Authentification group* » de *N* personnes est aléatoirement pris à partir d'une population plus large. De ce groupe d'authentification de *N* personnes, une personne est à son tour arbitrairement sélectionnée pour authentification, et différenciée des autres membres du groupe. Notons que ce scénario général inclut tout aussi bien les cas de la vérification de personnes que les cas d'identification. Tout système général de biométrie douce effectue la détection en employant λ signes caractéristiques (couleur des cheveux, couleur de peau, etc.) où chaque signes i (avec $i=1,2,..., \lambda$) peut être subdivisé en *instances* de signes μ_i, c'est-à-dire que chaque signe i peut prendre une des valeurs μ_i. Ainsi, nous appelons *catégorie* tout ensemble de λ éléments. Les éléments étant ici les instances des différents signes. Soit $\Phi=\{\varphi_i\}_{i=1}^{\rho}$ l'ensemble de ρ catégories c'est-à-dire l'ensemble de toutes les ρ combinaisons d'instances de signes de biométrie douce. Le nombre de catégories ρ, s'exprime comme suit :

$$\rho = \prod_{i=1}^{\lambda} \mu_i \qquad [1]$$

Moyennant un léger abus de langage, nous dirons qu'*un sujet appartient à une catégorie φ* si ses instances de signes sont un ensemble (de λ éléments) appartenant à la catégorie φ. Notons que pour avoir une authentification concluante d'un sujet, et par conséquent sa différenciation par rapports aux autres sujets du groupe d'authentification, nous devons être dans le cas où ledit sujet n'appartiendrait pas à la même catégorie que les autres membres du groupe d'authentification.

Etant donné un groupe d'authentification spécifique, la règle d'optimisation du maximum de vraisemblance (MV) de détection de la catégorie la plus probable à laquelle un sujet appartient s'écrit comme suit :

$$\widehat{\varphi} = arg\max_{\varphi\epsilon\Phi} P(\varphi)\, P(y/\varphi) \qquad [2]$$

où y est le vecteur d'observation, $P(\varphi)$ la fonction de densité de probabilité de l'ensemble des catégories pour une population donnée et $P(y/\varphi)$ la probabilité d'observer y sachant que le sujet appartient à la catégorie φ. Rappelons au passage que $\sum_{\varphi=1}^{\rho} \mathrm{P}(\varphi) = 1$.

Nous tenons à souligner que tout comme λ, μ_i et ρ, les paramètres, tels que la taille, les statistiques du groupe d'authentification, sont des paramètres importants pour un système de biométrie douce. Dans la suite, nous détaillerons le fonctionnement du système de biométrie douce introduit précédemment. L'analyse qui suit n'est en aucun cas exhaustive et s'attachera notamment à fournir des détails sur les paramètres tels que :

– La propagation des catégories effectives pour un groupe d'authentification donné. La propagation est utilisée ici comme une mesure de la convenance de Φ pour l'authentification de sujets d'un certain groupe d'authentification.

– La relation entre N et sa probabilité d'interférence correspondante en fonction de Φ (la probabilité que deux utilisateurs partagent la même catégorie et soit indissociable)

– La probabilité de l'erreur d'authentification due aux interférences doit aussi être considérée comme mesure de la fiabilité du système.

2.1. Propagation de la catégorie *Φ*

Nous considérons ici le cas où un système de biométrie douce est conçu pour la distinction parmi ρ catégories distinctes. Dans le scénario considéré, le groupe d'identification occupe de façon arbitraire un petit nombre de catégories ; ces catégories étant substantiellement corrélées entre-elles. Définissons l'ensemble de *catégories effectives* Φ_e comme étant l'ensemble des catégories présentes (c'est-à-dire non vides) dans un groupe d'authentification spécifique. Dans ce contexte, la cardinalité $\rho_e = | \Phi_e |$ est une mesure qui nous renseigne sur la diversité et les performances du système. Rappelons néanmoins que Φ_e et ρ_e sont des variables aléatoires dont les réalisations peuvent avec chacune des occurrences du groupe d'authentification.

Afin de mieux comprendre le caractère aléatoire ci-dessus, nous considérons le cas où les groupes d'authentification sont tirés d'une population générale qui est un ensemble de K = 646 sujets pris dans la base de données FERET [6], avec ρ = 1152 catégories, correspondant à une densité de probabilité P(φ) illustrée sur la figure 1. Cette densité de probabilité correspond aux signes et instances de signes du système proposé.

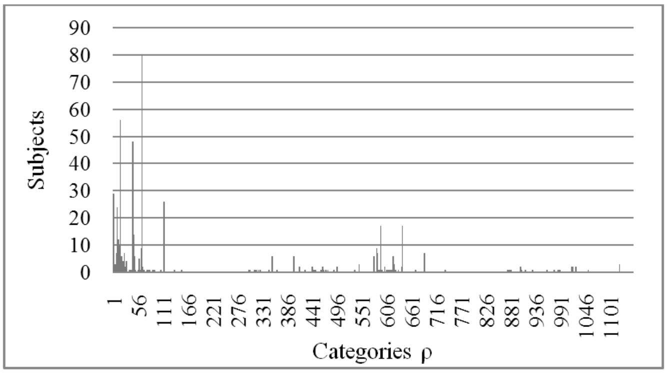


Figure 1. P(φ) *correspondant à la distribution de FERET et au système proposé.*

Etant donné la description précédente, la figure 1. décrit l'équation du nombre moyen de catégories vides :

$$\rho - E[\rho_e](N) \quad [3]$$

en fonction de N et où l'espérance est effectuée sur différentes réalisations du groupe d'authentification.

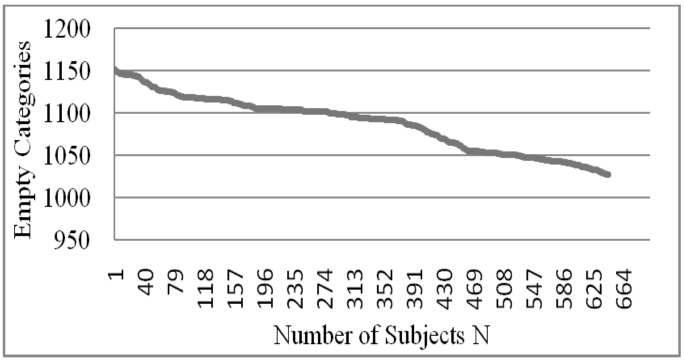


Figure 2. Nombre moyen de catégories vides en function de N (FERET).

Il devient évident que la solution naturelle pour augmenter E[ρ_e] consiste à augmenter ρ, ce qui revient naturellement à se poser la question de savoir si l'augmentation de ρ devrait se traduire par une augmentation des nombres de traits ou par une augmentation du nombre d'instances de traits. Nous nous intéresserons à ce problème d'allocation de ressources sous l'hypothèse simplificatrice de symétrie où $\mu_i = \mu$, pour tout $i = 1, \dots \lambda$. Avec cette hypothèse de symétrie, nous posons :

$$\rho = \mu^{\lambda} \qquad [4]$$

c'est-à-dire que ρ augmente de façon polynomiale avec μ et de façon exponentielle avec λ. Une comparaison des deux dérivées$\frac{d\rho}{d\mu}, \frac{d\rho}{d\lambda}$, permet d'identifier la région *limite des signes* d'un système de biométrie comme étant la région définie par

$$\lambda < \mu \ln \mu \qquad [5]$$

où ρ augmente plus rapidement avec λ qu'avec μ et où l'accent doit être mis sur l'augmentation de λ plutôt que de μ.

Exemple – Augmentation pratique d'un système pour ρ *croissant:* Nous proposons la structure de « bag » d'un système augmenté, système dans lequel l'augmentation des ressources (amélioration de la résolution des capteurs, augmentation de la capacité de calcul) peut être alloué de façon à aussi inclure une augmentation de l'ensemble des signes et d'instances de signes tel que décrite dans le Tableau 1 et entraînant ainsi des valeurs de ρ de l'ordre de quatre-vingt millions et qui conviendrait à diverses applications.

Couleur de peau	Couleur des cheveux	Couleur des yeux	Présence de lunettes	Présence de barbe	Présence de moustache	Age	Genre

3	8	6	2	2	2	3	2

Maquillage	Forme de visage	Formes des caractéristiques du visage	Mesures faciales	Mesures des caractéristiques du visage	Marques et grain de beauté du visage	Longueur des cheveux
4	3	3	3	6	6	3

Tableau 1. *Ensemble des signes du visage de la biométrie douce et leur nombre d'instances.*

2.2. *Limites de N pour une probabilité d'interférence donnée*

Dans cette partie, nous allons notamment décrire la relation entre N et la probabilité d'interférence correspondante en fonction de Φ. Nous définirons clairement l'évènement de collision ou interférence.

Définition: on parle de *collision* ou *interférence* lorsque deux ou plusieurs sujets *quelconques* appartiennent à la même catégorie. En parlant d'un sujet précis, nous dirons qu'un sujet subit une interférence s'il/elle appartient à une catégorie qui contient aussi des sujets du groupe d'authentification.

Au regard de ceci, nous nous intéressons à deux mesures de probabilité. La première est la probabilité $p(\mathrm{N};\rho)$ que le groupe d'authentification de taille N, choisi arbitrairement parmi une grande population de sujets, soit telle qu'il comporte deux sujets en collision. Rappelons brièvement la relation de $p(\mathrm{N};\rho)$ et du fameux paradoxe de l'anniversaire. Pour les autres mesures de la fiabilité du système, nous considérons le cas où un groupe d'authentification de taille N est aléatoirement choisi parmi une large population de sujets et où un sujet quelconque de ce groupe d'authentification serait en collision avec un autre membre de ce même groupe. Nous dénotons cette probabilité $\mathrm{q}(\mathrm{N})$et rappelons également que $\mathrm{q}(\mathrm{N}) < p(\mathrm{N})$. Pour résumer, $p(\mathrm{N})$ décrit la probabilité que l'interférence existe, bien qu'elle puisse induire des erreurs, tandis que $\mathrm{q}(\mathrm{N})$ décrit la probabilité d'erreur causée par les interférences.

Exemple: Dans un groupe de 10 sujets, $p(\mathrm{N})$ décrit la probabilité que deux sujets arbitrairement choisis parmi les 10 sujets appartiennent à la même catégorie φ_x. $\mathrm{q}(\mathrm{N})$ représente la probabilité qu'un sujet particulier entre en interférence avec un ou plusieurs des 9 sujets restant. La probabilité que n'importe quelle collision se produise est donc supérieure à la probabilité qu'un sujet précis entre en collision : ce qui se traduit par $\mathrm{q}(\mathrm{N}) < p(\mathrm{N})$

Nous nous attacherons d'abord à calculer et représenter $p(\mathrm{N})$ sous l'hypothèse simplificatrice d'uniformité statistique des catégories. La forme fermée de l'expression de cette probabilité est démontrée (voir [DAS 05]) comme étant

$$\mathrm{p}(\mathrm{N};\rho) = \begin{cases} 1 - \prod_{\mathrm{k}=1}^{\mathrm{N}-1}\left(1 - \frac{\mathrm{k}}{\rho}\right) & \mathrm{N} \leq \rho \\ 1 & \mathrm{N} > \rho \end{cases} \qquad [6]$$

et peut équivalemment être développée comme suit

$$p(\mathrm{N};\rho) = 1 - \left(1 - \frac{1}{\rho}\right)\left(1 - \frac{2}{\rho}\right)\ldots\left(1 - \frac{\mathrm{N}-1}{\rho}\right) = 1 - \frac{\rho!}{\rho^{n}(\rho-\mathrm{N})!}. \qquad [7]$$

Notons que, sous l'hypothèse d'uniformité, la probabilité $p(\mathrm{N};\rho)$ ci-dessus constitue une limite inférieure de cette même probabilité (ici en absence de l'hypothèse d'uniformité). De manière équivalente, nous pouvons déterminer la valeur maximale de N pour une probabilité de collision donnée. Selon l'expression en forme fermée, ce calcul s'appuie sur l'approximation suivante,

$$p(\mathrm{N};\rho) \approx 1 - \left(\frac{\rho-1}{\rho}\right)^{\frac{N(N-1)}{2}} \qquad [8]$$

issue de [AHM 00] et dont la résolution suivant N donne :

$$N(p;\rho) \approx \sqrt{2\rho \cdot ln\left(\frac{1}{1-p}\right)} \qquad [9]$$

Ce qui correspond à la valeur de N pour laquelle le système aura une probabilité d'interférence p. A titre d'exemple, pour $\rho = 1152$ et p = 0.5 on obtiendrait $\mathrm{N} = 39$. En d'autres termes, étant donné un système de biométrie douce doté de 1152 catégories, une répartition uniforme des sujets dans ces catégories et la probabilité de 50% qu'une collision survienne est vérifiée pour tout groupe de 39 sujets.

L'expression en forme fermée de $\mathrm{q}(\mathrm{N})$ est la suivante :

$$\mathrm{q}(\mathrm{N}) = 1 - \left(\frac{\rho-1}{\rho}\right)^{\mathrm{N}}. \qquad [10]$$

A titre d'exemple, étant donné $\rho = 1152$ et $\mathrm{q} = 0.5$ et toujours sous hypothèse d'uniformité, on a N>700, ce qui, comme on pourrait s'y attendre, est largement supérieur à sa valeur pessimiste correspondant à $p(\mathrm{N};\rho)$.

Dans un souci de généralisation, nous nous éloignerons de l'hypothèse d'uniformité afin de considérer un scénario plus réaliste où la distribution des catégories provient d'une base de données de la vie courante. Dans ce cas, la probabilité que tous les N sujets soient dans des catégories différentes est la somme des produits des évènements de non-collision [AHM 00] :

$$p(\mathrm{N};\rho) = 1 - \sum_{\alpha\neq\beta\neq\cdots\neq\omega} \mathrm{P}(\varphi_\alpha)\mathrm{P}(\varphi_\beta)\ldots\mathrm{P}(\varphi_\omega), \quad [11]$$

où l'indice de sommation correspond aux catégories non-vides selon le groupe d'authentification. Cette probabilité est schématisée en figure 3. Cette figure nous montre que cette probabilité dépasse, bien que ce soit de peu, la probabilité obtenue sous l'hypothèse d'uniformité.

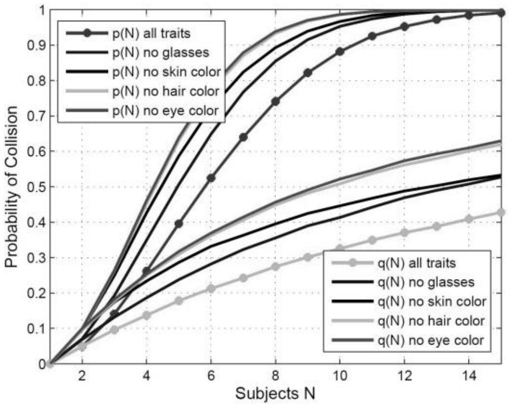


Figure 3. $q(\mathrm{N})$ *et* $p(\mathrm{N})$ *pour une distribution réelle et uniforme.*

2.3. Simulation d'évaluation du système dont les interférences sont limitées par des capteurs de très hautes résolutions

Dans ce qui suit, nous simulerons la probabilité d'erreur d'identification pour le scénario d'intérêt et sous l'hypothèse que les erreurs ne sont dues qu'aux interférences, c'est-à-dire sous l'hypothèse que les erreurs se produisent si et seulement si le sujet choisi partage la même catégorie qu'un autre quelconque sujet du groupe d'authentification. Ceci correspond au cas où le système de biométrie ne pourrait fournir une authentification concluante. Dans cette simulation, la plus grande population était constituée de 646 personnes de la base de données FERET et la simulation a été effectuée pour différentes tailles N du groupe d'authentification. La probabilité d'erreur d'authentification est représentée dans la figure suivante.

La figure 3 montre la probabilité de collision lorsque différents traits sont enlevés comme mesure de l'importance de chaque signes. La présence de moustache et barbe semble avoir le moins d'influence sur les résultats de détection tandis que les couleurs des cheveux et des yeux ont une grande influence sur la différenciation des individus.

3. Probabilité d'erreur totale d'un système de biométries douces

Dans le scénario opérationnel décrit ci-dessus, la *fiabilité* d'un système de biométrie douce dépend de la probabilité de fausse identification d'une personne quelconque de l'ensemble des N personnes.

Dans un tel contexte, la fiabilité du système de biométrie douce est généralement liée :

– au nombre de catégories que le système peut identifier,

– au degré avec lequel les caractéristiques/catégories représentent l'ensemble choisi (de sujets) sur lequel porte l'identification,

– à N sachant que les valeurs élevées de N reviennent à identifier une personne parmi de plus larges ensembles de personnes similaires,

– à la robustesse avec laquelle ces catégories peuvent être détectées.

Nous procédons ici à l'étude de la probabilité d'erreur générale du système de biométrie douce [DAN 11a], incluant ici, en plus des facteurs cités ci-dessus, les probabilités d'erreur de catégorisation algorithmique. En d'autres termes, nous examinons la probabilité d'erreur et ce, quelque soit la source d'erreur. Rappelons que les erreurs de notre système peuvent être dues à une mauvaise classification ou à l'interférence. Le premier aspect est lié aux statistiques de la population étudiée et le second au comportement de l'erreur des différents algorithmes de classification. Nous notons :

$$p = \left[p_1, p_2, \ldots p_\rho\right]^t \qquad [12]$$

qui définit entièrement les statitisques de la population.

En termes d'erreur de comportement, nous regardons le système de biométrie douce comme un système capable de classer un sujet de la catégorie φ_i à la catégorie estimée $\hat{\varphi}$ ou de classer ledit sujet dans la mauvaise catégorie, voir figure 3. Nous définissons ci-après :

$$\varepsilon_{ij} = P\big(\hat{\varphi}(v) = \varphi_j : v \in \varphi_i\big) \qquad [13]$$

la probabilité que le système de biométries douces mette les sujets de la $i^{\text{ème}}$ classe φ_i dans la $j^{\text{ème}}$ classe φ_{jj} (voir figure 4 pour l'illustration graphique). Plus simplement, ε_{ij} ≔ est l'élément de la $i^{\text{ème}}$ ligne et de la $j^{\text{ème}}$ colonne de ce que l'on nomme dans la littérature matrice de confusion et que nous notons E :

$$E = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1\rho} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2\rho} \\ \vdots & \ddots & \ddots & \vdots \\ \varepsilon_{\rho 1} & \varepsilon_{\rho 2} & \cdots & \varepsilon_{\rho\rho} \end{bmatrix} \qquad [14]$$

Toujours en lien avec ces paramètres, définissons :

$$\varepsilon_f = \sum_{i=1, i\neq f}^{\rho} \varepsilon_{fi} \qquad [15]$$

Pour dénoter la probabilité qu'un membre de la catégorie φ_f soit mal classé.

Finalement, nous utilisons la notation :

$$e = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_\rho] \qquad [16]$$

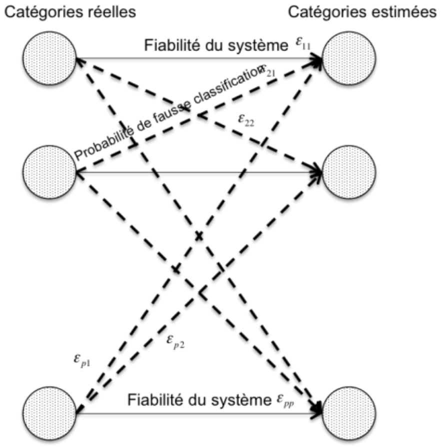

Catégories réelles
Catégories estimées
Fiabilité du système ε_{11}
ε_{21}
Probabilité de fausse classification
ε_{22}
ε_{p1}
ε_{p2}
Fiabilité du système ε_{pp}

Figure 4. *Paramètres de confusion.*

Un système de biométrie douce de matrice de confusion ***E*** et de vecteur d'erreur **e** qui opère sur une population dont les statistiques sont données par p a une probabilité d'erreur P_{err}:

$$P_{err} = p^t e \qquad [17]$$

3.1. *Probabilité d'erreur d'un système de biométries douces en cas de re-identification frontale-latérale*

Pour quantifier et mieux analyser P_{err}, nous présentons ici un réel système de biométrie douce, de matrice de confusion ***E*****,** employé sur la population de la base de données de FERET. Pour cette expérience, nous prenons une fois de plus, un groupe d'authentification de N sujets. Un sujet de ces N sujets est ensuite arbitrairement comme sujet-cible de l'authentification. Nous procédons ensuite à l'entraînement des algorithmes de classification (voir [DAN 11b]) afin d'extraire un vecteur de caractéristiques contenant les traits suivants :

Contraste : mesure sur toute l'image de l'intensité du contraste entre un pixel et ses voisins. Le contraste d'une image est lié à sa variance et à son inertie et s'exprime comme suit :

$$x_1 = \sum_{i,j} |i - j|^2 g(i,j) \qquad [18]$$

où *i* et *j* représente le niveau de gris de deux pixels, g se réfère au niveau de gris de la matrice de co-occurrence. La matrice de co-occurrence décrit la co-occurrence de niveaux de gris entre deux images : chaque élément *(i,j)* indiquant alors le nombre de fois que le pixel de valeur *i* est horizontalement adjacent au pixel de valeur *j*.

Corrélation : mesure la corrélation des pixels voisins est notée :

$$x_2 = \frac{\sum_{i,j}(i-\mu_i)(j-\mu_j)g(i,j)}{\sigma_i \sigma_j} \qquad [19]$$

où μ_i et μ_j représentent, respectivement, les valeurs moyennes des voisinages de *i* et *j* tandis que σ_i et σ_j représentent leurs variances respectives.

Energie: La somme des carrés des éléments ou moment angulaire d'ordre 2. Une énergie de 1 correspond à une image de couleur uniforme.

$$x_3 = \sum_{i,j} g(i,j)^2 \qquad [20]$$

Homogénéité : mesure de la proximité de la distribution des éléments :

$$x_3 = \sum_{i,j} \frac{g(i,j)}{1+i-j} \qquad [21]$$

– distance : en conjonction avec les informations de couleur et de texture, nous intégrons une mesure de relation simple en notre classificateur. Cette mesure s'intéresse à la divergence entre la densité de probabilité de l'intensité des parcelles relative à un sujet. Autrement dit, nous exprimons les trois relations entre les différentes intensités du sujet : cheveux—peau, peau—habits et cheveux—habits. Nous nous attendrions par exemple à avoir une distance plus élevée avec une personne ayant des cheveux couleur châtain et une peau claire qu'avec une personne ayant des cheveux blonds et une peau claire. Nous convertissons ensuite les parcelles en niveau de gris et évaluons trois fois par personne la distance L1 et ce pour toutes les relations entre les parcelles. Pour deux distributions *r* et *s* de caractères aléatoires, cette mesure est donnée par :

$$D = \|r - s\|_1 = \sum_{k=1}^{255} |r(k) - s(k)| \qquad [22]$$

où *k* représente un point parmi les 255 points d'intensité pour une image en niveau de gris.

Un tel vecteur de caractéristiques est extrait de chaque base d'images. Après l'étape d'entraînement, nous re-identifions le sujet-cible en faisant correspondre son/ses vecteurs de caractéristiques avec les vecteurs de caractéristiques des N sujets fournis par de l'étape d'entraînement. Dans cette expérimentation, les images de la base de donnée sont des portraits de face de sujets et pour l'étape de tests, les sujets sont photographiés de profil.

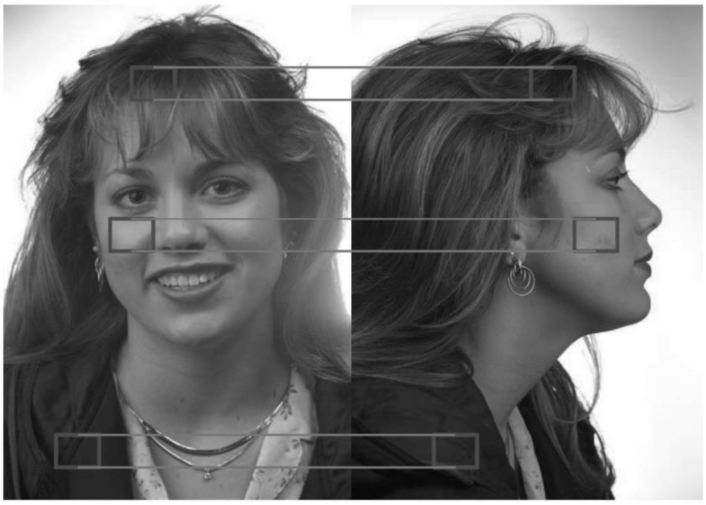


Figure 5. *Portrait de face et de profil pour un sujet de la galérie avec correspondance entre les régions d'intérêt pour les cheveux, la couleur de peau et la couleur d'habits.*

Cette procédure est répétée afin de moyenner la probabilité d'erreur sur toutes les itérations pour toutes les valeurs de N. La classification correspondante est effectuée par l'algorithme AdaBoost et la probabilité d'erreur P_{err} est représentée en figure 6.

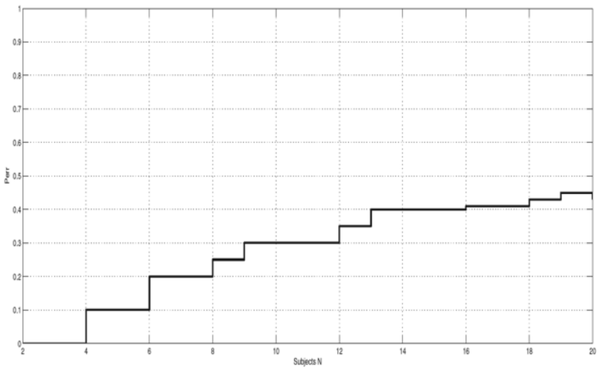


Figure 6. *Probabilité d'erreur P_{err} d'un système de biométries douces basée sur le classificateur à entrées multiples Adaboost.*

Le système de biométrie douce qui « *booste* » tous les signes décrits, couleur, texture et intensité fonctionne pour un scénario où la différence entre les 2 poses est d'environ 90 degrés. Notons ici que les algorithmes classiques de reconnaissance faciale, qu'ils soient holistiques ou basés sur les caractéristiques faciales, sont capables de gérer des différences de pose de moins de 15 degrés. Néanmoins le système de biométrie douce présenté a une probabilité d'erreur de 0.1 pour un groupe d'authentification de 4 sujets, ce qui ne serait pas suffisant pour en faire un système d'identification robuste. Ces performances limitées sont liées à la corrélation entre traits i.e. couleur de cheveux—couleur de peau ou couleur de peau—texture de peau (voir [DAN 10b]). Une solution pour améliorer ces performances consiste à augmenter le nombre de catégories. Dans ce cas, il faudrait aussi prendre en compte l'erreur ajoutée et évaluer cette probabilité par rapport au gain apportée par la nouvelle catégorie.

4. Conclusion

Dans ce chapitre, nous avons exploré l'utilisation de système de biométries douces à plusieurs traits pour l'identification de personnes. Nous avons notamment étudié la relation entre le groupe d'authentification, sa taille N, les catégories ρ et les catégories effectives Φ_e.

Il devient évident que la surveillance affectera de plus en plus notre quotidien (qualités de vie, sécurité). Pour cette raison, la place occupée par des systèmes de sécurité utilisant la biométrie va être de plus importante. Nous voyons le domaine des biométries douces avoir une position de plus en plus importante dans de tels systèmes.

Dans cette optique, nous devons avoir une meilleure compréhension des composantes telles que les systèmes biométries douces sans toute fois négliger la compréhension d'algorithmes de classification de nouveau traits et de nouvelles combinaisons de ceux-ci. Notre objectif est donc d'améliorer l'efficacité des systèmes de biométries douces tout en développant une meilleure compréhension de ses capacités et de ses limites.

Bibliographie

[AIL 04] AILISTO H., LINDHOLM M., MÄKELÄ S.-M., VILDJIOUNAITE E., *Unobtrusive user identification with light biometrics*. In Proceedings of NordiCHI, 2004.

[SAM 08] SAMANGOOEI S., GUO B., NIXON, M. S., *The use of semantic human description as a soft biometric.* In Proceedings of BTAS, 2008.

[KUM 09] KUMAR N., BERG A. C., BELHUMEUR P. N., NAYAR S. K., Attribute and simile classifiers for face verification. In Proceedings of IEEE ICCV, 2009.

[VAQ 09] VAQUERO D., FERIS R., TRAN D., BROWN L., HAMPAPUR, A., TURK, M., *Attribute based people search in surveillance environments*. In Proceedings of WACV, 2009.

[JAI 04] JAIN A.K., DASS S.C., NANDAKUMAR K., *Soft biometric traits for personal recognition systems*, In Proceedings of ICBA, 2004.

[KUM 08] KUMAR N. AND BELHUMEUR P. N. AND NAYAR S. K., *FaceTracer: a search engine for large collections of images with faces*, In Proceedings of ECCV, 2008.

[KUM 09] KUMAR N.AND BERG A. C., BELHUMEUR P. N., NAYAR S. K., *Attribute and simile classifiers for face verification*, In Proceedings of IEEE ICCV, 2009.

[GIV 97] GIVENS G., BEVERIDGE J. R., DRAPER B.A., BOLME D., *A statistical assessment of subject factors*, 1997.

[NEW 95] NEWHAM E., *The biometric report*, SJB Services, New York, 1995.

[WOL 06] WOLF F., POGGIO T., SINHA P., *Bag of words*, Citeseer, 2006.

[JOA 98] JOACHIMS T., *Text categorization with support vector machines: Learning with many relevant features*, Proceedings of ECML, 1998.

[LAZ 06] LAZEBNIK S., SCHMID C., PONCE J., *Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories*, Proceedings of ICPR, 2006.

[STA 10] STARK L., BOWYER K. W., SIENA S., *Human perceptual categorization of iris texture patterns*, In Proceedings of IEEE BTAS, 2010.

[LEE 99] LEE J.-E., JAIN A. K., JIN R., *Scars, marks and tattoos (SMT): Soft biometric for suspect and victim identification*, In Proceedings of BCC, 2008.

[RHO 56] RHODES, H., *Alphonse Bertillon: Father of scientific detection*. Pattern Recognition Letters, 1956.

[NII 10] NIINUMA K., PARK U., JAIN A. K., *Soft Biometric Traits for Continuous User Authentication*, IEEE Transactions on Information Forensics and Security}, vol. 5, 4, pages 771-780, 2010.

[DEN 09] DENMAN S., FOOKES C., BIALKOWSKI A., S. SRIDHARAN, *Soft-Biometrics: Unconstrained Authentication in a Surveillance Environment*, Proceedings of DICTA, pages 196--203, 2009.

[FOO 10] FOOKES C., DENMAN S., LAKEMOND R., RYAN D., SRIDHARAN S., PICCARDI M., *Semi-supervised intelligend surveillance system for secure environments*, Proceedings of IEEE ISIE,2010.

[MEL 10] MELTEM D., KSHITIZ G., SADIYE G., *Automated person categorization for video surveillance using soft biometrics*, In Proceedings of SPIE, pages 76670P-76670P-12, 2010.

[ZEW 04] ZEWAIL R., ELSAFI A., SAEB M., HAMDY N., *Soft and hard biometrics fusion for improved identity verification* , In Proceedings of MWSCAS, Vol. 1, pages I - 225-8, year = 2004.

[PAR 10] PARK U., JAIN A. K., *Face Matching and Retrieval Using Soft Biometrics*, IEEE Transactions on Information Forensics and Security, vol. 5, 3, pages 406-415, 2010.

[CHH 11] CHHAYA N., OATES T., *Integrating soft biometrics to extract text descriptors from triage images in mass disaster situations*, In Proceedings of HST, 2011.

[JAI 11] JAIN A. K. , KLARE B., PARK U., *Face recognition: Some challenges in forensics*, In Proceedings of IEEE FG, pages 726-733, 2011.

[ADJ 10] ADJEROH D., CAO D., PICCIRILLI M., ROSS A., *Predictability and Correlation in Human Metrology*, In Proceedings of the WIFS, 2010.

[DAN 10a] DANTCHEVA A., VELARDO C., D'ANGELO, A., DUGELAY, J.-L., *Bag of soft biometrics for person identification: New trends and challenges,* Mutimedia Tools and Applications, Springer, October 2010.

[DAN 11a] DANTCHEVA A., SINGH A., ELIA P., DUGELAY J.-L., *Search pruning in video surveillance systems: Efficiency-reliability tradeoff,* In Proceedings of ICCV 2011, IWITINCVPR Workshop, 2011.

[DAN 11b] DANTCHEVA A., DUGELAY, J.-L., *Frontal-to-side face re-identification based on hair, skin and clothes patches,* In Proceedings of AVSS 2011, 2011.

[DAN 10b] DANTCHEVA A., DUGELAY J.-L., ELIA P., *Person recognition using a bag of facial soft biometrics (BoFSB),* In Proceedings of MMSP, 2010.

[DAS 05] DASGUPTA A.*, The matching, birthday and the strong birthday problem: A contemporary review* Journal of Statistical Planning and Inference, vol. 130 (1-2), pp. 377-389, March 2005.

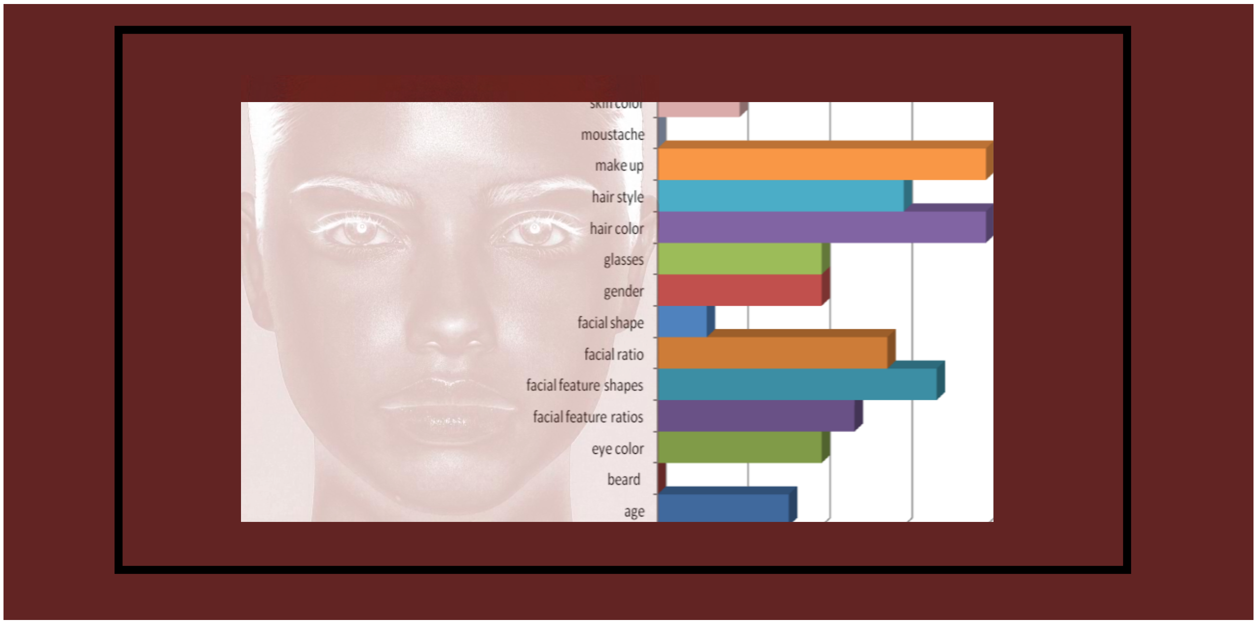
[AHM 00] AHMED S. E., MCINTOSH R. J., *An asymptotic approximation for the birthday problem*, *Crux Mathematicorum*, vol. 26, pp. 151-155, Apr. 2000.

FACIAL SOFT BIOMETRICS

METHODS, APPLICATIONS AND SOLUTIONS

Doctoral Thesis

Summary



Author:

Antitza DANTCHEVA

Supervisor:

Prof. Dr. Jean-Luc DUGELAY

EURECOM
Sophia Antipolis

TELECOM ParisTech

Reviewers:

Prof. Dr. Abdenour HADID, University of Oulu, Finland

Prof. Dr. Mark NIXON, University of Southampton, United Kingdom

Examiners:

Prof. Dr. Arun ROSS, West Virginia University, USA

Prof. Dr. Bernadette DORIZZI, Telecom SudParis, France

Dr. Sami ROMDHANI, Morpho, France

Abstract

This dissertation studies soft biometrics traits, their applicability in different security and commercial scenarios, as well as related usability aspects. We place the emphasis on human *facial soft biometric traits* which constitute the set of physical, adhered or behavioral human characteristics that can partially differentiate, classify and identify humans. Such traits, which include characteristics like age, gender, skin and eye color, the presence of glasses, moustache or beard, inherit several advantages such as ease of acquisition, as well as a natural compatibility with how humans perceive their surroundings. Specifically, soft biometric traits are compatible with the human process of classifying and recalling our environment, a process which involves constructions of hierarchical structures of different refined traits.

This thesis explores these traits, and their application in *soft biometric systems* (SBSs), and specifically focuses on how such systems can achieve different goals including database search pruning, human identification, human re–identification and, on a different note, prediction and quantification of facial aesthetics. Our motivation originates from the emerging importance of such applications in our evolving society, as well as from the practicality of such systems. SBSs generally benefit from the non-intrusive nature of acquiring soft biometric traits, and enjoy computational efficiency which in turn allows for fast, enrolment–free and pose–flexible biometric analysis, even in the absence of consent and cooperation by the involved human subjects. These benefits render soft biometrics indispensable in applications that involve processing of real life images and videos.

In terms of security, we focus on three novel functionalities of SBSs: pruning the search in large human databases, human identification, and human re–identification.

With respect to *human identification* we shed some light on the statistical properties of pertinent parameters related to SBSs, such as employed traits and trait–instances, total categories, size of authentication groups, spread of effective categories and correlation between traits. Further we introduce and elaborate on the event of interference, i.e., the event where a subject picked for identification is indistinguishable from another subject in the same authentication group.

Focusing on *search pruning*, we study the use of soft biometric traits in pre-filtering large human image databases, i.e., in pruning a search using soft biometric traits. Motivated by practical scenarios such as time–constrained human identification in biometric-based video surveillance systems, we analyze the stochastic behavior of search pruning, over large and unstructured data sets which are furthermore random and varying, and where in addition, pruning itself is not fully reliable but is instead prone to errors. In this stochastic setting we explore the natural tradeoff that appears between pruning gain and reliability, and proceed to first provide average–case analysis of the problem and then to study the atypical gain-reliability behavior, giving insight on how often pruning might fail to substantially reduce the search space. Moreover we consider actual soft biometric systems (nine of them) and the corresponding categorization algorithms, and provide a number of experiments that reveal the behavior of such systems. Together, analysis and experimental results, offer a way to quantify, differentiate and compare the presented SBSs and offer insights on design aspects for improvement of such systems.

With respect to *human re–identification* we address the problem of pose variability in surveillance videos. Despite recent advances, face-recognition algorithms are still challenged when applied to the setting of video surveillance systems which inherently introduce variations in the pose of subjects. We seek to provide a recognition algorithm that is specifically suited to a frontal-to-side re-identification setting. Deviating from classical biometric approaches, the proposed method considers color- and texture- based soft biometric traits, specifically those taken from patches of

hair, skin and clothes. The proposed method and the suitability of these patch-based traits are then validated both analytically and empirically.

Deviating from security related themes, we focus on a completely different application: employing soft biometrics in evaluation of *female facial aesthetics*. This approach is novel in that, in the context of female facial aesthetics, it combines soft biometrics with previous approaches on photo quality and beauty assessment. This study helps us to understand the role of this specific set of features in affecting the way humans perceive facial images. Based on the above objective parameters, we further construct a simple linear metric that suggests modifiable parameters for aesthetics enhancement, as well as tunes systems that would seek to predict the way humans perceive facial aesthetics. Moreover using the designed metric we evaluate beauty indices with respect to aging, facial surgery and females famous for their beauty. We simulate an automatic tool for beauty prediction with both realistic accuracy and performance.

Remaining in the realm of human perception, we also provide a comparative study of different access control systems based on fingerprint, PIN, soft biometrics and face recognition. Towards comparing these systems, we design real–life access control interfaces, each based on the above mentioned methods, and then proceeded to empirically evaluate the degree of usability for each of these interfaces. Our study is based on the recorded assessments of a set of users who rated their interaction with each interface, in terms of privacy, ease of use, user-friendliness, comfort and interaction time. The results reinforce, from a usability point of view, the employment of novel biometric authentication methods as viable alternatives to the currently predominant PIN based methods for access control.

Overall this dissertation has contributed the following:

- identification and introduction of novel applications for soft biometrics, such as human identification (bag of soft biometrics), re–identification as well as aesthetics prediction
- development of theoretical framework for SBSs in the applications: pruning the search and human identification
- application of the developed theoretical framework on existing SBSs
- construction of a novel image processing tool for classification of soft biometric traits and employing such a tool in challenging scenarios
- obtaining evidence for the high user friendliness of soft biometric based control access systems.

This work was conducted in part within the European Project ACTIBIO [ACT11] and was supported in part by the European Commission under contract FP7-215372.

Soft Biometrics

Traditional biometrics offer a natural and reliable solution for establishing the identity of an individual, and for this reason, the use of human physical and behavioral characteristics has been increasingly adopted in security applications. With this approach maintaining various advantages such as universality, robustness, permanence and accessibility, it is not surprising that current intrusion detection and security mechanisms and systems include by default at least one biometric trait.

Building on this progress, the latest addition of soft biometrics builds and adds on the main advantages of classical biometrics.

The beginnings of soft biometric science were laid by Alphonse Bertillon in the 19th century, who firstly introduced the idea of a person identification system based on biometric, morphological and anthropometric determinations, see [Rho56]. In his effort, Bertillon considered traits like colors of eye, hair, beard and skin; shape and size of the head, as well as general discriminators like height or weight and also indelible marks such as birth marks, scars or tattoos. These descriptors mainly comprise what is now referred to as the family of *soft biometrics*, a term first introduced by Jain et al. [JDN04b] to describe the set of characteristics that provide (some) information about an individual, but that are not generally sufficient for fully describing and identifying a person, mainly due to the lack of distinctiveness and permanence of such traits. As stated later [JDN04a], such soft biometrics traits can be inexpensive to compute, can be sensed at a distance, do not require the cooperation of the surveillance subjects, and can be efficiently used to narrow down a search for an individual from a large set of people. Along the lines of *semantic annotation* ([SGN08] and [RN10]) we here note the human compliance of soft biometrics as a main difference between soft biometrics and classical biometrics - a difference that renders soft biometrics suitable for many applications. The terms *light biometrics* see in [ALMV04], *similes* see in [KBBN09] and *attributes* see in [VFT$^+$09] have been describing traits we associate to soft biometrics. The following definition clarifies what is considered here as soft-biometric traits.

Definition: Soft biometric traits are physical, behavioral or adhered human characteristics, classifiable in pre–defined human compliant categories. These categories are, unlike in the classical biometric case, established and time–proven by human experience with the aim of differentiating individuals. In other words soft biometric traits are created in a natural way, used by people to characterize other people.

Our interest in this thesis is in understanding the role that soft biometrics can play in security and commercial systems of the future. In brief we begin by specifying soft biometric traits that adhere to the above definition. After an overview of related work, we proceed to explore different applications that benefit from soft biometric systems (SBSs), focusing on surveillance related person identification, and on pruning of large surveillance related searches. We also consider the specific scenario of applying soft biometrics for human frontal-to-side re-identification. We then change gear and deviate from security related applications to the more commercially oriented application of employing soft biometrics in quantifying and predicting female facial aesthetics. The above approaches are then complemented by a more practical automatic soft biometric classification tool that we present. Finally, motivated by human acceptance issues, we proceed to provide a usability study relating to soft biometrics.

Achievements of the dissertation

We proceed with an explicit description of the scenarios and applications of interest in the thesis.

Soft biometrics: characteristics, advantages and related work

We illustrate in Table 1) a range of facial characteristics which accept the definition stated in for soft biometrics. In a first attempt to differentiate between soft biometric traits we firstly identify the affiliation to *face* or *accessory* categories. We abuse slightly annotation and include hair color in the group of facial soft biometrics. The presented traits list is not exhaustive and will naturally increase with technological progress.We here note that even though classically *accessories* do not belong to biometrics, the new stated definition clearly incorporates such traits in the class of soft biometrics. The motivation for including accessories to soft biometrics lays in the associated highly descriptiveness and discrimination of attributes such as clothes color, e.g. “the person in the red shirt". Further significant factors for classifying soft biometric traits are *distinctiveness* and *permanence*. *Distinctiveness* is the strength with which a trait is able to distinguish between individuals. As an example 'beard' has a low distinctiveness, since it can only be applied to the male part of the population and furthermore possesses only two sub–categories (present or not). This example points out a certain correlation between *distinctiveness* and *nature of value*. Traits with continuous sub-categories are in general more distinctive than traits with discrete and moreover binary sub-categories. In this context the difference between *nature of value* and human labeling of traits is the following: while hair color has principally different nuances and is thus of continuous character, humans tend to discrete labeling. We adopt this human approach for developed soft biometric estimation algorithms, detecting for example hair color in categories such as black, blond, brown, rather than RGB values.

The *permanence* of a trait plays a major role for the application for which a SBS is employed. As an example an application, where identification within a day is required, will accept low permanence traits like age, weight or clothing color (inter vs. intra session observation).

The final subdivision *subjective perception* refers to the degree of ambiguity associated in identifying or labelling specific soft biometric traits sub-categories. We note the relation of subjective perception to the nature of value, where an increased amount of subcategories leads to a more difficult classification. In fact subjectivity lays even in the decision of the nature of value. In other words, colors for example can be argued to be continuous, due to the huge variance in nuances blending into each other, or to be discrete due to the fact that colors can be described by discrete RGB values.

We note that soft biometrics can be classified by additional aspects such as accuracy and importance, which are deducible from the named classification classes, depending on the cause for specification (e.g. suitability for a specific application).

Characteristics, advantages and limitations

Soft biometrics has carried in some extent the attributes of classical biometrics over, as the general idea of identification management based on *who you are* is still being pursuit. The traits provide weak biometrical information about the individual and correspondingly have inherited the predicates to be *universal*, *measurable* and *acceptable*; furthermore the trait's classification algorithm(s) *performance* should be able to meet the application's requirements. To a certain degree also the aspects *uniqueness*, *permanence* and *circumvention* play a role for soft biometrics, but are treated to a greater extent flexible.

Initially, soft biometric traits have been employed to narrow down the search of a database, in order to decrease the computational time for the classical biometric trait. An additional application is the fusion of soft biometrics and classical biometric traits to increase overall system performance. Soft biometrics impart systems substantial advantages: they can be partly derived from main detected classical biometric identifier, their acquisition is non intrusive and does not require enrolment; training can be performed in advance on individuals out of the specific identification group. Summarizing soft biometric traits typically are:

- Human compliant: Traits conform with natural human description labels.
- Computationally efficient: Sensor and computational requirements are marginal.
- Enrolment free: Training of the system is performed off–line and without prior knowledge of the inspected individuals.
- Deducible from classical biometrics: Traits can be partly derived from images captured for primary (classical) biometric identifier (e.g. eye color from eye images).
- Non intrusive: Data acquisition is user friendly or can be fully imperceptible.
- Classifiable from a distance: Data acquisition is achievable at long range.
- Classifiable pose flexible: Data acquisition is feasible from a number of poses.

Table 1: Table of soft biometric traits

Soft Biometric trait	**Face / Accessory**	**Nature of value**	**Permanence**	**Distinctiveness**	**Subjective perception**
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
Moustache	Face	Binary	Low/Medium	Low	Medium
Facial measurements	Face	Continuous	High	Medium	Medium/High
Facial shapes	Face	Discrete	High	High	High
Facial feature measurements	Face	Continuous	High	High	Medium/High
Facial feature shapes	Face	Discrete	High	High	High
Make–up	Face	Discrete	Low	Low	Medium
Ethnicity	Face	Discrete	High	Medium	Medium
Marks	Face	Discrete	High	Medium/High	Low
Gender	Face	Binary	High	Low	Low
Age	Face	Continuous	Low/Medium	Medium	Medium
Glasses	Accessory	Binary	Low/Medium	Low	Low
Hat	Accessory	Binary	Low	Medium	Low
Scarf	Accessory	Binary	Low	Medium	Low

- Not requiring the individual's cooperation: Consent and contribution from the subject are generally not needed.
- Preserving human privacy: The stored signatures are visually available to everyone and serve in this sense privacy.

The plethora or utilities has motivated an increasing number of research activities related to soft biometrics. We give an overview of scientific work gaining from the benefits related to soft biometrics.

Related work We here outline work, pertinent to soft biometrics. This overview does not claim to be an exhaustive state of the art, but rather a highlight selection on performed scientific studies.

Soft biometrics is a relatively novel topic and related work enfolds over several research fields. Recent work can be mainly classified in three research fields:

1. The first and largest field includes the study and identification of traits and associated image processing algorithms for classification and detection of such.
2. The second fast growing field identifies operational scenarios for the aforementioned algorithms and provides experimental results for such scenarios.
3. The third and smallest field comprises of the global and theoretical investigation of the employment of soft biometrics applications and related studies.

Scientific works belonging to the first field cover algorithms for traits such as iris pattern, see in [SBS10], or facial marks, see in [JP09].

The second field can be sub-classified in subgroups which differentiate the way soft biometrics are employed, as stand–alone systems, as pre-filtering mechanisms of bigger systems, or as fused parallel systems. Related scenarios include continuous authentication [NPJ10], video surveillance see [DFBS09], [FDL$^+$10], [MKS10], person verification [ZESH04] and moreover person identification [PJ10]. An interesting recent associated scenario for SBS based person identification is the recognition of faces in triage images of mass disaster situations [CO11].

Finally the third field involves studies on the placement of soft biometrics in applications such as forensics [JKP11] and human metrology [ACPR10].

Bag of facial soft biometrics for human identification

We consider the case where a SBS can distinguish between a set of traits (categories), which set is large enough to allow for the classification that achieves human identification. The concept of person identification based on soft biometrics originates in the way humans perform face recognition. Specifically human minds decompose and hierarchically structure complex problems into fractions and those fractions into further sub-fractions, see [Ley96], [Sim96]. Consequently face recognition performed by humans is the division of the face in parts, and subsequent classification of those parts into categories. Those categories can be naturally of physical, adhered or behavioral nature and their palette includes colors, shapes or measurements, what we refer to here as soft biometrics. The key is that each individual can be categorized in terms of such characteristics, by both humans or by image processing algorithms. Although features such as hair, eye and skin color, facial hair and shape, or body height and weight, gait, cloth color and human metrology are generally non distinctive, a cumulative combination of such features provides an increasingly refined and explicit description of a human. SBSs for person identification have several advantages over classical biometric systems, as of non intrusiveness, computational and time efficiency, human compliance, flexibility in pose- and expression-variance and furthermore an enrolment free acquirement in the absence of consent and cooperation of the observed person. Soft biometrics allow for a reduced complexity determination of an identity. At the same time though, the named reduced computational complexity comes with restrictions on the size of an authentication group. It becomes apparent that a measure of performance must go beyond the classical biometric equal error rate of the employed detectors and include a different and new parametrization. Our general interest here is to provide insightful mathematical analysis of reliability of general soft biometric systems, as well as to concisely describe the asymptotic behavior of pertinent statistical parameters that are identified to directly affect performance. Albeit its asymptotic and mathematical nature, the approach aims to provide simple expressions that can yield insight into handling real life surveillance systems.

We introduce the setting of interest, which corresponds to the general scenario where, out of a large population, an authentication group is randomly extracted as a random set of n people, out of which one person is picked for identification (and is different from all the other members of the authentication group). We note that this general scenario is consistent with both, the case of person verification as well as of identification. A general soft-biometric system employs detection that relates to λ soft biometric traits (hair color, skin color, etc), where each trait i, $i = 1, 2, \ldots, \lambda$, is subdivided into μ_i *trait instances*, i.e., each trait i can take one of μ_i values. We henceforth denote as category to be any λ-tuple of different trait-instances, and we let $\Phi = \{\phi_i\}_{i=1}^{\rho}$ define a set of all ρ categories, i.e., the set of all ρ combinations of soft-biometric trait-instances. The number of ρ, that the system is endowed with, is given by

$$\rho = \Pi_{i=1}^{\lambda} \mu_i \tag{1}$$

In this setting we elaborate on pertinent factors, such as those of the authentication group, traits, traits instances, overall categories and their interrelations. We then proceed to introduce and explain the event of *collision*, which is of significant character when employing SBSs for person identification. event where any two or more subjects belong in the same category ϕ. Focusing on a specific subject, we say that this subject experiences interference if he/she belongs in a category which also includes other subjects from the authentication group. In regards to this, we are interested in gaining insight on two probability measures. The first measure is the probability $p(n; \rho)$ that the authentication group of size n, chosen randomly from a large population of subjects, is such that there exist two subjects within the group that collide. We briefly note the relationship of $p(n; \rho)$ to the famous *birthday paradox*. For the other measure of system reliability, we consider the case where an authentication group of size n is chosen randomly from a large population of subjects, and where a randomly chosen subject from within this authentication group, collides with another member of the same group. We denote this probability as $q(n)$, and note that clearly $q(n) < p(n)$. To clarify, $p(n)$ describes the probability that interference exists, even though it might not cause error, whereas $q(n)$ describes the probability of an interference induced error. *Example:* In a group of N subjects $p(n)$ would describe the probability that any two subjects will belong to the same category ϕ_x. On the other hand $q(n)$ reflects the probability that a specific subject will interfere with one or more of the $N-1$ remaining subjects. In the following we provide a simulation of the probability of identification error, in the setting of interest, under the assumption that the errors are due to interference, i.e., under the assumptions that errors only happen if and only if the chosen subject shares the same category with another person from the randomly chosen authentication group. This corresponds to the setting where the soft-biometric approach cannot provide conclusive identification. In the simulation, the larger population consisted of 646 people from the FERET database, and the simulation was run for different sizes n of the authentication group. The probability of identification error is described in the following figure.

As a measure of the importance of each trait, Figure 1 describes the collision probability when different traits are removed. The presence of moustache and beard seem to have the least influence on the detection results, whereas hair and eye color have the highest impact on distinctiveness.

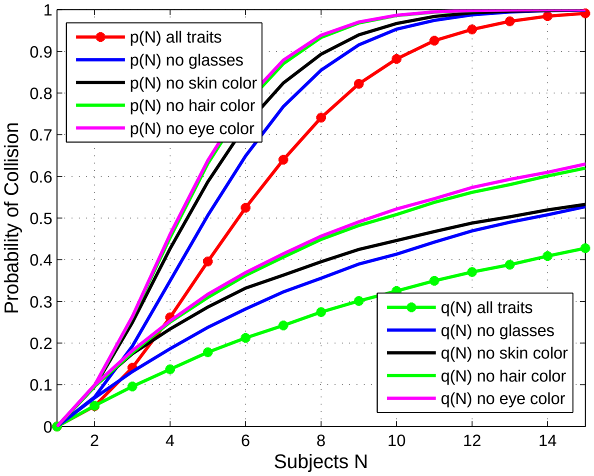


Figure 1: Collision probability in an n sized authentication group.

Furthermore we introduce the *number of effective categories* F which we identify as an important parameter related to collision, and is shown to directly affect the overall performance of an SBS. We analyze the statistical distribution and mean of F and furthermore offer an insight regarding the bounds of the statistical behavior of F over large populations. These bounds address the following practical question: if more funds are spent towards increasing the quality of an SBS, then what reliability gains do we expect to see?

We examine the influence of algorithmic estimation errors and give an example on the overall performance of a realistic SBS. We improve the performance by a study of the distribution between population in the overall categories. We then proceed to elaborate on the human compliant aspect of soft biometrics in re–identification, hereby specifically on the quantification of traits and on the human interaction view of an SBS.

Search pruning in video surveillance systems

We explore then the application using soft-biometric related categorization-based pruning to narrow down a large search.

In recent years we have experienced an increasing need to structure and organize an exponentially expanding volume of images and videos. Crucial to this effort is the often computationally expensive task of algorithmic search for specific elements placed at unknown locations inside large data sets. To limit computational cost, soft biometrics pruning can be used, to quickly eliminate a portion of the initial data, an action which is then followed by a more precise and complex search within the smaller subset of the remaining data. Such pruning methods can substantially speed up the search, at the risk though of missing the target due to classification errors, thus reducing the overall reliability. We are interested in analyzing this speed vs. reliability tradeoff, and we focus on the realistic setting where the search is time-constrained and where, as we will see later on, the environment in which the search takes place is stochastic, dynamically changing, and can cause search errors. In our setting a time constrained search seeks to identify a subject from a large set of individuals. In this scenario, a set of subjects can be pruned by means of categorization that is based on different combinations of soft biometric traits such as facial color, shapes or measurements. We clarify that we limit our use of "pruning the search" to refer to the categoriza-

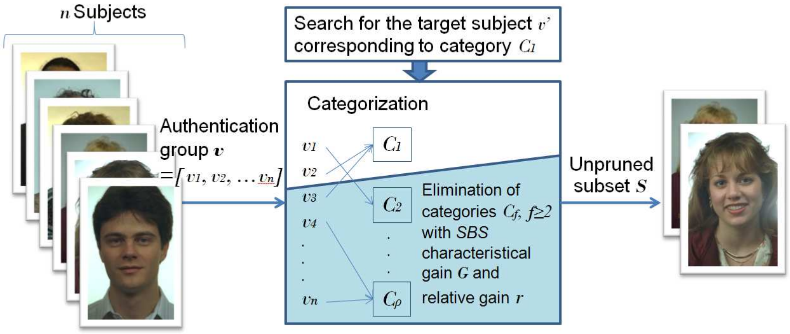

n Subjects
Search for the target subject v' corresponding to category C_1
Categorization
Authentication group $\boldsymbol{v}$ $=[v_1, v_2, \ldots v_n]$
v_1
v_2
v_3
v_4
v_n
C_1
C_2
C_ρ
Elimination of categories C_f, $f \geq 2$ with *SBS* characteristical gain G and relative gain r
Unpruned subset S

Figure 2: System overview.

tion and subsequent elimination of soft biometric-based categories, within the context of a search within large databases (Figure 2). In the context of this work, the elimination or filtering our of the employed categories is based on the soft biometric characteristics of the subjects. The pruned database can be subsequently processed by humans or by a biometric such as face recognition.

Towards analyzing the pruning behavior of such SBSs, we introduce the concept of *pruning gain* which describes, as a function of pruning reliability, the multiplicative reduction of the set size after pruning. For example a pruning gain of 2 implies that pruning managed to halve the size of the original set. We provide average case analysis of the pruning gain, as a function of reliability, but also moreover an atypical-case analysis, offering insight on how often pruning fails to be sufficiently helpful. In the process we provide some intuition through examples on topics such as, how the system gain-reliability performance suffers with increasing confusability of categories, or on whether searching for a rare looking subject renders the search performance more sensitive to increases in confusability, than searching for common looking subjects.

We then take a more practical approach and present nine different soft biometric systems, and describe how the employed categorization algorithms (eye color detector, glasses and moustache detector) are applied on a characteristic database of 646 people. We furthermore provide simulations that reveal the variability and range of the pruning benefits offered by different SBSs. We derive concise closed form expressions on the measures of pruning gain and goodput, provide simulations, as well as derive and simulate aspects relating to the complexity costs of different soft biometric systems of interest.

Frontal-to-side person re-identification

Typically biometric face-recognition algorithms are developed, trained, tested and improved under the simplifying assumption of frontal-to-frontal person recognition. Such algorithms though are challenged when facing scenarios that deviate from the training setting, such as for example in the presence of non-constant viewpoints, including the frontal-to-side scenario. Most person recognition algorithms, whether holistic or based on facial features, only manage to optimally handle pose differences that are less than about 15 degrees. As a result, a variation in the pose is often a more dominant factor than a variation of subjects. This aspect of pose variation comes to the fore in video surveillance, where a suspect may be pictured firstly frontal, whereas the corresponding test images could be captured from the side, thus introducing a *frontal-to-side recognition problem*.

Towards handling this problem, we employ multiple soft biometrics related traits. One of our tasks here is to get some insight into the significance of these traits, specifically the significance of using hair, skin and clothes patches for frontal-to-side re-identification. We are working on the color FERET dataset [Fer11] with frontal gallery images for training, and side (profile) probe images for testing. Towards achieving re-identification, the proposed algorithm first analyzes the color and furthermore texture of the three patches, see Figure 3. Then we study the intensity correlations between patches. This analysis is then followed by the construction of a single, stronger classifier that combines the above measures, to re-identify the person from his or her profile, see Figure 4.

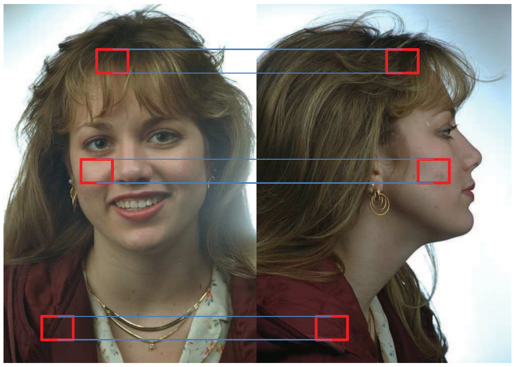


Figure 3: Frontal / gallery and profile / probe image of a subject. Corresponding ROIs for hair, skin and clothes color.

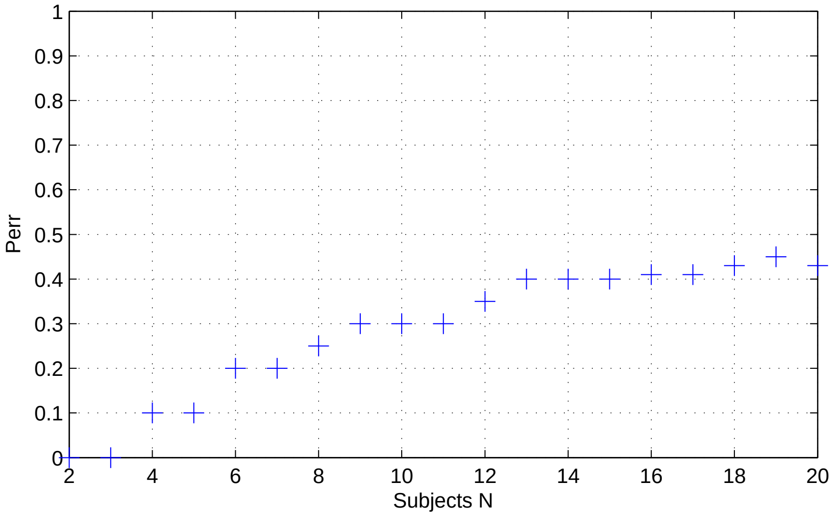


Figure 4: Overall-classifier obtained by boosting color, texture and intensity differences.

Deviating from the above security related applications, we consider then an application closer to entertainment, and specifically consider the application of soft biometrics in analyzing and quantifying facial aesthetics.

Soft biometrics for quantifying and predicting facial aesthetics

With millions of images appearing daily on Facebook, Picasa, Flickr, or on different social and dating sites, photographs are often seen as the carrier of the first and deciding impression of a person. At the same time though, human perception of facial aesthetics in images is a priori highly subjective.

We related among others soft biometric traits with this subjective human perception. In the provided study we quantify insight on how basic measures can be used to improve photographs for CVs or for different social and dating websites. This helps create an objective view on subjective efforts by experts / journalists when retouching images. We use the gained objective view to examine facial aesthetics in terms of aging, facial surgery and a comparison of average females relatively to selected females known for their beauty. Specifically we provide intuition on the role of features, image quality and facial features, see Figure 5, in human perception. We use these accumulated conclusions to construct a basic linear model that predicts attractiveness in facial photographs using different facial traits as well as image properties. We then examine and validate the designed metric. We employ the developed metric to conduct experiments and answer questions regarding the beauty index in three cases: for famous attractive females, for aging females and in case of facial surgery. Finally we proceed to simulate, based on both, the presented metric, as well as state of the art algorithmic accuracies an automatic tool for beauty prediction.

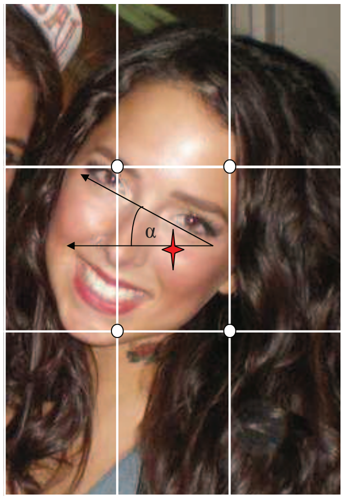


Figure 5: Example image of the web site HOTorNOT, $MOS = 9.8$. The white disks represent the stress points, the red cross the image center.

Practical implementation of soft biometrics classification algorithms

Towards practical implementation of the related concepts and ideas we develop a tool (concatenation of classification algorithms) for classification of facial soft biometric traits, where we specifically emphasize on the most obvious facial identifiers, primarily mentioned by humans, when portraying an unknown individual. The constructed tool is streamlined to achieve reliability of identification at reduced complexity, and hence focuses on simple yet robust soft-biometric traits, including hair color, eye color and skin color, as well as the existence of beard, moustache and glasses, see Table 2. We then specifically focus on extraction and categorization of eye color, and present an additional study where we illustrate the influence of surrounding factors like illumination, eye glasses and sensors on the appearance of eye color.

We create based on those traits a bag of six facial soft biometrics, see Figure 6, for which

Table 2: Table of Facial soft biometric traits

SB trait	Algorithm	Database
Skin color	Deduced from [KMB]	FERET
Hair color	Deduced from [ZSH08]	FERET
Eye color	Own developed	UBIRIS2
Beard	Own developed	FERET
Moustache	Own developed	FERET
Glasses	Deduced from [JBAB00]	FERET

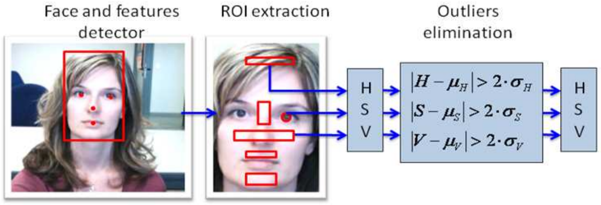

Face and features detector
ROI extraction
Outliers elimination
H
S
V
$|H-\mu_H|>2\cdot\sigma_H$
$|S-\mu_S|>2\cdot\sigma_S$
$|V-\mu_V|>2\cdot\sigma_V$
H
S
V

Figure 6: ROI for the set of facial soft biometrics. Outlier filtering was a function of the standard deviation σ and the mean μ for each of the H,S and V parameters.

estimation algorithms are featured, along with the related experimental results. We then proceed to focus on eye color as a soft biometric trait and examine an automatic eye color classifier in challenging conditions, such as changing illumination, presence of glasses and camera sensors.

User acceptance study relating to soft biometrics

Finally we conclude with a usability study that verifies the user acceptance of SBSs, specifically when compared to existing PIN or fingerprint access control systems.

The pervasiveness of biometric systems, and the corresponding growth of the biometric market see [usa11a], has successfully capitalized on the strength of biometric-based methods in accurately and effectively identifying individuals. As a result, modern state-of-the-art intrusion detection and security systems include by default at least one biometric trait. It is the case though that little emphasis has been given to better understanding user-acceptance and user-preference regarding such systems. Existing usability related works, such as in [CAJ03] and [LBCK03], focus on establishing functional issues in existing ATM machines, or on studying the influence of user interaction on the performance of fingerprint based systems (see [KED11]) and interfaces (see [RJMAS09]). Other interesting works (see [usa11b], [CG05], [CJMR09]), analyze possible methods that improve interface design. Our emphasis here is on providing insight on the attitudes and experiences of users towards novel and emerging biometric verification methods, and to explore whether such novel biometric technologies can be, in terms of user acceptance, valid alternatives to existing prevalent PIN based systems. Our focus, in addition to considering the traditional PIN-based method, is to explore the usability aspects of systems based on classical biometrics such as fingerprint and face recognition, and to then proceed to study the usability of systems based on the emerging class of soft-biometric methods. Our evaluation is based on having the users rate and rank their experiences with different access methods.

We briefly describe the user test setting, as well as the conditions and the performed test procedures. We then proceed to elaborate on the chosen verification methods and on the designed interfaces. We present the results obtained from the user study, in terms of evaluation and quantification of the different usability measurement characteristics. We provide the user test outcomes of direct comparisons between the four presented methods. Finally we draw connections to other significant traits such as cost efficiency, accuracy and processing speed, see Figure 7.

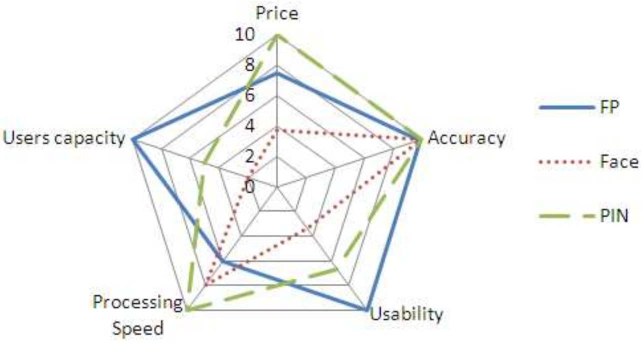


Figure 7: Comparison of fingerprint (FP), face recognition and PIN based access control systems.

We finally note that this dissertation is supported by different journal and conference publications, which are not cited throughout the thesis, but which are listed in full in Appendix **??**.

Future Work

It is becoming apparent that surveillance will increasingly affect our quality of life and security. Research in this area has been embraced by both academia and industry. For this reason, security related biometric systems will become larger and more dynamic. We see the area of soft biometrics having from now on a solid position in such systems. Towards this we will need better understanding of the component parts of such SBSs, and a corresponding better understanding of novel trait classification algorithms, as well as novel ways of combining and analyzing such algorithms. Our aim will be to allow for more efficient SBSs, but also develop a rigorous understanding of the capabilities and limits of such systems.

Our aim in the future will also be, in addition to developing novel algorithms for SBSs, to also identify and develop new commercial applications that can benefit by the power of soft biometrics.

Chapter 1

Publications

The featured list spans over all published and to be published documents of the author. None of these publications appear in the Bibliography.

Journals

A. Dantcheva and C. Velardo and A. D'Angelo and J.-L. Dugelay, "Bag of soft biometrics for person identification. New trends and challenges," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 739 - 777, 2011.

A. Dantcheva and J.-L. Dugelay, "Perception of Female Facial Beauty based on Anthropometric, Non Permanent and Acquisition Characteristics," to be submitted.

A. Dantcheva, P. Elia and J. L. Dugelay, "Human-like person re–identification using soft biometrics," to be submitted.

Conference Papers

A. Dantcheva, J.-L. Dugelay, and P. Elia, "Person recognition using a bag of facial soft biometrics (BoFSB),"*in Proc. of IEEE MMSP*, 2010.

A. Dantcheva and J.-L. Dugelay and P. Elia, "Soft biometric systems: reliability and asymptotic bounds," *in Proc. of BTAS*, 2010.

A. Dantcheva and N. Erdogmus and J.-L. Dugelay, "On the reliability of eye color as a soft biometric trait," *in Proc. of WACV*, 2011.

A. Dantcheva and J.-L. Dugelay, "Female facial aesthetics based on soft biometrics and photo-quality," *in Proc. of ICME*, 2011.

A. Dantcheva and J.-L. Dugelay, "Frontal-to-side face re–identification based on hair, skin and cloths patches," *in Proc. of AVSS*, 2011.

A. Dantcheva, A. Singh, P. Elia, J. L. Dugelay, "Search pruning video surveillance systems: Efficiency-reliability tradeoff," *in Proc. of ICCV Workshop IWITINCVPR, 1st IEEE Workshop on Information Theory in Computer Vision and Pattern Recognition in the International Conference on Computer Vision*, 2011.

A. Dantcheva, P. Elia and J. -L. Dugelay, "Gain, reliability and complexity measures in biometric search pruning based on soft biometric categorization," submitted to ICME 2011.

A. Dantcheva, J. -L. Dugelay, "User Acceptance of Access Control based on Fingerprint, PIN, Soft Biometrics and Face Recognition," submitted to ICB 2011.

M. Ouaret, A. Dantcheva, R. Min, L. Daniel, J. -L. Dugelay, “BIOFACE, a biometric face demonstrator,” ACMMM 2010, ACM Multimedia 2010, October 25-29, 2010, Firenze, Italy , pp 1613-1616.

Book Chapter

C. Velardo, J. -L. Dugelay, L. Daniel, A. Dantcheva, N. Erdogmus, N. Kose, R. Min, X. Zhao, “Introduction to biometry Book chapter of "Multimedia Image and Video Processing"” (2nd edition); CRC Press; 2011

Bibliography

[ACPR10] D. Adjeroh, D. Cao, M. Piccirilli, and A. Ross. Predictability and correlation in human metrology. In *Proceedings of the WIFS*, 2010.

[ACT11] European project actibio, 2011.

[ALMV04] H. Ailisto, M. Lindholm, S.-M. Mäkelä, and E. Vildjiounaite. Unobtrusive user identification with light biometrics. In *Proceedings of NordiCHI*, 2004.

[CAJ03] L. Coventry, A. De Angeli, and G. Johnson. Usability and biometric verification at the atm interface. In *Proceedings of ACM CHI*, 2003.

[CG05] L. F. Cranor and S. Garfinkel. *Security and usability*. O'Reilly Media, Inc., 2005.

[CJMR09] L. Coventry, G. Johnson, T. McEwan, and C. Riley. Biometrics in practice: What does hci have to say? In *Proceedings of INTERACT*, pages 920–921, 2009.

[CO11] N. Chhaya and T. Oates. Integrating soft biometrics to extract text descriptors from triage images in mass disaster situations. In *Proceedings of HST*, 2011.

[DFBS09] S. Denman, C. Fookes, A. Bialkowski, and S. Sridharan. Soft-biometrics: Unconstrained authentication in a surveillance environment. In *Proceedings of DICTA*, pages 196–203, 2009.

[FDL$^+$10] C. Fookes, S. Denman, R. Lakemond, D. Ryan, S. Sridharan, and M. Piccardi. Semi-supervised intelligend surveillance system for secure environments. In *Proceedings of IEEE ISIE*, 2010.

[Fer11] NIST database FERET, 2011.

[JBAB00] X. Jiang, M. Binkert, B. Achermann, and H. Bunke. Towards detection of glasses in facial images. *Pattern Analysis and Applications*, 3:9–18, 2000.

[JDN04a] A.K. Jain, S. C. Dass, and K. Nandakumar. Can soft biometric traits assist user recognition? In *Proceedings of SPIE*, volume 5404, pages 561–572, 2004.

[JDN04b] A.K. Jain, S.C. Dass, and K. Nandakumar. Soft biometric traits for personal recognition systems. In *Proceedings of ICBA*, 2004.

[JKP11] A. K. Jain, B. Klare, and U. Park. Face recognition: Some challenges in forensics. In *Proceedings of IEEE FG*, pages 726–733, 2011.

[JP09] A. K. Jain and U. Park. Facial marks: Soft biometric for face recognition. In *Proceedings of ICIP*, volume 1, pages 37–40, 2009.

[KBBN09] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *Proceedings of IEEE ICCV*, 2009.

[KED11] E. P. Kukula, S. J. Elliott, and V. G. Duffy. The Effects of Human Interaction on Biometric System Performance. *Lecture Notes in Computer Science*, 4561:904–914, 2011.

[KMB] P. Kakumanua, S. Makrogiannis, and N. Bourbakis. A survey of skin-color modeling and detection methods. In *Proceedings of ICPR*, volume 40.

[LBCK03] L. Little, P. Briggs, L. Coventry, and D.J. Knight. *Attitudes Towards Technology Use in Public Areas: The Influence of External Factors on ATM use*, volume 2. Lawrence Erlbaum Associates: NJ, 2003.

[Ley96] M. Leyton. *The architecture of complexity: Hierarchic systems, Symmetry, Causality, Mind*. Cambridge, MA: MIT Press, 1996.

[MKS10] D. Meltem, G. Kshitiz, and G. Sadiye. Automated person categorization for video surveillance using soft biometrics. In *Proceedings of SPIE*, pages pp. 76670P–76670P–12, 2010.

[NPJ10] K. Niinuma, U. Park, and A. K. Jain. Soft Biometric Traits for Continuous User Authentication. *IEEE Transactions on Information Forensics and Security*, 5(4):771–780, 2010.

[PJ10] U. Park and A. K. Jain. Face Matching and Retrieval Using Soft Biometrics. *IEEE Transactions on Information Forensics and Security*, 5(3):406–415, 2010.

[Rho56] H.T.F. Rhodes. Alphonse bertillon: Father of scientific detection. *Pattern Recognition Letters*, 1956.

[RJMAS09] C. Riley, G. Johnson, H. McCracken, and A. Al-Saffar. Instruction, feedback and biometrics: The user interface for fingerprint authentication systems. In *Proceedings of INTERACT*, pages 293–305, 2009.

[RN10] D. Reid and M. Nixon. Imputing human descriptions in semantic biometrics. In *Proceedings of ACM MM, Workshop on Multimedia in Forensics, Security and Intelligence*, 2010.

[SBS10] L. Stark, K. W. Bowyer, and S. Siena. Human perceptual categorization of iris texture patterns. In *Proceedings of IEEE BTAS*, 2010.

[SGN08] S. Samangooei, B. Guo, and Mark S. Nixon. The use of semantic human description as a soft biometric. In *Proceedings of BTAS*, 2008.

[Sim96] H. A. Simon. *The sciences of the artificial*. Cambridge, MA: MIT Press, 1996.

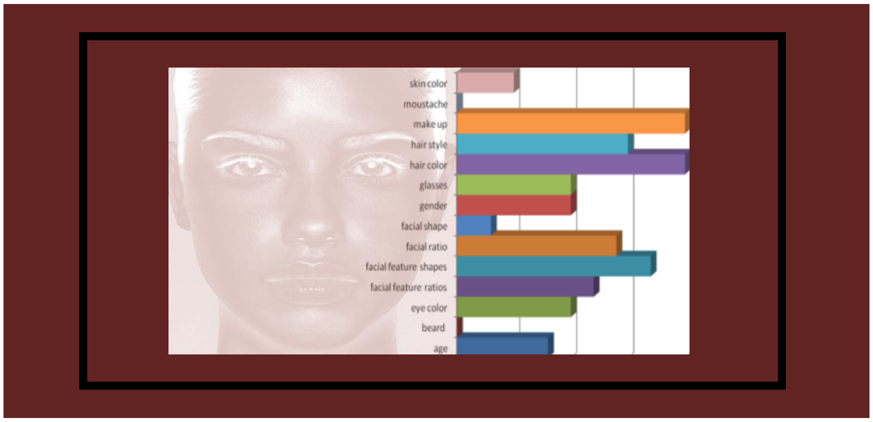
[usa11a] BCC research market forecasting, the global biometrics market, 2011.

[usa11b] NIST usability and biometrics, 2011.

[VFT$^+$09] D. Vaquero, R. Feris, D. Tran, L. Brown, A. Hampapur, and M. Turk. Attribute-based people search in surveillance environments. In *Proceedings of WACV*, 2009.

[ZESH04] R. Zewail, A. Elsafi, M. Saeb, and N. Hamdy. Soft and hard biometrics fusion for improved identity verification. In *Proceedings of MWSCAS*, volume 1, pages I – 225–8, 2004.

[ZSH08] M. Zhao, D. Sun, and H. He. Hair-color modeling and head detection. In *Proceedings of WCICA*, pages 7773–7776, 2008.



BIOMETRIES FACIALES DOUCES

METHODES, APPLICATIONS ET SOLUTIONS

Antitza DANTCHEVA

Cette thèse s’intéresse aux biométries dites douces, et notamment à leurs utilisations en termes de sécurité, dans le cadre de différents scénarii commerciaux, incluant des aspects usage. L'accent sera ainsi porté sur les caractéristiques faciales qui constituent un jeu de traits significatifs de l’apparence physique mais aussi comportementale de l’utilisateur permettant de différencier, classer et identifier les individus.

Ces traits, qui sont l'âge, le sexe, les cheveux, la peau et la couleur des yeux, mais aussi la présence de lunettes, de moustache ou de barbe, comportent plusieurs avantages notamment la facilité avec laquelle ils peuvent être acquis, mais également du fait qu’ils correspondent à la façon dont les êtres humains perçoivent leurs environnements.

Plus précisément, les traits issus de la biométrie douce sont compatibles avec la manière dont l’humain tend à catégoriser son entourage, une démarche impliquant une structuration hiérarchique des différents traits.

Cette thèse explore ces différents traits et leurs applications dans les systèmes de biométries douces (SBS), et met l’accent sur la manière dont de tels systèmes peuvent atteindre des buts différents, y compris la recherche accélérée dans des bases de données, l'identification et la ré-identification d’individus, mais également la prédiction et la quantification de l'esthétique d’un visage. Ce travail est motivé notamment par l'importance croissante de ces applications dans notre société en constante évolution, mais aussi par le côté peu contraignant du système. En effet, les SBS sont généralement non-intrusifs, et nécessitent le plus souvent de faibles temps de calculs, permettant ainsi une analyse biométrique rapide, sans imposer obligatoirement l’accord et la coopération de l’individu. Ces atouts rendent la biométrie douce indispensable dans les applications qui ont besoin de traitement d'images ou de vidéos en temps réel.

EURECOM
Sophia Antipolis

TELECOM
ParisTech