Ph.D. in Electronic and Computer Engineering
Dept. of Electrical and Electronic Engineering
University of Cagliari

Security of Multimodal Biometric Systems against Spoof Attacks

By

Zahid Akhtar

Advisor
Prof. Fabio Roli

Co-Advisors
Dr. Giorgio Fumera
Dr. Gian Luca Marcialis

XXIV Cycle
March 2012

Ph.D. in Electronic and Computer Engineering
Dept. of Electrical and Electronic Engineering
University of Cagliari

Security of Multimodal Biometric Systems against Spoof Attacks

By

Zahid Akhtar

A DISSERTATION

Submitted to
University of Cagliari
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Electronic and Computer Engineering

XXIV Cycle
March 2012

To my parents

ACKNOWLEDGEMENTS

First and above all, I would like to express my heartfelt gratitude to God, the merciful and the guardian, for blessing me with the physical, mental, and financial strength to conclude my PhD studies successfully.

I owe my deepest gratitude to my advisor, mentor, and motivator Prof. Fabio Roli for providing me the research opportunity in challenging and exciting fields of pattern recognition and biometrics. His dedication, acute perception, attention to details, commitment to quality, quest for perfection, and thirst for novelty have been true inspirations to my research and personal life. His frequent invitation of the eminent researchers for the talks and engagement in the brainstorming sessions were of great help in widening the horizons and honing my research skills. I am deeply indebted to him for guidance and suggestions that made my Ph. D. studies a memorable experience.

I am extremely grateful to Dr. Giorgio Fumera for his uncomparable coguidance and his continuous encouragement, support and help throughout all the obstacles and setbacks. He has enriched my view in research topics and taught me different aspects like how to conduct effective and efficient research, how to present and write a technical article, just to name a few. I am obliged to him for his tremendously thoughtful suggestions and guidance for my work and successful completion of this thesis. Many thanks to Dr. Gian Luca Marcialis for his timely guidance, suggestions, discussions and assistance in writing research papers. I would also thank Dr. Giorgio Giacinto and Dr. Luca Didaci for their support and aid whenever requested.

I would like to express my sincere gratitude to Dr. (Mrs.) D. C. Gharpure, University of Pune, for her timely guidance, motivation and giving the first time opportunity for me to carry out research in image processing and pattern recognition. I am eternally grateful to Dr. D. G. Kanhere, University of Pune, for motivating and helping me to come to Italy for higher studies.

I would like to thank Dr. Davide Ariu, Dr. Battista Biggio, Dr. Igino Corona, Dr. Luca Piras, Dr. Ignazio Pillai, Dr. Roberto Tronci, Dr. Biagio Freni, Daniele Muntoni, Massimiliano Dibitonto, Riccardo Satta, Zisis Bimpisidis, Dr. Jose Antonio Iglesias, Simone Sini, Paolo Denti, Luca Ghiani, Valerio Mura, Alessandro Fanti, Abdel Wassie, Shahzad Barkati, Sandeep Kale, Mohammad Rizwan, Dr. Arunkumar Walunj, Sudip Chakraborty and all my colleagues in India and Italy for their excellent support, discussion and jaunts during this period.

I would like to give very special thanks to Nasir Alfarid for his support throughout the thick and thin times and giving me advice, inspiration and motivation to achieve the best in personal and professional life, and Dr. Ajita Rattani for numerous interesting and entertaining discussions, suggestions, insights, and help.

Finally, I would like to thank my parents and all my family members for their unconditional love, support, encouragement, timely counsel and prayers. It is to them that I dedicate, everything that I am and all that is to come.

Abstract

A biometric system is essentially a pattern recognition system being used in *adversarial* environment. Since, biometric system like any conventional security system is exposed to malicious adversaries, who can manipulate data to make the system ineffective by compromising its integrity. Current theory and design methods of biometric systems do not take into account the vulnerability to such adversary attacks. Therefore, evaluation of classical design methods is an open problem to investigate whether they lead to design secure systems. In order to make biometric systems secure it is necessary to understand and evaluate the threats and to thus develop effective countermeasures and robust system designs, both technical and procedural, if necessary. Accordingly, the extension of theory and design methods of biometric systems is mandatory to safeguard the security and reliability of biometric systems in adversarial environments. In this thesis, we provide some contributions towards this direction.

Among all the potential attacks discussed in the literature, spoof attacks are one of the main threats against the security of biometric systems for identity recognition. Multimodal biometric systems are commonly believed to be intrinsically more robust to spoof attacks than systems based on a single biometric trait, as they combine information coming from different biometric traits. However, recent works have question such belief and shown that multimodal systems can be misled by an attacker (impostor) even by spoofing *only one* of the biometric traits. Therefore, we first provide a detailed review of state-of-the-art works in multimodal biometric systems against spoof attacks. The scope of state-of-the-art results is very limited, since they were obtained under a very restrictive "worst-case" hypothesis, where the attacker is assumed to be able to fabricate a perfect replica of a biometric trait whose matching score distribution is identical to the one of genuine traits. Thus, we argue and investigate the validity of "worst-case" hypothesis using large set of *real* spoof attacks and provide empirical evidence that "worst-case" scenario can not be representa-

tive of real spoof attacks: its suitability may depend on the specific biometric trait, the matching algorithm, and the techniques used to counterfeit the spoofed traits. Then, we propose a security evaluation methodology of biometric systems against spoof attacks that can be used in real applications, as it does not require fabricating fake biometric traits, it allows the designer to take into account the different possible qualities of fake traits used by different attackers, and it exploits only information on genuine and impostor samples which is collected for the training of a biometric system. Our methodology evaluates the performances under a simulated spoof attack using model of the fake score distribution that takes into account explicitly different degrees of the quality of fake biometric traits. In particular, we propose two models of the match score distribution of fake traits that take into account all different factors which can affect the match score distribution of fake traits like the particular spoofed biometric, the sensor, the algorithm for matching score computation, the technique used to construct fake biometrics, and the skills of the attacker. All these factors are summarized in a single parameter, that we call "attack strength". Further, we propose extension of our security evaluation method to *rank* several biometric score fusion rules according to their relative robustness against spoof attacks. This method allows the designer to choose the most robust rule according to the method prediction. We then present empirical analysis, using data sets of face and fingerprints including real spoofed traits, to show that our proposed models provide a good approximation of fake traits' score distribution and our method thus providing an adequate estimation of the security[1] of biometric systems against spoof attacks. We also use our method to show how to evaluate the security of different multimodal systems on publicly available benchmark data sets without spoof attacks. Our experimental results show that robustness of multimodal biometric systems to spoof attacks strongly depends on the particular matching algorithm, the score fusion rule, and the attack strength of fake traits. We eventually present evidence, considering a multimodal system based on face and fingerprint biometrics, that the proposed methodology to rank score fusion rules is capable of providing correct ranking of score fusion rules under spoof attacks.

[1] In this thesis, we will use both "security" and "robustness" terms interchangeably, to indicate performance of biometric systems against attacks.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

We human beings have an innate ability to recognize, identify, and categorize objects in a seemingly efficient, fast and effortless fashion. For instance, a child can recognize easily his best friend in a picture without experiencing any problem. Since the recognition process occurs *subliminally*, hence it is hard even for the computer scientists in conventional research paradigms to translate this process into a computer algorithm as *accurate* as human being. In other words, it is neither possible to explain nor to perceive meticulously how the recognition process works. However, Alan Turing (1912-1954), who is widely considered to be the father of modern computer science and artificial intelligence, thought that future had already arrived and in a couple of years machines would be able to think and act automatically such as understanding verbal languages or reading handwritten character (letter or number) and so forth. In a point of fact, these are still open research issues and very challenging tasks for researchers and computer scientists in the areas of pattern recognition and machine learning.

Pattern recognition or *pattern classification* can be defined as "the act of taking in raw data and taking an action based on the category of the pattern" [17]. Pattern recognition techniques are currently used in several security applications such as biometrics based person recognition, spam filtering, and intrusion detection in computer networks, with the goal to discriminate between a 'legitimate' and a 'malicious' pattern class. For example, genuine or impostor users in biometric systems. However, these tasks are different from classical pattern recognition tasks, since intelligent and adaptive adversaries (human beings) can manipulate their samples to defeat the system. For instance, biometric spoof attack using fake fingerprints. Since, classical pattern recognition techniques

do not take into account the adversarial nature of classification problems like the one mentioned above, they therefore exhibit significant performance degradation when used in adversarial settings, namely under attacks. For instance, in the following we quote a sentence from [18], related to the security of biometric systems.

> *"It is possible to have a system that is extremely accurate at distinguishing the correct person under normal conditions, but which is highly vulnerable to simple methods used to circumvent the security".*

Therefore, pattern recognition techniques have to take into account the presence of malicious adversaries, explicitly at design phase, to improve the security of the systems.

1.1 Adversarial Pattern Classification

Pattern classification is the scientific discipline whose goal is to classify the objects (samples) into a number of classes or categories. Depending on the type of application, these objects (commonly referred as *patterns*) may be any type of measurements, images or signal waveforms that need to be classified. In pattern classification, typically a set of patterns (the raw data), whose class is unknown, is given. The objective is then to devise an algorithm that assigns such patterns to one of the (possibly predetermined) classes, using some prior information. In addition, proper actions can be taken based on the outcome of the pattern classification. For instance, in fingerprint based high security access control system, when the impostor is detected, the system may decided to ring the alarm bell. A wide variety of pattern recognition, typically known as classification algorithms or *classifiers*, have been proposed for many classification tasks.

Traditionally, a classifier is designed by training it on a set of patterns (samples or feature vectors) whose true class is known– referred also as *training set* or *design set*, to find a *classification function*. When a pattern has to be classified, the classifier utilizes the acquired knowledge to assign the class to a given input pattern. The capability of the classifier, designed using the training data set, to operate satisfactorily with data outside training set is known as classifier's *generalization capability*.

The classification function can be estimated by either supervised (classification) learning or unsupervised (clustering) learning, the first one involves only labeled data (training patterns with known class labels) while the latter involves

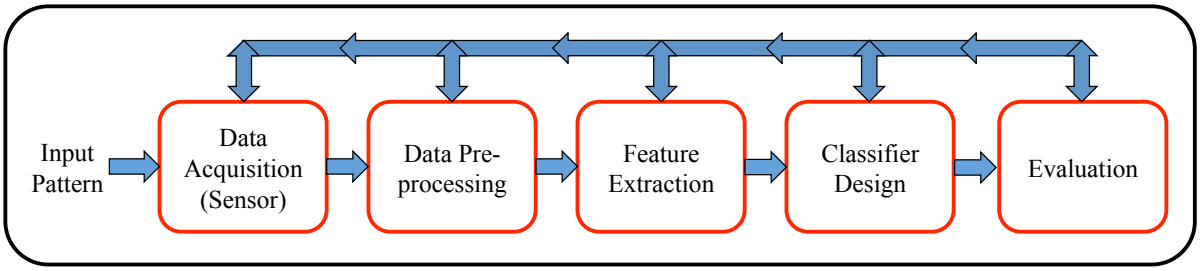

Figure 1.1: The basic stages involved in the design of a pattern classification system.

only unlabeled data. To date, many classification algorithms have been proposed in the literature, such as bayesian classifiers, neural networks, support vector machines (SVMs), decision trees and k-nearest neighbor classifiers, just to name a few. Indeed, it is clear from the literature that there is no best classifier for all types of problem. However, the simplest strategy could be to select the best performing classifier on the task at hand.

Figure 1.1 shows the various stages followed for the design of a pattern classification system. The first step is to collect a pre-processed set of training samples. The role of data pre-processing module is therefore to segment the pattern of interest from the background, remove noise and any other operation which will contribute in defining a compact representation of the pattern. Features are then extracted from each training sample. In practice, a larger than necessary number of feature candidates is generated and then the best of them is adopted. The classifier, which is chosen among different algorithms, is trained on appropriate features. Finally, once the classifier has been designed (trained), one can evaluate the performance of the designed classifier (i.e., what is the classification error rate) on test set, namely prediction of classifier's behavior on the *unseen* samples that were not present in the training set, and whose class labels are unknown as well. The feedback path allows one to go back, depending on the results, to redesign the preceding stages in order to improve the overall performance.

We have already pointed out that pattern classification techniques have been greatly implicated in several security application (e.g. biometrics) to overcome the shortcomings of classical security systems. The current surge of interest in pattern classification for security applications, however, raises a vital issue: "*are pattern classification techniques themselves secure?*". Pattern classification systems themselves in principle can be circumvented by a malicious adversary. In particular, attacks can be devised at any stage of the system (see Figure 1.2). For instance, a biometric recognition system can be attacked by an ac-

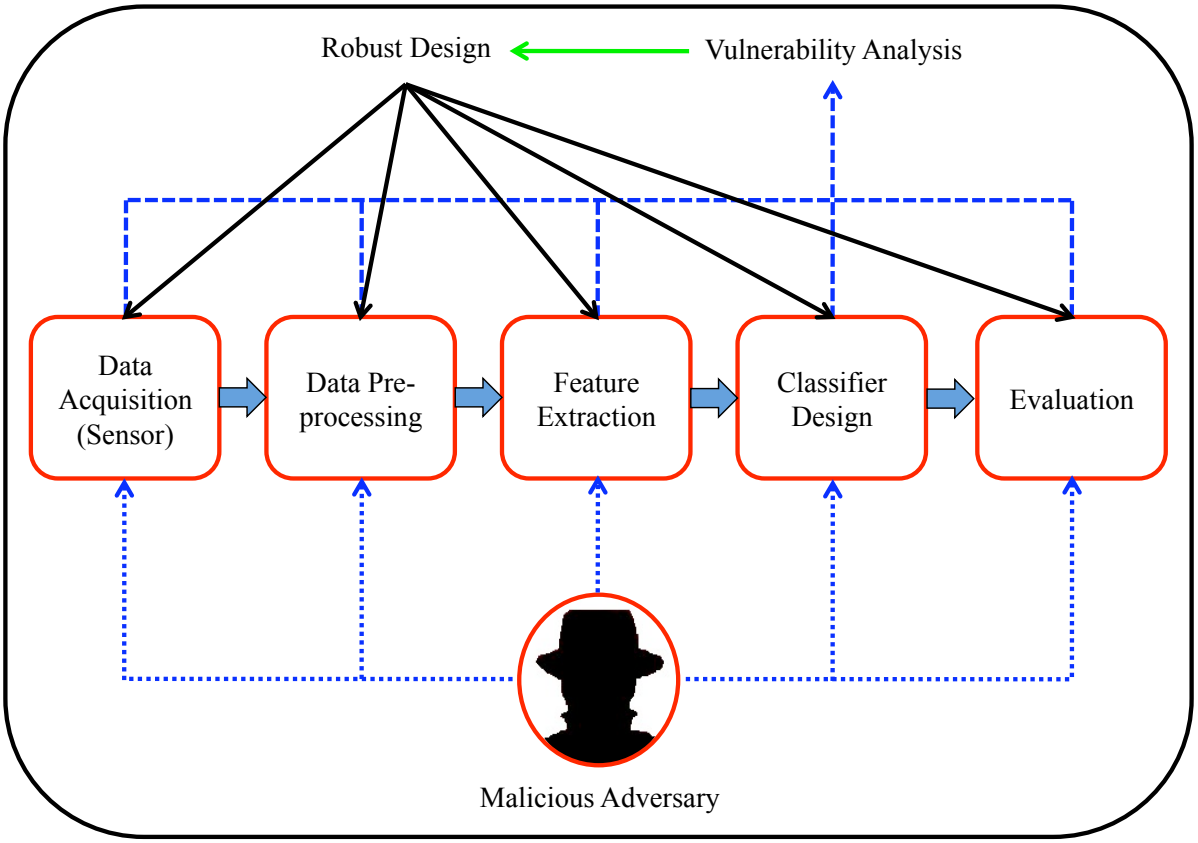

Figure 1.2: Adversaries may exploit different vulnerabilities at any stage of a pattern recognition system (adapted from [7]). Thus, the system designers should look for such vulnerabilities in advance, and propose specific countermeasures.

curate three-dimensional model of a fake fingerprint belonging to a legitimate user. In general, it is necessary to identify and understand the threat (attack) points of a pattern classification system when used in adversarial environments, so that effective technical and procedural countermeasures can be proposed.

Another significant issue that emerged as a consequence of the application of classification techniques in security tasks is related to the performance evaluation of classifiers. The traditional approaches of classifier performance evaluation may give a too *optimistic* estimate of the real performance, since they are carried out using data set that does not contain attack samples at all, or which may contain attack samples which however were not targeted against the system. Thereby, they do not provide *robustness* estimate of a classifier under attacks. For instance, a biometric recognition system is generally not tested against spoof attacks. Hence, we can say that the system which is more robust under attacks may be a better choice with respect to the system which is more accurate according to the standard performance evaluation methods, since

the performance of latter one may drop faster than former one due to attacks. In general, the design of robust pattern classification systems, namely designing of systems having optimal trade-off between accuracy and robustness under attacks, is itself an open issue.

Since, pattern classification techniques were not originally thought to be operating in adversarial settings, thus in classical approaches they are not explicitly designed from the scratch to be *secure* to address the issues like vulnerability identification, performance evaluation, and design of robust classifiers. At least in principle, the design process of a pattern classification system should consider explicitly the presence of malicious adversaries at any stage: ranging from data acquisition to classification, including feature extraction and selection, and performance evaluation. This approach, as shown in Figure 1.2, is usually referred to as *security by design* in security engineering.

In this thesis, we thoroughly investigate the open issues raised from the application of pattern recognition systems in adversarial environments in the context of biometric recognition systems.

1.2 Biometrics

We all human beings use our natural abilities to recognize an individual through their face, voice, gait and other characteristics. Whereas, computers require programming algorithms in order to recognize an individual using the same observable information. Technological advances are promising to close the gap between human perception and computer recognition. Especially, *biometrics* which relies on measuring a variety of anatomical, physiological and behavioural characteristics and matching to measurements that were previously collected from the person, thus recognizing him using distinctive personal traits such as fingerprint, face, voice and so forth. For instance, fingerprint based recognition system, the person must place his finger on a fingerprint sensor whenever he wants to log in. The sensor will capture the fingerprint image he provides and will then match it to previously collected fingerprint measurements. If the latest fingerprint measurement matches closely enough, the system acknowledges that the genuine person is present and logs him in or grants him access. It is worth noting that the person has no device to lose or password to forget: he can authenticate himself as long as his fingerprint (biometrics characteristics) hasn't been badly injured or degraded. Biometrics thus can provide one of the most substial benefits to the security arena in guarding against attempts to establish fraudulent multiple identities or prevent identity fraud. As,

biometrics can not be shared due to being an intrinsic property of an individual, therefore making it possible to know automatically who did 'what', 'where' and 'when'. Anyway, the principal goal of the use of biometrics is to attain the capability of accurately recognizing individuals with greater reliability, convenience, speed and lower cost.

Depending on the context, a biometric system can be used either in a verification mode or an identification mode. In verification (*Am I who I claim I am?*) mode, a person's claimed identity is confirmed based upon validating a sample collected against a previously collected biometric sample for that individual. On the other hand, in identification (*Am I who I claim I am?* or *Who am I?*) mode, the system has to recognize a person based upon comparison of biometrics collected against a database of previously collected samples of N individuals.

Unimodal biometric systems [48] perform person recognition based on a single source of biometric information, e.g., single fingerprint or face. These systems are contended with problems like noise in sensed data, non-universality, lack of individuality of the chosen biometric trait, absence of an invariant representation for the biometric trait and susceptibility to circumvention. Some of the limitations imposed by unimodal biometric systems can be alleviated by using *multimodal* biometric systems [76] that consolidate evidence from multiple biometric sources. For example, a multimodal biometric system could be a combination of iris recognition and fingerprint recognition to confirm the identity of a user. The integration of two or more types of biometric information sources, which basically takes advantage of the proficiency of each individual biometric, helps to meet stringent performance requirements, reliability and robustness against attacks. It is worth noting that multimodal biometrics systems are progressively becoming common in biometric applications deployment.

In a multimodal biometric system information fusion can take place at different levels: sensor level, feature level, score level, rank level and decision level. These levels of fusion can be broadly categorized into fusion prior to matching and fusion after matching. Fusion at the score level, which is also adopted in this thesis, is the most common approach since it offers the best trade-off between the information content and the ease in fusion.

As a general rule, if a biometric system is made by human beings, it can be defeated by human beings. Since, biometric traits are not secret, hence they can be captured, copied and replayed. For instance, we human beings often leave measurable traces of our biometric traits wherever we go, like fingerprints on surfaces, the recorded sound of voice, or even video records of face and body.

This "latency" provides a way for attackers to generate a bogus biometric trait and use it to trick the system into thinking that the genuine user is actually present. Moreover, it may be possible to intercept a legitimate biometric trait collected from genuine user and replay it later. This act of submitting an artifact to a biometric system, where a person pretends to be another person, in order to fool the system is known as *spoof attack* [36, 78].

Among all the possible threats, spoof attack is a fatal threat for biometric authentication systems. Spoof attacks have a great practical relevance because they don't require advanced technical skills and, therefore, the potential number of attackers is very large. The classical countermeasure against spoof attacks is "*liveness*" or *"vitality"* detection [43, 78], which aims at detecting physiological signs of life in the biomsetric trait, using hardware devices embedded into the sensor or signal processing algorithms, to ensure that a sample is being acquired from live human being and not simply a prosthetic one.

Multimodal biometric systems, apart from liveness detection methods, are also considered a natural anti-spoofing technique. It is commonly believed that multimodal systems are more robust to spoof attacks than unimodal systems. This claim was done on the basis of the intuitive assumption, rather than on any theoretical or empirical evidence, that the intruder needs to spoof *all* fused biometric traits *simultaneously* to evade them [78, 34].

However, recent works have questioned the security of multimodal biometric systems against spoof attacks. It has been shown empirically that multimodal systems can be cracked by spoofing *only one* of the biometric traits. However, This result was obtained under a restrictive "worst-case" scenario, where the attacker is able to perfectly replicate a genuine biometric trait whose matching score distribution is identical to the one of genuine traits. Thereby, this raises the issue of investigating more thoroughly the security of multimodal biometric systems against spoof attacks and devising new methods to design robust systems against them, which is still an open issue.

To sum up, in this section we introduced the biometrics and it's major issue of security against spoof attacks, which will be described in the next two chapters in detail. In this thesis, we thoroughly investigate the open issues raised from the application of biometric recognition systems in such adversarial environments. We in fact propose different techniques to address the problem of performance evaluation of biometric systems under adversarial settings. We summarize our contributions in the next section.

1.3 Outline and Goals of this Thesis

In the previous sections, we provided an overview of pattern classification, adversarial pattern classification and biometrics. We also pointed out the security issues of pattern recognition techniques in adversarial environments, which are thoroughly discussed and investigated in the following chapters in the context of multimodal biometric systems. The main contributions of this thesis are summarized below.

- We first provide a state-of-the-art of works on multimodal biometric systems in adversarial settings. In particular, we present a review of spoof attacks on multimodal biometric systems according to the main issues, namely,
  - identifying vulnerabilities of multimodal biometric systems to spoof attacks, which can be exploited by an attacker to make them ineffective;
  - evaluating the security of multimodal biometric systems against spoof attacks;
  - designing of robust multimodal biometric systems against spoof attacks.

  Multimodal biometric systems are commonly believed to be more robust to spoof attacks than unimodal systems. However, the review shows, contrary to a common belief, that a multimodal biometric system can be evaded by an impostor even by spoofing *only one* biometric trait. Nevertheless, this conclusion was obtained under a stringent hypothesis of a "worst-case" attack, where the attacker is able to replicate perfectly the genuine biometric traits, which was obtained by *simulation* under consideration that the matching score distribution of fake traits is identical to the one of genuine users. This hypothesis also allows one to evaluate the robustness of multimodal biometric systems against spoof attacks without the need of actual fabrication of fake biometric traits. A substantial increase of the false acceptance rate (FAR) of multimodal systems under spoof attacks was indeed highlighted. Thus, these results raise the issue of investigating more thoroughly the security of multimodal systems against real spoof attacks, namely their performance degradation in the non-worst case scenario when the attacker is not able to fabricate exact replica of genuine

biometric traits and devising accordingly new methods to design robust systems against attacks, which is still an open issue.

- We then investigate the vulnerability of multimodal biometric systems to real spoof attacks in accordance with the common beliefs about its robustness to attacks.

  Such empirical evaluation allow us to address open issues such as to what extent the drop of performance under the "worst-case" attack scenario is representative of the performance under real spoof attacks, whether multimodal systems can be more robust than each corresponding unimodal systems, even in the case when all biometric traits are spoofed and whether multimodal systems can be cracked by spoofing all the fused traits, even when the attacker is not able to fabricate exact replicas of the genuine user's traits.

- Next, we propose a general methodology to evaluate the security of biometric systems against spoof attacks.

  A straightforward approach to evaluate the security of a biometric system against spoof attacks could be to fabricate fake biometric traits and present them to the system. However, constructing fake biometric traits exhibiting various quality degrees is cumbersome and impractical task. We thus propose a security evaluation method that does not require fabrication of fake traits. This method evaluates the security under a simulated spoof attack using models of the fake score distribution that take into account explicitly also scenarios more realistic than the worst-case one, namely different degrees of the quality of fake biometric traits. The main feature of the method is that it can be applied to any multimodal system, namely, to any set of matchers combined with any score fusion rule, and it allows to simulate a spoof attack against any subset of the component matchers.

- We eventually propose extension of our security evaluation method to rank the biometric fusion rules according to their robustness against attacks.

  The main objective is to rank the several state-of-the-art score level fusion rules according to their relative robustness against attacks. This can provide information to the designer of a multimodal system to choose a score fusion rule, taking also into account its robustness to attacks.

This thesis is structured as follows. In the next chapter, we introduce biometrics together with its challenges and limitations. In addition, we also present

an insight to unimodal and multimodal biometric systems. In Chapter 3, we closely analyze the problem of attacks on biometric systems, and report a critical review of state-of-the-art works on multimodal biometric systems against attacks. In Chapter 4, we experimentally evaluated the performance of multimodal biometric systems against real spoof attacks, namely its vulnerability to spoof attacks. In Chapter 5, we describe the proposed methodologies, first to evaluate the performance of biometric systems against spoof attacks and second to rank biometric fusion rules under attacks. Subsequently, an extensive experimental analysis is also reported. Concluding remarks and new open research issues are eventually discussed in Chapter 6.

Chapter 2

Biometric Systems

2.1 Introduction

The need for dependable identity management system has expanded in the wake of enhanced concerns about security and to combat the exponential growth in identity theft. Conventional methods to establish the identity of a persons are knowledge-based security: "what you know", such as password, personal identification, and token-based (possession-based) security: "what you have", like ID card, physical key. However, surrogate representation of identity like passwords can be easily divulged using dictionary attacks [44] and social engineering [60], and ID cards can be misplaced, shared or stolen [41] by an impostor to gain unauthorized access, thus defeating the system security. In addition, significant authentication tasks such as detecting multiple enrollments and non-repudiation can't be attained by password- or ID card-based authentication systems. Thus, reliable and user-friendly identity management systems with robust authentication techniques based on "who you are", namely biometrics, are needed.

Biometric systems substantiate the identity of a person using his biological or behavioral characteristics such as face, fingerprint, iris, palmprint, hand geometry, voice, signature, and gait etc., which are also known as biometric modalities or traits. Some of the most widely used biometric traits by the identity management systems are shown in Figure 2.1.

The prime advantage of biometric systems compared to conventional identification methods is replacing "what the person carries" and "what the person remembers" paradigms with "who the person is", thus preventing identity fraud by using biometrics patterns that are claimed to be unique, permanent and hard

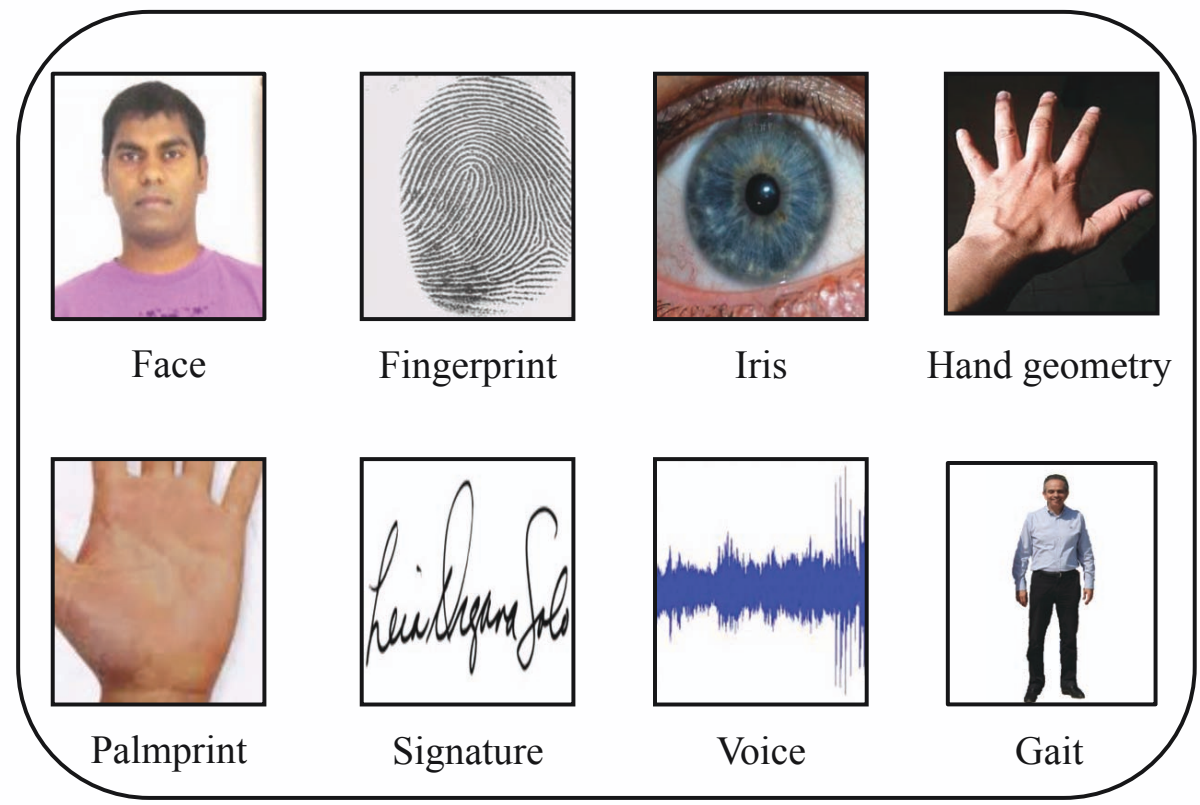

Figure 2.1: Examples of the physiological and behavioural body traits that can be used for biometric recognition.

to forge. Since, the person is required to be present at the time of the authentication by the biometric system, thus also preventing false repudiation claims. In order to combat increasing security threats in information era, governments, industries and academies have largely adopted and encouraged research on biometric authentication systems. The most well-known large scale biometric systems are US-VISIT program [85], the Schiphol Privium scheme at Amsterdam's Schiphol airport [88], the fingerprint based biometric system at Disney World, Orlando [31] and Atuomated Fingerprint Identification System (IAFIS) [87].

2.2 Basic Structure of a Biometric System

A generic biometric system consists of five main modules: a sensor module; a quality assessment and feature extraction module; a matching module; Decision making module; a system database module, as described below.

1. **Sensor module:** A suitable biometric sensor/scanner is applied to attain the raw biometric data of an individual. For instance, an optical sensor can be used to acquire the fingerprint images by capturing the friction ridge

structure of the finger. Since, sensor module determines the interaction of the human with the system, thus playing a pivotal role in the performance of the biometric systems.

2. **Quality assessment and feature extraction module:** A quality assessment algorithm is used, in order to determine the suitability of the biometric data for the subsequent processing. If the quality is inadequate then the biometric sample is rejected and reacquired. If the quality assessment algorithm is not incorporated then the acquired data is subject to signal enhancement algorithm to improve it's quality. The biometric sample is then processed to glean a set of salient discriminatory features. The extracted feature set procured during enrollment is stored in the database, referred as *template*, thereby constituting the identity of an individual.

3. **Matching module:** To verify the identity of an individual, the extracted feature set from the biometric sample (known as query or input or probe) is compared against the enrolled template to generate the match score, which determines the amount of similarity (similarity score) or distance (distance score) between the two feature sets. The system conducts a one-to-one comparison to verify a claimed identity, while the comparison is one-to-many to determine an identity.

4. **Decision making module:** Decision making module uses match score, to validate a claimed identity in the verification task or to provide a ranking of the enrolled identities to identify an individual in the identification task. In order to determine the authenticity of an individual, generally, the match score is compared to a predefined threshold. The identity of an individual is verified successfully, if the match score of the query is equal or higher than the threshold for "similarity score", while equal or lower for "distance score".

5. **System database module:** The system database, which acts as the depository of biometric information, is used to store the extracted feature set from the raw biometric sample (i.e., *template*), along with some biographic information (such as name, Personal Identification Number (PIN), address, etc.) characterizing an individual.

2.3 Verification and Identification

Depending on the application context, a biometric system may operate in the following two modes: verification mode or identification mode.

2.3.1 Verification mode

In the verification mode, the system validates the authenticity of a claimed identity by comparing the captured biometric data with her own biometric template stored in the system database, thus conducting a one-to-one comparison to determine whether the claim is true or not. Verification can be posed as the following two class classification problem: if the degree of similarity betwen the input sample and template of the claimed identity is above a predefined threshold, then the claim is classified as "genuine". Otherwise, the claim is rejected and the user is considered an "impostor". Commonly, verification is used for positive recognition, which aims to prevent multiple people from using the same identity.

2.3.2 Identification mode

In the identification mode, the system recognizes an individual by comparing the user's biometric input with the templates of all the persons enrolled in the database, thus conducting a one-to-many comparison to establish an individual's identity. The system outputs either the identity of the user having highest degree of similarity between his templates and the input sample or the decision that the user presenting the input is not an enrolled user. Identification is used for negative recognition, which aims to establish whether the person is who he/she denies to be [36], therefore preventing a single person from using multiple identities. Figure 2.2 shows the enrollment, and authentication stages of a biometric system operating in the verification and identification modes.

2.4 Performance Evaluation of a Biometric System

Due to the several factors such as imperfect sensing conditions (e.g., sensor noise), changes in the user's biometric characteristic (e.g., face aging), variations in ambient conditions (e.g., inconsistent illumination levels in iris recognition) and improper interaction with the sensor (e.g., occluded face), the two extracted feature sets originated from the same biometric trait of an individual seldom correspond to each other. Thus, the biometric sample matching is never

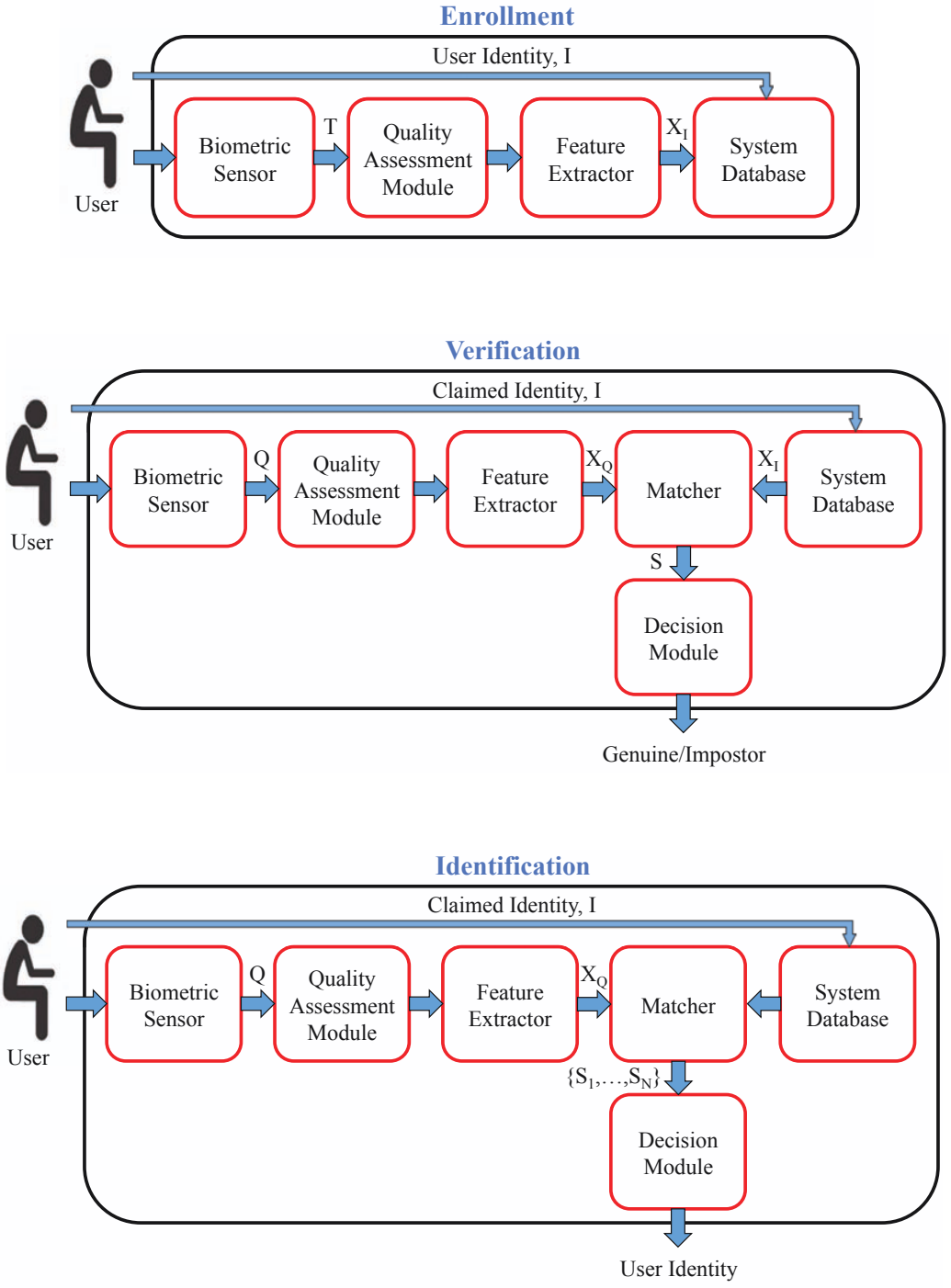

Figure 2.2: Enrollment, verification, and identification stages in a biometric system. Here, T represents the biometric sample obtained during enrollment, Q is the query biometric sample obtained during recognition, while X_I and X_Q are the template and query feature sets, respectively. S represents the match score and N is the number of users enrolled in the database.

100% perfect, unlike conventional authentication methods, such as password-based systems, where a perfect match between two alphanumeric strings is

mandatory to validate a user's identity. As a result, the performance of a biometric system is never 100% accurate. The variation evidenced in the biometric feature set of a user is known as intra-class variation, while it is called inter-class variation when observed in the feature sets originating from two different users [38].

A genuine match score is obtained by comparing two samples of the same biometric trait of a user, whereas the impostor match score is acquired by comparing two biometric samples originating from different users [36]. A biometric verification system can make two types of errors: Type I and Type II. Type I error, also known as false reject or false non-match, occurs when a genuine match score falls below the operating threshold (s^*), therefore the genuine user is rejected. Type II error, also referred as false accept or false match, occurs when an impostor match score exceeds the operating threshold (s^*), therefore the impostor is wrongly accepted.

The performance of the system, at specified operating threshold value, is evaluated in terms of the false acceptance rate (FAR) (or, the False Match Rate (FMR)): the fraction of impostor scores exceeding the threshold, and false rejection rate (FRR) (or, the False Non Match Rate (FNMR)): the proportion of genuine scores falling below the threshold.

Let S, $f_{genuine}(s) = p(S = s|genuine)$ and $f_{impostor}(s) = p(S = s|impostor)$ be the similarity match score, and probability density functions of the genuine and impostor scores, respectively. The FAR and FRR of the biometric system are estimated as follows:

$$FAR(s^*) = p(S \geq s^*|impostor) = \int_{s^*}^{\infty} f_{impostor}(s)ds \tag{2.1}$$

$$FRR(s^*) = p(S < s^*|genuine) = \int_{-\infty}^{s^*} f_{genuine}(s)ds \tag{2.2}$$

If the operating threshold is increased, FAR will decrease but the FRR will increase and vice versa. The Genuine Accept Rate (GAR) is the percentage of genuine scores exceeding the operating threshold (s^*). Therefore,

$$GAR(s^*) = p(S \geq s^*|genuine) = 1 - FRR(s^*) \tag{2.3}$$

The overall performance of a biometric system can be summarized by regulating the different values of threshold and computing the corresponding FAR and FRR. Plots like Detection Error Tradeoff (DET Curve) [58], where FRR is plotted as a function of FAR on a normal deviate scale, and Receiver Operating Characteristic (ROC) curve [19], where the GAR is plotted as a function of

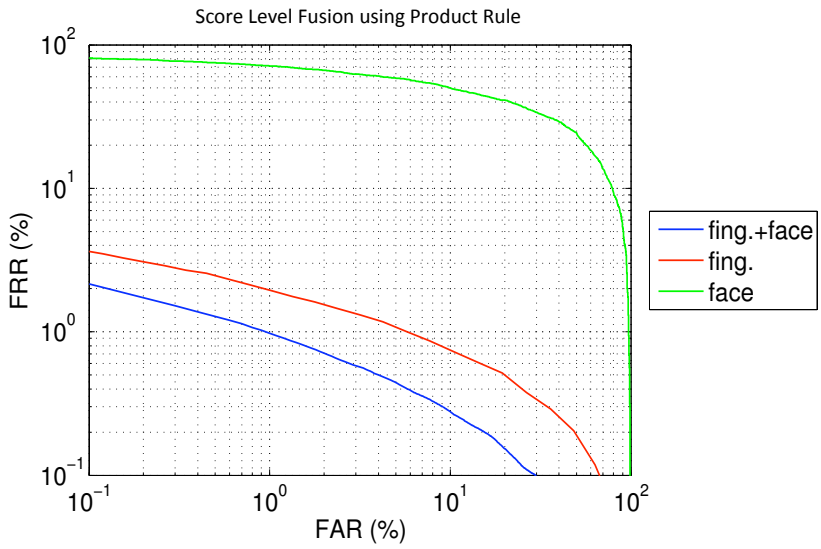

Figure 2.3: DET curves obtained on a face and a fingerprint matcher and using score level product fusion rule.

FAR on linear, logarithmic or semi-logarithmic scale, are used to represent the complete performance curve. Equal Error Rate (EER) is the location on a DET or ROC curve where the FAR equals the FRR. A lower EER value indicates better performance. Figure 2.3 shows an example of DET Curve.

2.5 Limitations of Unimodal Biometric System

Some of the main factors affecting the accuracy of the unimodal biometric systems [40] are as follows:

1. **Noise in sensed data:** Noise in the acquired biometric sample may result from defective and improperly maintained sensors or unfavorable ambient conditions. For instance, accumulation of dirt or the residual remains on a fingerprint sensor may result in a noisy fingerprint image as shown in Figure 2.4. Noisy biometric sample may not be successfully matched, for genuine users, with their respective templates in the database or may be incorrectly matched with the impostors, thus leading to a significant reduction in the performance of the system [27, 92].

2. **Intra-class variations:** Intra-class variations in biometric samples are typically produced by the user's inappropriate interaction with the sensor

Figure 2.4: Example of a noisy fingerprint image.

(e.g., incorrect facial pose - see Figure 2.5), changes in the environmental conditions (e.g., illumination changes), use of different sensors during enrollment and verification, or temporal variation in the biometric traits such as aging [47]. Large intra-class variations usually decrease the genuine acceptance rate (GAR) of a biometric system.

3. **Inter-class similarities:** Inter-class similarity is defined as the overlap of the biometric samples, in the feature space, corresponding to multiple classes or individuals. The lack of uniqueness in the biometric feature set leads to an increase in the false acceptance rate (FAR) of the system. Hence, there is an upper bound on the number of unique individuals that can be accommodated by the biometric system.

4. **Non-universality:** Universality means that every person using a biometric system is able to present the respective biometric trait. The biometric system may not be able to extract meaningful biometric data from a subset of users. For example, the National Institute of Standards and Technology (NIST) has reported that it is not possible to extract correct minutia features from the fingerprints of two percent of the population (manual workers with many cuts and bruises on their fingertips, people with hand-related disabilities etc.), due to the poor quality of the ridges [89]. This contributes to an increase in the failure to enroll (FTE) rate. Hence, no biometric trait is truly universal.

Figure 2.5: Intra-class variation associated with an individual's face images. Although the images belong to the same individual, an appearance-based facial recognition system is unlikely to match these three images successfully, because of change in pose [35].

5. **Interoperability issues:** Most biometric systems are designed and operated under the assumption that the biometric sample to be compared are obtained using the same sensor and, hence, are restricted in their ability to match or compare biometric samples originating from different sensors.

6. **Spoof attacks:** Biometric spoof attack is the deliberate attempt to manipulate one's biometric traits in order to avoid recognition, or the creation of physical biometric artifacts in order to take on the identity of another person.

Thereby, unimodal biometric system are not sufficient to meet the variety of requirements, including matching performance, imposed by several large-scale authentication systems. The limitations of the unimodal biometric system can be mitigated by one of the standard solutions known as multimodal biometric systems.

2.6 Multimodal Biometric Systems

Systems that consolidate the evidences from two or more biometric traits, in order to more reliably determine the identity of an individual, are known as multimodal biometric systems [37]. Since, different biometric traits usually compensate for the inherent limitations of the other traits, therefore multimodal biometric systems can alleviate many limitations of the traditional unimodal systems [34]. Multimodal biometric systems offer the following advantages over unimodal systems:

1. Significant improvement in the overall accuracy can be attained by combining the biometric evidences obtained from different sources using an

effective fusion technique. The use of multiple biometric sources increases the dimensionality of the feature space, thus reducing inter-class similarities.

2. The effect of noisy input data causing performance degradation can be mitigated by the fusion of multiple biometric sources. For instance, in a face and a fingerprint matcher based system; if the face sample obtained is not of sufficient quality during a particular acquisition, then fingerprint samples may still provide sufficient discriminatory information to enable reliable decision-making.

3. Multimodal biometric systems are also able to address the non-universality problem and helps to decrease the failure to enroll rate (FTER) and failure to capture rate (FTCR). For example, if an individual is not able to enroll in a fingerprint system due to burns or cuts etc., he can still be identified using other biometric traits like iris etc.

4. Multimodal biometric systems can also proffer "degrees-of-freedom" in user authentication. Such as, during enrollment face, fingerprint and iris were captured; later, during authentication any combination of these traits may be acquired, depending on the nature of the application or the convenience of the user.

Therefore, a properly designed multimodal biometric systems can improve accuracy and reliability of unimodal systems, with the increase in population coverage. Extensive empirical evidences have shown that they are effective to this aim [76, 40, 71, 72, 77, 68].

One of the fundamental issues in designing of a multimodal biometric system is to determine the type of information that should be fused. The information fusion can be carried at various levels: sensor level, feature level, score level, rank level and decision level, as described below. Conventionally, the availability of the information content decreases from the sensor level to the decision level.

1. **Sensor level:** The raw data acquired from multiple sensors are combined in sensor level fusion before they are subjected to feature extraction [20]. In this type of fusion, the multiple cues must be compatible; hence usually fusion of the same biometric trait, obtained either using a single sensor or different compatible sensors, is carried out. For example, the fingerprint impressions obtained from optical and solid state sensors can be combined

to form a single image to be input to the feature extraction and matching modules.

2. **Feature level:** Feature level fusion refers to consolidating the evidence presented by two biometric feature sets of the same individual. The two feature sets are concatenated to form a single feature set to compare with the enrollment template in the system database, which itself is a concatenated feature set.

3. **Score level:** In score level fusion, feature sets are extracted independently by each subsystem, which are later compared with separately stored respective templates. Depending on the proximity of feature set and the template, each subsystem computes its own match score. The individual scores are finally fused to produce a single match score for decision-making process.

4. **Rank level:** This type of fusion is conducted in identification mode, where each subsystem associates a rank with each enrolled identity. Thus, the rank level fusion schemes consolidate the ranks produced by the individual subsystems in order to derive a consensus rank for each identity in order to establish the final decision.

5. **Decision level:** A decision level, also known as abstract level, fusion is carried out by combining the authentication decision made by individual biometric matchers. Fusion at the decision level is too rigid, since only limited information is available at this level.

As mentioned above one of the most fundamental issues in the multimodal biometric systems is to determine the type of information that should be consolidated by the fusion module. Since, the amount of informations goes on decreasing as one proceeds from sensor level to decision level, therefore mutlimodal biometric systems that fuse information at at early stages of processing are expected to yield more promising results than the systems that fuse the information at later stage. There has been a proliferation of works discussing different fusion schemes to integrate multiple sources of biometric information at different levels. Usually, the benefits of fusion technique are exploited when individual sources of information show complementary nature. Large performance disparity between component sources may dilute the performance of the "stronger" source [14].

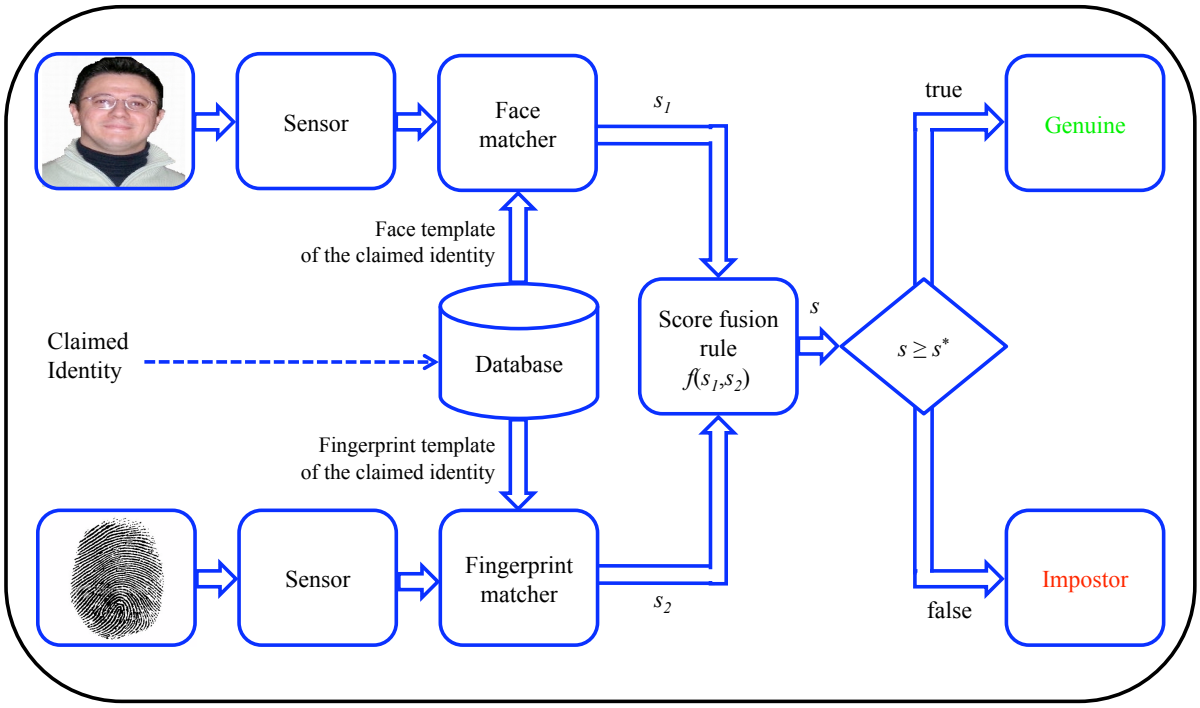

Figure 2.6: A multimodal biometric system made up of a fingerprint and a face sensor, whose match scores are combined through a fusion rule.

Fusion at the match score level has been extensively studied in the literature [76, 22, 55, 75, 67, 32, 83, 91, 74, 73, 42] and is also the dominant level of fusion in multimodal biometric systems, due to ease in accessing and combining the match scores. Therefore, in this thesis we also adopted fusion at the match score level. Figure 2.6 illustrates the architecture of a multimodal biometric system, with reference to the one considered in this thesis, namely a multimodal system composed of a face and a fingerprint matcher. Such systems operates as follows: At the design phase, genuine users are enrolled into the system, by storing their biometric traits (templates) in a database together with the corresponding identities. At authentication phase, the user provides his face and fingerprint to the respective sensors, and claims his identity. Then, the biometric traits are individually processed and compared with the correspondent template of the claimed identity, and provides a real-valued match score (denoted here as s_1 and s_2, respectively, for the face and the fingerprint matcher): higher the score, higher is the similarity. Lastly, the match scores are combined according to a given fusion rule which outputs a new real-valued score $f(s_1,s_2)$: the claimed identity is accepted and the person is classified as a genuine user, if $f(s_1,s_2) \geq s^*$; otherwise, it is classified as an impostor. The term s^* is an acceptance threshold that must be set during design process according to ap-

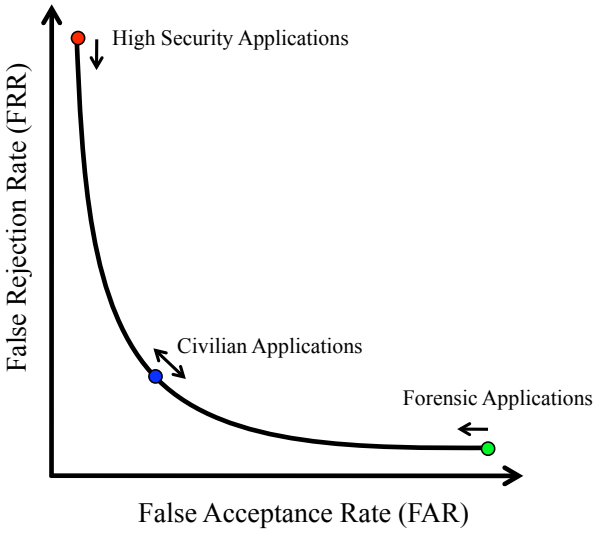

Figure 2.7: Typical operational regions of different biometric applications.

plication requirements in terms of false acceptance (FAR) and false rejection (FRR) rates. Figure 2.7 shows a typical DET curve for a multimodal biometric system. The highlighted points on the curve show the operational regions for different application scenarios.

2.7 Summary

Rapid advancements in computer networking, communication, and mobility together with enhanced concerns about security and identity theft has resulted in a pronounced need for reliable authentication systems. Conventional identity management schemes based on passwords or ID cards are limited in their ability to meet the security requirements. Some of the limitations of conventional authentication methods can be addressed by biometrics, that uses the physical and behavioral characteristics of a person, such as fingerprint, face, iris etc., to establish the identity. As a result, biometric systems are being deployed in various applications including travel and transportation, financial institutions, health care, law enforcement agencies and border crossing, thus enhancing security and reducing identify fraud. Unimodal biometric systems (system using only one biometric trait) suffer from several factors such as noisy data, large intra-class variations, and improper user interaction etc. Some of the limitations of unimodal biometric systems can be mitigated by multimodal biometric systems (systems using more than one biometric trait). A systematically designed multimodal biometric system can increase matching accuracy and population coverage in comparison to the unimodal system. In multimodal systems, the ev-

idence presented by multiple biometric sources can be consolidated at various levels: sensor level, feature level, score level, rank level and decision level. Fusion at the score level has received the maximum attention from the biometrics research community, due to ease in accessing and combing the match scores. Score fusion in a multimodal biometric verification system can be formulated as a two-class classification problem: genuine and impostor classes. Also, in this thesis, score level fusion based multimodal biometric system, made up of a face and a fingerprint matcher, has been adopted to study the issues of robustness against spoof attacks.

Chapter 3

Security of Biometric Systems

3.1 Introduction

In spite of many advantages, biometric systems like any other security applications are vulnerable to a wide range of attacks. An attack on a biometric system can take place for three main reasons:

1. A person may wish to disguise his own identity. For instance, An individual/terrorist attempting to enter a country without legal permission may try to modify his biometric trait or conceal it by placing an artificial biometric trait (e.g. a synthetic fingerprint, mask, or contact lens) over his biometric trait. Recently, in January 2009, the Japanese border control fingerprint system was deceived by a woman who used tape-made artificial fingerprints on her true fingerprints [90].

2. An attack on a biometric system can occur because an individual wants to attain privileges that another person has. The impostor, in this case, may forge biometric trait of genuine user in order to gain the unauthorized access to systems such as person's bank account or to gain physical access to a restricted region.

3. A benefit to sharing biometric trait may be the cause to attack the biometric systems. Someone, for instance, can establish a new identity during enrollment using a synthetically generated biometric trait. Thus, sharing the artificial biometric trait leads to sharing that fraudulent identity with multiple people.

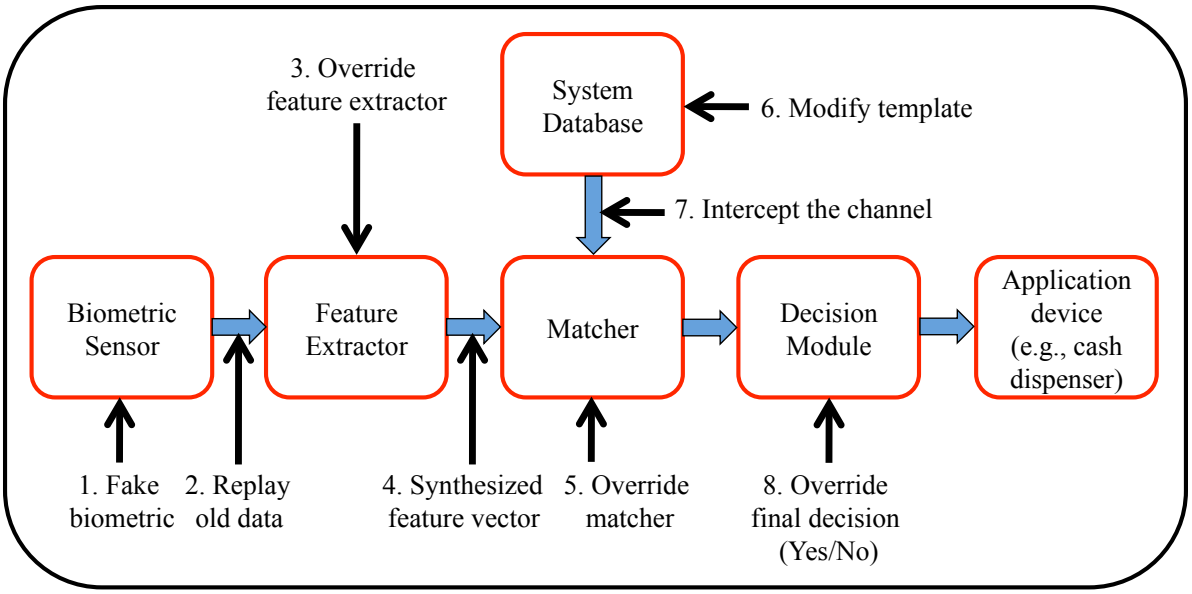

Figure 3.1: Points of attack in a generic biometric system.

3.2 Attacks against Biometric Systems

Adversary attacks exploit the system vulnerabilities generally at one or more modules or interfaces. Eight possible different points where security of biometric systems can be compromised have been identified in [70], (see Figure 3.1), as described below:

1. A fake biometric trait may be presented at the sensor such as a fake finger, a copy of a signature, or a face mask.

2. Digitally stored biometric data may be resubmitted to the system. In this kind of attack, a previously recorded biometric data is replayed into the system bypassing the sensor, thus also called as “replay attack". For instance, presenting a digital copy of fingerprint image or recorded audio signal of a speaker.

3. The feature extractor may be attacked with a Trojan horse program that produces predetermined feature sets.

4. Legitimate feature sets extracted from the biometric input may be replaced with synthetic feature sets. For example, if minutiae of a fingerprint are transmitted to a remote matcher (say over the Internet) than this threat is very real.

5. The matcher may be attacked with a Trojan horse program that always directly produce a specified result - match, no match, or a score.

6. The enrolled templates in the database may be modified or removed, or new templates may be introduced in the database, which could result in authorization for a fraudulent individual, or at least denial of service for the person associated with the corrupted template.

7. The enrolled templates in the stored database are sent to the matcher through a communication channel which could be attacked to change the contents of the templates before they reach the matcher.

8. The final decision output by the biometric system may be overridden with the choice of result from the hacker. Even if the feature extraction and matching modules had an excellent performance characteristics, it has been rendered useless by the simple exercise of overriding the result.

3.3 Spoof Attacks

Among the potential attacks discussed in the literature, the one with the greatest practical relevance is "spoof attack", which consists in submitting a stolen, copied or synthetically replicated biometric trait to the sensor to defeat the biometric system security in order to gain unauthorized access. Recently, it has been shown that spoof attacks can be carried against many types of biometrics, like fingerprint, face, and iris [28, 59, 10, 82, 79, 33]. This kind of attack is also known as "direct attack", since it is carried out directly on the biometric sensor. The feasibility of a spoof attack is much higher than other types of attacks against biometric systems, as it does not require any knowledge on the system, such as the feature extraction or matching algorithm used. Digital protection techniques like hashing, encryption, and digital signature, are not useful due to the nature of spoofing attacks, which are done in the analogical domain, outside the digital limits of the system.

"Liveness" testing (vitality detection) methods have been suggested among feasible counteractions against spoof attacks by several researchers. Liveness testing, which aims to detect whether the submitted biometric trait is live or artificial[1], is performed by either software module based on signal processing or hardware module embedded into the input device itself [6, 10, 12, 13, 15, 43,

[1] In this thesis, we will use both "fake" and "spoofed" terms interchangeably, to indicate an artificial replica of genuine client's biometric trait.

Figure 3.2: Spoofed fingers: Silicone (left) and Gelatin (right).

45, 51, 52, 94]. But, the literature review states that so far no effective method exists. Moreover, the collateral effect when biometric systems are coupled with liveness detection methods is the increase of false rejection rate (FRR): the percentage of genuine users rejected by the system. Other spoof detection algorithms have been proposed for other biometric traits, but their performances are not satisfactory as well.

3.3.1 Fingerprint spoofing

It is worth noting that the idea of fooling fingerprint recognition systems using a replicated fake fingertip is not a novelty. For the very first time, the idea of fingerprint spoofing was described by the mystery writer R. A. Freeman in his book "The Red Thumb Mark" [23], published in 1907. James Bond, more recently, was able to spoof a fingerprint recognition system with a thin layer of latex glued on his fingertip, in the film "Diamonds are Forever" (1971) [30].

Real-life fingerprint spoofing is also quite an old exercise. Alert Wehde carried out the very first endeavor to spoof fingerprints in the 1920s [11]. He, then an inmate at a Kansas penitentiary, used his expertise in photography and engraving to produce a gummy fingerprint from a latent fingerprint. The latent fingerprint was highlighted using forensic methods, and a photograph was taken. The photograph was then used to etch the print onto a copper plate, which was later used to generate fake latent fingerprints on surfaces.

In recent years, several research studies have been conducted to investigate how spoofed fingerprints can circumvent state-of-the-art fingerprint recogni-

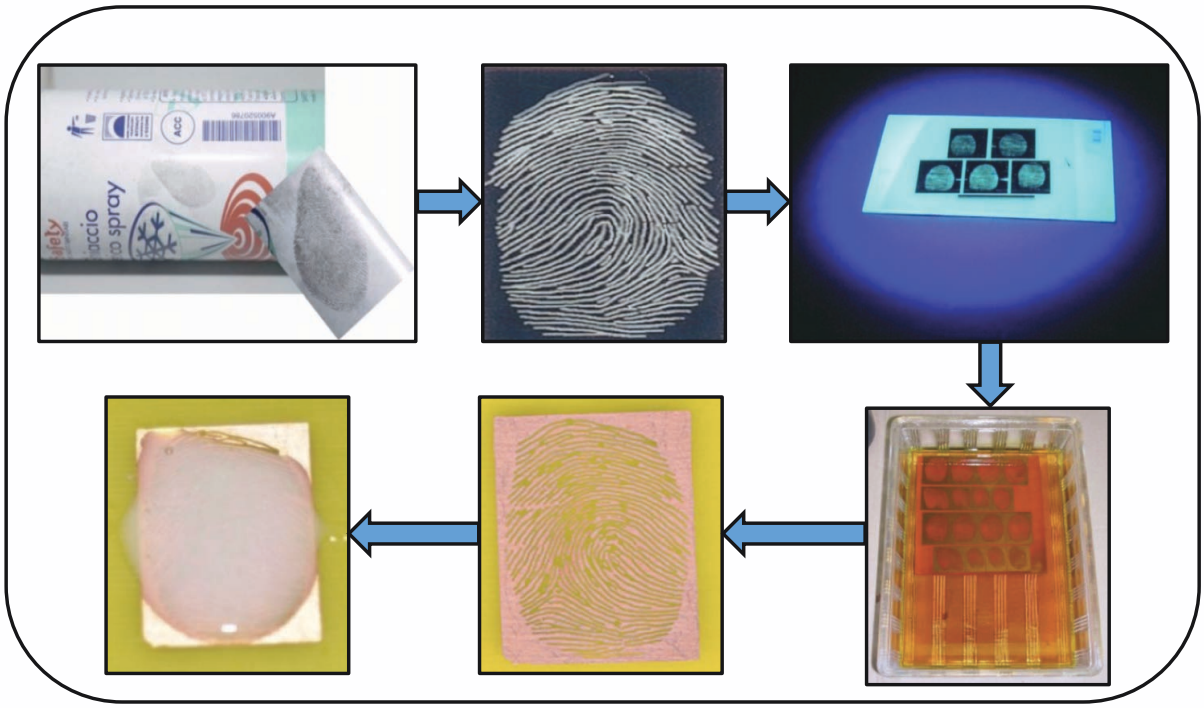

Figure 3.3: Spoofed fingerprint reproduction without cooperation.

tion systems. Authors in [69] studied the susceptibility of different biometric fingerprint sensors to fake fingerprints synthesized with silicone and plasticine. Results on six optical and solid-state commercial sensors were reported. Five sensors permitted the unauthorized access into the system on the first attempt, while the remaining one was spoofed on the second attempt.

Matsumoto et al. [59] reported, by conducting similar experiments as in [69], that fake fingerprints fabricated with gelatin are more effective. The authors tested eleven commercial fingerprint sensors, with a success rate higher than 60%, even also when the fake fingerprints were replicated from the latent fingerprint. A similar evaluation of robustness of different sensors to fake fingerprints fabricated with several spoofing techniques can be found in [84, 43]. In [43], the authors extended the experiments reported in [59] to test new sensors embedded with fake detection measures. The authors concluded that such fake detection measures were able to reject spoofed fingerprints replicated using non-conductive materials such as silicone, while were not able to detect fake fingerprints fabricated using conductive materials like gelatin (see Figure 3.2).

Similarly, in [25] systems with different well-known sensors (including devices produced by Biometrika, Digital Persona, Fujitsu, Identix, Siemens or Precise Biometrics) were tested by gummy fingers replicated by several spoofing materials and techniques.

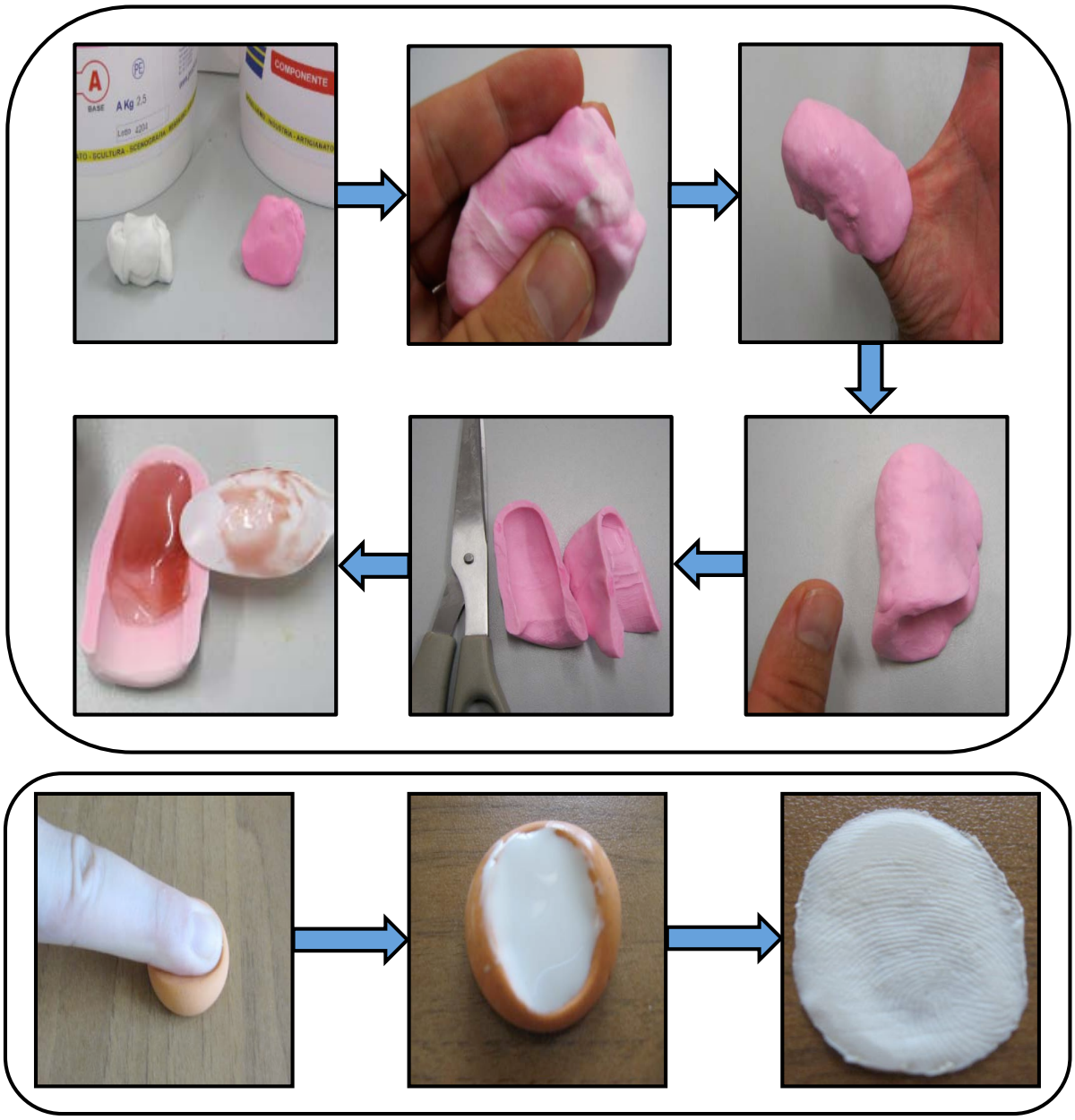

Figure 3.4: Two examples of spoofed fingerprint reproduction with cooperation.

Recently, in [24] the possibility of fabricating fake fingerprints from standard minutiae templates was studied. A two-stage process was carried out to create the spoofed fingerprints. In the first stage, fingerprint images were reconstructed from the genuine user's minutiae template. This stage was termed as "from the template to the image". In the second stage, called "from the image to the gummy fingerprint", the reconstructed images were utilized to produce fake fingerprints. In spite of some errors that were accumulated during the reconstruction process, more than 70% of the fake fingerprints were accepted by the system.

The existing literature, as described above, suggests that fingerprint spoofing methods can be classified into two broad categories: "consensual/cooperative/direct casts" and "non-consensual/non-cooperative/indirect casts". In the consensual method, the fake fingerprints are created with the consent and collaboration of the fingerprint owner. In the non-consensual method the cooperation of the user is not required, since the latent finger-marks, that the user has unnoticeably left on some surface, are used to fabricate the spoofed fingerprint using a very similar procedure to that mentioned in [11]: today, printed circuit board etching is a successful molding technique for producing "gummy" and other soft material artificial fingers [80] (see Figure 3.3). It is worth noting that most of the research studies have been conducted using spoofed fingerprints fabricated with the consensual method. Also, the consensual method was adopted to create high quality fake fingerprints in the First and Second Editions of the International Fingerprint Liveness Detection Competition (LivDet09, LivDet11) [52, 94]. The method consists of the following steps (see Figure 3.4):

1. The user presses his finger on a soft material such as wax, play doh, dental impression material, or plaster;
2. The negative impression of the fingerprint is fixed on the surface to form a mold;
3. A casting material such as liquid silicon, wax, gelatin, moldable plastic, plaster or clay, is poured in the mould;
4. When the liquid is hardened, the fake/spooed fingerprint is formed.

3.3.2 Face spoofing

In spite of the fair amount of advancement in biometric face recognition systems, face spoofing, also known as "copy attack", still poses a serious threat to the system security. Face spoofing methods may vary according to the targeted face recognition system. Face recognition systems can be broadly classified into two groups: 2D (two-dimensional) and 3D (three-dimensional) systems. A biometric 2D face recognition system takes into consideration only the two dimensional image of the face. 3D systems are clearly more complex, and recognize faces on the basis of features extracted from the 3D shape of the whole face, using methods such as paraxial viewing, or patterned illumination light [29]. Conventionally, face recognition systems can be spoofed by presenting (i) a photograph, (ii) a video, or (iii) a 3D face model/mask of a legitimate user.

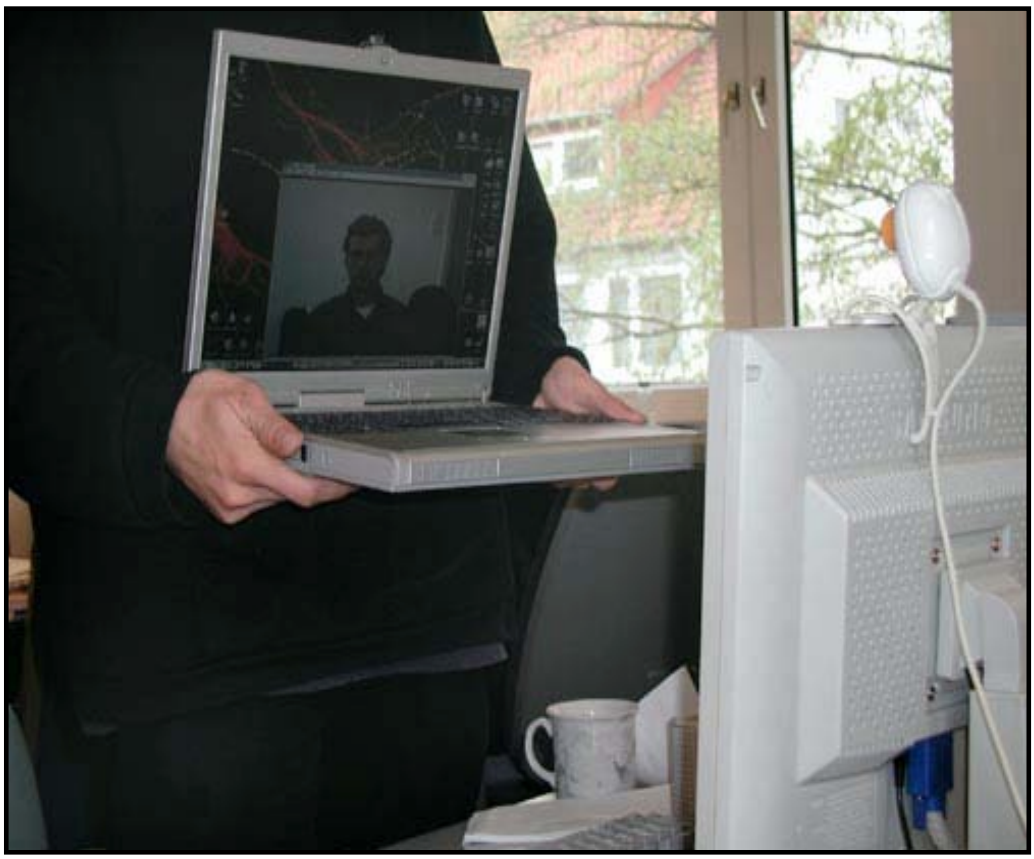

Figure 3.5: An example of face spoofing using “photo-attack” method.

Face spoof attack through photograph or video is the most common, cheapest and easiest method to circumvent face recognition systems [10, 95]. Spoof attacks through photograph, known as “photo-attacks”, consist in submitting a photograph of a legitimate user to the face recognition system, displayed in hard copy or on the screen of a portable computer or smart phone [10, 95] (see Figure 3.5). Since face is not concealed like other biometric traits, it may be easier to spoof; for instance, a genuine user’s facial image can be simply captured using distant cameras, without their awareness and prior consent for spoofing purpose. Moreover, due to social image sharing and social networking websites, personal facial photographs of many users are usually accessible to the public. For instance, an impostor can obtain the photographs of genuine users from a social network, and submit them to a biometric authentication

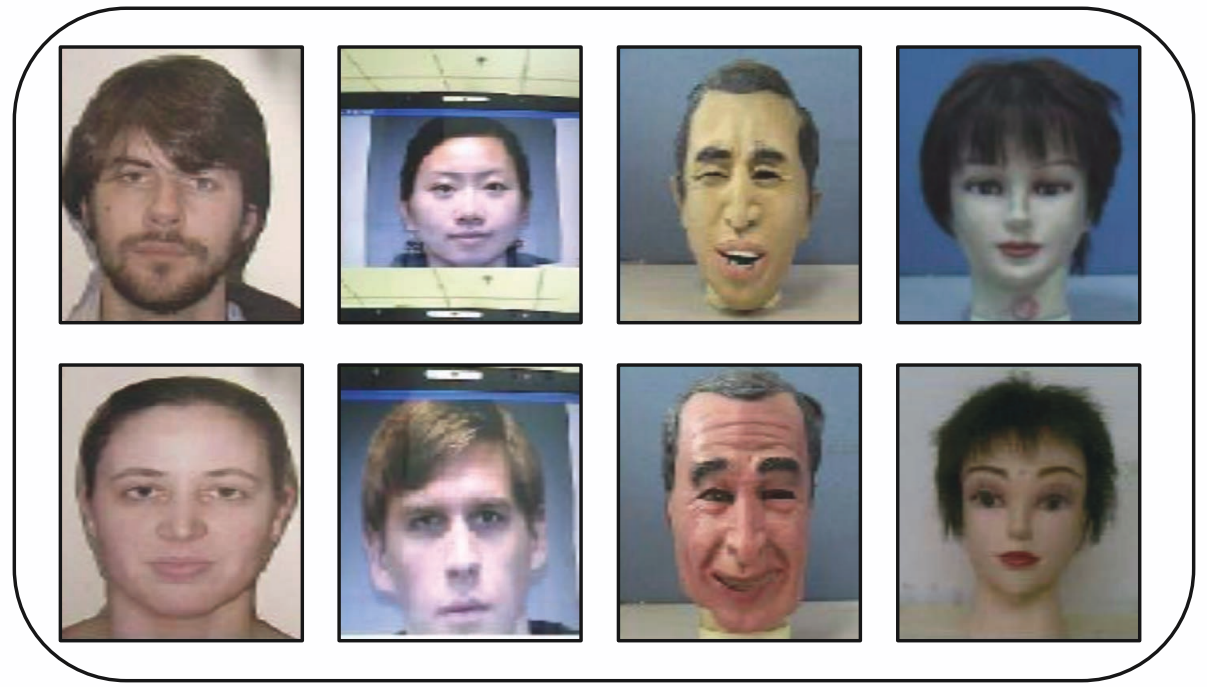

Figure 3.6: Some of spoofed face examples. Materials from left column to right are: photo, video replay, rubber and silica gel (adapted from [95]).

system to fool it. It is worth noting that the "photo-attack" method was used to assess the performance of face liveness detection systems at the competition on counter measures to 2D facial spoofing attacks, held in conjunction with the 2012 International Joint Conference on Biometrics (IJCB 2012) [10].

The advent of public video sharing websites and reduction in the cost of high quality tiny video cameras and portable devices have also made it easy to obtain or capture a genuine user's facial video samples without subject's consent and awareness, which can later be presented to the system using a portable device for spoofing purpose [95, 10, 45]. The likelihood of success of a video attack becomes higher due to the physiological clues in the displayed spoofed faces (e.g., facial expression, eye blinking, and head movement), which could also thwart liveness detection techniques based on these clues.

3D face recognition systems can be spoofed by 3D face model or face mask fabricated by rubber, plastic, silica gel [95] (see Figures 3.6 and 3.7). The prerequisite to spoof 3D face systems is the presented face to camera should be three dimensional, thus making 3D face spoofing more complicated than a 2D face spoofing. Due to ease in spoof fabrication, and to the wide adoption of 2D systems across the globe, "photo attacks" and "video attacks" are still the most common techniques to spoof faces.

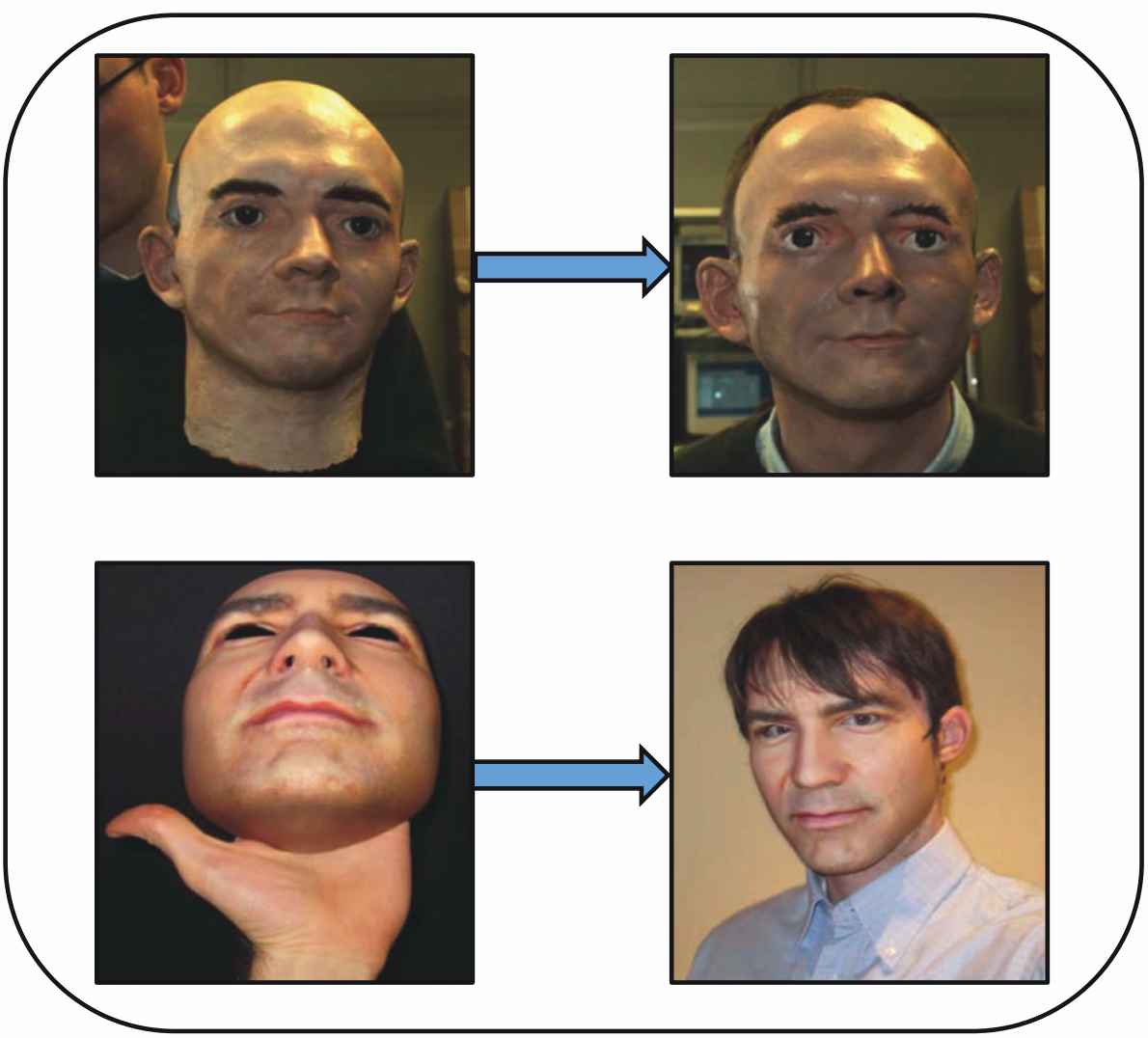

Figure 3.7: Examples of spoofed 3D faces (adapted from [21]).

3.4 Robustness of Multimodal Biometric Systems against Spoof Attacks

Besides ad hoc countermeasures, multimodal biometric systems are also considered as a natural defence mechanism against spoof attacks. Multimodal systems have been originally proposed to overcome the weaknesses and some inherent limitations of unimodal systems, in order to improve the identity recognition performance. Their effectiveness has been shown by extensive theoretical and empirical evidences [56, 61, 62]. In addition, they proved to be quite robust under stress conditions, namely, poor cooperation of the user (e.g., wearing glasses or beard) [55]. It is also commonly believed that multimodal systems are intrinsically more robust to spoof attacks than systems using a single biometric trait

[78, 34]. The belief on their intrinsic robustness against spoof attacks is based on the assumption that an intruder has to spoof all fused biometrics simultaneously to crack the multimodal system [78, 34]. Such an attack would require more effort than spoofing a single trait, and this could be a valid deterrent for discouraging the adversary from attempting the attack.

However, this assumption is not based on any theoretical findings or empirical evidences, but only on intuitive and qualitative arguments, which rely mainly on the higher performance of multi-modal systems with respect to unimodal ones. However, some recent works, contrary to the common belief, have provided experimental evidence that spoofing only one biometric trait can be sufficient to evade the system, even when more than two biometric traits are used [74, 73, 42]. They considered systems made up of two (face and fingerprint) [74, 73] or three matchers (face, fingerprint and iris) [42]. A substantial increase of the false acceptance rate (FAR) of these systems under spoof attacks was indeed highlighted. Most of the results however reported in [74, 73, 42] were obtained under the unrealistic and stringent hypothesis (known as "worst-case" scenario) that the attacker is able to fabricate a perfect replica of the genuine user's biometric trait whose matching score distribution is identical to the one of genuine traits; which was simulated by assuming that the matching score distribution of fake traits is identical to the one of genuine users. In [74], the authors have proposed two new score fusion rules to improve robustness against spoof attacks. Such rules (a fuzzy rule and a modification of the LLR rule) are based on a quality measure aimed at discriminating among fake and live traits, and (in the case of the modified LLR rule) to take into account the possibility of spoof attack at design phase, by simulating the presence of spoofed trait scores among training data set using "worst-case" assumption. While, this "worst-case" assumption may be reasonable for some biometric traits, like 2D faces (as discussed in [73]), its validity for other biometric traits like iris and fingerprint is questionable.

It is worth noting that in [73] some experiments have been carried out using a fraction of real spoofed fingerprints from the Fingerprint Liveness Detection Competition (LivDet09) [52]. The fake fingerprint score distribution obtained in the study was significantly different from the genuine score distribution, therefore additionally providing the proof that worst-case hypothesis is not realistic for all biometric traits and spoofing techniques.

Thus, the results in [74, 73, 42] raise the issue of investigating more thoroughly and systematically the robustness of multimodal systems against real spoof attacks, namely under non-worst case scenarios (when the fake traits are

not exact replicas of genuine ones), and devising new methods to design robust multimodal biometric systems against them.

3.5 Open Issues in Robustness of Multimodal Biometric Systems against Spoof attacks

Although, multimodal biometric systems offer several advantages such as better recognition accuracy, increased population coverage and greater flexibility, the problem of evaluating the performance of a multimodal biometric system, in terms of both generalisation capability and robustness against spoof attack, is an open research problem. Major open problems that are still unresolved to ensuring a secure multimodal biometric recognition system include:

(a) *Can multimodal biometric systems be actually cracked by attacking only one sensor via real attacks?*

Though, the robustness of multimodal biometric systems has been questioned recently in [74, 73, 42] by showing that, in some application scenarios, they can be cracked by spoofing only one of the biometric traits. However, the scope of those results are very limited, since they were obtained by *simulating* the scores of spoofed traits under the unrealistic hypothesis known as "worst-case" scenario, where it is assumed that the attacker is able to fabricate a perfect replica of a biometric trait whose matching score distribution is identical to the one of genuine traits. Nevertheless, if the results in [74, 73, 42] turned out to be held in more realistic application scenarios, this would imply that multimodal systems are just a deterrent rather than a real defense against spoof attacks. On the other hand, a more wide and realistic evaluation could point out the conditions under which multimodal biometric systems are a good defence countermeasure. In other words, it is still crucial to investigate whether the conclusion drawn from the results in [74, 73, 42], that the multimodal biometric systems are not intrinsically more robust against spoof attacks, also holds in realistic scenarios.

(b) *Is the "worst-case" scenario hypothesized in [74, 73, 42] for spoofing biometrics representative of real spoof attacks?*

Recent works [74, 73, 42] have shown that multimodal biometric systems can be highly vulnerable to spoof attacks, under a very restrictive "worst-case" working hypothesis. However, the "worst-case" hypothesis for spoof attacks may not be generalized for all biometric traits, as also pointed out in

[73]. It is thus interesting to further investigate whether and to what extent the "worst-case" scenario hypothesized in [74, 73, 42] is realistic, which is still itself an open issue. A more systematic and large-scale experimental analysis are indeed required. Such analysis will allow us to point out to what extent the drop in performance under the "worst-case" attack scenario is representative of the performance under real spoof attacks.

(c) *How can the security of multimodal systems be evaluated, under realistic attacks, without fabricating spoofed traits?*

A straightforward approach to evaluate the security of a multimodal biometric system against realistic spoof attacks is to fabricate fake biometric traits and submit them to the system to check its vulnerability. However, fabricating fake biometric traits is a time-consuming and cumbersome task, thus impractical for the system designer [52]. A potential alternative is to develop methods based on *simulating* the spoofed biometric traits. The state-of-the-art [74, 73, 42] solves this problem by simulating the effect of spoof attacks in terms of modelling the score distribution of the corresponding biometric trait. In particular, the fake score distribution is assumed to be coincident to the genuine users one, thus drawing a "worst-case" scenario. Therefore, an interesting open issue is the development of realistic models which are efficient in simulating the score distribution of the real spoof attacks. This is possible by considering different degrees of the quality of the fake traits, which in real scenarios can be due to different forgery skills of the attackers, thus overcoming the intrinsic limits of the "worst-case" assumption. The two-fold aim of the development should be (a) assessing the robustness of multimodal systems, and (b) to design novel fusion rules, robust to spoof attacks. In particular, it is crucial to develop evaluation methodologies that can be applied to any multimodal system, namely, to any set of matchers combined with any fusion rule, and allowing to simulate a spoof attack against any subset of the component matchers. Moreover, it can also be useful to compare the robustness of different score fusion rules applied to a given multimodal system. However, till date no systematic research effort has been carried out towards this direction.

3.6 Summary

In this chapter we first discussed the rationale behind the attacks on biometric authentication systems. We further mentioned the eight points of

vulnerabilities that an adversary may exploit to mislead a biometric recognition system. Among all the attacks, spoof attacks i.e. presenting fake biometric traits to the sensor are gaining much attention and we accordingly defined and discussed them at length. In fact, the comprehensive review of the state-of-the-art spoofing methods for face and fingerprint biometrics is presented. Then, in Section 3.4, the literature review on the evaluation of the robustness of multimodal biometric systems against spoof attacks is provided. Various existing studies concluded, contrary to a common belief, that multimodal biometric systems are vulnerable to spoof attacks, and can be cracked even by spoofing only one trait. We then listed representative open issues identified from the existing literature related to direct attacks for multimodal systems.

To conclude, we can say that the state-of-the-art presented in this chapter clearly highlights the need for a more systematic security analysis of multimodal systems against real spoof attacks. Indeed, in the following chapters of this thesis we further focused and explored above mentioned open issues to advance the state-of-the-art, namely, evaluating the performance of multimodal systems against real spoof spoof attacks, checking the validity of "worst-case" scenario hypothesized in the literature of biometric spoof attacks and the designing of methods to analyze the security/robustness of multimodal systems against real spoof attacks without fabricating spoofed traits.

Chapter 4

Real Spoof Attacks against Multimodal Biometric Systems

4.1 Introduction

In this chapter, we address the problem of evaluating the robustness of multimodal biometric systems against real spoof attacks. We argue that existing studies [74, 73, 42] in the literature to assess the security of multimodal biometric systems against spoof attacks do not provide a complete and exhaustive view of the performance degradation of a multimodal system under real spoof attacks, since they only focus on assessing the performance using simulated spoof attacks in "wost-case" scenario, and mainly disregard the issue of robustness of the multimodal biometric systems against real spoof attacks.

The need for evaluating the robustness of multimodal biometric systems under real spoof attacks was also highlighted in Chapter 3, where we reviewed several works which assessed the robustness of multimodal biometric systems under "wost-case" scenario based spoof attacks.

As discussed in Section 3.5 of Chapter 3 that, until a few years ago, multimodal biometric systems were commonly believed to be not only more accurate but also intrinsically more robust to spoof attacks than unimodal systems [78, 34]. The belief on their intrinsic robustness against spoof attacks was based on the assumption that an attacker has to spoof all fused biometric traits *simultaneously* to crack the system. However, recently this belief has been questioned and shown empirically that spoofing only one biometric trait can be adequate to crack the multimodal biometric systems [74, 73, 42]. But, most of the results were attained under the "worst-case" scenario, which was obtained by *simulating* the fake biometric trait's scores under the hypothesis that their distribution is

identical to the genuine user's one, namely, that the attacker is able to fabricate a perfect replica of the attacked genuine biometric trait. However, a scenario in which the attacker can always replicate the exact replica of genuine user's biometric traits may not be realistic. For instance, a scenario in which the attackers always capturing the exact face of genuine clients as the ones used into the system as templates may not be feasible. Thus, it is difficult to generalize the findings in [74, 73, 42]: a more systematic and large-scale empirical investigation is indeed required.

In other words, it is still necessary to test the robustness of multimodal biometric systems under various scenarios of real spoof attacks, namely under "non-worst" case scenarios. Such analysis would allow to check the validity of results in [74, 73, 42], that is to investigate experimentally the open issues (a) and (b) pointed out in Section 3.5 of previous chapter 3, viz., *"Can multimodal biometric systems be actually cracked by attacking only one sensor via real attacks?"* and *"Is the "worst-case" scenario hypothesized in [74, 73, 42] for spoofing biometrics representative of real spoof attacks?"*

To address these open issues, in this chapter we analyze the robustness of multimodal biometric systems, made up of a face and a fingerprint matcher, against real spoof attacks by carrying out a substantial set of experiments with large data sets containing real spoof attacks. We created and collected, both for fingerprints and faces, several data sets consisting of real spoof attacks. In order to get a very large set of fake fingerprints, we fabricated the spoofed fingerprints using four materials with state-of-the-art methods. It is worth noting that these fake fingerprints were also used in the First and Second Fingerprint Liveness Detection Competition [52, 94]. We also built three different face data sets corresponding to different attack scenarios depending on how the attacker fabricates the fake faces. One of them is public data set recently used in the Competition on Countermeasures to 2D Facial Spoofing Attacks [10]. We conducted many experiments on a large number of score fusion rules, including the one proposed in [74], which is explicitly designed to be robust to spoof attacks. In addition, we also simulated "worst-case" spoof attacks, based on the same hypothesis in [74, 73, 42], in order to compare the corresponding results with those attained by real spoof attacks.

Below, we first describe the data sets in Section 4.2. Section 4.3 summarizes well-known and widely used fusion rules, which are adopted in this study, with experimental protocol. While Section 4.4 discusses the achieved experimental results with concluding remarks.

Figure 4.1: Original template image of a fingerprint of our data set (Left). A spoof of the same fingerprint obtained by using latex (middle), and silicon (right).

4.2 Data Sets

The size and the characteristics of the data sets described in the following sections are reported in Table 4.1.

4.2.1 Fingerprint

The fingerprint data set used in this study is composed of of 142 distinct users[1]. For each "live" finger and its corresponding fake replica, twenty different impressions were acquired in two different sessions, separated by about two weeks interval. Only four fingers were considered in this case: the left and right index and thumb fingers. To create the fake fingerprints, we adopted the "consensual" method described in Chapter 3, Section 3.3.1. We used a plasticine-like material as the mold, while the spoofs were created with four basic materials: silicon, latex, gelatin and alginate as the casts. These four materials are commonly adopted for replicating fingerprints, and have been used for assessing the performance of fingerprint liveness detection systems at First and Second International Competition on Fingerprint Liveness Detection (LivDet) [52, 94]. The fingerprint images were acquired using the well-known Biometrika FX2000 optical sensor, which has a resolution of 569 dpi, and a sensing area of 13.2 mm.

Some sample images, showing the average quality of provided spoofs, from our data sets are shown in Figure 4.1. This figure shows the original, "live" client image, beside a replica made up of latex, and a replica made up of silicone. As it can be seen, the latex image is very similar to the original one, whilst the second one is characterized by some artifacts. The fake fingerprints used in

[1]By "users", here, we mean a distinct finger, even if it belongs to the same person

Figure 4.2: Left: original template image of one of the users of our live face data set. Middle: spoofed face of the *Photo Attack* data set, obtained by a "photo-attack" method. Right: spoofed face of the *Personal Photo Attack* data set, obtained by a personal photo voluntarily provided by the same user.

this study represent the state-of-the-art in fingerprint spoofing, thus providing a reasonable set of realistic scenarios.

4.2.2 Face

We collected and built three face data sets. The first two data sets include the same users but two different kinds of face spoof attacks: the *Photo Attack* and the *Personal Photo Attack* data sets. The "live" face images of each user were collected into two sessions, with a time interval of about two weeks between them, under different lighting conditions and facial expressions.

We then created the spoofed face images for the *Photo Attack* data set using the "photo attack" method described in [10, 95, 51, 45]. It consists in displaying a photo of the targeted user on a laptop screen, which is then put in front of the camera. In particular, the testing "live" face images of the clients were used to this end. This simulates a scenario in which the attacker can obtain photos of the targeted user under a setting similar to the one of the verification phase.

To build the *Personal Photo Attack* data set of spoofed faces, we used a set of personal photos voluntarily provided by 25 of the 50 users in our data set. On average, we were able to collect 5 photos per client. These photos were taken in different times and under different environmental conditions than those of the live templates. This simulates a scenario where the attacker may be able to collect a photo of the targeted client from the Web; for instance, from a social network or from an image search engine.

Figure 4.2 shows an example of the original template image of one of the users, a spoof obtained by the photo attack, and a spoof obtained from an image

Data set	Number of users	Number of spoofs per user	Number of live per user
Silicon	142	20	20
Latex	80	3	5
Gelatin	80	3	5
Alginate	80	3	5
Photo Attack	40	60	60
Personal Photo Attack	25	3 (avg.)	60
Print Attack	50	12	16

Table 4.1: Characteristics of the fake fingerprint and fake face data sets used in the experiments.

voluntarily provided by the same user. These two spoofs reflect two different degrees of expected effectiveness, but also of realism. In fact, a photo attack based on one of the images in the data set appears to have, by visual inspection, more chances to be successful than a spoof obtained by personal photos, as the latter are often significantly different from the template images of a biometric system. On the other hand, the latter case may be more realistic, as it would be probably easier for an attacker to obtain a photo of the targeted client from the Web, than an image similar to his template. According to the above observations, we expect that the fake score distribution of our Photo Attack data set (provided by some matching algorithm) will be very similar to that of the genuine users (as verified in Section 4.4.2), whilst the effectiveness of a spoof attack based on personal photos will strongly depend on the ability of the attacker to obtain images similar to the templates used by the system.

The third face data set, we used is *Print Attack* database [10, 6]. After the Competition on Countermeasures to 2D Facial Spoofing Attacks, held in conjunction with the International Joint Conference on Biometrics, in 2011, the Print Attack database was made publicly available. It consists of 200 video clips of printed-photo attack attempts to 50 clients, under different lighting conditions, and of 200 real-access attempts from the same clients. As we need to operate on images, we extracted the "live" and spoofed face images from the corresponding videos. In particular, for each client, we extracted 12 "live" face images and 16 spoofed face images from each video clip, as summarized in Table 4.1.

4.3 Experimental Setup

In this section, we describe our experimental setup, including the multimodal data sets and score fusion rules used.

In this study, we used a multimodal biometric system based on face and fingerprint, like the one of Figure 2.6, with different score fusion rules.

4.3.1 Fusion rules

Score fusion rules can be subdivided into fixed and trained. The difference between them is that the latter include a set of parameters to be estimated from training data. We describe here the most widely used rules, which will also be used in the experiments of Section 4.4. In the following, s_1 and s_2 denote scores provided respectively by face and fingerprint matchers, and $s = f(s1, s2)$ is the score fusion rule.

Fixed rules

Sum. The fused score is obtained by simple addition of the individual score values:

$$f(s_1, s_2) = s_1 + s_2 \ . \tag{4.1}$$

Product. The product rule computes the fused score as:

$$f(s_1, s_2) = s_1 \times s_2 \ . \tag{4.2}$$

Trained rules

Weighted sum by Linear Discriminant Analysis (LDA). The individual scores are linearly combined as:

$$f(s_1, s_2) = w_0 + w_1 s_1 + w_2 s_2 \ . \tag{4.3}$$

The weights w_0, w_1 and w_2 are set as the ones that maximize the Fisher distance (FD) between the score distributions of genuine and impostor users. In the case of two matchers, FD is defined as follows:

$$FD = \frac{(\mu_\mathrm{I} - \mu_\mathrm{G})^2}{\sigma_\mathrm{I}^2 + \sigma_\mathrm{G}^2} \ , \tag{4.4}$$

where μ_I and μ_G are the means respectively of the impostor and genuine score distributions, while σ_I^2 and σ_G^2 are their variances.

Likelihood ratio (LLR). This is a trained rule which corresponds to the so-called Neyman-Pearson test:

$$f(s_1, s_2) = \frac{p(s_1, s_2|\mathrm{G})}{p(s_1, s_2|\mathrm{I})} \,. \tag{4.5}$$

Conditional independence between s_1 and s_2, given that they come either from an impostor or a genuine user, is often assumed such that $p(s_1, s_2|\cdot) = p(s_1|\cdot)p(s_2|\cdot)$, where $p(\cdot|\mathrm{G})$ and $p(\cdot|\mathrm{I})$ are the matching scores probability density function (PDF) of genuine and impostor users, respectively. In general, parametric (e.g., Gaussian, Gamma, Beta) or non-parametric models (e.g., Parzen windows) can be used to fit the genuine and impostor distributions. Note that, in this case, $f(s_1, s_2)$ is not obtained as a matching score between two biometric traits, but as a ratio between likelihoods. Nevertheless, the decision rule is the same as above.

Extended LLR (ExtLLR). This is a variation of the LLR, which was proposed in [74] to make it robust against spoof attacks. The basic idea was to explicitly take into account the probability distribution of spoof attacks when modeling the probability distribution of the impostor class. To this end, the following model was proposed in [74][2]. Let the random variable $U \in \{\mathrm{G}, \mathrm{I}\}$ denote whether a user is a genuine or an impostor. Given a multimodal system made up of M matchers, their scores are assumed to be conditionally independent, given U: $p(s_1, s_2, \ldots, s_M|U) = p(s_1|U)p(s_2|U)\cdots p(s_M|U)$. M binary random variables T_i are then introduced, to denote whether a user is attempting a spoof attack against the i-th matcher ($T_i = 1$), or not ($T_i = 0$). Genuine users are assumed to always submit a real biometric trait, namely: $P(T_1 = 0, \ldots, T_M = 0|U = \mathrm{G}) = 1$. Furthermore, it is assumed that each of the $2^M - 1$ possible combinations of attacks against one or more matchers are equiprobable. Denoting with α the prior probability of a spoofing attack, this implies:

$$P(T_1, \ldots, T_M|U = \mathrm{I}) = \begin{cases} 1 - \alpha & \text{if } T_i = 0, i = 1, \ldots, M, \\ \frac{\alpha}{2^M - 1} & \text{otherwise.} \end{cases} \tag{4.6}$$

Finally, M binary random variables F_i are further introduced, to denote whether a spoof attack carried out by an impostor against the i-th matcher (i.e., when $U = \mathrm{I}$ and $T_i = 1$) is "successful" ($F_i = 1$) or not ($F_i = 0$), in the sense defined below. Clearly, $F_i = 0$ when the i-th matcher is not under attack, which implies

[2]The availability of a quality score for each matcher was considered in [74], together with the matching score. In the description of the ExtLLR rule we omit the quality score, since it was not used in our experiments, and also because we mainly intend to evaluate the contribution of the Extended LLR rule to the robustness of the LLR rule, due only on its capability to model the presence of spoofed samples.

$P(F_i = 0|T_i = 0) = 1$. The probability of success $P(F_i = 1|T_i = 1)$ is denoted with c_i. Its value was related to the "security" ot the corresponding matcher. In [74] it is pointed out that evaluating the security of a matcher is a very difficult problem, if not impossible to solve, and thus c_i must be evaluated based on general knowledge about the biometrics at hand. In the experiments of [74], such value was manually set (see below).

It is now possible to derive the conditional distribution of scores that is needed by the standard LLR rule (see Eq. 5.4), by marginalising over the $2M$ random variables T_i and F_i:

$$\begin{aligned} p(s_1,\ldots,s_M|U) &= \textstyle\sum_{T_1,\ldots,T_M}\sum_{F_1,\ldots,F_M} p(s_1,\ldots,s_M,T_1,\ldots,T_M,F_1,\ldots,F_M|U) \\ &= \textstyle\sum_{T_1,\ldots,T_M}\sum_{F_1,\ldots,F_M} P(T_1,\ldots,T_M|U)\times\prod_{i=1}^{M}[P(F_i|T_i)P(s_i|F_i,U)]\,. \end{aligned} \tag{4.7}$$

To evaluate the above probability, it is necessary to know the M distributions $P(s_i|F_i,U)$. Given the above assumptions, for genuine users ($U = \mathrm{G}$) we have $F_i = 0$, and thus $P(s_i|F_i = 0, U = \mathrm{G})$ can be learnt from genuine training samples, as in the standard LLR rule. For impostor users ($U = \mathrm{I}$) two assumptions are made in [74]. First, in the case of unsuccessful attacks, the conditional score distribution $P(s_i|F_i = 0, U = \mathrm{I})$ is identical to the one of impostors users that do not attempt spoof attacks, also called "zero-effort" impostors in [42]. Therefore, this distribution can be learnt from traning data as well. Second, the score distribution of successful spoofing attacks is identical to the one of genuine scores: $P(s_i|F_i = 1, U = \mathrm{I}) = P(s_i|F_i = 0, U = \mathrm{G})$. The latter assumption corresponds to the "worst-case" scenario mentioned above.

It immediately follows that, for a bimodal system ($M = 2$) as the one considered in [74] and in this work, the expression of the joint likelihood in (4.7) is:

$$\begin{aligned} p(s_1,s_2|I) &= \frac{\alpha}{3}(1-c_1)(1+c_2)p(s_1|G)p(s_2|I) && (4.8)\\ &+ \frac{\alpha}{3}(1+c_1)(1-c_2)p(s_1|I)p(s_2|G) && (4.9)\\ &+ \frac{\alpha}{3}(1-c_1)(1-c_2)p(s_1|G)p(s_2|G) && (4.10)\\ &+ \frac{\alpha}{3}(c_1+c_2+c_1c_2)p(s_1|I)p(s_2|I) && (4.11)\\ &+ (1-\alpha)p(s_1|I)p(s_2|I)\,, && (4.12) \end{aligned}$$

where the terms (4.8) and (4.9) are related to successful spoofing attempts against one trait (respectively, 1 and 2), (4.10) corresponds to a successful spoof

attempt against both of them, (4.11) accounts for unsuccessful spoof attempts against both traits, and (4.12) corresponds to "standard" impostor attempts without spoof attacks, namely "zero-effort" impostors [42].

LLR and ExtLLR require the estimation of individual likelihood. To this aim, we fit the available genuine and impostor match scores with a parametric distribution [74, 73, 42]. A Gamma distribution was used, as done in [74], because it turned out to provide a good approximation of our data. Moreover, the parameters, α, c_1 and c_2 were set to the same values used in [74], respectively 0.01, 0.3, and 0.7. Note that in a real application the values of these parameters, namely the probability of a spoofing attempt, and the probability that a spoofing attempt against each of the considered biometric does not succeed, can not be estimated from training data, and can only be hypothesized.

4.3.2 Experimental protocol

We used similar experimental protocol as in [74, 42], described in the following:

- Since no multimodal data sets including spoof attacks are available publicly. Hence, to built multimodal data sets (i.e., data sets in which each user has a face and a fingerprint trait), we randomly associated face and fingerprint images of pairs of users of the available face (Photo Attack, Personal Photo Attack and Print Attack) and fingerprint (silicon, latex, gelatin and alginate spoofs) data sets, thus obtaining twelve "chimerical" data set. We carried out this procedure because the face and fingerprint data sets did not contain the same users. Note that building *chimerical* data sets is a widely used approach in experimental investigations on multimodal biometrics [66, 78, 26, 16].

- To carry out more runs of the experiments, each of the twelve chimerical data set was randomly subdivided into five pairs of training and testing sets. For each training set we used 40% of the "virtual" users,[3] while the remaining 60% were used to build the testing set. Furthermore, all the above procedure was repeated five times, for different random associations of face and fingerprint images of pairs of users (namely, creating different "virtual" users). In each run, the parameters of the trained fusion rules have been estimated on the training set. The presented results are average testing set performance over the resulting twenty-five runs.

[3]The clients of a chimerical data set are usually referred to as "virtual" users [26, 16], since they do not correspond to a real person or identity.

- The fake matching scores were computed by comparing each fake image of a given user with the corresponding template image.
- The performance was assessed by computing DET curves (FAR vs. FRR). Note that, in the evaluation of spoof attacks, the FAR corresponds to the percentage of spoof attempts that were accepted as genuine. To avoid confusion with the classical interpretation of the FAR (referred to the "zero-effort" impostors), the term SFAR (spoof false acceptance rate) was introduced in [42]. The SFAR represents the probability that an impostor *attempting a spoof attack* is wrongly accepted as a genuine user. For instance, if the "equal error rate" (EER) operational point, defined as the point where the FRR equals the false acceptance rate (FAR), should be chosen according to application requirements, the alternative choice suggested in [42] to improve robustness is to choose the point where the FRR equals the SFAR. Similar choices can be made for other application requirements. In practice, this allows one to improve robustness against spoofing attacks (namely, reducing the SFAR), at the expense of a higher FRR. However, we do not adopt the term SFAR here, as it will be clear from the context whether we refer to the percentage of accepted zero- or non-zero-effort impostors.

The NIST Bozorth3 matching algorithm [64] was used for fingerprint verification. It is based on matching the fingerprint minute details, called "minutiae". The Elastic Bunch Graph Matching (EBGM) algorithm was used for face verification [93, 86]. It is based on representing a face with a graph whose nodes are the so-called face "landmarks" (centred on the nose, eyes, and other points detected on the face), are labelled by a feature vector, and are connected by edges representing geometrical relationships among them. Face and fingerprint matchers produced match scores in the range $[0,1]$ and $[0,990]$, respectively. Since score normalization is necessary before using fusion rules at the score-level, the fingerprint scores were normalized into the range $[0,1]$ using the hyperbolic tangent method [78].

We investigated three attack scenarios: (a) only fingerprints are spoofed, (b) only faces are spoofed, (c) both fingerprints and faces are spoofed (bimodal or double spoofing). We also evaluated the "worst-case" attacks as defined in [74, 73, 42]. Accordingly, the fake scores are fictitiously generated by randomly drawing a set of match scores from those associated to the genuine users in the testing set. We considered the fusion rules described in the Section 4.3.1: sum, product, weighted sum (LDA), LLR, and Extended LLR.

4.4 Experimental results

In this section, we discuss experimental analysis and results, whose main goals are to investigate whether attacking only one sensor using real spoof attacks allows the attacker to crack the multimoal biometric systems and to verify if the "worst-case" hypothesis made in [74, 73, 42] holds for real spoof attacks, namely, if it can be reliably exploited for predicting the performance drop under spoof attacks.

4.4.1 Analysis of robustness against real spoof attacks

We, first, investigate: is the attacker able to evade the multimodal biometric systems by forging at only one sensor via real spoof attacks?

The results are reported in Figures 4.3–4.8, in terms of Detection Error Trade-off (DET) curves attained on the test set. A DET curve reports the false rejection rate (FRR) as a function of the false acceptance rate (FAR), both computed parametrically being equally the decision threshold (s^*) on the fused score. Note that the FAR under a spoof attack is defined as the percentage of spoof attempts that got accepted as genuine, which is also referred to as Spoof FAR (SFAR) in [42].

We report the results on chimerical data sets, used in the experiment, as follows: latex spoofed fingerprints and photo attack spoofed faces (Figure 4.3); silicon spoofed fingerprints and personal photo attack spoofed faces (Figure 4.4); latex spoofed fingerprints and personal photo attack spoofed faces (Figure 4.5); silicon spoofed fingerprints and photo attack spoofed faces (Figure 4.6); gelatin spoofed fingerprints and print attack spoofed faces (Figure 4.7); alginate spoofed fingerprints and print attack spoofed faces (Figure 4.8). Each plot of Figures 4.3–4.8 refers to a different score fusion rule, indicated in the title of each plot.

From the Figures 4.3–4.8, it is easy to see that the multimodal biometric systems can be cracked, using real spoof attacks, by attacking only one sensor, even if the attacker does not fabricate a perfect replica of genuine user's trait. For instance, in Figure 4.3, using the weighted product rule (LDA), if the 1% FAR operational point is chosen on the training set, an average FAR of 0.73% is obtained on testing samples under normal operation. When only fingerprints is spoofed using latex instead, the FAR increases to 64.91%; while in case of only face spoofing using photo attacks, the FAR grows to 2.17%. This leads to the validation of the results obtained in [74, 73, 42], though those results were obtained under "worst-case" scenario, where the attacker was able to fab-

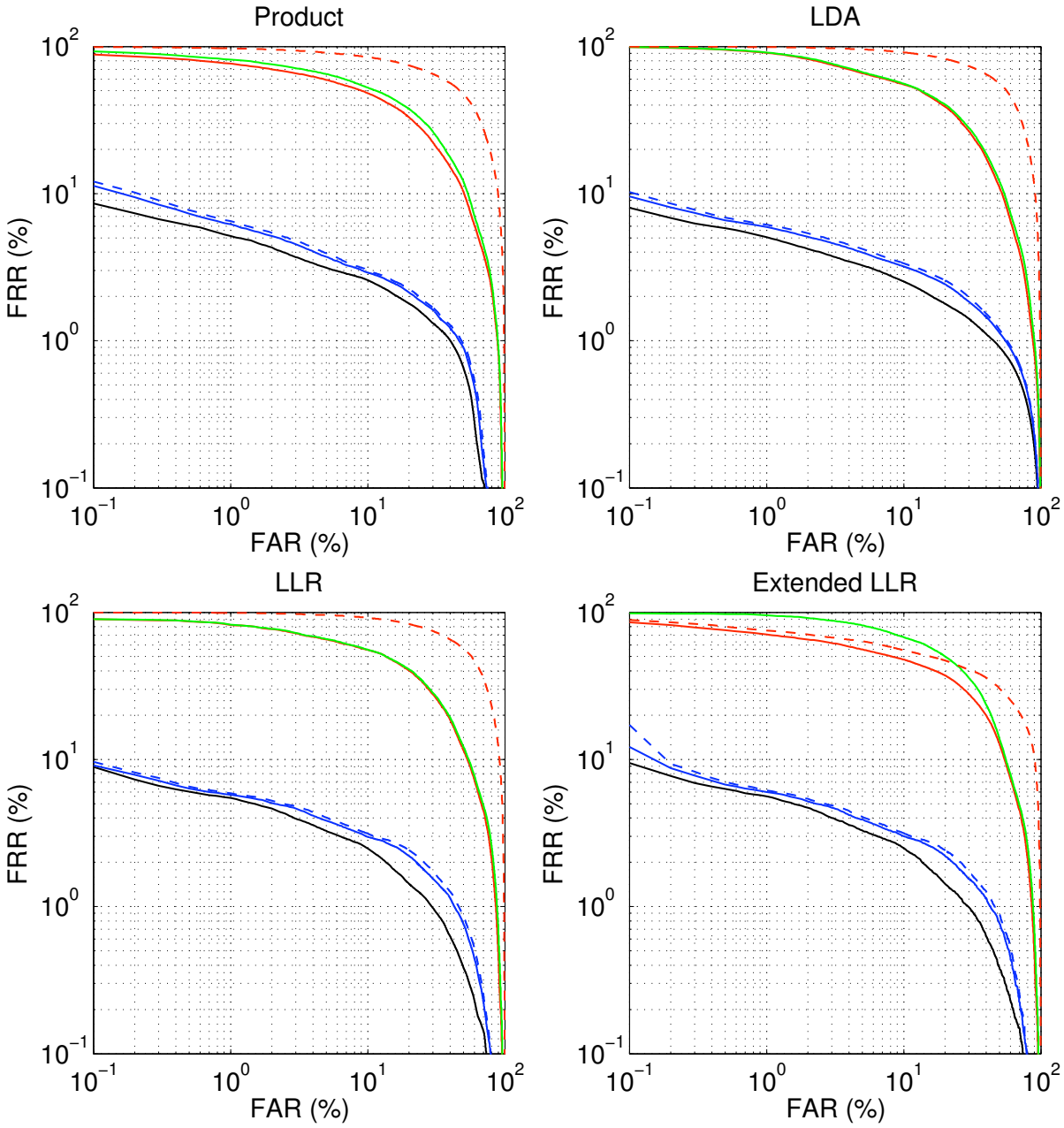

Figure 4.3: Average DET curves attained on test set using latex spoofed fingerprints and photo attack spoofed faces. Each plot refers to a different score fusion rules, indicated in the tittle of each plot. Each plot contains the DET curves attained with no spoof attacks (black), under real spoof attacks (solid curves) and under simulated "worst-case" spoof attacks (dashed curves). Red: fingerprint spoofing only. Blue: face spoofing only. Green: both face and fingerprint spoofing.

ricate exact replica of spoofed traits. Similar phenomenon can be noted in other plots of Figures 4.3–4.8 corresponding to different fusion rules. Thus, we can conclude that multimodal biometric systems are not intrinsically robust against spoof attacks, as they can be cracked by spoofing only one biometric.

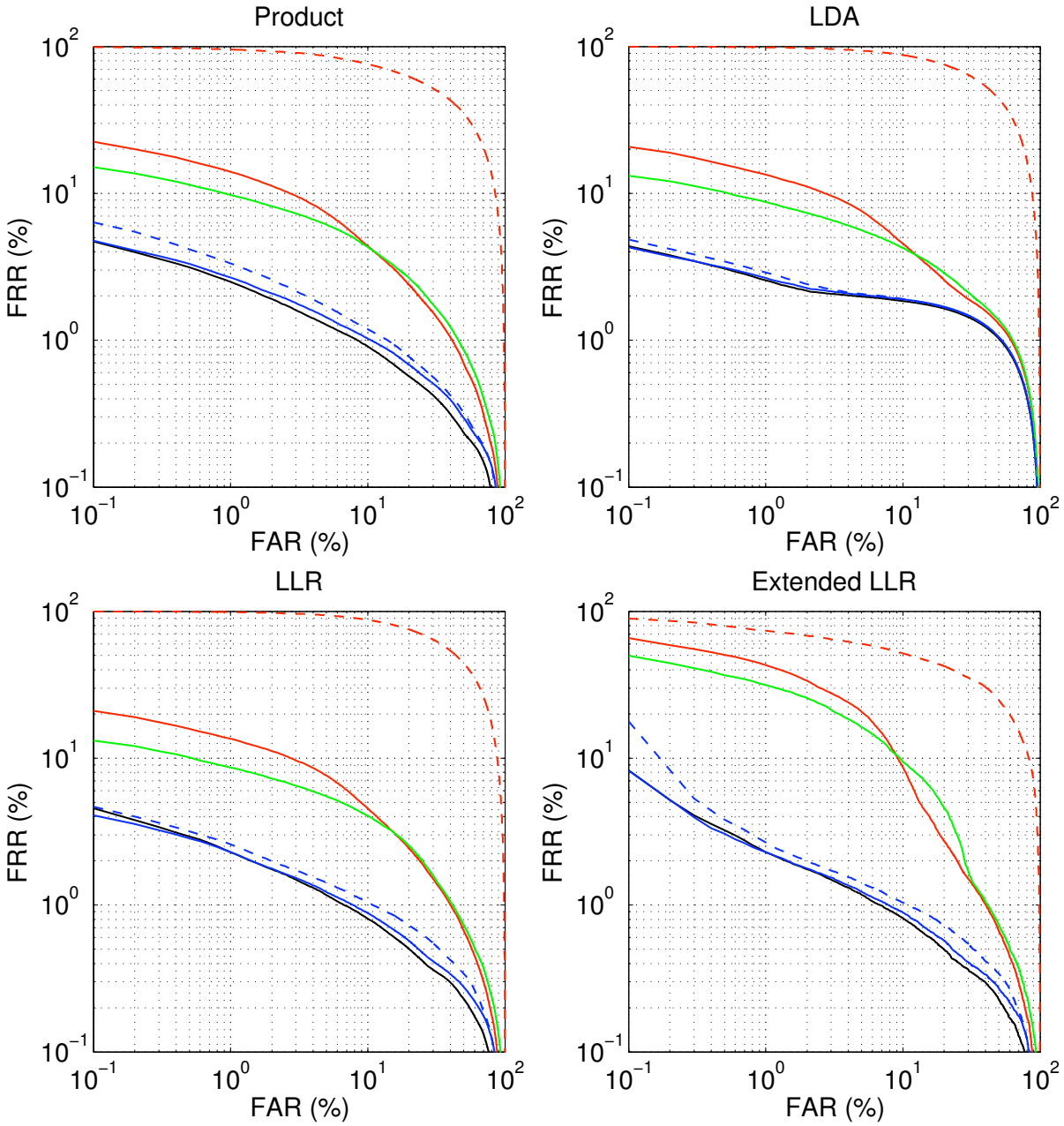

Figure 4.4: Average DET curves attained on test set using silicon spoofed fingerprints and personal photo attack spoofed faces. Each plot refers to a different score fusion rules, indicated in the tittle of each plot. Each plot contains the DET curves attained with no spoof attacks (black), under real spoof attacks (solid curves) and under simulated "worst-case" spoof attacks (dashed curves). Red: fingerprint spoofing only. Blue: face spoofing only. Green: both face and fingerprint spoofing.

Additionally, in Tables 4.2 and 4.3, we report the performance attained on data set used for Figures 4.3 and 4.4, respectively, by all fusion rules, including the sum rule, for different operating points (i.e., decision thresholds). This allows us to better understand the above discussion and to compare more directly

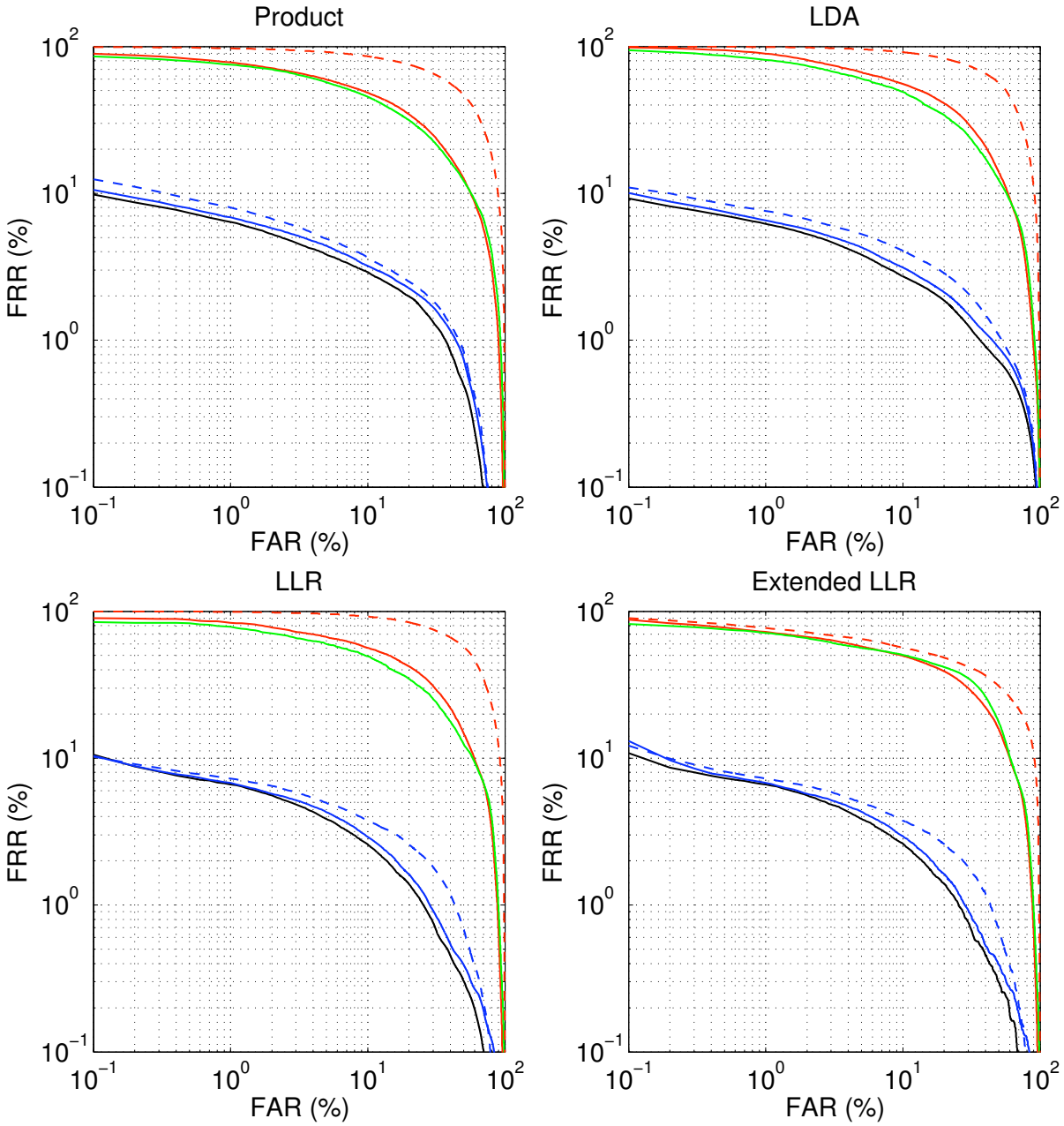

Figure 4.5: Average DET curves attained on test set using latex spoofed fingerprints and personal photo attack spoofed faces. Each plot refers to a different score fusion rules, indicated in the tittle of each plot. Each plot contains the DET curves attained with no spoof attacks (black), under real spoof attacks (solid curves) and under simulated "worst-case" spoof attacks (dashed curves). Red: fingerprint spoofing only. Blue: face spoofing only. Green: both face and fingerprint spoofing.

performance (in terms of FAR and FRR) and robustness to spoof attacks (in terms of SFAR) of the different fusion rules, besides making the results better accessible. Furthermore, the tables also give information about the standard deviation of FRR, FAR and SFAR, which is not provided by the DET curves.

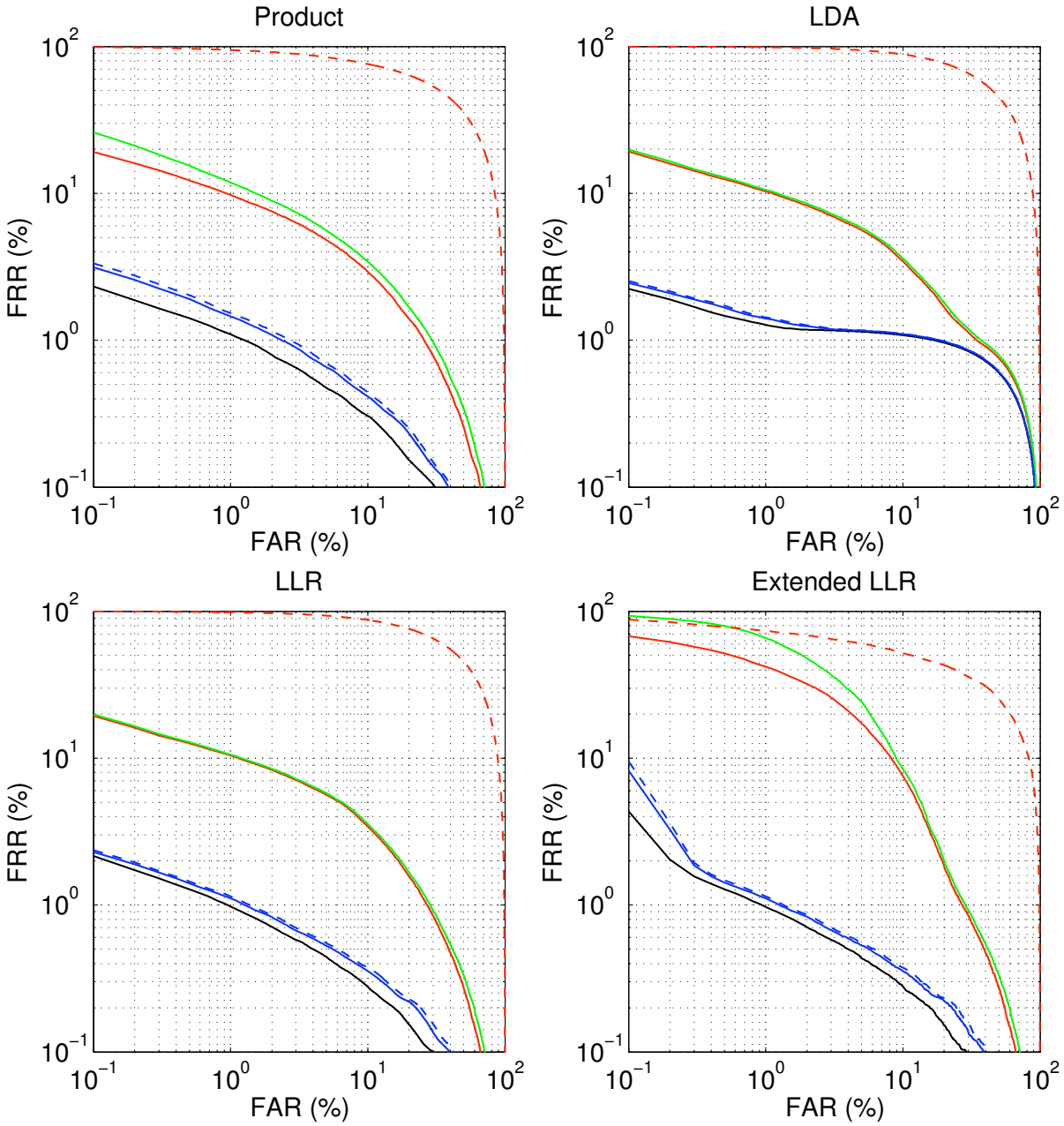

Figure 4.6: Average DET curves attained on test set using silicon spoofed fingerprints and photo attack spoofed faces. Each plot refers to a different score fusion rules, indicated in the tittle of each plot. Each plot contains the DET curves attained with no spoof attacks (black), under real spoof attacks (solid curves) and under simulated "worst-case" spoof attacks (dashed curves). Red: fingerprint spoofing only. Blue: face spoofing only. Green: both face and fingerprint spoofing.

We considered the following three operating points: EER (when FAR=FRR), FAR=1%, FAR=0.1%. Each operating point was fixed on the DET curve obtained without spoof attacks, namely, the one attained by considering genuine users and zero-effort impostors. The FRR at each selected operating point is

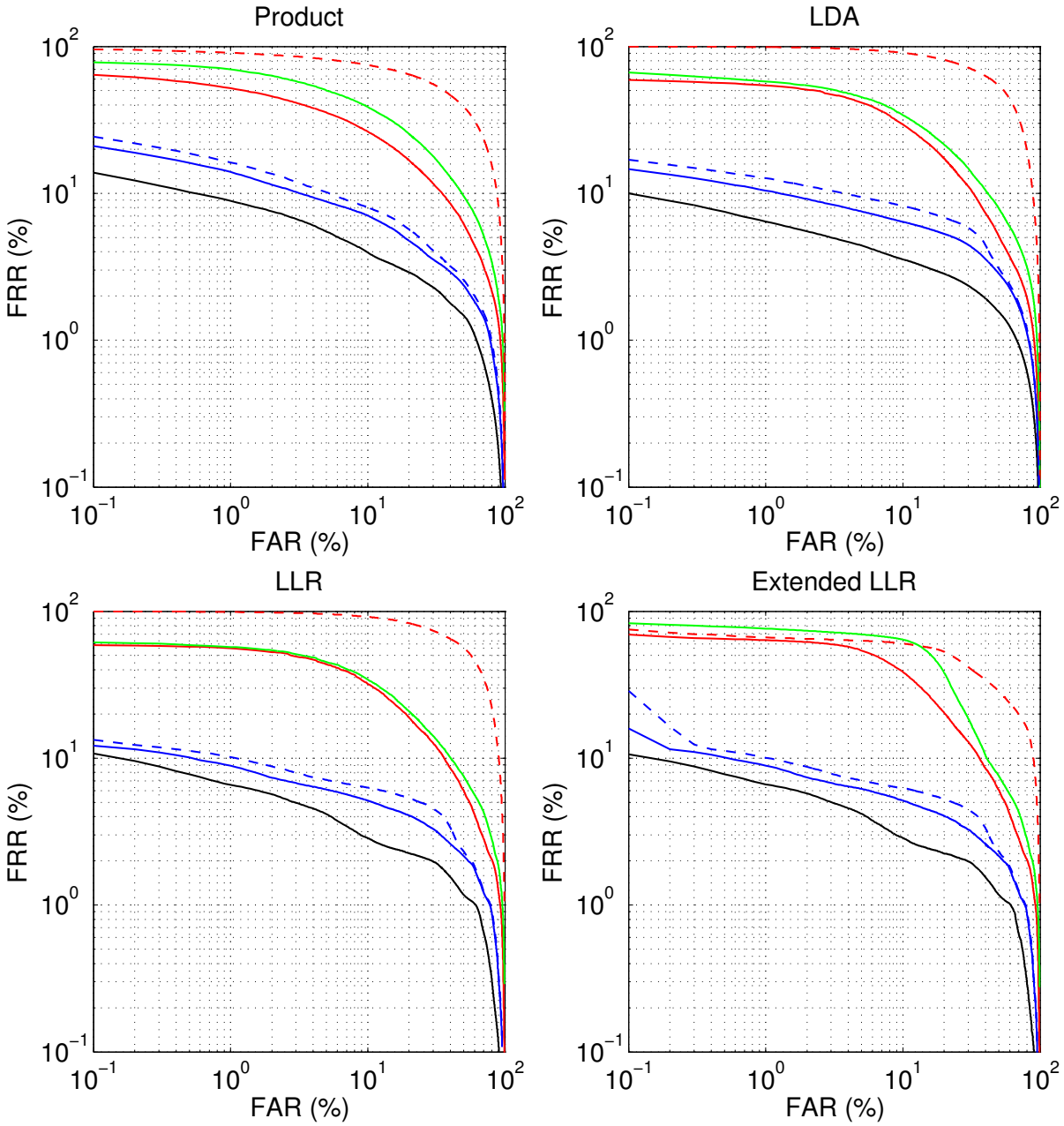

Figure 4.7: Average DET curves attained on test set using gelatin spoofed fingerprints and print attack spoofed faces. Each plot refers to a different score fusion rules, indicated in the tittle of each plot. Each plot contains the DET curves attained with no spoof attacks (black), under real spoof attacks (solid curves) and under simulated "worst-case" spoof attacks (dashed curves). Red: fingerprint spoofing only. Blue: face spoofing only. Green: both face and fingerprint spoofing.

reported in the first column of Tables 4.2 and 4.3 (*no spoof*). Then, we computed the SFAR attained by the different spoof attacks at the same operating point (reported in the remaining columns). This indeed provides a complete understanding of performance and robustness of each fusion rule: once the op-

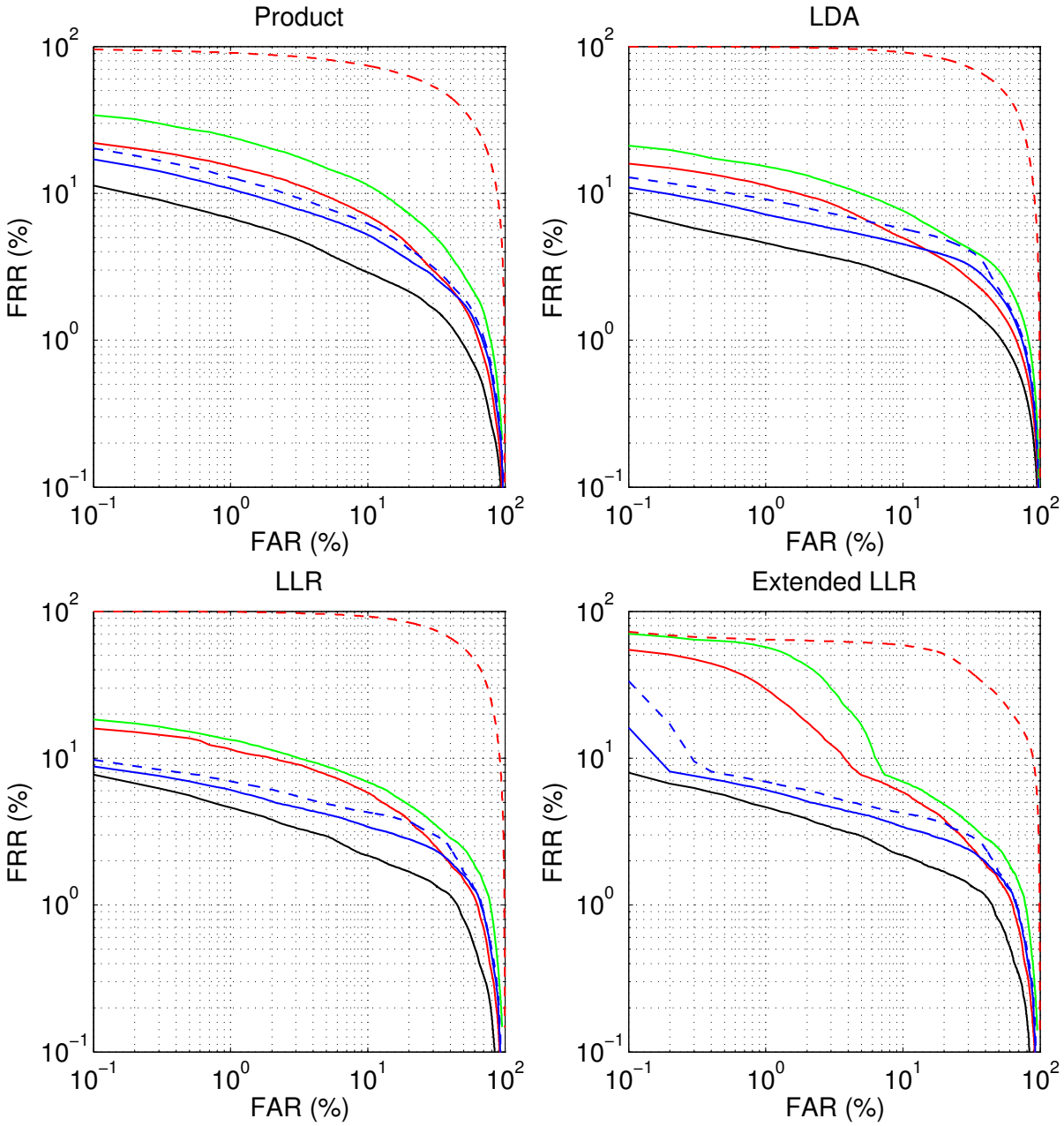

Figure 4.8: Average DET curves attained on test set using alginate spoofed fingerprints and print attack spoofed faces. Each plot refers to a different score fusion rules, indicated in the tittle of each plot. Each plot contains the DET curves attained with no spoof attacks (black), under real spoof attacks (solid curves) and under simulated "worst-case" spoof attacks (dashed curves). Red: fingerprint spoofing only. Blue: face spoofing only. Green: both face and fingerprint spoofing.

erating point is fixed, the effect of spoofing is only to increase the FAR (actually, the SFAR) as it only affects impostor matching scores, while the FRR remains constant.

Let us now compare the different fusion rules used in these experiments. It

Rule	*no spoof*	*face*	*w-face*	*fing.*	*w-fing.*	*both*
	EER %	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	9.98 ± 2.1	33.25 ± 3.9	37.82 ± 3.8	44.07 ± 4.8	79.85 ± 3.8	60.89 ± 2.9
Product	3.49 ± 1.4	5.72 ± 2.1	6.43 ± 2.2	70.06 ± 5.4	96.11 ± 1.8	73.10 ± 4.9
LDA	3.32 ± 1.5	8.39 ± 4.3	9.87 ± 4.8	70.79 ± 5.6	96.36 ± 2.2	74.09 ± 5.4
LLR	3.60 ± 1.4	5.58 ± 2.8	6.36 ± 3.2	71.41 ± 5.1	96.46 ± 2.2	73.47 ± 5.1
Ext. LLR	3.61 ± 1.4	5.64 ± 2.7	6.40 ± 3.1	71.49 ± 5.0	96.38 ± 2.2	73.57 ± 5.1
	FRR % at FAR=1%	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	17.41 ± 3.2	15.58 ± 1.6	20.46 ± 1.9	28.38 ± 5.2	69.00 ± 6.1	46.00 ± 3.8
Product	5.15 ± 2.7	1.93 ± 0.4	2.28 ± 0.5	63.22 ± 4.7	94.37 ± 3.1	66.57 ± 4.4
LDA	5.05 ± 2.6	2.17 ± 0.5	2.73 ± 0.7	64.91 ± 4.7	95.12 ± 3.1	67.83 ± 4.6
LLR	5.46 ± 2.6	1.22 ± 0.4	1.43 ± 0.5	64.94 ± 4.7	95.22 ± 3.1	66.38 ± 4.7
Ext. LLR	5.63 ± 2.8	1.17 ± 0.4	1.38 ± 0.5	64.68 ± 4.8	94.94 ± 3.3	66.03 ± 4.8
	FRR % at FAR=0.1%	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	22.76 ± 3.5	8.84 ± 1.1	12.68 ± 1.5	21.65 ± 4.9	62.68 ± 6.7	37.30 ± 3.9
Product	8.59 ± 4.1	0.30 ± 0.1	0.36 ± 0.1	53.41 ± 5.5	90.62 ± 4.1	56.79 ± 4.9
LDA	7.99 ± 3.7	0.26 ± 0.1	0.32 ± 0.2	56.32 ± 5.3	92.45 ± 3.8	58.66 ± 5.2
LLR	8.91 ± 4.1	0.14 ± 0.1	0.17 ± 0.1	56.23 ± 6.0	92.39 ± 3.9	57.48 ± 6.0
Ext. LLR	9.46 ± 5.2	0.16 ± 0.1	0.19 ± 0.1	56.13 ± 5.7	90.97 ± 5.5	57.27 ± 6.0

Table 4.2: EER, FRR at FAR=1%, and FRR at FAR=0.1% for the considered fusion rules on latex spoofed fingerprints and photo attack spoofed faces (*no spoof*). The SFAR corresponding to the same operating points is reported for real spoofing of fingerprint (*fing.*), face (*face*), and both traits (*both*), and under simulated worst-case spoofing of fingerprint (*w-fing.*), and face (*w-face*). Results are averaged over 25 runs and reported as mean and standard deviation.

shows that standard rules (Sum, Product, LDA and LLR) did not exhibit appreciable performance differences, both in the case of spoof attacks against only one biometric (either face or fingerprint), and of double spoofing. In other words, they exhibited a similar robustness. It is worth noting that the Extended LLR exhibited a significantly worse performance than standard rules, despite it was specifically designed to be robust under spoofing attacks. This behavior seems due to the fact that the worst-case assumption behind this rule turned out to be too pessimistic. Note also that, as pointed out in Section 4.3.1, another problem of Extended LLR is that setting its parameters (α, c_1 and c_2 for a multimodal systems composed of two traits) is not trivial, as their values can only be hypothesized in advance (and, for instance, not tuned on some validation data).

Further, we extend our investigation to evaluate the two common beliefs about the robustness of multimodal biometric systems: First, multimodal biometric systems can be more robust than each corresponding unimodal system, even in the case when all biometric traits are spoofed. Second, multimodal sys-

Rule	*no spoof*	*face*	*w-face*	*fing.*	*w-fing.*	*both*
	EER %	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	15.14 ± 3.2	19.80 ± 6.0	43.97 ± 5.2	18.56 ± 2.7	66.45 ± 6.9	21.75 ± 6.9
Product	1.89 ± 0.7	2.55 ± 1.5	3.88 ± 1.3	24.32 ± 5.2	96.82 ± 1.1	25.61 ± 15.1
LDA	1.70 ± 0.7	1.50 ± 0.9	2.31 ± 1.7	22.21 ± 6.3	96.80 ± 1.7	20.91 ± 14.7
LLR	1.78 ± 0.7	1.96 ± 1.2	2.67 ± 1.1	25.57 ± 5.7	97.46 ± 1.0	24.45 ± 14.8
Ext. LLR	1.79 ± 0.7	1.95 ± 1.2	2.60 ± 1.0	25.50 ± 5.6	97.44 ± 1.1	24.43 ± 14.8
	FRR % at FAR=1%	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	28.82 ± 6.2	1.68 ± 0.7	21.47 ± 2.5	2.25 ± 0.8	42.40 ± 11.7	2.27 ± 2.0
Product	2.49 ± 1.2	1.25 ± 0.5	2.16 ± 0.3	19.54 ± 4.7	95.95 ± 1.7	20.51 ± 13.6
LDA	2.56 ± 1.3	0.98 ± 0.4	1.47 ± 0.6	19.40 ± 5.2	96.42 ± 1.8	18.52 ± 13.2
LLR	2.29 ± 1.1	1.02 ± 0.4	1.43 ± 0.2	21.32 ± 5.1	96.82 ± 1.6	20.40 ± 14.0
Ext. LLR	2.29 ± 1.1	1.02 ± 0.4	1.43 ± 0.2	21.32 ± 5.1	96.75 ± 1.5	20.36 ± 13.9
	FRR % at FAR=0.1%	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	35.35 ± 6.5	0.18 ± 0.2	14.04 ± 1.9	0.55 ± 0.4	33.61 ± 11.6	0.64 ± 1.2
Product	4.69 ± 1.8	0.12 ± 0.1	0.34 ± 0.1	9.25 ± 3.6	92.56 ± 2.7	9.05 ± 8.5
LDA	4.36 ± 1.8	0.10 ± 0.1	0.18 ± 0.1	10.17 ± 3.9	94.18 ± 2.4	9.30 ± 8.9
LLR	4.56 ± 1.9	0.08 ± 0.1	0.13 ± 0.0	10.03 ± 4.5	94.01 ± 2.6	9.17 ± 8.8
Ext. LLR	8.28 ± 7.6	0.10 ± 0.1	0.17 ± 0.0	9.61 ± 3.9	84.95 ± 16.4	9.04 ± 8.4

Table 4.3: EER, FRR at FAR=1%, and FRR at FAR=0.1% for the considered fusion rules on silicon spoofed fingerprints and personal photo attack spoofed faces (*no spoof*). The SFAR corresponding to the same operating points is reported for real spoofing of fingerprint (*fing.*), face (*face*), and both traits (*both*), and under simulated worst-case spoofing of fingerprint (*w-fing.*), and face (*w-face*). Results are averaged over 25 runs and reported as mean and standard deviation.

tems can be cracked by spoofing all the fused traits, even when the attacker is not able to fabricate an exact replica of the genuine user's traits.

For this purpose, in Figure 4.9 we show the DET curves of the unimodal and the multimodal systems under normal operation (i.e., with no spoof attack), using solid curves, and the performance under a spoof attack against one trait (both for the unimodal and multimodal systems) and both traits (for the multimodal systems), using dashed curves. For reference, we also report the DET curve corresponding to a "worst-case" attack against both traits, which was simulated as in [74, 73, 42]. Note that the latter DET curve corresponds to the line FAR = FRR, as the match score distribution of fake traits is assumed to be identical to the one of genuine users. Note also that the red (fingerprint individual system) and green curves (face individual system), as well as the black dashed curve ("worst-case" attack against both traits of the multi-modal system) are the same in all plots. Therefore, only the two blue curves corresponding to multi-

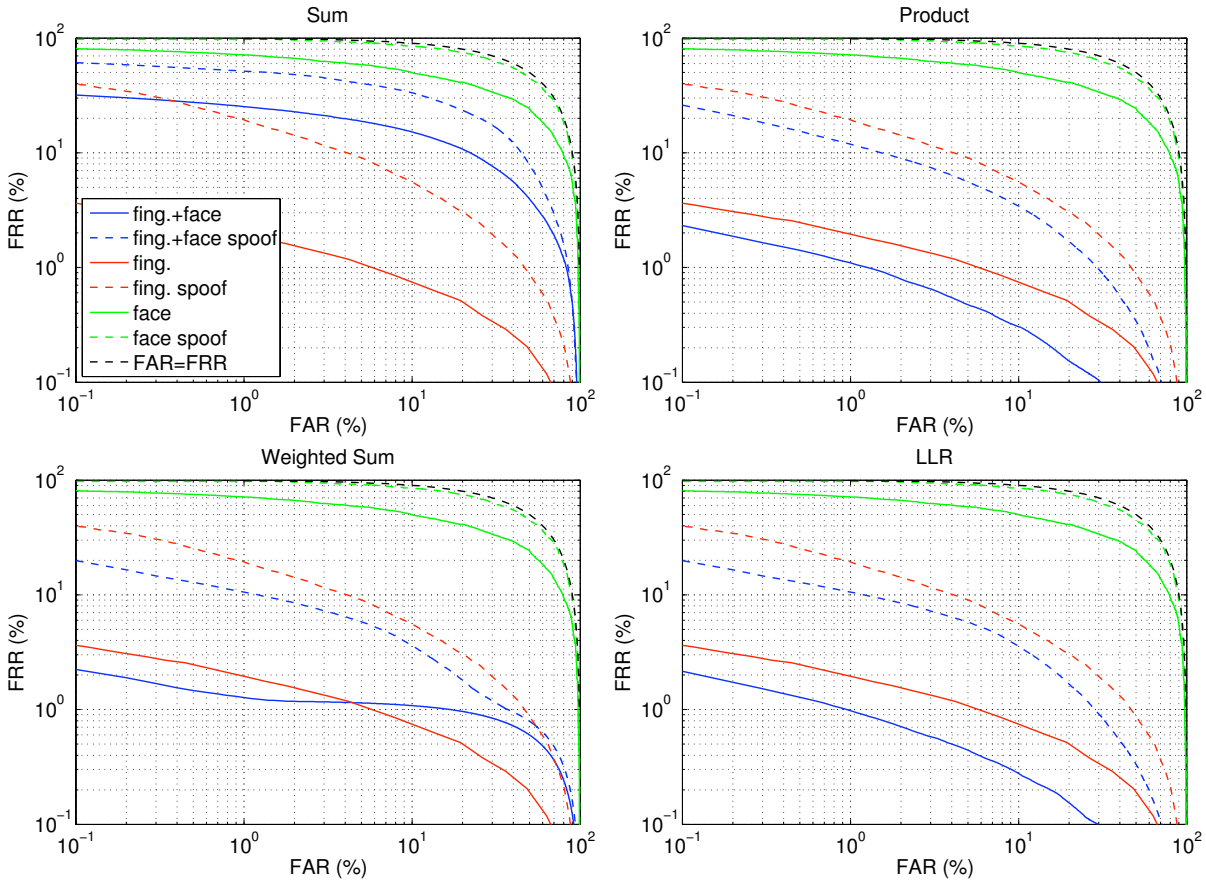

Figure 4.9: Average DET curves obtained in our experiments on the testing set using silicon spoofed fingerprints and photo attack spoofed faces. Each plot refers to a different fusion rule, and contains DET curves of the systems under normal operation (solid curves) and under spoof attacks (dashed curves). Green: unimodal face system. Red: unimodal fingerprint system. Blue: multimodal face and fingerprint system (a spoof attack against both traits is considered). Black: multimodal system under a simulated "worst-case" spoof attack against both traits.

modal systems without spoof attacks, and with realistic spoof attacks against both traits, change depending on the fusion rule.

We can first observe that, under normal operation (i.e., no spoof attacks), information fusion has improved the performance of the corresponding unimodal systems, except using sum and weighted sum score fusion rules, at high FAR vales. This behavior is due to the fact that the genuine and impostor score distributions of the face matcher in the considered data set turned out to be more overlapping than the ones produced by the fingerprint matcher (see Figures 4.11 and 4.12), and it can be noted from the worse DET curves obtained by the individual face system with respect to the individual fingerprint system. In other words, generally the benefit of fusion are exploited when the fused matchers show complementary nature. However, since in this case the performance

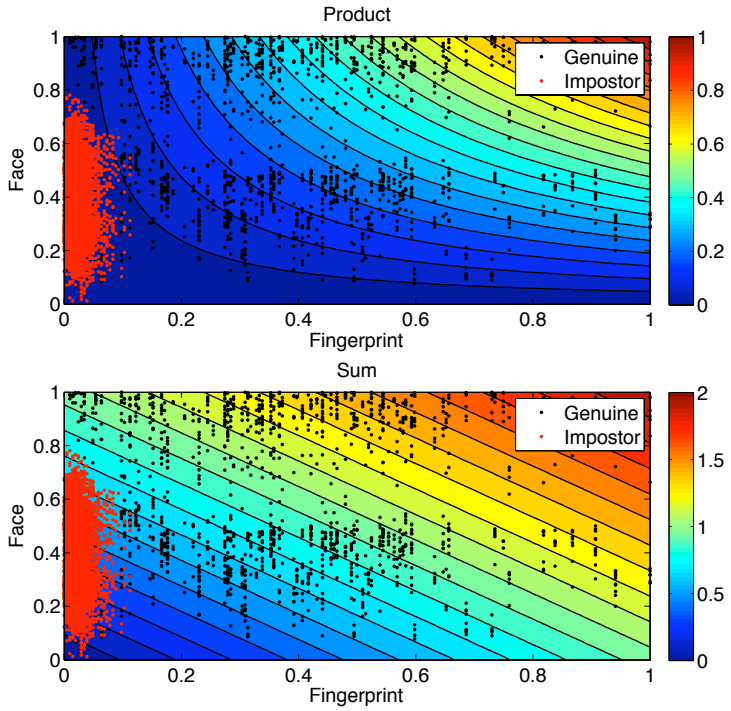

Figure 4.10: Fusion of face and fingerprint matching scores through product (top) and sum (bottom). The values attained by the two fusion rules are shown in different colors. Genuine and impostor scores for fingerprint spoof attacks and face spoof attacks are also reported to highlight how the product rule may outperform the sum rule.

of the face matcher was considerably worse over all the DET curve than that of the fingerprint matcher, and this performance imbalance clearly degraded the performance of the multimodal system using sum rule below the best performing matcher. But, this performance imbalance phenomenon did not affect product rule (although one may think that the product rule should be similarly affected), as exemplified in Figure 4.10. In fact, the (hyperbolic) decision functions provided by the product rule correctly assigned a very low match score to the majority of impostors, biased by the very low output of the fingerprint matcher. Conversely, on average, the sum rule increased their score, worsening the performance.

Let's now compare the DET curves related to the real spoof attacks against unimodal and multimodal systems in Figure 4.9, thus providing the evidence to the common claim that multimodal biometric systems are more robust to spoof attacks than unimodal ones. As it is easy to see that the performance of the multimodal system under a real spoof attacks against both traits is better than the one attained by both unimodal systems under attack, for all score fusion rules, though the performance under attack considerably worsen both for the unimodal face and fingerprint system. In other words, an attacker has lower chances to

evade the multimodal systems considered in our experiments, when he spoofs both traits, than to evade each single unimodal system. Accordingly, we can say that the multimodal systems considered in our experiments exhibited a higher robustness to real spoof attacks against both traits, than the corresponding unimodal systems against the same attacks.

However, a comparison of the solid blue (normal operation) and dashed blue (real spoof attacks) curves clearly shows that the performance of the multimodal systems under a real spoof attacks against both traits is significantly worse than under normal operation, for all the fusion rules considered in this study. This indicates that the probability of an impostor evading the multimodal systems is high, even if the attacker does not fabricate a perfect replica of the spoofed trait, namely under non-worst scenarios. For instance, using the LLR score fusion rule, in Figure 4.9, at ZeroFAR operational point on the training set (namely, the lowest decision threshold s^*, which provides a zero percent "zero-effort" impostor acceptance rate on training samples) under normal operation, an average FAR of 0.31% is obtained on test sets. When both face and fingerprints are spoofed ("non-zero" effort impostor) instead, the FAR increases to 55.01%.

To sum up, we provided experimental evidences that multimodal biometric systems are not intrinsically robust against spoof attacks as believed so far. They can be fooled by spoofing only one biometric, even when the attacker is not able to fabricate the exact replica a genuine user's trait. However, multimodal systems are more robust to spoof attacks than corresponding unimodal systems that compose them.

4.4.2 "Worst-case" hypothesis validation

As mentioned in the introduction of this chapter that state-of-the-art method to evaluate the security of the multimodal biometric systems against spoof attacks is to hypothesize spoof attacks under the "worst-case" scenario, where the attacker is able to replicating exactly the targeted biometric. However, this "worst-case" hypothesis may not be always true for all biometrics in real-life scenarios, as also pointed out in [42]. Therefore, in this section, we evaluate to what extend the "worst-case" scenario is realistic, and thus analyzing to what extent the drop of performance under "worst-case" attack scenario is representative of the performance under real spoof attacks.

From Figure 4.3, it is easy to see that the "worst-case" assumption is realistic when faces are spoofed by the photo attack and print attack data sets, using an image similar to the template: the fake score distributions are very close to the

ones of genuine users (see Figure 4.12). This is also true for the corresponding values in Table 4.2 (*face* and *w-face* columns). Thus, modelling fake score distributions as genuine ones, as proposed in [73], seems acceptable in this scenario. The same does not hold however for latex-based fake fingerprints, which are nevertheless the highest quality (and most effective) fake fingerprints obtained in our data sets: as it can be seen by plots in Figure 4.3 and from the values in Table 4.2 (*fing.* and *w-fing.* columns), the corresponding FAR is clearly overestimated by the "worst-case" assumption (being equal the FRR), with the only exception of the Extended LLR rule.

A similar behaviour to that described above can be noted in Figure 4.4. In particular, in this case, the spoofed traits were less effective than in the previous case, resulting in a stronger violation of the "worst-case" assumption. As the differences between the FAR attained under the "worst-case" assumption, and the one observed on our data sets, is even higher, both for spoofed faces and for spoofed fingerprints. In the case of face spoofing, the performance is very close to the one attained without a spoof attack. In the case of fingerprint spoofing, the performance is remarkably far both from the one attained in the "worst-case" scenario, and the one attained without spoof attacks.

Similarly, it can also be seen in Figure 4.9 that the performance attained by the multimodal systems under real spoof attacks against both traits is much better (although still not suitable for the requirements of security applications) than the one predicted under the "worst-case" assumption (black dashed line).

For the sake of completeness, and to further confirm the results of the evaluation carried out in this chapter, we also report here the matching score distributions of the genuine, impostor, and fake traits, for each fingerprint and face data set, obtained by the Bozorth3 and EBGM matching algorithms, respectively, in Figures 4.11 and 4.12.

The "worst-case" scenario hypothesized in [74, 73, 42] amounts to assuming that the distribution of the fake traits corresponds to that of the genuine users. But, above we have already pointed out that this hypothesis can be violated, leading to a too pessimistic evaluation of the FAR of multimodal biometric systems under spoof attacks. The score matching distributions in Figures 4.11 and 4.12 confirm the above discussed results, which are summarized as follows:

1. The "worst-case" assumption is too pessimistic and unrealistic in the case of fingerprint spoofing, even when the fake fingerprints are constructed with the consensual method, as in all our data sets (Figure 4.11). The reason is that the fingerprint image obtained by a fake fingerprint often

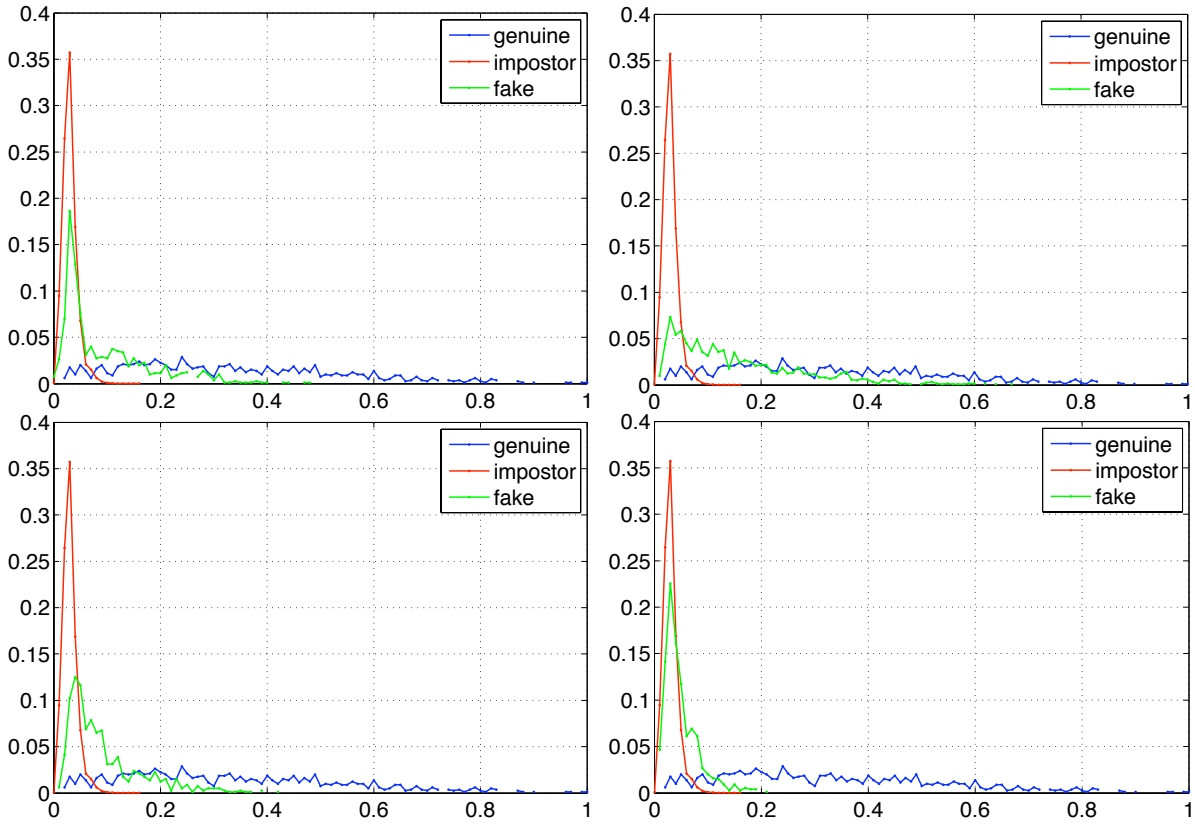

Figure 4.11: Score matching distributions for the fingerprint data sets: Top (left): fake fingerprints obtained by using silicon. Top (right): fake fingerprints obtained by using latex. Bottom (left): fake fingerprints obtained by using gelatin. Bottom (right): fake fingerprints obtained by using alginate.

presents artifacts which affect the matching algorithm; for instance, not all minutiae points can be perfectly replicated from the source image. Nevertheless, the distributions of the fake matching scores may still significantly worsen the performance with respect to the "zero-effort" impostor distribution, although not to the extent predicted by the "worst-case" hypothesis in [74, 42]; in particular, this is true when gelatin and latex are used (Figure 4.11).

2. Conversely, the "worst-case" assumption is well suited to face spoofing, provided that the fakes are constructed with images that are very similar to the stored templates, as in the case of the Photo Attack and Print Attack data sets (see Figure 4.12). The reason is that printing a face image on paper, or displaying it on a laptop screen, does not generate any particular artifact which affects the matching algorithm. However, this does not exclude that some particular artifacts may exist (e.g., printing failures

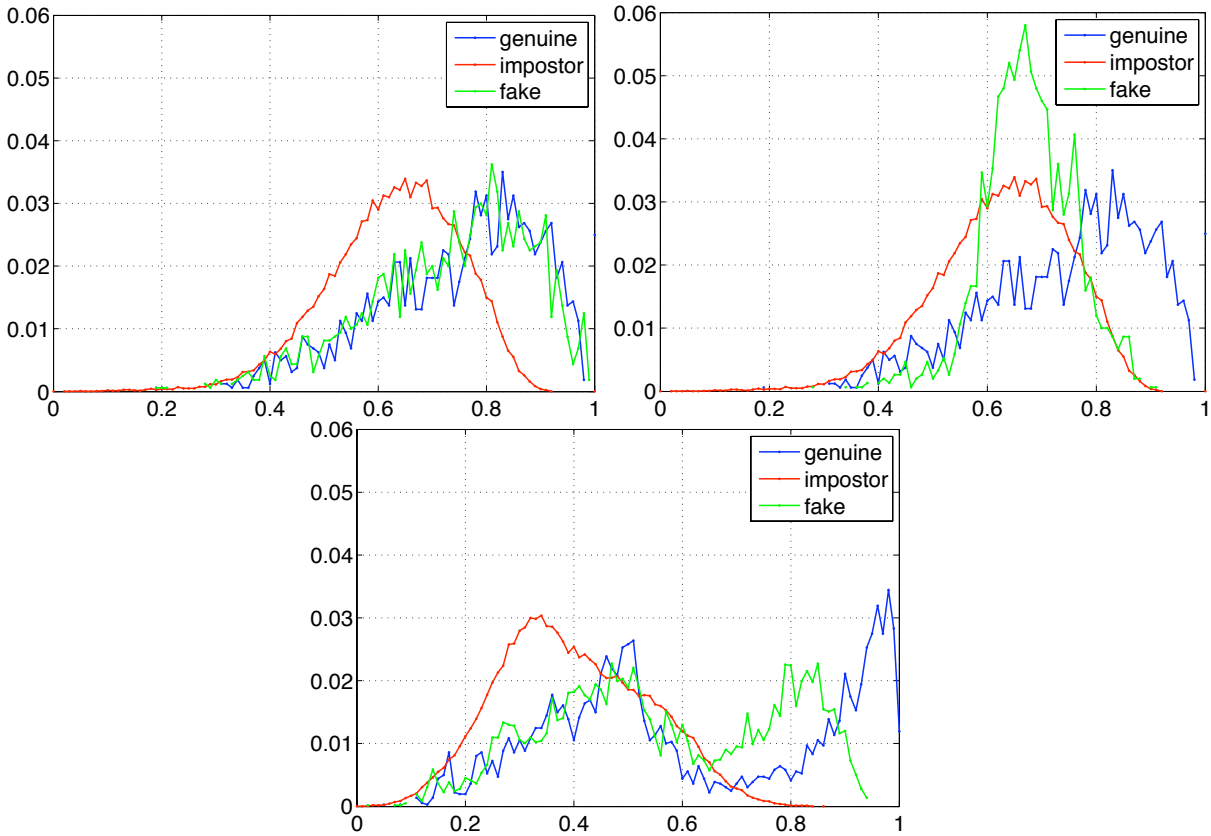

Figure 4.12: Score matching distributions for the face data sets. Top (left): fake faces obtained by a photo attack (Photo Attack data set). Top (right): fake faces obtained by a personal photo voluntarily provided by the user (Personal Photo Attack data set). Bottom: fake faces obtained by a print attack (Print Attack data set).

or blurring), and, indeed, they can be successfully exploited for liveness detection [10, 82, 95]. This is however not the case, when face images significantly different than the stored templates are used, e.g., when they are collected through the Web, as in the Personal Photo Attack (see Figure 4.12).

To conclude this section, we can state that above results provide evidence that modelling the matching score distribution of spoof attacks using the "worst-case" assumption of [74, 73, 42] is not always suitable for evaluating the robustness of multimodal systems, and for developing robust score fusion rules. Moreover, they also showed that producing very effective fake faces may be relatively easier for an attacker. This is in agreement with the results of the Competition on Countermeasures to 2D Facial Spoofing Attacks [6, 10], and further highlights the need for effective liveness detection techniques against face spoofing.

4.5 Summary

In this chapter, we investigated the problem of spoof attacks against multimodal biometric verification systems. In particular, we focused on a multimodal system consisting of a fingerprint and a face matcher.

The state-of-the-art method to assess the security of a multimodal biometric system against spoof attacks is a "worst-case" scenario in which the attacker is able to replicate perfectly the genuine biometric traits, and thus the fake score distribution is identical to the one of genuine users [74, 73, 42]. This lead to the conclusion that multimodal biometric systems are not intrinsically robust against spoof attacks, as they can be cracked by spoofing only one biometric. However, "worst-case" scenario may not be always true for all biometrics. Thus, we presented an extensive experimental analysis involving real spoof attacks to provide the further evidence to the concussion drawn in [74, 73, 42]. In addition, our experiments on real spoof attacks also provided evidence of two common beliefs about the robustness of multimodal biometric systems. First, they can be more robust than each corresponding unimodal system, even when all the fused biometric traits are spoofed. Second, their performance under a spoof attack against all traits is still unacceptable for security applications. In other words, they can be cracked by spoofing all the fused traits, even when the attacker is not able to fabricate an exact replica of the genuine user's traits.

Through experiments carried out on several face and fingerprints real spoof attacks and using five different score fusion rules (including the one of [74]), we found that a "worst-case" scenario can not be representative of real spoof attacks: its suitability may depend on the specific biometric trait, the matching algorithm, and the techniques used to fabricate the spoofed traits. In particular, we found that the above "worst-case" assumption can be too pessimistic, thus resulting in a significant overestimation of the FAR that a multimodal system may incur under a real spoof attack. This can also undermine the effectiveness of score fusion rules based on such assumption, like the one of [74], that turned out to be less effective than standard rules like Product, LDA and LLR. More details of the work presented in this chapter can be found in the following publications [2, 8, 9].

Our empirical analysis suggests an interesting open issue of investigating the development of more realistic models of the score distribution produced by spoof attacks, aimed both at assessing the robustness of multimodal systems, and designing novel and robust score fusion rules. In the next chapter we address this open issue.

Chapter 5

Method for Evaluating the Security of Multimodal Biometric Systems against Spoof Attacks

5.1 Introduction

As described in previous chapter, state-of-the-art method to evaluate the security of multimodal systems against spoof attacks is "worst-case" scenario, where the fake distribution is simulated by assuming that attacker is able to replicate exactly the targeted biometric: in other words, the fake score distribution coincides with that of genuine users. But, this methods suffer from the main drawback of inability to reasonably approximate performance of the system under real spoof attacks, due to operation at stringent "worst-case" hypothesis. Moreover, our empirical results in chapter 4 showed that the drop of performance predicted under spoof attacks with the "worst-case" scenario is not always realistic.

A straightforward approach could be to fabricate fake traits to evaluate the security of the system under design. However, constructing reliable fake replicas is a cumbersome task, and fabricating fakes exhibiting different "quality" degrees is even more difficult and impractical [52, 94, 57, 50]. A potential alternative is to develop methods based on simulating the distribution of fake biometric traits. Thus, it is crucial to develop evaluation methodologies that allow assessing the robustness of multimodal biometric systems depending on the particular biometric traits, the score fusion rule used, and the "quality" of the fake traits used by the attacker, which in real scenarios are likely to be not perfect replicas of the genuine traits. Due to complication in fabricating the real spoof attacks, it is strongly desirable that the use of such methodolo-

gies requires only samples of genuine and impostor users, which are the unique samples commonly available to the designer of a biometric system. The need for a methodology to evaluate the robustness of a multimodal biometric systems against spoof attacks without actual fabrication of spoof attacks was also highlighted in Chapter 3, where we reviewed several works which assessed the robustness of multimodal systems using *simulated* spoof attacks, and in Chapter 4, where we evaluated the performance of the systems against real spoof attacks. To the best of our knowledge, no systematic research effort has been carried out toward this direction yet in the literature.

Thus, in this chapter our gaol is to develop a general methodology to evaluate the robustness of a multimodal biometric system in adversarial environments at the design phase, where the concept of "security" is related to the performance degradation a biometric system incurs when it is under spoof attack. We first present in Section 5.1.1 the rationale underlying our methodology, which is then described in Sections 5.2–5.3. Eventually, in Sections 5.4–5.5, we present experimental results on data set comprised of real spoof attacks to evaluate the capability of our methods to approximate score distributions produced by real spoof attacks, to reliably assess the robustness of multimodal systems under attacks, and then accordingly to predict the relative robustness of several score-level fusion rules, namely the ranking of their performance under attacks.

5.1.1 Goal and scope of the proposed method

Based on above mentioned motivations in section 5.1 and experimental evidences of Chapter 4, we propose a method to evaluate the security (robustness) of multimodal systems against spoof attack, which is aimed at addressing the open issue (c) stated in Chapter 3 Section 3.5, i.e., *How can the security of multimodal systems be evaluated, under realistic attacks, without fabricating spoofed traits?*

In this chapter, our goals are: (i) To evaluate the capability of our proposed models for fake score distribution to approximate real fake score distributions.; (ii) To investigate whether our proposed performance evaluation method provides a good estimate of the performance of biometric systems under spoof attacks.; (iii) To predict the FAR of different score-level fusion rules under spoof attacks, namely the *ranking* of their performance under attacks.

In order to avoid the straightforward but cumbersome solution of constructing spoofed biometric traits to test the system, our method is based on simulating the effects of a spoof attack on the distribution of corresponding match

scores, as in [74, 73, 42]. However, differently from these works, our aim is to take into account also more realistic, non-worst case scenarios, in which the fake score distribution can be different than the genuine one. When a fake trait is submitted to a biometric matcher, several factors may affect the resulting output score distribution, like the particular biometric trait spoofed, the particular matching algorithm used, the forgery technique used by the attacker, and the skills of the attacker, etc. However, at the state-of-the-art their effect on position and shape of score distribution is unknown. Thus, we make substantive assumptions on the potential form and shape it can get and thus propose to model such distribution by assuming the effect of the above factors that they can exhibit different shapes, and in particular, that it can be identical either to the impostor or to the genuine score distributions, or lies between them. We model the fake score distribution as a function of the genuine and impostor distributions, on the basis of a single parameter, that we call "attack strength". This parameter controls the degree of similarity of the fake and genuine scores, with respect to the impostor scores. The "attack strength" quantifies the effect of several factors mentioned above, and allow to figure out more possible scenarios than that of the only worst-case one in [74, 73, 42]. The "attack strength" parameter quantifies the ability of the attacker to fabricate a fake biometric trait that mimics the true trait: the higher the similarity, the higher the "attack strength" of the spoof attack. To evaluate the robustness of a system under spoof attacks using our method, the testing impostor scores of the matcher under attack have to be replaced with simulated fake scores generated as mentioned above. Repeating this procedure for different values of the "attack strength" parameter allows the designer to evaluate robustness under attack scenarios characterised by different degrees of similarity of the fake and genuine score distributions, namely, to evaluate robustness against different attackers which are able to fabricate more or less good replicas of true biometric traits, thus providing an estimation of the system performance under potential spoof attacks of different strength.

In summary, the approach we propose consists in simulating the effect of a spoof attack at the matcher's output, by modelling the distribution of the scores produced by submitting fake traits to the biometric system. This approach can be used to analyze the robustness of biometric authentication systems against spoof attacks, under one or more possible forms that the fake score distribution may take, in the following scenarios:

1. Analysis of the performance of a given individual matcher.

2. Analysis of the performance of a given multimodal system, namely, anal-

ysis of a given set of matchers and a given score fusion rule.

3. Comparison of different multimodal systems made up by the same set of matchers, using different score fusion rules, namely, ranking of score fusion rules according to their robustness to spoof attacks.

5.2 Models of the Match Score Distribution produced by Spoof Attacks

In the following, we describe our proposed two models of fake scores distributions to simulate the spoof attacks at match score level: Non-Parametric model and Parametric model.

5.2.1 Non-parametric model

In real-life frameworks it is reasonable to assume that the score distribution of fake traits may be different than the genuine one, and it may take different forms, depending on factors like the one mentioned in Section 5.1.1.

To develop a model of the fake scores distribution it would be very useful to start from empirical data. However, although several data sets of live and fake biometric traits currently exist, but they do not provide useful information on the score distribution of fake traits, since they have not been indexed by the users' identity. To our knowledge, the only exception is the data set of [1], which however is rather small and not publicly available.

The solution we propose is to make a working assumption on the possible forms that the fake scores distribution may exhibit, due to the possible effects of the different factors mentioned in Section 5.1.1. In the following we denote with s the score of a biometric matcher, and with G and I the events that the input biometric is true and comes respectively from a genuine user and an impostor, while the event that it is a fake biometric will be denoted as F. The corresponding score distributions will thus be denoted as $p(s|\mathrm{G}), p(s|\mathrm{I})$ and $p(s|\mathrm{F})$. Our working assumptions on the form of $p(s|\mathrm{F})$ are the following:

1. In the worst case for the system (and it is the best case for the attacker), the attacker is able to fabricate exact replicas of the targeted biometric trait, and thus the distribution of fake scores is identical to the one of genuine user: $p(s|\mathrm{F}) = p(s|\mathrm{G})$. This is the only scenario considered in [74, 73, 42].

2. In the best case for the system (worst case for the attacker), the fake trait is very different from the one of the targeted genuine user, such that the

attacker does not get a better result than if he submitted his own original trait. Accordingly, in tis case $p(s|\mathrm{F}) = p(s|\mathrm{I})$. We consider this as the case of fakes with the worst quality. We don't consider the case of fakes exhibiting an even lower quality, leading to a score distribution worse (for the attacker) than the impostor one: although it may be possible in practice, it is obviously of no interest to the purpose of robustness evaluation.

3. In "intermediate" cases, we assume that $p(s|\mathrm{F})$ lies between $p(s|\mathrm{I})$ and $p(s|\mathrm{G})$, and model its possible shapes as discussed below.

In absence of more specific information on the possible shapes that the fake score distribution may exhibit, we propose to simulate "intermediate" cases as follows: we replace each impostor score s_I with a fictitious score s_F given by

$$s_\mathrm{F} = (1-\alpha)s_\mathrm{I} + \alpha s_\mathrm{G}\ , \tag{5.1}$$

where s_G is a randomly drawn genuine score, and $\alpha \in [0,1]$ is a parameter which controls the degree of similarity of the distribution of fake scores to the one of genuine scores, namely, a parameter that controls the relative distance of the fake distribution to the ones of impostor and genuine users. The resulting distribution of fictitious fake scores $p(s|\mathrm{F})$ is thus "intermediate" between the ones of $p(s|\mathrm{I})$ and $p(s|\mathrm{G})$. By using different values of α, one gets different possible distributions: higher the α value, the closer $p(s|\mathrm{F})$ to the genuine score distribution $p(s|\mathrm{G})$, and thus the more effective the spoof attack. Accordingly, we name α as "attack strength". This parameter, α, and related Eq. (5.1), are aimed not to model the physical fake generation process, but only its effect on the corresponding distribution $p(s|\mathrm{F})$, which depends on several causes like the spoof attacks carried out using different techniques, or attempted by different attackers with different forgery skills, etc. For example, in the case of fingerprint, its "similarity" to the impostors distribution will be caused by several factors as artefacts in the replica, the image distortion from the mould to the cast, the good/bad pressure of the attacker on the sensor surface when placing the spoofed fingerprint, whilst its "similarity" to the genuine users one is given by the fact that several important features, as the ridge texture and minutiae locations, will be the same of the correspondent subject.

5.2.2 Parametric model

We propose the second kind of simulation of the fake score distribution using a parametric model as follows. Based on the same working assumptions

mentioned in Section 5.2.1 for non-parametric model, here we model $p(s|\mathrm{F})$ in "intermediate" cases using a parametric model based on a given distribution, like a Gaussian, Gamma or Beta. We assume that $p(s|\mathrm{F})$ has the same form as $p(s|\mathrm{G})$ and $p(s|\mathrm{I})$, and that the value of each of its parameters is between the values of the corresponding parameter in $p(s|\mathrm{G})$ and $p(s|\mathrm{I})$. For instance, if $p(s|\mathrm{G})$ and $p(s|\mathrm{I})$ are modeled as Gaussians with mean and variance denoted as $\mu_\mathrm{G}, \mu_\mathrm{I}, \sigma^2_\mathrm{G}$ and σ^2_I, we are assuming that $p(s|\mathrm{F})$ is Gaussian as well, and that its mean and variance satisfy the constraints:

$$\begin{aligned} \mu_\mathrm{F} &\in [\min\{\mu_\mathrm{G}, \mu_\mathrm{I}\}, \max\{\mu_\mathrm{G}, \mu_\mathrm{I}\}], \\ \sigma_\mathrm{F} &\in [\min\{\sigma_\mathrm{G}, \sigma_\mathrm{I}\}, \max\{\sigma_\mathrm{G}, \sigma_\mathrm{I}\}]. \end{aligned} \tag{5.2}$$

In other words, we model the possible fake score distributions as a "morphing" of the impostor distribution toward the genuine one. To simplify this model, we further constrain the parameters of $p(s|\mathrm{F})$ to satisfy a linear proportionality constraint with respect to their range, with the same value of the coefficient. For instance, in the case of Gaussian distributions this amounts to assume that the mean and variance of $p(s|\mathrm{F})$ is given by:

$$\begin{aligned} \mu_\mathrm{F} &= \alpha\mu_\mathrm{G} + (1-\alpha)\mu_\mathrm{I}, \\ \sigma_\mathrm{F} &= \alpha\sigma_\mathrm{G} + (1-\alpha)\sigma_\mathrm{I}, \end{aligned} \tag{5.3}$$

for some α ("attack strength") $\in [0, 1]$. Note that $\alpha = 1$ and $\alpha = 0$ lead respectively to the worst and best cases of assumptions 1 and 2 of Section 5.2.1.

Through the α parameter, our model allows one to take into account in the simplest possible way all the above mentioned different factors which can affect the fake scores distribution, in the absence of more precise information on their impact. In the next section we show how to apply this model to assess the robustness of a biometric system against spoof attacks, and also how to use to predict the ranking of different score fusion rules of a multimodal biometric system, according to their robustness to spoof attacks.

5.3 Security Evaluation Method for Multimodal Biometric Systems against Spoof Attacks

Based on the models of the fake score distribution described above, we propose the procedure summarized as Algorithm 1 to evaluate the security (robustness) of a multimodal biometric system, under spoof attacks against one or more of the component matchers.

Algorithm 1 Procedure for evaluating the security of multimodal biometric systems against spoof attacks

Inputs:

- A multimodal system made up of N matchers;
- A training set $(G_{\text{tr}}, I_{\text{tr}})$ and a testing set $(G_{\text{ts}}, I_{\text{ts}})$ made up of N-dimensional matching score vectors coming from genuine and impostor users;
- $f(\mathbf{s};\theta_f) \in \{\text{G},\text{I}\}$: a score fusion rule with parameters θ_f (including the decision threshold), where $\mathbf{s}$ is an input score vector and G and I denote the labels corresponding to the 'genuine' and 'impostor' decision;
- $\hat{P}(\cdot|\theta)$: a parametric model of the class-conditional genuine and impostor score distributions (for parametric model only);
- $\alpha_1,\ldots,\alpha_n$: a set of attack strength values for the n matchers subject to a simulated spoof attack.

Output: The system's performance under a simulated spoof attack to matchers $1,\ldots,n$, with attack strength values $\alpha_1,\ldots,\alpha_n$.

1: Set the parameters of $f(\mathbf{s};\theta_f)$ (if any) with the decision threshold, on training data $(G_{\text{tr}}, I_{\text{tr}})$, according to given performance requirements. For parametric model, also fit the model $\hat{P}(\cdot|\theta)$ to testing data $(G_{\text{ts}}, I_{\text{ts}})$, to approximate the genuine and impostor score distributions $\hat{P}(\mathbf{S}|\text{G};\theta_{\text{G}})$ and $\hat{P}(\mathbf{S}|\text{I};\theta_{\text{I}})$.

2: For parametric model: Compute the fake score distribution $\hat{P}(\mathbf{S}|\text{F};\theta_{\text{F}})$ according to our model, using $\hat{P}(\mathbf{S}|\text{G};\theta_{\text{G}})$ and $\hat{P}(\mathbf{S}|\text{I};\theta_{\text{I}})$, and then randomly draw a set F_{ts} of scores from $\hat{P}(\mathbf{S}|\text{F};\theta_{\text{F}})$, and replace the scores I_{ts} of the matcher under attack with F_{ts}.
For non-parametric model: Replace the scores I_{ts} of the matcher under attack with a same number of fictitious fake scores F_{ts} generated by our model for given α values.

3: Evaluate the system's performance on the scores $(G_{\text{ts}}, F_{\text{ts}})$, using the score fusion rule $f(\mathbf{s};\theta_f)$.

First, the threshold of the score fusion rule (and its parameters, if any) has to be estimated from training data, using the genuine and impostor score distributions, G_{tr} and I_{tr}, according to application requirements (for instance, setting a desired false acceptance rate (FAR) or genuine acceptance rate (GAR) value. This defines the so-called operational point of the biometric system). To evaluate the robustness (security) of the multimodal biometric system against a spoof attack, we propose to replace the impostor scores I_{tr} corresponding to the matcher under attack with a set of fictitious fake scores F_{tr}, obtained from any of our above proposed models. This procedure can be repeated for different α values in the range $[0,1]$, to get a complete picture of the system's performance as a function of the "attack strength".

For any decision threshold value the FAR evaluated under a simulated spoof

attack is likely to be higher than the FAR evaluated in the standard way, without spoof attacks. The GAR remains unchanged instead, as spoof attacks do not affect genuine scores. Accordingly, as the value of α increases, the corresponding FAR is likely to increase from the values attained for $\alpha = 0$, corresponding to the absence of attacks, to the worst-case corresponding to $\alpha = 1$. Namely, the system's performance is likely to decrease from the value attained for $\alpha = 0$ to the $\alpha = 1$. Therefore, the above procedure allows one to assess *how* the system's performance degradation as the attack strength increases. The more gracefully the performance degrades (namely, the higher the α value for which the FAR drops below some value of interest), the more robust a system is. In particular, it can be useful to figure out the amount of the relative "shift" (the corresponding α value) of the impostor score distribution toward the genuine one, such that the system's performance (the FAR) drops below some given value.

Note that the above procedure can be carried out analitycally or numerically for some fusion rules and some parametric model of the score distributions. For instance, in the case of the LLR rule with Gaussian score distributions, the expression of the FAR as a function of the decision threshold can be obtained analitically (in integral form), and its evaluation can be done numerically, as shown later in the this section. In general, the evaluation can always be carried out empirically.

It is also worth noting that this performance evaluation procedure follows basically the standard procedure used in the biometric field. In fact, the standard procedure evaluates the performance of a multimodal biometric system using a training set $(G_{\text{tr}}, I_{\text{tr}})$ and a testing set $(G_{\text{ts}}, I_{\text{ts}})$ made up of N-dimensional match score vectors coming from genuine and impostor users. The peculiarities of performance evaluation implemented by Algorithm 1 concern the use of the attack strength parameters (the parameters α) and the fact that the impostor scores of the testing set are replaced by fictitious scores obtained from our models of the fake score distribution in order to simulate the effects of a spoof attack.

The performance prediction under attack provided by the above method (Algorithm 1), using our proposed models of fake score distribution in Sections 5.2.1 and 5.2.2, is useful, only if one can give a reasonable approximation of the distribution of fake scores that a system will incur, which in practice is very difficult. Accordingly, we extend our proposed Algorithm 1 proposing to apply it to a different, possibly more useful, aim: to predict the *relative* robustness of several score level fusion rules, namely the *ranking* of their performance under attack for a range of different simulated distributions of fake scores. The main goal of extended method is thus not to predict the FAR of different score fusion

Algorithm 2 Prediction of the ranking of score fusion rules, based on their robustness against spoof attacks.

- A multimodal system made up of N matchers;
- A training set $(G_{\text{tr}}, I_{\text{tr}})$ and a testing set $(G_{\text{ts}}, I_{\text{ts}})$ made up of N-dimensional matching score vectors coming from genuine and impostor users;
- A set of score fusion rules;
- $\hat{P}(\cdot|\theta)$: a parametric model of the class-conditional genuine and impostor score distributions (for parametric model only);
- $\alpha_1, \ldots, \alpha_n$: a set of attack strength values for the n matchers subject to a simulated spoof attack.

Output: The ranking of score fusion rules according to their predicted robustness to spoofing attacks, as a function of the parameter α.

1: Set the parameters of $f(\mathbf{s};\theta_f)$ (if any) with the decision threshold, on training data $(G_{\text{tr}}, I_{\text{tr}})$, according to given performance requirements. For parametric model, also fit the model $\hat{P}(\cdot|\theta)$ to testing data $(G_{\text{ts}}, I_{\text{ts}})$, to approximate the genuine and impostor score distributions $\hat{P}(\mathbf{S}|\text{G};\theta_{\text{G}})$ and $\hat{P}(\mathbf{S}|\text{I};\theta_{\text{I}})$.
2: **for each** α value **do**
3: For parametric model: Compute the fake score distribution $\hat{P}(\mathbf{S}|\text{F};\theta_{\text{F}})$ according to our model, using $\hat{P}(\mathbf{S}|\text{G};\theta_{\text{G}})$ and $\hat{P}(\mathbf{S}|\text{I};\theta_{\text{I}})$, and then randomly draw a set F_{ts} of scores from $\hat{P}(\mathbf{S}|\text{F};\theta_{\text{F}})$, and replace the scores I_{ts} of the matcher under attack with F_{ts}.
For non-parametric model: Replace the scores I_{ts} of the matcher under attack with a same number of fictitious fake scores F_{ts} generated by our model for given α values.
4: Evaluate the FAR of each score fusion rule on the scores $(G_{\text{ts}}, F_{\text{ts}})$.
5: Rank the score fusion rules according to their FAR.
6: **end for**

rules under a specific spoof attack, but their *ranking* with respect to a range of potential attacks. Therefore, it can provide useful information to the designer of a multimodal system about the *relative* robustness of different score fusion rules to spoof attacks characterized by a different "effectiveness" (namely by a fake score distribution more or less close to the one of genuine scores), thus allowing the designer to choose the most robust one according to the model predictions. The procedure is summarized in Algorithm 2. Using this method, the corresponding FAR of each score fusion rule can be computed for each α value, and the different rules can be ranked in terms of their predicted FAR. Finally, the ranking of score fusion rules can be analyzed as a function of α.

5.3.1 Case study: a bi-modal biometric system using LLR fusion rule with Gaussian distributions

In this section, we show how to analitically/numerically evaluate the robustness of a bi-modal biometric system against a spoof attack, when the score distributions are modelled as Gaussians, and the LLR rule is used. We assume that application requirements are given in terms of a desired FAR value.

The logarithm of the likelihood ratio for a system with two independent scores with class-conditional Gaussian distributions, denoted with $z(s_1, s_2)$, is given by:

$$z(s_1, s_2) = \log \frac{p(s_1|\mathrm{G})p(s_2|\mathrm{G})}{p(s_1|\mathrm{I})p(s_2|\mathrm{I})} = \log\left(\frac{\sigma_{\mathrm{I}_{s1}}\sigma_{\mathrm{I}_{s2}}}{\sigma_{G_{s1}}\sigma_{G_{s2}}}\right) + \frac{1}{2}\left[\frac{(s_1-\mu_{\mathrm{I}_{s1}})^2}{\sigma^2_{\mathrm{I}_{s1}}} + \frac{(s_2-\mu_{\mathrm{I}_{s2}})^2}{\sigma^2_{\mathrm{I}_{s2}}} - \frac{(s_1-\mu_{G_{s1}})^2}{\sigma^2_{G_{s1}}} - \frac{(s_2-\mu_{G_{s2}})^2}{\sigma^2_{G_{s2}}}\right]. \tag{5.4}$$

The decision function is given by $\mathrm{sign}\left(z(s_1, s_2) - \log s^*\right)$, where the value $+1$ means that the user is accepted as genuine, while a value of -1 means that he is rejected as an impostor. The decision threshold s^* has to be set so that the desired FAR is attained. The region of the score space (s_1, s_2) corresponding to genuine users, denoted as G, can be found analitically by solving the quadratic inequality $z(s_1, s_2) - \log s^* \geq 0$. The left-hand side of such inequality can be rewritten by re-arranging the terms of Eq. 5.4, as:

$$z(s_1, s_2) - \log s^* = As_1^2 + Bs_1s_2 + Cs_2^2 + Ds_1 + Es_2 + F, \tag{5.5}$$

where the threshold s^* is included in the term F. Depending on the value of $B^2 - 4AC$, the solution of $z(s_1, s_2) - \log s^* = 0$ corresponds to:

- $B^2 - 4AC < 0$: an ellipse;
- $B^2 - 4AC = 0$: a parabola;
- $B^2 - 4AC > 0$: an hyperbola.

This allows to find analitycally the region G.

The FAR for a given s^* value is defined as:

$$FAR(s^*) = \int\int_{\mathrm{G}} p(s_1|\mathrm{I})p(s_2|\mathrm{I})\mathrm{d}s_1\mathrm{d}s_2. \tag{5.6}$$

The above integral can be computed numerically.

Now the threshold s^* can be set to the value s^{**} which gives the desired FAR on training data. Assuming that s_1 corresponds to the matcher subject to a spoof attack, the corresponding FAR on testing data can be found as:

$$FAR(s^{**}) = \int\int_{\mathrm{G}} p(s_1|\mathrm{F})p(s_2|\mathrm{I})\mathrm{d}s_1\mathrm{d}s_2, \tag{5.7}$$

where $p(s_1|\mathrm{F})$ and $p(s_2|\mathrm{I})$ are now obtained from testing data as described in the previous section. The above integral can be computed numerically as well.

5.3.2 Case study: a bi-modal system using Sum, Weighted sum and Product fusion rules with Gaussian distributions

In last section, we have presented in detail how to compute the FAR of a bi-modal biometric systems under attacks using LLR fusion rule. While, in this section we show how to analytically/numerically evaluate the robustness of a bi-modal systems against spoof attacks, when score distributions are modelled as Gaussians, and the Sum/Weighted sum/Product score fusion rule is used. Without loss of generality, let s, s^{**} be fused score and the decision threshold that has to set on training data so that desired operating FAR/GAR is obtained, respectively.

Sum:

The FAR for a given s^{**} is defined as:

$$\begin{aligned} FAR(s^{**}) &= \int_{s^{**}}^{+\infty} p(s|\mathrm{I})\mathrm{d}s \\ &= 1 - \int_{-\infty}^{s^{**}} p(s|\mathrm{I})\mathrm{d}s \\ &= 1 - \left[\frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{s^{**} - (\mu_{I_{s1}} + \mu_{I_{s2}})}{\sqrt{\sigma_{I_{s1}}^2 + \sigma_{I_{s2}}^2}\sqrt{2}}\right)\right] \\ &= \frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{s^{**} - (\mu_{I_{s1}} + \mu_{I_{s2}})}{\sqrt{\sigma_{I_{s1}}^2 + \sigma_{I_{s2}}^2}\sqrt{2}}\right) \end{aligned} \tag{5.8}$$

where $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp[-t^2]\mathrm{d}t$; which is so-called error function.

Weighted sum:

The weighted sum of two statistically independent and normally distributed random variables is normal with mean equal to sum of weighted means and standard deviation equal to root of sum of squared weighted standard deviations. The FAR of weighted sum fusion score rules could be defined as:

$$
\begin{aligned}
FAR(s^{**}) &= \int_{s^{**}}^{+\infty} p(s|\mathrm{I})\mathrm{d}s \\
&= 1 - \int_{-\infty}^{s^{**}} p(s|\mathrm{I})\mathrm{d}s \\
&= 1 - \left[\frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{s^{**} - (\omega_{s1}\mu_{I_{s1}} + \omega_{s2}\mu_{I_{s2}})}{\sqrt{\omega_{s1}^2\sigma_{I_{s1}}^2 + \omega_{s2}^2\sigma_{I_{s2}}^2}\,\sqrt{2}}\right)\right] \\
&= \frac{1}{2} - \frac{1}{2}\,\mathrm{erf}\left(\frac{s^{**} - (\omega_{s2}\mu_{I_{s1}} + \omega_{s2}\mu_{I_{s2}})}{\sqrt{\omega_{s1}^2\sigma_{I_{s1}}^2 + \omega_{s2}^2\sigma_{I_{s2}}^2}\,\sqrt{2}}\right) \qquad (5.9)
\end{aligned}
$$

Any optimal weight search algorithm can be opted to attain the best values of weight ω_{s1} and ω_{s2}.

Product:

The FAR of Product fusion score rule could be defined as:

$$
\begin{aligned}
FAR(s^{**}) &= \int_{s^{**}}^{+\infty} p(s|\mathrm{I})\mathrm{d}s \\
&= \left\{\int_0^{+\infty} p(s_2|\mathrm{I}_{s2})\mathrm{d}s_2 \int_{\frac{s^{**}}{s_2}}^{+\infty} p(s_1|\mathrm{I}_{s1})\mathrm{d}s_1\right\} + \left\{\int_{-\infty}^{0} p(s_2|\mathrm{I}_{s2})\mathrm{d}s_2 \int_{-\infty}^{\frac{s^{**}}{s_2}} p(s_1|\mathrm{I}_{s1})\mathrm{d}s_1\right\} \\
&= \left\{\int_0^{+\infty} p(s_2|\mathrm{I}_{s2})\mathrm{d}s_2 \left[1 - \int_{-\infty}^{\frac{s^{**}}{s_2}} \frac{1}{\sigma_{\mathrm{I}_{s1}}\sqrt{2\pi}}\, e^{-\frac{(s_1-\mu_{\mathrm{I}_{s1}})^2}{2\sigma_{\mathrm{I}_{s1}}^2}}\,\mathrm{d}s_1\right]\right\} + \\
&\quad \left\{\int_{-\infty}^{0} p(s_2|\mathrm{I}_{s2})\mathrm{d}s_2 \left[\int_{-\infty}^{\frac{s^{**}}{s_2}} \frac{1}{\sigma_{\mathrm{I}_{s1}}\sqrt{2\pi}}\, e^{-\frac{(s_1-\mu_{\mathrm{I}_{s1}})^2}{2\sigma_{\mathrm{I}_{s1}}^2}}\,\mathrm{d}s_1\right]\right\} \\
&= \left\{\int_0^{+\infty} p(s_2|\mathrm{I}_{s2})\mathrm{d}s_2 \left[1 - \left(\frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{\frac{s^{**}}{s_2} - \mu_{\mathrm{I}_{s1}}}{\sigma_{\mathrm{I}_{s1}}\sqrt{2}}\right)\right)\right]\right\} + \\
&\quad \left\{\int_{-\infty}^{0} p(s_2|\mathrm{I}_{s2})\mathrm{d}s_2 \left[\left(\frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{\frac{s^{**}}{s_2} - \mu_{\mathrm{I}_{s1}}}{\sigma_{\mathrm{I}_{s1}}\sqrt{2}}\right)\right)\right]\right\} \\
&= \left\{\frac{1}{2}\int_0^{+\infty} p(s_2|\mathrm{I}_{s2})\mathrm{d}s_2 - \frac{1}{2}\int_0^{+\infty} p(s_2|\mathrm{I}_{s2})\,\mathrm{erf}\left(\frac{\frac{s^{**}}{s_2} - \mu_{\mathrm{I}_{s1}}}{\sigma_{\mathrm{I}_{s1}}\sqrt{2}}\right)\mathrm{d}s_2\right\} +
\end{aligned}
$$

$$
\begin{aligned}
& \left\{\frac{1}{2}\left[1-\int_0^{+\infty} p(s_2|\mathrm{I}_{s2})\mathrm{d}s_2\right]+\frac{1}{2}\int_{-\infty}^{0} p(s_2|\mathrm{I}_{s2})\,\mathrm{erf}\left(\frac{\frac{s^{**}}{s_2}-\mu_{\mathrm{I}_{s1}}}{\sigma_{\mathrm{I}_{s1}}\sqrt{2}}\right)\mathrm{d}s_2\right\} \\
= & \left\{\frac{1}{2}\left[\frac{1}{2}\pm\frac{1}{2}\,\mathrm{erf}\left(\frac{\pm\mu_{\mathrm{I}_{s1}}}{\sigma_{\mathrm{I}_{s1}}\sqrt{2}}\right)\right]-\frac{1}{2}\int_0^{+\infty} p(s_2|\mathrm{I}_{s2})\,\mathrm{erf}\left(\frac{\frac{s^{**}}{s_2}-\mu_{\mathrm{I}_{s1}}}{\sigma_{\mathrm{I}_{s1}}\sqrt{2}}\right)\mathrm{d}s_2\right\}+ \\
& \left\{\frac{1}{2}\left[\frac{1}{2}+\frac{1}{2}\,\mathrm{erf}\left(\frac{-\mu_{\mathrm{I}_{s1}}}{\sigma_{\mathrm{I}_{s1}}\sqrt{2}}\right)\right]+\frac{1}{2}\int_{-\infty}^{0} p(s_2|\mathrm{I}_{s2})\,\mathrm{erf}\left(\frac{\frac{s^{**}}{s_2}-\mu_{\mathrm{I}_{s1}}}{\sigma_{\mathrm{I}_{s1}}\sqrt{2}}\right)\mathrm{d}s_2\right\} \\
= & \left\{\left[\frac{1}{4}\pm\frac{1}{4}\,\mathrm{erf}\left(\frac{\pm\mu_{\mathrm{I}_{s1}}}{\sigma_{\mathrm{I}_{s1}}\sqrt{2}}\right)\right]-\frac{1}{2}\int_0^{+\infty} p(s_2|\mathrm{I}_{s2})\,\mathrm{erf}\left(\frac{\frac{s^{**}}{s_2}-\mu_{\mathrm{I}_{s1}}}{\sigma_{\mathrm{I}_{s1}}\sqrt{2}}\right)\mathrm{d}s_2\right\}+ \\
& \left\{\left[\frac{1}{4}+\frac{1}{4}\,\mathrm{erf}\left(\frac{-\mu_{\mathrm{I}_{s1}}}{\sigma_{\mathrm{I}_{s1}}\sqrt{2}}\right)\right]+\frac{1}{2}\int_{-\infty}^{0} p(s_2|\mathrm{I}_{s2})\,\mathrm{erf}\left(\frac{\frac{s^{**}}{s_2}-\mu_{\mathrm{I}_{s1}}}{\sigma_{\mathrm{I}_{s1}}\sqrt{2}}\right)\mathrm{d}s_2\right\} \\
= & 1-\int_0^{\infty} p(s_2|\mathrm{I}_{s2})\,\mathrm{erf}\left(\frac{\frac{s^{**}}{s_2}-\mu_{\mathrm{I}_{s1}}}{\sigma_{\mathrm{I}_{s1}}\sqrt{2}}\right)\mathrm{d}s_2
\end{aligned}
\tag{5.10}
$$

In order to compute the FAR under spoof attacks using Sum (Equation 5.8), Weighted sum ((Equation 5.9) and Product (Equation 5.10) rules, the imposotor distribution/parameters of corresponding matcher subject to spoof attack should be replaced with distribution/parameters $p(\cdot|\mathrm{F})$ obtained from testing data set as described in Section 5.3.

5.4 Evaluation of the Capability of proposed Models in Approximating Score Distributions of Real Spoof Attacks

In this section, we present a preliminary validation of our proposed models of the fake score distribution of a single matchers, using the the data sets described in Chapter 4 Section 4.2. We used silicon spoofed fingerprints and photo attack spoofed faces. The fingerprint and the face recognition systems used in the experiments were implemented using the minutiae-based Neurotechnologs VeriFinger 6.0 [63] and the elastic bunch graph matching (EBGM) [93, 86], respectively. More precisely, our aim here is to investigate whether our proposed models of the fake score distribution can reasonably approximate score distributions of real spoof attacks, for some α value.

In Figure 5.1 the histograms of genuine, impostor and fake scores computed with the above mentioned data sets are shown. Notice in particular that the histogram of scores associated to spoofed fingerprints is close to the impostor

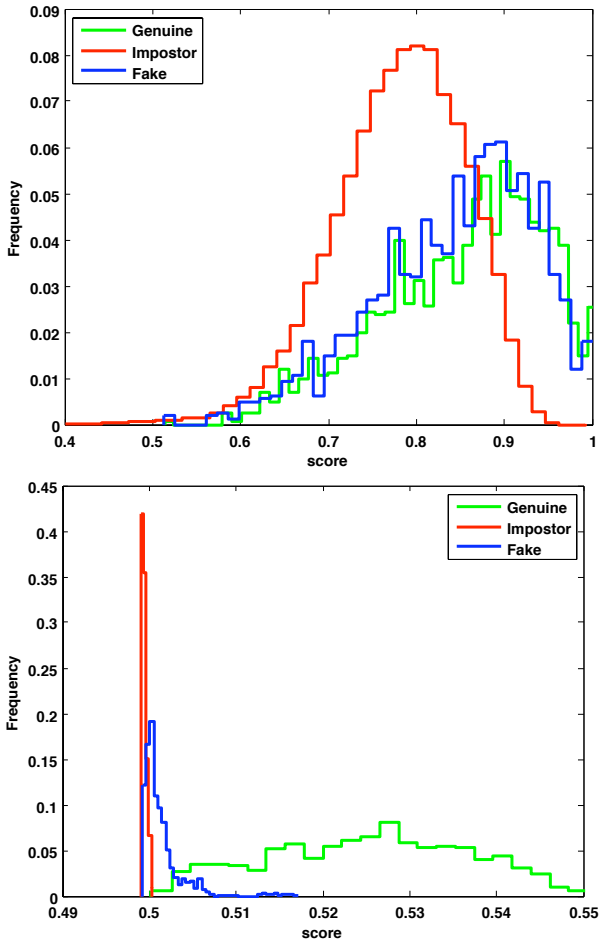

Figure 5.1: Histograms of genuine, impostor and fake scores computed with photo attack spoofed faces (top) and silicon spoofed fingerprints (bottom) data sets.

distribution, while the histogram of scores associated to faces is very close to the genuine distribution, thus these distributions exhibiting two very different degrees of "attack strength". This provides a first, qualitative support to the assumption behind our model, namely that different real fake score distributions can lie at different relative "distances" from the genuine and impostor ones.

To investigate whether the score distributions of real spoof attacks shown in Figure 5.1 can be reasonably approximated by our models, for some α value; we evaluated the dissimilarity between them and the ones provided by our models, as a function of the "attack strength" (α). Our goal is to check if a value of parameter α exists that minimizes the dissimilarity between the two distributions. The simulated fictitious fake scores were obtained as described in Algorithm 1 using respective models. To assess the dissimilarity between the two distributions, we used the L1-norm Hellinger distance [49], also called "Class Separation Statistic" [39]. The L1-norm Hellinger distance between two probability

Data set	Hellinger distance	α
Face	0.0939	0.9144
Fingerprint	0.4397	0.0522

Table 5.1: Minimum values of the Hellinger distance between the score distribution of real spoof attacks and the one obtained by (Algorithm 1) using non-parametric model, as a function of α, for the face and fingerprint data sets. The corresponding α value is also shown.

distribution functions $f(x)$ and $g(x), x \in \mathscr{X}$ can be measured as:

$$\int_{\mathscr{X}} |f(x) - g(x)| \mathrm{d}x.$$

It takes values in the range $[0,2]$, where the values of 0 and 2 correspond to identical and to fully separated distributions, respectively. Since this is a non-parametric class separation statistic, it can be used for all possible distributions.

Table 5.1 reports the values of the parameter α which minimize the dissimilarity between the score distributions of real spoof attacks shown in Figure 5.1 and the one obtained by our method (Algorithm 1) using non-parametric model (Equation 5.1). The corresponding distributions are depicted in Figure 5.2.

We can see in Figure 5.2 and Table 5.1 that our approximation is rather good for the face data set. It is less good for the fingerprint data set instead, but it could be still acceptable to the aim of evaluating the relative robustness of different score fusion rules in a multimodal biometric system, which is the final aim of our models. Another way to say this is that the designer of a biometric system in practice can not know in advance the shapes and positions of score distributions of spoof attacks that will occur. Accordingly, the robustness (security) of a multimodal system must be evaluated for several α values. The above results provide a preliminary evidence that the simulated distributions one obtained using our model, for different α values, can actually give reasonable approximations of possible score distributions of real spoof attacks.

For the sake of completeness, we also evaluated the accuracy of our model of fake score distribution in approximating the performance of the *individual* matcher under attack, for the values of Table 5.1 that give the best approximation of the score distributions of real spoof attacks, although this is not the final aim of this model as explained above.

To investigate the accuracy of the FAR approximated by our model, for all possible values of the threshold; we compared the FAR of the unimodal system attained under real spoof attacks with the FAR provided by our model. We

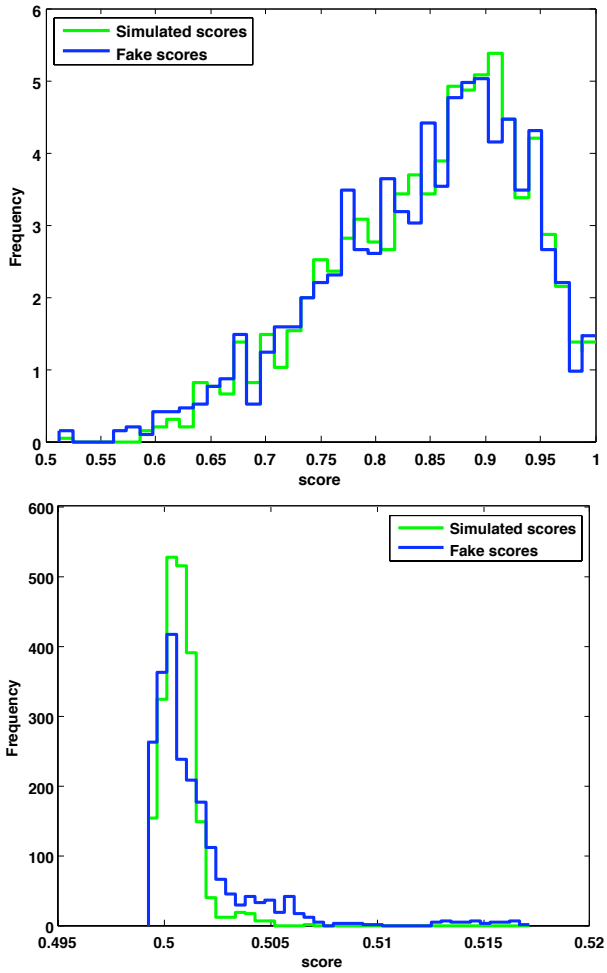

Figure 5.2: Probability distributions of the scores of fake faces (top) and of fake fingerprints (bottom) obtained from our data sets (blue), and obtained by our method for fake score simulation (green), for the α value of Table 5.1.

selected FAR as performance measure because only the False Acceptance Rate (FAR) value changes when the system is under a spoof attack, while the genuine acceptance rate (GAR) remains unchanged, since it does not depend on the match scores of the impostors, as also pointed out in Section 5.3.

Figure 5.3 depicts the FAR as a function of the threshold for the unimodal biometric system when no spoof attack is included in the data set (i.e., using only the genuine and impostor data; the "no attack" curve), under a real spoof attack against the face (fingerprint) matcher (using the fake biometric traits of our data set; the "real spoof attack" curve), and by a simulated spoof attack (using the fake scores provided by our method with the α values of Table 5.1; the "simulated attack" curve).

From Figure 5.3 (top), it is easy to see that our non-parametric model, in the case of face spoofing, provides a quite accurate approximation of the FAR;

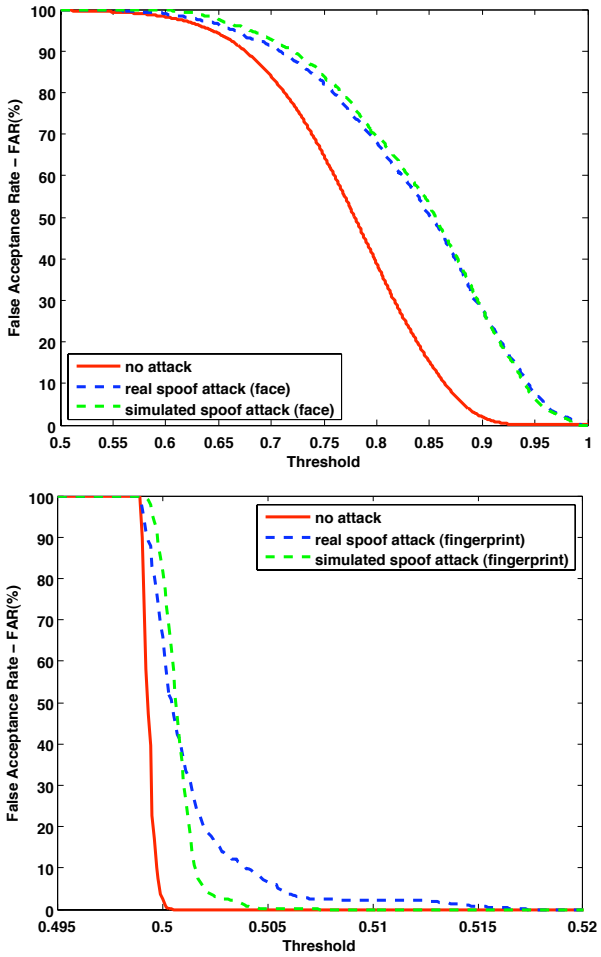

Figure 5.3: FAR of the uni-modal biometric systems as a function of the threshold applied to the score, when the data set does not contain spoof attacks ("no attack" curve), under a real spoof attack against the face (top) or fingerprint (bottom) matcher ("real spoof attack" curve), and under a spoof attack simulated with our method ("simulated attack" curve).

the maximum absolute difference between the real and the approximated FAR is 0.02%. The FAR by our model, in the case of fingerprint spoofing (Figure 5.3, bottom), is overestimated by an amount of up to 0.03% for threshold values lower than 0.502, while underestimated up to a larger amount for threshold values greater than 0.502. This is due to the positive skewness of the real fake fingerprint scores, as can bee seen in Figure 5.2.

However, the decision threshold corresponding to the zeroFAR operational point (i.e. operational point such that the threshold leads to a zero FAR value on training data, and maximizes the correspondent GAR value) is 0.500 (see Figure 5.1). Therefore, threshold values more than this one are out of the designer interest and can be neglected. This also means that threshold values where the real FAR is underestimated by our model can be neglected as well, since they

	Operational point	Real FAR	Approximated FAR (our model)	Approximated FAR (worst-case assumption)
Face System	zeroFAR	0.048	0.042	0.114
	1%FAR	0.235	0.233	0.243
Fingerprint System	zeroFAR	0.506	0.625	0.948
	1%FAR	0.600	0.808	0.951

Table 5.2: Comparison between the FAR attained at the zeroFAR and 1% FAR operational points by the unimodal biometric system under a real spoof attack ("real FAR") and the FAR approximated by our model ("approximated FAR").

are localized for threshold values higher than 0.502.

Further we accordingly focused on high security operational points, in particular the zeroFAR and 1% FAR, which are very crucial in order to assess the system security. The corresponding FAR attained by the fake score distribution in our data set ("Real FAR") and the approximated ones using our model is reported in Table 5.2. In addition, for comparison we also report the approximated FAR obtained using the "worst-case" assumption of [74, 73, 42]. The reported results show that our method, at these operational points, provides a good approximation of the performance under spoof attacks of the two considered unimodal biometric systems. The overestimation of the values for the fingerprint system is in some sense beneficial, since it puts the biometric system designer in the position to expect a performance decrease higher than that occurring in the real case. In addition, it is worth noting that our model is more flexible and appropriate for match score distributions of real spoof attacks quite far from the "worst-case" one, as happens for fingerprints.

Similar results were obtained using parametric model (Equation 5.3) with Gaussian (normal) distribution, as shown in Table 5.3 and Figure 5.4. Re-

Data set	Hellinger distance	α
Face	0.2110	1.0000
Fingerprint	0.4013	0.0356

Table 5.3: Minimum values of the Hellinger distance between the score distribution of real spoof attacks and the one obtained by (Algorithm 1) using parametric model with Gaussian distribution, as a function of α, for the face and fingerprint data sets. The corresponding α value is also shown.

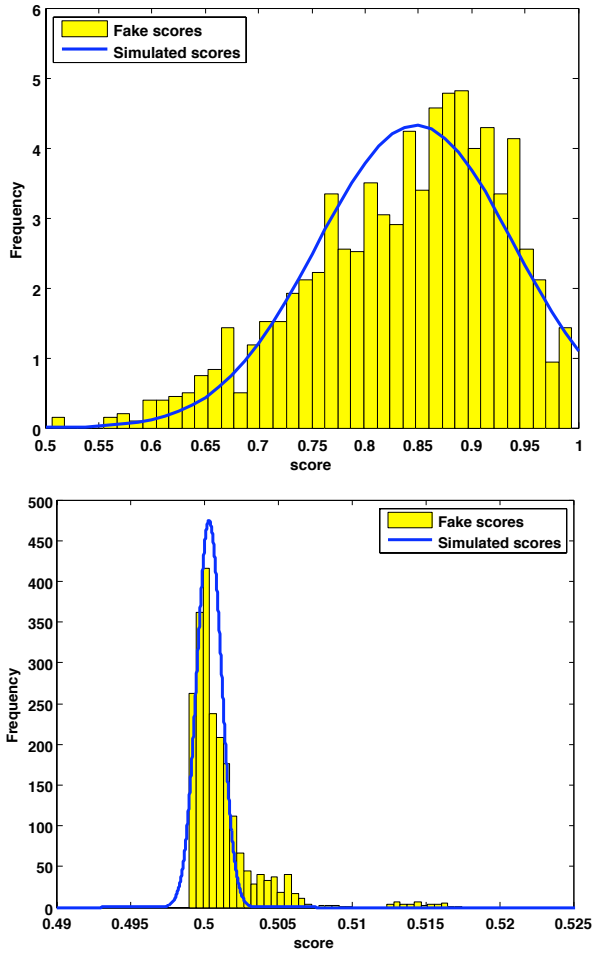

Figure 5.4: Probability distributions of the scores of fake faces (top) and of fake fingerprints (bottom) obtained from our data sets (yellow), and obtained by our method for fake score simulation (blue), for the α value of Table 5.3.

sults with parametric model using Gaussian is less approximate than with non-parametric model, since the real distributions are skewed, a better approximation could be obtained by using a Gamma or Beta distribution.

To sum up, our preliminary results provide some evidence that our model are able to reasonably approximate score distributions of real spoof attacks.

5.5 Evaluation of proposed Security Evaluation Method on Multimodal Biometric Systems

In the last Section, we showed that our proposed models of the fake score distributions are capable of approximating reasonably the score distributions produced by real spoof attacks. In this section, we examine the accuracy of our models and methods in approximating the performance of the multimodal bio-

metric systems under spoof attacks.

We will begin with investigation to check whether our proposed security evaluation method (Algorithm 1) provides a good estimate of the performance of a multimodal biometric system when any one of the biometric trait is spoofed. Then further we will analyze in detail the impact of spoof attacks on LLR fusion rule. Finally, we will carry out the empirical study using Algorithm 2 to predict the *relative* robustness of several score-level fusion rules, namely the ranking of their performance under spoof attack.

5.5.1 Performance estimation of multimodal biometric systems under spoof attack

In the previous Section, we have shown that our method using non-parametric model allows to approximate well a given score distribution of fake biometric traits and the performance of a unimodal biometric system under real spoof attacks. While in this section we evaluate whether the Algorithm 1 proposed in Section 5.3 gives a good approximation of performance of a multimodal system made up of the face and fingerprints matchers, when any one of the two traits is spoofed. We carried out the experiments using Algorithm 1 with parametric model (Equation 5.3) using Gaussian distribution. Note that when the score distribution of real spoof attack is unknown, the designer does not have any knowledge of the attack strength and, therefore, he/she should use our Algorithm 1 to evaluate performance as a function of the fake quality parameter α. But, for the data set used we also have the fake score distribution, and we know from Table 5.3, what are the α values that provide the best approximation of such distribution, and, therefore, we used such values in this experiment. It is worth noting that this does not make our results lose any generality.

In these experiments we used the LLR as the score fusion rule, and assumed that the face and fingerprint scores are conditionally independent. The interest on the LLR rule is motivated by three main reasons: it is widely used in multimodal systems; it is the optimal rule when the score distributions are exactly known (in the sense that it gives the minimum false rejection rate, FRR, for any given FAR value, and vice-versa); its robustness to spoof attacks has already been questioned in previous works [74, 73].

The performance under a spoof attack was evaluated by replacing all the impostor scores with the fake scores associated to the spoof attacks in our data sets. The performance estimated by our model was evaluated according to the Algorithm 1, namely, by replacing all the impostor scores with fictitious fake

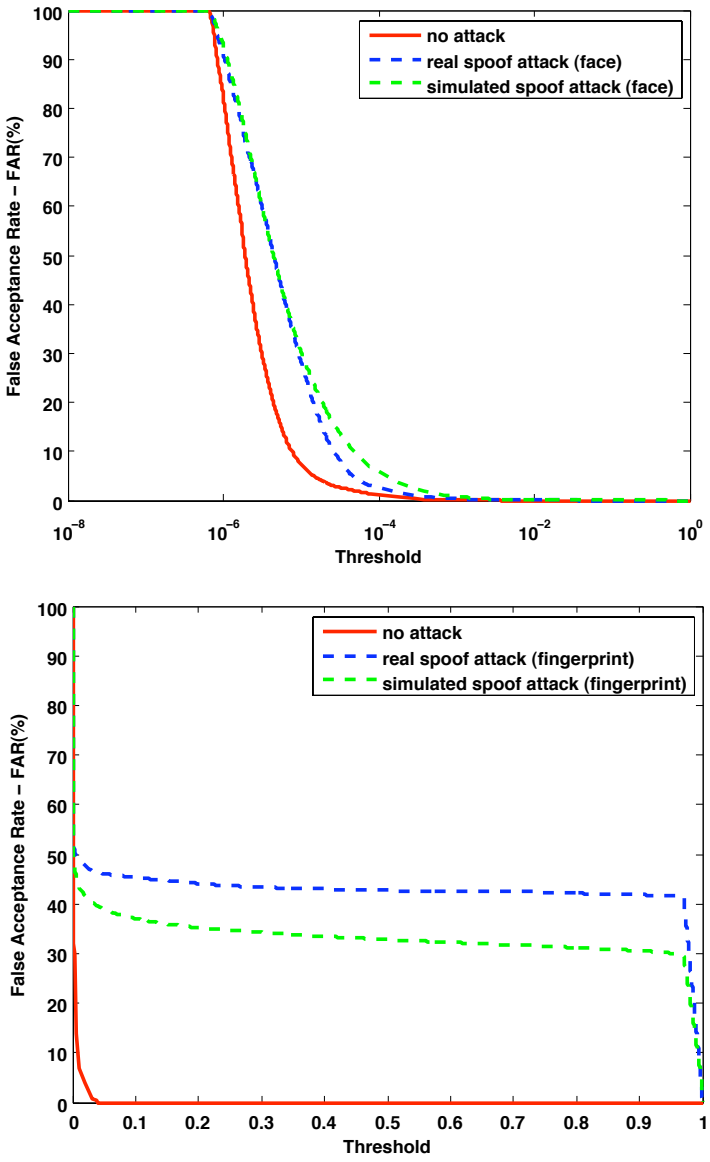

Figure 5.5: FAR of the multimodal biometric system as a function of the threshold applied to the fused score, when the data set does not contain spoof attacks ("no attack" curve), under a real spoof attack either against the face (top) of fingerprint (bottom) matcher ("real spoof attack" curve), and under a spoof attack simulated with our method ("simulated attack" curve). The LLR score fusion rule is used.

scores randomly drawn from the distributions provided by our model (the score distributions of fake samples are shown in Figure 5.4), and using the α values of Table 5.3.

Figure 5.5 shows the FAR as a function of threshold for the multimodal biometric system when no spoof attack is included in the data set (i.e., using only the genuine and impostor data; the "no attack" curve), under a real spoof attack either against the face or the fingerprint matcher (using the fake biometric

	Operational point	Real FAR	Approximated FAR
Face spoofing	zeroFAR	0.025	0.027
	1%FAR	0.051	0.089
Fingerprint spoofing	zeroFAR	0.052	0.048
	1%FAR	0.062	0.063

Table 5.4: Comparison between the FAR attained at the zeroFAR and 1%FAR operational points by the bi-modal biometric system under a real spoof attack ("real FAR") and the FAR approximated by our model ("approximated FAR").

traits of our data set; the "real spoof attack" curve), and by a simulated spoof attack (using the fake scores provided by our model with the α values of Table 5.3; the "simulated attack" curve).

It can be seen that our model provides a good approximation of the FAR vs Threshold curves under a spoof attack, for almost all Threshold values. In the case of face spoofing (Figure 5.5, top), the difference between the real and the approximated FAR never exceeds 0.05. In the case of fingerprint spoofing (Figure 5.5, bottom), our method underestimates the FAR by an amount of about 0.10, for threshold values up to about 0.95. This is due to the skewed distribution of the real fake fingerprint scores, as can be seen in Figure 5.4 (bottom), which can be better approximated by a Gamma or Beta distribution rather than a Gaussian. Note that the FAR in Figure 5.5 of the "no attack" curves is always very close to zero, except for very low threshold values (this can be better appreciated in the plot at the top of Figure 5.5, thanks to a logarithmic scale for the threshold values). The reason is that the fingerprint matcher, and thus the LLR rule, allows to discriminate very well between the genuine and impostor distributions, as can be seen from Figure 5.1.

We further evaluated the approximation provided by our model, when the threshold is set on training data in order to attain very low FAR values, which are the most relevant ones in biometric systems' security evaluation. We considered the zeroFAR and 1%FAR operational points, namely the lowest threshold values which lead to a FAR on training data equal respectively to zero and to 0.01. The results are reported in Table 5.4. It can be seen our model provides a very good approximation of the performance of the multimodal system in these operational points. Note that, in the case of fingerprint spoofing, the zeroFAR and 1%FAR operational points correspond to threshold values above 0.95, where the approximation of the FAR provided by our model is much better than the one attained for lower threshold values, as can be seen from Figure 5.5,

bottom.

The above results show that our method allows providing a good approximation of the performance of a multimodal system under spoof attacks. This is a relevant result considering that: our method makes use only of the information on the genuine and impostor score distributions, which is the only information usually available for a system designer; it is based on a single parameter attack strength (α).

5.5.2 Robustness analysis of likelihood ratio score fusion rule

In the previous Sections, we have shown that our method allows to approximate well a given score distribution of fake biometric traits and it can be used to estimate the performance of a multimodal biometric system under attack. However, in real applications, the designer does not have a training set of fake traits used by the attacker.

Therefore, in this Section, we put ourselves in a real scenario, where no information about spoof attacks is available, but only genuine and impostor distributions are known. In this case, the designer can use our method to evaluate the performance of a multimodal biometric system and to analyze for which values of the "attack strength" his system is robust enough. In particular, we carried out a case study involving the evaluation of the robustness of multimodal systems with the LLR fusion rule. We used Algorithm 1 with parametric model (Equation 5.3) to evaluate analitically/numerically the robustness of multimodal system against spoof attacks, when the score distributions are modelled as Gaussians, and the LLR rule is used, as explained in Section 5.3.1.

We used for experiments the NIST Biometric Score Set 1 (NIST BSSR1) [65], which is a multimodal data set of match scores obtained from three biometric recognition systems: two face recognition systems, named 'G' and 'C', respectively, and one fingerprint system applied to two different fingers, namely, the left index and the right index. Related groups of fingerprint match scores are called LI (left index) and RI (right index). Thus, match scores of four verification systems are contained in this data set: face system C, face system G, fingerprint system LI, and fingerprint system RI. For each individual, one genuine score and 516 impostor scores are available for each matcher and each modality, on a set of 517 users.

We considered four different multimodal systems by pairing in all possible ways the scores of the face and fingerprint matchers of the same individual. The resulting systems are therefore (Face G, Fingerprint LI), (Face G, Fingerprint

RI), (Face C, Fingerprint LI), and (Face C, Fingerprint RI). In the following they will be denoted for short with the corresponding symbols: G-LI, G-RI, C-LI and C-RI. The scores were normalized using the hyperbolic tangent method [78].

To be able to evaluate the performance degradation due only to a spoof attack, without any component due to a mismatch between the training and testing sets, in these experiments we used all the available data both as the training and as the testing set, which corresponds to the ideal situation in which the score distributions are exactly known. To this aim, we applied Algorithm 1 with $G_{ts} = G_{tr}$, and $I_{ts} = I_{tr}$. In Algorithm 1 we used parametric model with Gaussian distribution to model the genuine and impostor score distributions, and computed the FAR values under a simulated spoof attack as described in Section 5.3.1.

Three different operational points were considered: 0.01%, 0.1% and 1% FAR. The values we used for the "attack strength" α range form 0 to 0.1 with steps of 0.01, plus the values from 0.1 to 1 with steps of 0.1. Note that in this setting the simulated spoof attacks affect only the FAR, while the FRR remains unchanged. Accordingly, the robustness of the considered multimodal systems can be evaluated in terms of the behaviour of their FAR as a function of the "attack strength".

Results are reported in Figures 5.6–5.9. These figures show the FAR attained by the four multimodal systems, for each operational point, as a function of the "attack strength" (namely, "fake strength", α), under a simulated spoof attack. They can be considered a "prediction of the impact, in terms of FAR, of attacks using fake biometric traits of different quality. It is worth noting that such a prediction of the impact of a spoof attack can not be obtained with state-of-the-art performance evaluation methods. Note that the FAR attained for $\alpha = 0$ is the one corresponding to the absence of attacks, while the one for $\alpha = 1$ corresponds to the "worst-case" considered in [74, 73, 42]. In each plot, we report the results attained under a simulated spoof attack either on the fingerprint or the face matcher. Note in particular that using our method it is not possible to compare the FAR attained under spoof attacks against *different* biometrics, being equal the α value, as there is no relationship among the α values related to fake score distributions of different biometrics.

In all the considered systems, the FAR under spoof attack increases as the "attack strength" increases, namely as the simulated score distribution of spoof attack approaches to genuine score distribution.

In particular, we can see, in all plots, that FAR increases very quickly as

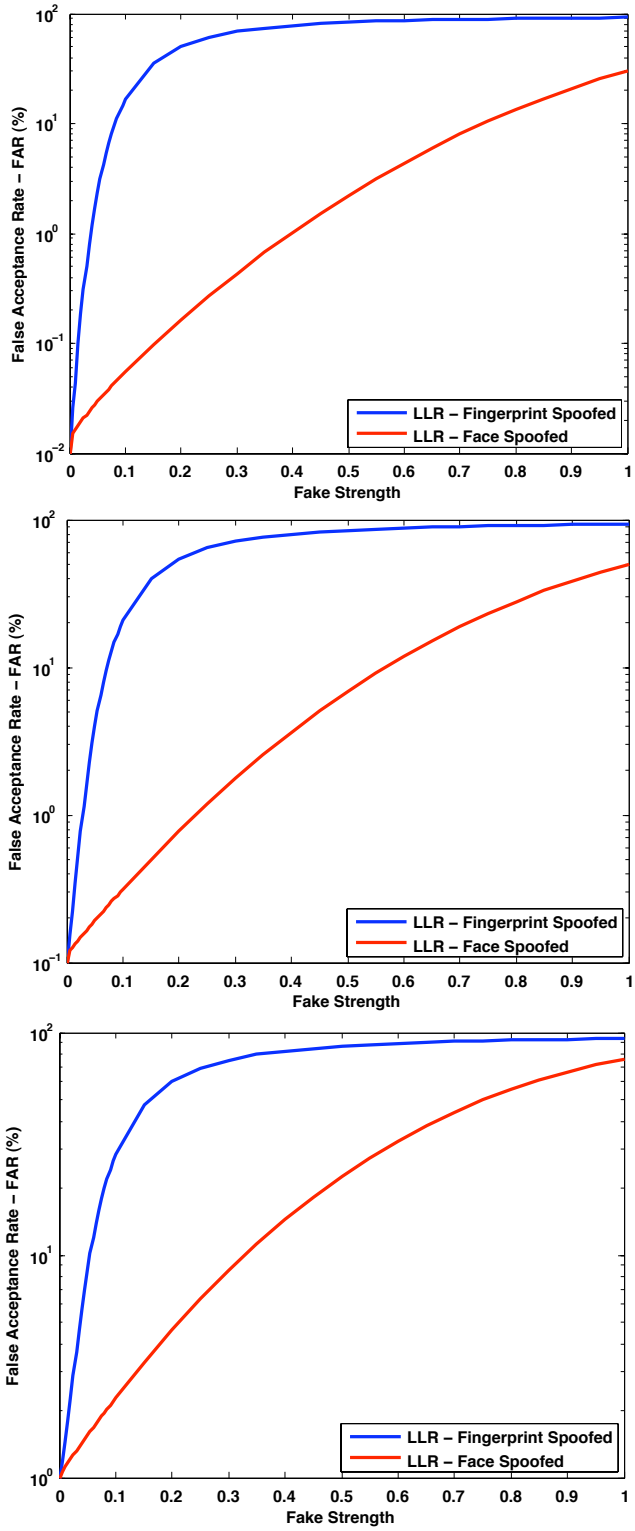

Figure 5.6: FAR (%) of the G–RI system at 0.01 % (top) , 0.1 % (middle) and 1 % FAR (bottom), as function of the fake strength α, when either the fingerprint (blue curve) or the face (red curve) is spoofed.

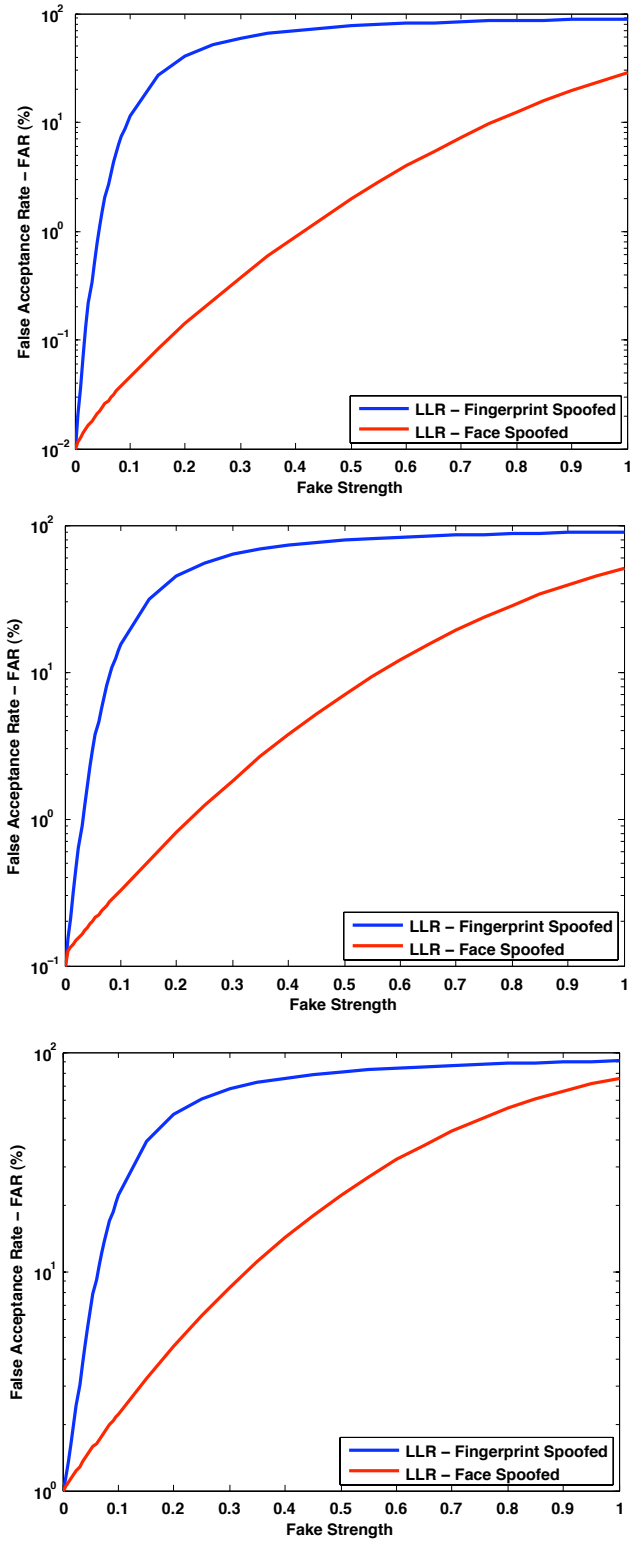

Figure 5.7: FAR (%) of the G–LI system at 0.01 % (top) , 0.1 % (middle) and 1 % FAR (bottom), as function of the fake strength, when either the fingerprint (blue curve) or the face (red curve) is spoofed.

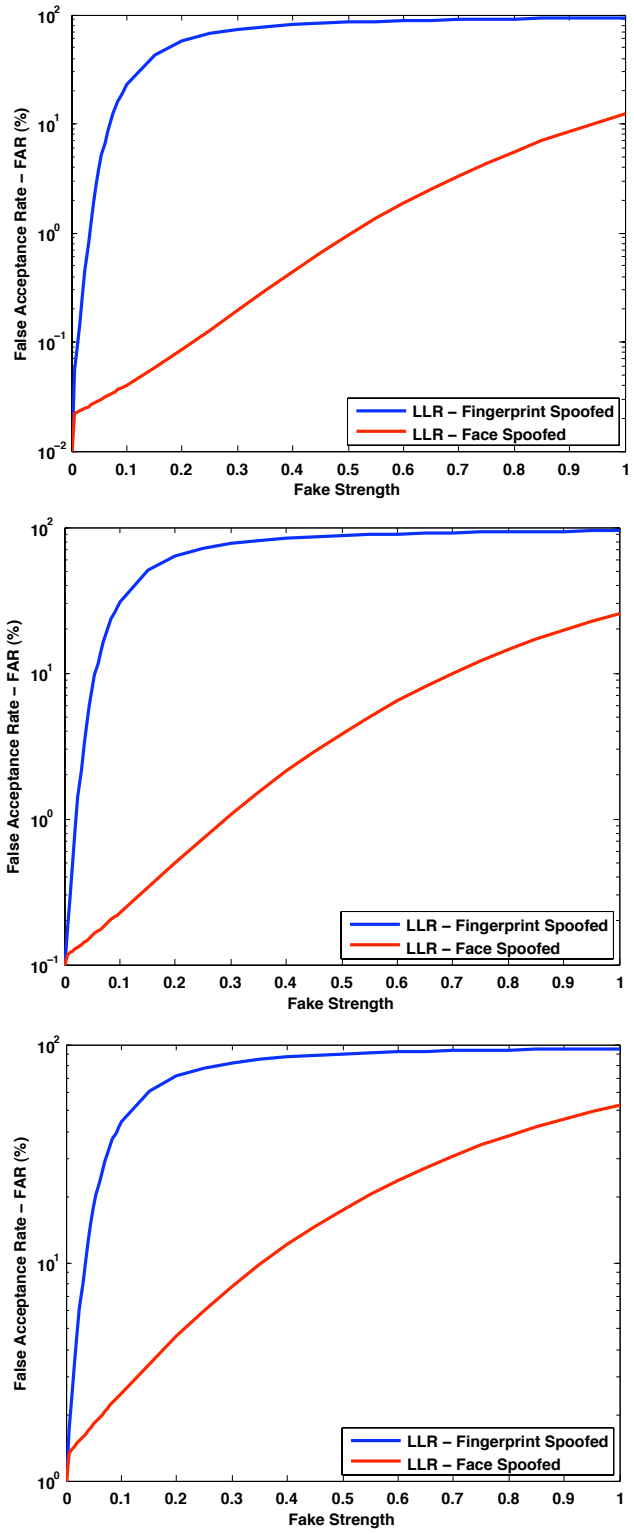

Figure 5.8: FAR (%) of the C–RI system at 0.01 % (top) , 0.1 % (middle) and 1 % FAR (bottom), as function of the fake strength, when either the fingerprint (blue curve) or the face (red curve) is spoofed.

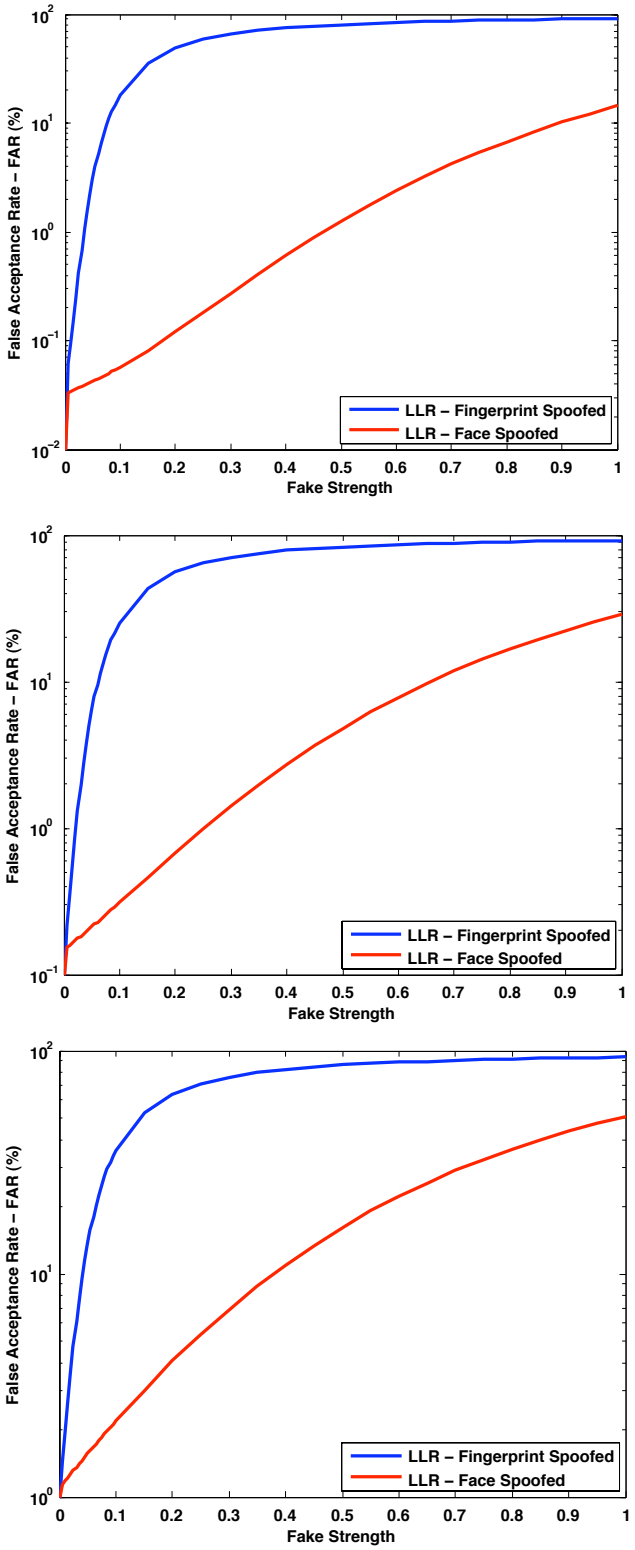

Figure 5.9: FAR (%) of the C–LI system at 0.01 % (top) , 0.1 % (middle) and 1 % FAR (bottom), as function of the fake strength, when either the fingerprint (blue curve) or the face (red curve) is spoofed.

a function of the "attack strength", when only fingerprints are spoofed, to the extent that the FAR becomes unacceptably high even for low α values. For instance, in the G–RI system with 1% FAR operational point (Figure 5.6, bottom), the FAR under attack exceeds 50% as the fake strength is above 0.15. This means that 1% of the impostors are erroneously recognised as genuines, when they provide their real fingerprint and face. Instead, when impostors provide a spoofed fingerprint of a genuine user together with their real face, 50% of them would be recognised as genuine users, as long as the mean and variance of the score distribution of the spoofed fingerprints is shifted from the one of the real impostors' fingerprints of just 15% towards the corresponding parameters of the genuine score distribution. This means, for the attacker, that improving the quality of the fabrication of fake fingers is very worth, as a slight improvement of such quality provides a substantial increase of the probability of cracking the multimodal biometric system.

On the other hand, face spoofing causes instead a relatively more graceful increase of FAR as a function of the "attack strength", in all the considered systems. Nevertheless, it always leads to FAR values exceeding 10%, for a sufficently high fake strength. In our experiments, this is due to the fact that the genuine and impostor score distributions of the face matchers in the considered data sets turn out to be more overlapping than the ones produced by the fingerprint matchers. Consequently, for a same value of the fake strength α, when the face is spoofed, the fingerprint matcher allows to detect a higher fraction of impostors than vice versa.

All in all, we can conclude that face spoofing is not the best attack strategy against the considered multimodal biometric systems. The effort of the attacker for improving the quality of the fake faces does not provide a substantial increase of FAR, namely, a substantial increase of the probability of cracking the system.

Thus, from the point of view of "attacker", our results show that the knowledge of which is the most accurate biometric of a multimodal system is a strategic asset. Attacking the most accurate biometric is a very effective and opportunist strategy. A slight improvement of the quality of the fake traits used can provide a substantial increase of the probability of cracking the biometric system, therefore, the effort of the adversary is compensated very well. On the other hand, if the attacker does not have this knowledge, she could be obliged to spoof all the biometrics. In this case, a multimodal system can be a valid deterrent.

Reported results in particular showed that the LLR rule may be very vul-

nerable to spoof attacks; despite its theoretical optimality when the genuine and impostor score distributions are exactly known and there is no attack. The vulnerability increases as the targeted matcher provides less overlapping score distributions.

5.5.3 Ranking of score fusion rules under spoof attacks

In Sections 5.5.1–5.5.2, we have shown validation of our proposed models of the match score distribution produced by spoof attacks on real spoof attacks, on uni and multimodal systems' performance estimation under spoof attacks using our data sets consisting of real spoof attacks, and consequently thus applied our method to evaluate the robustness of LLR rule in real scenario where only genuine and impostor distributions are known, but no information about spoof attacks is available. However, the performance prediction under attack provided by the above models is useful, only if one can give a reasonable approximation of the distribution of fake scores that a system will incur, which in practice is very difficult.

Accordingly, based on the above results, we further proposed to apply our models to a different, possibly more useful aim: not to predict the FAR of different score fusion rules under a specific spoof attack, but to predict their *ranking* with respect to a range of potential attacks, namely different attack strength (α) values in the range $[0, 1]$, that lead to different (simulated) potential distributions of fake scores. This can give the designer of a multimodal system useful information about the relative robustness of different score fusion rules to spoof attacks characterised by a different "effectiveness", namely by a fake score distribution more or less close to the one of genuine scores. The procedure is summarised in Algorithm 2.

We used Algorithm 2 with non-parametric model (Equation 5.1) to assess whether our method can accurately predict the ranking of score fusion rules in terms of their FAR, when the α value that best fits a real fake score distribution is known, and how a choice between different rules can be made, based on their predicted robustness across the whole range of attack strength (α) values.

In following experiments we utilized the same data sets used in experiments of Sections 5.4–5.5.1: silicon spoofed fingerprints and photo attack spoofed faces data sets. For the investigation we used three fixed score fusion rules (sum, product and Bayesian) and five trained ones (weighted sum, weighted product, perceptron, likelihood ratio (LLR), and the extended likelihood ratio (ExtLLR) of [74]). The Sum, Product, Weighted sum, LLR, and Extended LLR

(ExtLLR) rules were described in Chapter 4 Section 4.3.1. We describe here the other rules used in our experiments.

Bayesian [81]**.** The fused score produced is

$$f(s_1, s_2) = \frac{s_1 \times s_2}{(1-s_1)(1-s_2)+(s_1 \times s_2)};$$

where s_1 and s_2 denote scores provided respectively by face and fingerprint matchers, and $s = f(s_1, s_2)$ is the score fusion rule.

Weighted product [46]**.** The fused score is obtained as

$$f(s_1, s_2) = s_1^w \times s_2^{1-w}.$$

The the value of weights in Weighted sum and Weighted product were set by maximising the system performance on the chosen operational point, namely, by minimising the false rejection rate (FRR) on the available data, given the chosen false acceptance rate (FAR).

Perceptron [54, 53]**.** The score produced by the Perceptron-based fusion rule is given by

$$f(s_1, s_2) = \frac{1}{1+exp[-(s_0+w_1 s_1+w_2 s_2)]};$$

where weights w_0, w_1, and w_2 were computed by maximizing the Fisher distance (Equation 4.4) between the score distributions of genuine and impostor users.

Since no parameter tuning is required by our method, we used all the available data both as training and testing sets.

We first report, in Table 5.5, the *ranking* of the eight score fusion rules considered, according to the FAR attained on testing data under real spoof attacks in our data set, at two high security operational points: zeroFAR and 1% FAR chosen on training data (in absence of spoof attacks).

In order to evaluate the accuracy of our method with non-parametric model in approximating the FAR of multimodal system under attack, as a function of α, we report in Table 5.6 the FAR attained by the LLR rule for α values $0.1, 0.2, \ldots, 1.0$, and the ones of Table 5.1 (which minimize the dissimilarity between the score distributions of real spoof attacks shown in Figure 5.1 and the one obtained by our method). We can see that our model provides a good approximation of the real FAR under attack, provided that the optimal α value is used, in the case of face spoofing. The approximation is not as accurate in case of fingerprint spoofing: the real FAR is overestimated (when the optimal

Face Spoofing			
zeroFAR		1% FAR	
FAR(%)	Rules	FAR(%)	Rules
0.04	ExtLLR	2.26	ExtLLR
0.05	LLR	2.29	LLR
0.27	W. Product	10.72	W. Product
0.48	W. Sum	18.37	W. Sum
1.30	Perceptron	20.95	Perceptron
6.75	Bayesian	23.47	Bayesian
6.80	Sum	23.49	Sum
6.82	Product	23.57	Product

Fingerprint Spoofing			
zeroFAR		1% FAR	
FAR(%)	Rules	FAR(%)	Rules
0.00	Bayesian	1.05	Bayesian
0.00	Sum	1.15	Sum
0.00	Product	1.33	Product
24.56	W. Sum	42.59	W. Sum
27.73	Perceptron	44.11	Perceptron
34.87	W. Product	51.10	W. Product
50.42	ExtLLR	60.31	ExtLLR
50.43	LLR	60.32	LLR

Table 5.5: Ranking of fusion rules according to their FAR under real spoof attacks, when either the face (top) or the fingerprint is spoofed (bottom), at two operational points.

α value is used) by an amount of about 16-18%. The FAR predicted by the worst-case assumption of [74, 73, 42] (corresponding to $\alpha = 1$) is as accurate as the one provided by the considered model, for face spoofing, but is much more inaccurate for fingerprint spoofing, where it overestimates the FAR of about 40 to 50%. This is an evidence that our model is more appropriate than the one based on the worst-case assumption. Qualitatively similar results were obtained with the other fusion rules.

It is worth noting that in our data sets fingerprint spoofing leads to a higher increase of FAR than face spoofing as in experiments, Section 5.5.2, using NIST BSSR1 biometric benchmark data sets. The is due to the same fact explained there: genuine and impostor score distributions of the face matcher turned out to be more overlapping than the ones produced by the fingerprint matcher.

We can see that FAR increases, unfortunately, very quickly as a function of α (attack strength) in the case of fingerprint spoofing, up to the extent that the performance drops considerably even for low α values. This suggests, as also

Face Spoofing		
$\alpha = 0$	zeroFAR	1% FAR
0.1	0.00	1.09
0.2	0.00	1.15
0.3	0.01	1.26
0.4	0.01	1.35
0.5	0.01	1.45
0.6	0.01	1.59
0.7	0.01	1.76
0.8	0.02	2.01
0.9	0.04	2.24
0.9144	**0.04**	**2.33**
1	0.06	2.68
Realistic attack	0.05	2.29

Fingerprint Spoofing		
$\alpha = 0$	zeroFAR	1% FAR
0.0522	**66.04**	**78.04**
0.1	86.94	91.28
0.2	95.06	97.28
0.3	97.90	99.08
0.4	99.12	99.43
0.5	99.34	99.70
0.6	99.63	99.79
0.7	99.73	99.84
0.8	99.75	99.86
0.9	99.76	99.87
1	99.85	99.89
Realistic attack	50.43	60.32

Table 5.6: FAR (%) attained by the multimodal system under a simulated spoof attack against the face (top) and the fingerprint matcher (bottom), as a function of α, using LLR rule, at two operational points. The FAR under the α value that best fitted the real fake score distributions (see Table 5.1) is shown in boldface.

explained in Section 5.5.2, that the LLR rule, which in principle is the optimal fusion rule under normal operational conditions, can be very vulnerable to spoof attacks. Similar results were observed for the other fusion rules.

As, we have seen that the our method can give a better prediction of the FAR than the one based on the worst-case assumption, but this prediction can nevertheless be quite inaccurate. Thus, to assess how well our method (Algorithm 2) predicts the ranking of score fusion rules, we first compared the real *ranking* of Table 5.5, and the one predicted by the Algorithm 2 with non-parametric

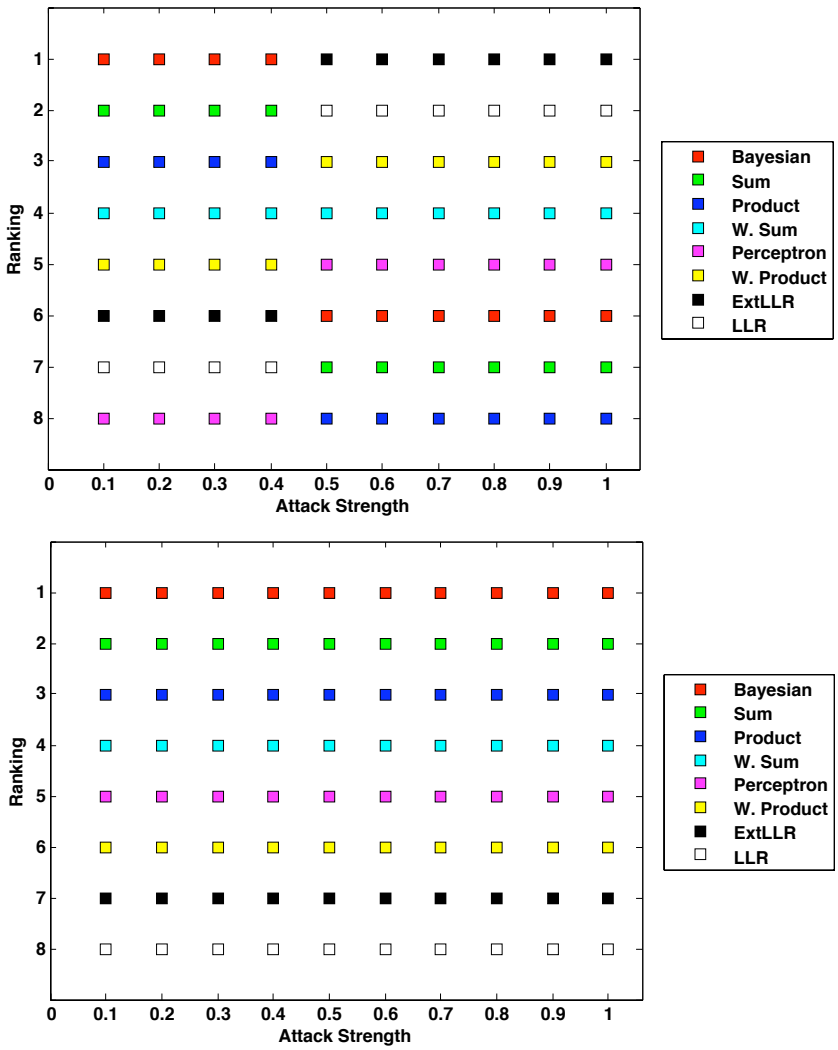

Figure 5.10: Ranking of the considered fusion rules as a function of attack strength α, when only face (top) or fingerprint (bottom) is spoofed, at the zeroFAR and 1% FAR (the ranking was identical for both operational points).

model for the optimal α value of Table 5.1. We found that our method always predicted the correct *ranking* corresponding to the optimal α value. This is an evidence that our method is capable to provide a reliable *ranking* prediction for any given α value, if the fake score distribution is best approximated by our model using such value.

Since in practice a multimodal system can be subject to different attacks, that are best approximated by different α values, and that such attack strength (α) values are unknown. Subsequently, we investigated whether the ranking of score fusion rules predicted by our method (Algorithm 2) changes smoothly, for smooth changes of the fake score distribution, namely, of the parameter α. The *ranking* predicted for the eight considered score fusion rules as a function

of attack strength α is reported in Figure 5.10 for the zeroFAR and 1% FAR operational points (the ranking was identical for both operational points). These results suggests that our method can nevertheless be used to choose the score fusion rule that exhibits the highest "average" robustness across the whole range of attack α values. To better understand this point, we can observe that under fingerprint spoofing the predicted ranking of each rule remains constant, and identical to the one under the real spoof attack (see Table 5.5, bottom). Note in Figure 5.10 (bottom) that Bayesian rule always exhibits the best ranking. This suggests that the Bayesian rule should be a good choice in terms of robustness, even if the designer does not know the optimal attack strength (α) value for any given spoof attack.

For face spoofing, two different rankings are predicted instead: one for $\alpha < 0.5$, and the other $\alpha \geq 0.5$. The latter corresponds to the one observed under a real spoof attacks (Table 5.5, top). We can see that Bayesian and ExtLLR rules are the top-ranking ones in the two intervals, respectively. However, except for the weighted sum and weighted product rules, that exhibit a constant or almost constant, and rather high ranking in both intervals, the ranking of the other rules drastically changes. This suggests that the weighted sum or the weighted product rule is a reasonable choice to avoid the risk of a very low performance under attack, unless the fake score distribution of possible face spoofing attacks is believed to be either close to the impostor one (namely, it is best approximated by a low α value), in which case the Bayesian rule is the best choice, or it is believed to be close to the genuine user distribution, and in this case the ExtLLR rule is the best choice.

All in all, we can conclude that even if the designer does not know the "optimal" attack strength (α) value for a given attack, and thus can not obtain an accurate prediction of the corresponding FAR, he could nevertheless obtain a reliable prediction of the ranking of different fusion rules. In other words, using our method the ranking of different score fusion rules under a range of spoof attacks can be predicted more reliably than their exact performance. This is a potentially useful information for the choice of a fusion rule based also on its robustness against spoof attacks.

5.6 Summary

Evaluating the robustness of multimodal biometric systems under spoof attacks is a crucial issue, do to the fact that replicating biometrics is a real menace. The state-of-the-art solves this problem by simulating the effect of a spoof attacks

in terms of fake score distribution modelling, for each individual matcher. In particular, the fake score distribution is assumed to be coincident to the genuine users one, thus drawing a "worst-case" scenario.

However, a more realistic modelling should take into account a larger set of cases. Unfortunately, the approach of fabricating fake biometric traits to evaluate the performance of a biometric system under different real spoof attacks is impractical. Thus, in this chapter we proposed a method for evaluating empirically or analitically/numerically the robustness of multimodal systems against spoof attacks, based on *simulating* the corresponding score distribution, and then consequently we proposed extension of our method to rank the different biometric score fusion rules according to their *relative* robustness against spoof attacks. Indeed, we proposed two models (non-parametric and parametric) of the score distributions produced by the spoof attacks, that accounts for different possible degrees of the quality of the fake traits, which in real scenarios can be due to different forgery skills of the attackers etc., thus resulting in different degrees of similarity between the genuine and the fake score distribution. Such factors are summarized in our models in a single parameter associated to the degree of similarity of the fake score distribution to the genuine one, which is named accordingly "attack strength" (α). A designer may use this method to generate several fake distributions for different α values, to analyze the robustness of the multimodal system under design.

We reported experimental results to provide some evidence that our models are capable to give reasonable approximations of score distributions produced by real spoof attacks, and also to be a good alternative to the model based on the "worst-case" scenario adopted so far. We then also presented an extensive experimental analysis involving unimodal and multimodal biometric systems based on data set of faces and fingerprints with real spoof attacks, to show how our proposed methodology can be used reliably to assess the systems' robustness against attacks. In particular, we applied our robustness evaluation method to a case study involving multimodal systems made up by a face and a fingerprint matcher, whose scores are fused using the well known LLR rule. Eventually, we present an experimental analysis, using our proposed method of ranking biometric score fusion rules in terms of their robustness against spoof attacks, to provide further evidence that our method is also capable of predicting dependably the *raking* of fusion rules. Although our work was mainly focused on performance analysis, and no countermeasures against spoof attacks were explicitly proposed, we believe that our findings may not only help system designers and researchers to evaluate the current impact of spoof attacks, but also

to devise more reliable robust fusion rules and multimodal biometric systems.

To sum up, from the point of view of the "attacker", our results show that the knowledge of which is the most accurate biometric of a multimodal system is a strategic asset. Attacking the most accurate biometric is a very effective and opportunist strategy. A slight improvement of the quality of the fake traits used can provide a substantial increase of the probability of cracking the biometric system, therefore, the effort of the adversary is compensated very well. On the other hand, if the attacker does not have this knowledge, she could be obliged to spoof all the biometrics. In this case, a multimodal system can be a valid deterrent. While from the point of view of the "designer" of a biometric system, our empirical results should make him aware that multimodal systems can be highly vulnerable when *only* a subset of biometric traits is spoofed, and therefore they can not be considered *intrinsically* robust to spoof attacks as commonly believed so far. In particular, the degree of vulnerability turns out to depend on factors such as the accuracy of the biometric trait subject to a spoof attack, score fusion rule and the quality of the fakes fabricated by the attacker; the success of the spoof attack can strongly depend on such quality. In addition, our results show that the designer should consider very carefully the choice of the score fusion rule, being aware that fusion rules which are optimal when the system is not under attack could be very vulnerable to spoof attacks, even if such attacks are carried out by non-professional attackers who use fake biometrics of low quality, namely low "attack strength" (α). This highlights the importance of tools for performance evaluation like the ones proposed in this chapter as well as developing defence strategies, for multimodal systems, against spoof attacks like the one is [74], i.e., devising *ad-hoc* score fusion rules. All work presented in this chapter can be found in detail in [4, 3, 5].

Chapter 6

Conclusions and Future Research

6.1 Conclusions

Biometric identity recognition is a pattern recognition application with a potential adversarial nature. Such adversarial nature comes from the fact that a malicious user (referred to as "impostor" in the biometric literature) may try to be recognized by a biometric system as a genuine one by purposely carrying out actions ("attacks") aimed at undermining the expected working of the system, possibly by exploiting some system's vulnerability. Security against attacks of malicious users (so-called "impostor") is a major problem for the widespread use of biometric systems for identity management. While biometric recognition systems are increasingly being applied to many real-world security applications, the research efforts on this topic are still at a very early stage. They need a systematization and a uniform treatment, toward the development of a clear theoretical framework of practical design methods for biometric recognition systems in adversarial environments. This thesis provides some contributions towards this direction.

Among the potential attacks discussed in the literature, the one with the greatest practical relevance consists in submitting a fake biometric trait to a biometric system. This attack is known as "spoof attack" (also named "direct attack" since it is carried out directly on the biometric sensor). Spoof attacks have a great practical relevance because they don't require advanced technical skills and, therefore, the potential number of attackers is very large. Although several potential counter measures have been proposed so far, like fingerprint "liveness" detection, no effective solution exists yet.

Besides, *ad hoc* countermeasures, multimodal biometric systems are also considered as a defense technique against spoof attacks. Multimodal systems

have been originally proposed to overcome the weaknesses of using any individual biometric trait, and to consequently improve the identity recognition performance. It is commonly believed that multimodal systems are more robust against spoof attacks than systems using a single biometric. The claimed superior robustness of multimodal systems against spoof attacks is based on the intuitive argument that their evasion would require to spoof *all* biometric traits *simultaneously*.

However, the robustness of multimodal biometric systems has been questioned very recently. It has been shown that, in some application scenarios, multimodal systems can be cracked by spoofing *only one* of the biometric traits. In Chapter 3, we made an in-depth analysis of several works in multimodal biometric system, according to their security in adversarial environments, namely, (1) performance evaluation, with particular focus on the concept of robustness under spoof attack, and (2) design of robust multimodal systems. This state-of-the-art works in multimodal systems under spoof attacks highlighted the need for a more thoroughly, systematic and unifying analysis and development of mutlimodal biometric systems under spoof attacks.

The scopes of state-of-the-art results of multimodal systems against spoof attacks are very limited, since they were obtained under the unrealistic hypothesis known as "worst-case" scenario, where the attacker is able to fabricate a perfect replica of a biometric trait whose matching score distribution is identical to the one of genuine traits. Therefore, we argued and investigated in Chapter 4 that the "worst-case" scenario may not hold in realistic case. We provided empirical evidence that a "worst-case" scenario can not be representative of real spoof attacks: its suitability may depend on the specific biometric trait, the matching algorithm, and the techniques used to fabricate the spoofed traits. In particular, we found that the "worst-case" assumption can be too pessimistic, resulting in a significant overestimation of performance drop that a multimodal system may incur under a real spoof attack. This can also undermine the effectiveness of robust score fusion rules based on such assumption. In addition, our analysis on real spoof attacks provided evidence of two common beliefs about the robustness of multimodal biometric systems. First, they can be more robust than each corresponding unimodal system, even in the case when *all* biometric traits are spoofed. Second, they can be cracked by spoofing *all* the fused traits, even when the attacker is not able to fabricate an exact replica of the genuine user's traits.

In Chapter 5, we then proposed a method to evaluate the security/robustness of a multimodal biometric system against spoof attacks. Since the straightfor-

ward approach fabricating fake biometric traits to test the security of a biometric system is impractical for the system designer, we proposed a method for the security evaluation that does not require to fabricate fake traits. In particular, we proposed two models of the match score distribution of fake biometric traits, that accounts for different possible realistic scenarios characterized by factors like different spoofing techniques and attackers' capability etc. Such factors are summarized in our models in a single parameter associated to the degree of similarity of the fake score distribution to the genuine one, which is named accordingly "attack strength". The proposed models exploit only information on genuine and impostor samples which is collected for the training of a biometric system. The main feature of our method is that it allows analyzing the performance of a multimodal system against several spoof attack distributions for different "attack strength" values, namely non-worst case scenarios. Our models allow to develop a method to empirically or analitically/numerically evaluate the security of biometric systems against attacks, by simulating their effect on the match scores. The proposed method can be applied to any multimodal system, namely, to any set of matchers combined with any score fusion rule, and it allows to simulate a spoof attack against any subset of the component matchers. Furthermore, we proposed extension of security evaluation method aimed at *ranking* several score-level fusion rules under attack, to allow the designer to choose the most robust one according to the method predictions.

We then provided empirical validation of our models of the fake score distribution on data sets of face and fingerprint images, including images of fake traits exhibiting a very different quality. Our models provided a good approximation of the real fake score distribution, and of the performance of unimodal and multimodal systems under a real spoof attack. We also carried out a set of experiments, on large and publicly data sets without spoof attack samples, aimed at showing some concrete examples of application of our method to assess the performance of multimodal system under spoof attacks. Eventually, we provided empirical evidence using data set containing real spoof attacks that our method can rank correctly score-level fusion rules under spoof attacks.

To sum up, in this thesis we highlighted the main problems related to security of multimodal biometric systems against spoof attacks, and proposed some possible techniques which can be exploited to evaluate robustness of biometric systems under design and also to select the most robust score fusion rule under attacks. In principle, the proposed techniques are completely general, that is to say, they are not tailored neither to a specific biometric trait nor kind of biometric system, although their specific implementation has to take into account,

clearly, particular application constraints and requirements. We argue that our contributions are a first step toward the systematisation of the problem of security of multimodal systems against spoof attacks, which is however far from being completely investigated yet. In other words, research in the security of multimodal biometric systems against attacks is still at an early stage.

6.2 Future Research Directions

A number of theoretical and empirical research directions arise from the work carried out in this Thesis. The following ones are of special interest:

1. We believe that the techniques proposed in this thesis can be further expanded and refined by constructing proper data sets containing spoof attacks, to analyze the behavior of the real distribution of fake scores under different conditions (different biometric traits, spoofing techniques, matchers, etc.). This would allow one to check whether the assumptions underlying our models provide good approximations of the fake score distributions, and to modify them if necessary. As a consequence, this would allow further to improve the effectiveness of the proposed method for security evaluation a practical tool for the designers of biometric systems, without requiring the actual implementation of spoof attacks.

2. Since the proposed method allows to point out the vulnerabilities of multimodal biometric score fusion rules to spoof attacks, it could be exploited to develop proper countermeasure to improve their robustness.

3. Our results also highlighted the need of methods, like the one proposed in this work, for evaluating and comparing the performance of different multimodal biometric systems against spoof attacks. Such methods will also play a crucial role in the development of new score fusion rules that aim to be robust to spoof attacks, as they will allow to evaluate their actual robustness.

4. The traditional design process of multimodal biometric systems consists in the following steps: collecting a set of labelled samples, choosing a set of features from respective biometric traits(possibly selecting or extracting a feature subset), choosing a fusion rule, training the chosen rule using a given learning algorithm, and estimating its performance on testing samples. In general, all these steps should be revisited to take into account the presence of spoof attacks, as we did in this thesis for performance

evaluation. For instance, feature selection should not be carried out only looking for the highest generalisation capability, but features should be selected also on the basis of their vulnerability to attacks. Indeed, it may be more difficult to modify the features to evade the system, if more robust features are selected. Analogously, feature extraction algorithms have to be designed from the scratch to be as secure as possible. As, the features quickly loose their discriminant capability due to spoof attacks. To counteract, the system should be constantly updated by the designers, typically by re-training or by adding new features (i.e., updating the feature extraction and matching algorithm). These procedures need to be fast and computationally efficient, and, if possible, they should be automated. For instance, some automatic technique may be exploited for detecting when the system performance may incur a significant degradation, and to consequently re-training the system, or demanding for new features selection algorithm. To this aim, concept drift, unsupervised learning, and active learning techniques may be exploited.

5. Future efforts should also be directed to design or model adversary aware biometric classifiers. Adversary aware classifiers utilize adversarial pattern classification techniques to remain robust to spoof attacks. The robustness may be attained by adapting the training set or fusion methods on injecting the spoof attacks to the system. The design of such an optimal and robust biometric systems will substancially advance the state-of-the-art in this field.

To conclude, we hope that the contributions of this thesis will capture the attention of the scientific community, since it is a great opportunity and challenge to improve the research in this field. Biometric recognition systems as security systems neither have been tried at such large scales nor have they dealt with such a wide use of sensitive personal information, and thus demand rises for a more secure implementation of the systems, as well as new techniques to address open issues like the ones mentioned above.

List of Publications Related to Thesis

Published papers

- Zahid Akhtar, Giorgio Fumera, Gian Luca Marcialis, and Fabio Roli, "Robustness Evaluation of Biometric Systems under Spoof Attacks", In *16th International Conference on Image Analysis and Processing (ICIAP)*, pp. 159–168, 2011.
- Zahid Akhtar, Battista Biggio, Giorgio Fumera, Gian Luca Marcialis, "Robustness of Multi-modal Biometric Systems under Realistic Spoof Attacks against All Traits", In *IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BioMS)*, pp. 5–10, 2011.
- Zahid Akhtar, Giorgio Fumera, Gian Luca Marcialis and Fabio Roli, "Robustness analysis of Likelihood Ratio score fusion rule for multi-modal biometric systems under spoof attacks", In *45th IEEE International Carnahan Conference on Security Technology (ICCST)*, pp. 237–244, 2011.
- Battista Biggio, Zahid Akhtar, Giorgio Fumera, Gian LucaMarcialis, Fabio Roli, "Robustness of multi-modal biometric verification systems under realistic spoofing attacks", In *International Joint Conference on Biometrics (IJCB)*, 2011.
- Zahid Akhtar, Giorgio Fumera, Gian Luca Marcialis and Fabio Roli, "Evaluation of multimodal biometric score fusion rules under spoof attacks", In *5th IAPR/IEEE International Conference on Biometrics (ICB)*, 2012.
- Battista Biggio, Zahid Akhtar, Giorgio Fumera, Gian LucaMarcialis, Fabio Roli, "Security evaluation of biometric authentication systems under realistic spoofing attacks", In *IET Biometrics*, In press, 2012.

For an updated list of publications by the author please visit http://prag.diee.unica.it/pra/eng/people/akhtar.

Bibliography

[1] A. Abhyankar and S. A. C. Schuckers. Integrating a wavelet based perspiration liveness check with fingerprint recognition. *Pattern Recognition*, 42(3):452–464, 2009. [cited at p. 68]

[2] Z. Akhtar, B. Biggio, G. Fumera, and G. L. Marcialis. Robustness of multimodal biometric systems under realistic spoof attacks against all traits. In *IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BioMS 2011)*, pages 5–10, Milan, Italy, 2011. [cited at p. 64]

[3] Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli. Robustness analysis of likelihood ratio score fusion rule for multi-modal biometric systems under spoof attacks. In *45th IEEE International Carnahan Conference on Security Technology (ICCST)*, pages 237–244, Barcelona, Spain, 2011. [cited at p. 101]

[4] Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli. Robustness evaluation of biometric systems under spoof attacks. In *16th International Conference on Image Analysis and Processing (ICIAP 2011)*, pages 159–168, 2011. [cited at p. 101]

[5] Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli. Evaluation of multimodal biometric score fusion rules under spoof attacks. In *The 5th IAPR International Conference on Biometrics (ICB)*. In press, 2012. [cited at p. 101]

[6] A. Anjos and S. Marcel. Counter-measures to photo attacks in face recognition: a public database and a baseline. In *International Joint Conference on Biometrics (IJCB 2011)*. In press, 2011. [cited at p. 27, 43, 63]

[7] B. Biggio. *Adversarial Pattern Classification*. PhD thesis, University of Cagliari, Cagliari (Italy), 2010. [cited at p. xiv, 4]

[8] B. Biggio, Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli. Robustness of multi-modal biometric verification systems under realistic spoofing attacks. In *International Joint Conference on Biometrics (IJCB 2011)*, Washington DC, USA, 2011. [cited at p. 64]

[9] B. Biggio, Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli. Security evaluation of biometric authentication systems under realistic spoofing attacks. *IET Biometrics*, In press, 2012. [cited at p. 64]

[10] M. M. Chakka, A. Anjos, S. Marcel, R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori, F. Roli, J. Yan, D. Yi, Z. Lei, Z. Zhang, S. Z. Li, W. R. Schwartz, A. Rocha, H. Pedrini, J. Lorenzo-Navarro, M. Castrillon-Santana, J. Maatta, A. Hadid, and M. Pietikainen. Competition on counter measures to 2-D facial spoofing attacks. In *International Joint Conference on Biometrics (IJCB 2011)*. In press, 2011. [cited at p. 27, 32, 33, 40, 42, 43, 63]

[11] S. A. Cole. *Suspect Identities - A History of Fingerprinting and Criminal Identification*. Harvard University Press, 2001. [cited at p. 28, 31]

[12] P. Coli, G. L. Marcialis, and F. Roli. Vitality detection from fingerprint images: A critical survey. In *Proceedings of the international conference on Advances in Biometrics*, pages 722–731, 2007. [cited at p. 27]

[13] P. Coli, G. L. Marcialis, and F. Roli. Fingerprint silicon replicas: Static and dynamic features for vitality detection using an optical capture device. *Int. J. Image Graphics*, 8(4):495–512, 2008. [cited at p. 27]

[14] J. Daugman. Combining multiple biometrics. Available at http://www.cl.cam.ac.uk/~jgd1000/combine/combine.html, 2000. [cited at p. 21]

[15] R. Derakhshani, S. A. C. Schuckers, L. A. Hornak, and L. O. Gorman. Determination of vitality from a noninvasive biomedical measurement for use in fingerprint scanners. *Pattern Recognition*, 36(5):383–396, 2003. [cited at p. 27]

[16] B. Dorizzi, S. Garcia-Salicetti, and L. Allano. Multimodality in biosecure: Evaluation on real vs. virtual subjects. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1089–1092, 2006. [cited at p. 47]

[17] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. Wiley, second edition, 2001. [cited at p. 1]

[18] T. Dunstone and G. Poulton. Vulnerability assessment. *Biometric Technology Today*, 2011(5):5–7, 2011. [cited at p. 2]

[19] T. Fawcett. An introduction to ROC analysis. *Pattern Recogn. Lett.*, 27:861–874, June 2006. [cited at p. 16]

[20] G. Feng, K. Dong, D. Hu, and D. Zhang. When faces are combined with palmprints: A novel biometric fusion strategy. In *First International Conference Biometric Authentication*, pages 701–707, 2004. [cited at p. 20]

[21] J. Fierrez, J. Galbally, A. Anjos, C. McCool, F. Alegre, N. Evans, A. Thiebot, A. Hadid, S. Z. Li, G. L. Marcialis, J. Carter, J. Bustard, and J. Acedo. TABULA RASA trusted biometrics under spoofing attacks (D2.3: Specifications of spoofing attacks). Technical report, The European Commission, 2011. [cited at p. xv, 34]

[22] J. Fierrez-Aguilar, J. Ortega-Garcia, D. Garcia-Romero, and J. Gonzalez-Rodriguez. A comparative evaluation of fusion strategies for multimodal biometric verification. In *Proceedings of the 4th international conference on Audio- and video-based biometric person authentication*, AVBPA'03, pages 830–837, Berlin, Heidelberg, 2003. Springer-Verlag. [cited at p. 21]

[23] R. A. Freeman. *The Red Thumb Mark*. Collingwood, London, U.K., 1907. [cited at p. 28]

[24] J. Galbally, R. Cappelli, A. Lumini, D. Maltoni, and J. Fiérrez-Aguilar. Fake fingertip generation from a minutiae template. In *International Conference on Pattern Recognition*, pages 1–4, 2008. [cited at p. 30]

[25] J. Galbally-Herrero, J. Fierrez-Aguilar, J. D. Rodriguez-Gonzalez, F. Alonso-Fernandez, J. Ortega-Garcia, and M. Tapiador. On the vulnerability of fingerprint verification systems to fake fingerprint attacks. In *Proc. IEEE Intl. Carnahan Conf. on Security Technology*, pages 130–136, October 2006. [cited at p. 29]

[26] S. Garcia-Salicetti, M. Mellakh, L. Allano, and B. Dorizzi. A generic protocol for multibiometric systems evaluation on virtual and real subjects. In *5th International Conference Audio- and Video-Based Biometric Person Authentication*, pages 295–315. 2005. [cited at p. 47]

[27] M. D. Garris, C. I. Watson, and C. L. Wilson. Matching performance for the US-Visit IDENT system using flat fingerprints. Technical report, 7110, National Institute of Standards and Technology (NIST), July 2004. [cited at p. 17]

[28] B. Geller, J. Almog, P. Margot, and E. Springer. A chronological review of fingerprint forgery. *Journal of Forensic Sciences*, 44(5):963–968, 1999. [cited at p. 27]

[29] A. Godil, Y. Ressler, and P. Grother. Face recognition using 3d facial shape and color map information: comparison and combination. In *Proceedings of the SPIE - The International Society for Optical Engineering*, pages 351–361, 2005. [cited at p. 31]

[30] G. Hamilton-(Director). Diamonds are forever (film). http://www.imdb.com/title/tt0066995/, 1971. [cited at p. 28]

[31] K. Harmel and L. Spadanuta. Disney world scans fingerprint details of park visitors. Available at http://www.boston.com/news/nation/articles/2006/09/03/disney_world_scans_fingerprint_details_of_park_visitors, September 2006. [cited at p. 12]

[32] M. He, S.-J. Horng, P. Fan, R.-S. Run, R.-J. Chen, J.-L. Lai, M. K. Khan, and K. O. Sentosa. Performance evaluation of score level fusion in multimodal biometric systems. *Pattern Recogn.*, 43:1789–1800, May 2010. [cited at p. 21]

[33] X. He, Y. Lu, and P. Shi. A fake iris detection method based on FFT and quality assessment. In *Proc. Chinese Conf. on Pattern Recognition*, pages 316–319, 2008. [cited at p. 27]

[34] L. Hong, A. Jain, and S. Pankanti. Can multibiometrics improve performance? In *AutoID*, pages 59–64, October 1999. [cited at p. 7, 19, 35, 39]

[35] R. L. Hsu. *Face Detection and Modeling for Recognition*. PhD thesis, Department of Computer Science and Engineering, Michigan State University, 2002. [cited at p. xiv, 19]

[36] A. K. Jain, P. Flynn, and A. A. Ross. *Handbook of Biometrics*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007. [cited at p. 7, 14, 16]

[37] A. K. Jain, L. Hong, and Y. Kulkarni. A multimodal biometric system using fingerprints, face and speech. In *Second International Conference on Audio- and Video-based Biometric Person Authentication*, pages 182–187, 1999. [cited at p. 19]

[38] A. K. Jain and D. Maltoni. *Handbook of Fingerprint Recognition*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2003. [cited at p. 15]

[39] A. K. Jain, S. Prabhakar, and S. Chen. Combining multiple matchers for a high security fingerprint verification system. *Pattern Recognition Letters*, 20(11-13):1371–1379, 1999. [cited at p. 78]

[40] A. K. Jain and A. Ross. Multibiometric systems. *Commun. ACM*, 47:34–40, January 2004. [cited at p. 17, 20]

[41] A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Trans. Circuits Syst. Video Techn.*, 14(1):4–20, 2004. [cited at p. 11]

[42] P. A. Johnson, B. Tan, and S. Schuckers. Multimodal fusion vulnerability to non-zero effort (spoof) imposters. In *IEEE Workshop on Information Forensics and Security (WIFS)*, pages 1–5, 2010. [cited at p. 21, 35, 36, 37, 39, 40, 46, 47, 48, 49, 57, 60, 61, 62, 63, 64, 67, 68, 82, 88, 96]

[43] H. Kang, B. Lee, H. Kim, D. Shin, and J. Kim. A study on performance evaluation of the liveness detection for various fingerprint sensor modules. In *7th International Conference Knowledge-Based Intelligent Information and Engineering Systems*, pages 1245–1253, 2003. [cited at p. 7, 27, 29]

[44] D. V. Klien. Foiling the cracker; a survey of, and improvements to unix password security. In *The Second USENIX Workshop on Security*, pages 5–14, August 1990. [cited at p. 11]

[45] K. Kollreider, H. Fronthaler, and J. Bigun. Verifying liveness by multiple experts in face biometrics. In *IEEE Computer Vision and Pattern Recognition Workshop on Biometrics*, pages 1–6, 2008. [cited at p. 27, 33, 42]

[46] A. Kumar, V. Kanhangad, and D. Zhang. A new framework for adaptive multimodal biometrics management. *IEEE Transactions on Information Forensics and Security*, 5(1):92–102, 2010. [cited at p. 95]

[47] A. Lanitis. A survey of the effects of aging on biometric identity verification. *Int. J. Biometrics*, 2:34–52, December 2010. [cited at p. 17]

[48] S. Latifi and N. Solayappan. A survey of unimodal biometric methods. In *Security and Management*, pages 57–63, 2006. [cited at p. 6]

[49] L. LeCam. *Asymptotic Methods in Statistical Decision Theory*. Springer series in statistics. Springer, New York, 1986. [cited at p. 78]

[50] E. C. Lee, Y. J. Ko, and K. R. Park. Fake iris detection method using purkinje images based on gaze position. *Optical Engineering*, 47(6):067204, 2008. [cited at p. 65]

[51] J. Li, Y. Wang, T. Tan, and A. K. Jain. Live face detection based on the analysis of fourier spectra. In *Biometric Technology for Human Identification, SPIE*, volume 5404, pages 296–303, 2004. [cited at p. 28, 42]

[52] G. L. Marcialis, A. Lewicke, B. Tan, P. Coli, D. Grimberg, A. Congiu, A. Tidu, F. Roli, and S. A. C. Schuckers. First international fingerprint liveness detection competition - LivDet 2009. In *International Conference on Image Analysis and Processing (ICIAP)*, pages 12–23, 2009. [cited at p. 28, 31, 35, 37, 40, 41, 65]

[53] G. L. Marcialis and F. Roli. High security fingerprint verification by perceptron-based fusion of multiple matchers. In *5th International Workshop on Multiple Classifiers Systems (MCS04)*, pages 364–373, 2004. [cited at p. 95]

[54] G. L. Marcialis and F. Roli. Fusion of multiple fingerprint matchers by single-layer perceptron with class-separation loss function. *Pattern Recognition Letters*, 26:1830–1839, 2005. [cited at p. 95]

[55] G. L. Marcialis and F. Roli. Score-level fusion of fingerprint and face matchers for personal verification under "stress" conditions. In *Proceedings of the 14th International Conference on Image Analysis and Processing*, pages 259–264, 2007. [cited at p. 21, 34]

[56] G. L. Marcialis, F. Roli, and L. Didaci. Personal identity verification by serial fusion of fingerprint and face matchers. *Pattern Recogn.*, 42:2807–2817, November 2009. [cited at p. 34]

[57] G. L. Marcialis, F. Roli, and A. Tidu. Analysis of fingerprint pores for vitality detection. In *20th International Conference on Pattern Recognition (ICPR)*, pages 1289–1292, 2010. [cited at p. 65]

[58] A. F. Martin, G. R. Doddington, T. Kamm, M. Ordowski, and M. A. Przybocki. The DET curve in assessment of detection task performance. In *EUROSPEECH*, 1997. [cited at p. 16]

[59] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. Impact of artificial "gummy" fingers on fingerprint systems. In *Op. Security and Counterfeit Deterrence Tech. IV, vol. 4677 of Proc. of SPIE*, pages 275–289, 2002. [cited at p. 27, 29]

[60] K. D. Mitnick and W. L. Simon. *The Art of Deception: Controlling the Human Element of Security*. John Wiley & Sons, Inc., New York, USA, 2003. [cited at p. 11]

[61] K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain. Likelihood ratio-based biometric score fusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2):342–347, 2008. [cited at p. 34]

[62] K. Nandakumar, Y. Chen, A. K. Jain, and S. C. Dass. Quality-based score level fusion in multibiometric systems. In *International Conference on Pattern Recognition*, pages 473–476, 2006. [cited at p. 34]

[63] Neurotechnology. Verifinger available at http://www.neurotechnology.com/verifinger.html. [cited at p. 77]

[64] NIST. Bozorth3. Available at http://www.nist.gov/itl/iad/ig/nbis.cfm. [cited at p. 48]

[65] NIST. BSSR1 available http://www.itl.nist.gov/iad/894.03/biometricscores/index.html. [cited at p. 87]

[66] N. Poh and S. Bengio. Can chimeric persons be used in multimodal biometric authentication experiments? In *2nd Int'l Machine Learning and Multimodal Interaction Workshop 2005 (MLMI'05)*, pages 87–100, 2005. [cited at p. 47]

[67] N. Poh, T. Bourlai, J. Kittler, L. Allano, F. Alonso-Fernandez, O. Ambekar, J. Baker, B. Dorizzi, O. Fatukasi, J. Fierrez, H. Ganster, J. Ortega-Garcia, D. Maurer, A. A. Salah, T. Scheidat, and C. Vielhauer. Benchmarking quality-dependent and cost-sensitive score-level multimodal biometric fusion algorithms. *Trans. Info. For. Sec.*, 4:849–866, December 2009. [cited at p. 21]

[68] N. Poh and J. Korczak. Hybrid biometric person authentication using face and voice features. In *Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication*, AVBPA '01, pages 348–353, London, UK, 2001. Springer-Verlag. [cited at p. 20]

[69] T. Putte and J. Keuning. Biometrical fingerprint recognition: Don't get your fingers burned. In *4th Working Conf. on Smart Card Research and Advanced Applications*, pages 289–303, 2000. [cited at p. 29]

[70] N. K. Ratha, J. H. Connell, and R. M. Bolle. An analysis of minutiae matching strength. In *Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication*, AVBPA '01, pages 223–228, London, UK, 2001. Springer-Verlag. [cited at p. 26]

[71] A. Rattani, D. R. Kisku, M. Bicego, and M. Tistarelli. Feature level fusion of face and fingerprint biometrics. In *First IEEE International Conference on Biometrics, Theory, Applications and Systems (BTAS 2007)*, pages 1–6, 2007. [cited at p. 20]

[72] A. Rattani and M. Tistarelli. Robust multi-modal and multi-unit feature level fusion of face and iris biometrics. In *International Conference on Biometrics*, pages 960–969, 2009. [cited at p. 20]

[73] R. N. Rodrigues, N. Kamat, and V. Govindaraju. Evaluation of biometric spoofing in a multimodal system. In *Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, pages 1–5, 2010. [cited at p. 21, 35, 36, 37, 39, 40, 47, 48, 49, 57, 61, 63, 64, 67, 68, 82, 84, 88, 96]

[74] R. N. Rodrigues, L. L. Ling, and V. Govindaraju. Robustness of multimodal biometric fusion methods against spoof attacks. *Journal of Visual Languages and Computing*, 20:169–179, June 2009. [cited at p. 21, 35, 36, 37, 39, 40, 45, 46, 47, 48, 49, 57, 61, 62, 63, 64, 67, 68, 82, 84, 88, 94, 96, 101]

[75] F. Roli, J. Kittler, G. Fumera, and D. Muntoni. An experimental comparison of classifier fusion rules for multimodal personal identity verification systems. In *Proceedings of the Third International Workshop on Multiple Classifier Systems*, pages 325–336. Springer-Verlag, 2002. [cited at p. 21]

[76] A. Ross and A. K. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24(13):2115–2125, 2003. [cited at p. 6, 20, 21]

[77] A. Ross, A. Rattani, and M. Tistarelli. Exploiting the doddington zoo effect in biometric fusion. In *3rd IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS'09)*, pages 1–7, September 2009. [cited at p. 20]

[78] A. A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiometrics*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. [cited at p. 7, 35, 39, 47, 48, 88]

[79] V. Ruiz-Albacete, P. Tome-Gonzalez, F. Alonso-Fernandez, J. Galbally, J. Fierrez, and J. Ortega-Garcia. Direct attacks using fake images in iris verification. In *Proc. Workshop on Biometrics and Identity Management*, pages 181–190, 2008. [cited at p. 27]

[80] M. Sandstrom. Liveness detection in fingerprint recognition systems. Linkopings Universitet: Linkoping, 2004. [cited at p. 31]

[81] C. Y. Suen and L. Lam. Multiple classifier combination methodologies for different output levels. In *Proceedings of the First International Workshop on Multiple Classifier Systems*, pages 52–66, 2000. [cited at p. 95]

[82] X. Tan, Y. Li, J. Liu, and L. Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *Proceedings of the 11th European conference on Computer vision: Part VI*, pages 504–517, 2010. [cited at p. 27, 63]

[83] Q. Tao and R. Veldhuis. Hybrid fusion for biometrics: Combining score-level and decision-level fusion. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Workshop on Biometrics*, pages 1–6, 2008. [cited at p. 21]

[84] L. Thalheim and J. Krissler. Body check: biometric access protection devices and their programs put to the test. *Computer Magazine*, pages 114–121, 2002. [cited at p. 29]

[85] Department of Homeland Security. Privacy impact assessment for the automated biometric identification system (IDENT). Available at http://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_usvisit_ident_final.pdf, July 2006. [cited at p. 12]

[86] Colorado State University. The EBGM matching algorithm. Available at http://www.cs.colostate.edu/evalfacerec/algorithms5.php. [cited at p. 48, 77]

[87] Federal Bureau of Investigation. Integrated automated fingerprint identification system. Available at http://www.fbi.gov/hq/cjisd/iafis.htm. [cited at p. 12]

[88] Luchthaven Schiphol. Privium: A select way to travel. Available at http://www.schiphol.nl/Travellers/AtSchiphol/Privium.htm. [cited at p. 12]

[89] NIST Report to the United States Congress. Summary of NIST standards for biometric accuracy, tamper resistance, and interoperability. Available at ftp://sequoyah.nist.gov/pub/nist_internal_reports/NISTAPP_Nov02.pdf, 2002. [cited at p. 18]

[90] Woman uses tape to trick biometric airport fingerprint scan. http://techcrunch.com/2009/01/02/woman-uses-tape-to-trick-biometric-airport-fingerprint-sc [cited at p. 25]

[91] M. Vatsa, R. Singh, and A. Noore. SVM fusion of multimodal biometric match scores with image quality metric. *International Journal of Neural Systems*, 17(5):880–893, 2007. [cited at p. 21]

[92] C. Wilson, A. R. Hicklin, M. Bone, H. Korves, P. Grother, B. Ulery, R. Micheals, M. Zoepfl, S. Otto, and C. Watson. Fingerprint vendor technology evaluation 2003: Summary of results and analysis report. Technical Report NISTIR 7123, National Institute of Standards and Technology (NIST), June 2004. [cited at p. 17]

[93] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. *Face recognition by elastic bunch graph matching*, pages 355–398. CRC Press, Inc., Boca Raton, FL, USA, 1999. [cited at p. 48, 77]

[94] D. Yambay, L. Ghiani, P. Denti, G. L. Marcialis, F. Roli, and S. Schuckers. LivDet2011 - fingerprint liveness detection competition 2011. In *International Conference on Biometrics (ICB 2012)*. In press, 2012. [cited at p. 28, 31, 40, 41, 65]

[95] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li. Face liveness detection by learning multispectral reflectance distributions. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 436–441, 2011. [cited at p. xv, 32, 33, 42, 63]