UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR
DEPARTAMENTO DE TECNOLOGÍA Y DE LAS COMUNICACIONES

BIOMETRIC SECURITY: A NEW MULTIMODAL HILL-CLIMBING ATTACK

–TRABAJO FIN DE MÁSTER–

SEGURIDAD BIOMÉTRICA: UN NUEVO ATAQUE HILL-CLIMBING MULTIMODAL

Author: Marta Gómez Barrero
Ingeniera Informática y Licenciada en Matemáticas,
Universidad Autónoma de Madrid

A Thesis submitted for the degree of:

Máster Oficial en Ingeniería Informática y de Telecomunicación
(Master of Science)

Madrid, June 2013

Colophon

This book was typeset by the author using LaTeX2e. The main body of the text was set using a 11-points Computer Modern Roman font. All graphics and images were included formatted as Encapsuled Postscript (^{TM} Adobe Systems Incorporated). The final postscript output was converted to Portable Document Format (PDF) and printed.

This Thesis was printed with the financial support from EPS-UAM and the Biometric Recognition Group-ATVS.

Departmento:	Tecnología Electrónica y de las Comunicaciones Escuela Poliétcnica Superior Universidad Autónoma de Madrid (UAM), SPAIN
Título:	Biometric Security: Hill-Climbing Attacks and Inverse Biometrics
Autor:	**Marta Gómez Barrero** Ingeniero de Informática y Licenciada en Matemáticas (Universidad Autónoma de Madrid)
Director:	**Dr. Javier Galbally Herrero** Doctor Ingeniero de Telecomunicación (Universidad de Cantabria) Universidad Autónoma de Madrid, SPAIN
Tutor:	**Prof. Julián Fiérrez Aguilar** Doctor Ingeniero de Telecomunicación (Universidad Politécnica de Madrid) Universidad Autónoma de Madrid, SPAIN
Fecha:	10 de Julio de 2013
Tribunal:	**Dr. Daniel Ramos Castro** Universidad Autónoma de Madrid, SPAIN **Prof. Julián Fiérrez Aguilar** Universidad Autónoma de Madrid, SPAIN **Dr. Manuel Sánchez-Montañés Isla** Universidad Autónoma de Madrid, SPAIN
Calificación:	

The research described in this Thesis was carried out within the Biometric Recognition Group – ATVS at the Dept. of Tecnología Electrónica y de las Comunicaciónes, Escuela Politécnica Superior, Universidad Autónoma de Madrid. The project was partially funded by a PhD scholarship from Spanish Ministerio de Educacion, Ciencia y Deporte.

Abstract

As any technology aimed to provide a security service, biometric systems are exposed to external attacks which could compromise their integrity. Thus, it is of special relevance to understand the threats to which they are subjected and to analyse their vulnerabilities in order to prevent possible attacks and increase their benefits for the users. This line of thought has encouraged some research studies in the last decade, as well as common projects in the European Union, which involve universities as well as private companies.

In this work, a new multimodal indirect attack has been proposed and its impact on the security offered by biometric systems studied. First, new indirect attacks based on hill-climbing algorithms to unimodal systems have been proposed. Their performance has been thoroughly analysed on systems based on face and iris, and compared to existing state-of-the-art algorithms. The performance of the multimodal attack has been also compared to the performance of the unimodal sub-algorithms.

During all the experiments, a rigorous and completely reproducible protocol has been followed. This allows us to establish a fair comparison between the attacking schemes proposed, regardless of the biometric system attacked. Furthermore, future studies, possibly carried out by other researchers, can follow the same protocol (also described in the present work) and their findings can be compared to ours.

The results show to what extent the proposed techniques affect the security offered by the biometric systems tested, and the necessity of new measures to counterfeit these types of threats. Some countermeasures have been also tested: score quantization has been applied to the face-based unimodal and the multimodal recognition systems, while for the iris-based system the use of the most consistent bits of the iriscode has been considered. Even though the countermeasures applied barely reduced the success chances of the unimodal sub-algorithms, in the case of the multimodal attack, score quantization successfully prevents an eventual impostor from gaining access to the system. [1]

[1] Part of the work of this thesis has been published in Pattern Recognition Letters, and presented in the 5th International Conference in Biometrics, in the 17th Iberoamerican Conference on Pattern Recognition and in the European Workshop on Biometrics and Identity Management, BioID. The accepted papers are included in Appendix A.

A mis padres.
A mi abuelo.
A mi familia.

A Fer.

A little step may be the beginning of a great journey
(Un pequeño paso puede ser el inicio de un gran viaje)
Unknown

Acknowledgements

This MSc Thesis summarises the work carried out during my Master studies with the Biometric Recognition Group - ATVS since 2011.

First of all, I would like to thank my advisors Dr. Javier Galbally and Prof. Julian Fierrez for their guidance and support since I started working in the group. They have trusted me from the beginning and encouraged me to keep going, no matter the obstacles in the path.

But they were not the first people to trust me and help me grow as a person: all the teachers and professors I have encountered during my years at IES Diego Velázquez and during my degree studies at UAM have also helped me in my way to this MSc. Special mention to Jose M Letona, Mario Lopez, Ezequiel Castellanos and Fernando Soria. I would not be here if it wasn't for them.

I would like to thank my colleagues (and friends) from the lab and "upstairs": Pedro Tome, Ruben Vera, Ruifang Wang, Javier Franco, Javier Gonzalez and Daniel Ramos. Also the new guys: Alvaro, Alicia, Ester... and the ones who were here when I arrived, but now are working somewhere else: Miriam Moreno, Maria Puertas and Ignacio Gonzalez. Thanks for all the jokes, for all the good meals and cañas together... and for making those hot and cold days at the lab so much easier!!

But without my family (my parents, my grandpa) and friends (Landecho, Patri,...), without Fer, nothing would have been the same: thanks for sharing the best moments, and for being there during the most difficult times.

Gracias.

Marta Gómez Barrero
Madrid, July 2013

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Automatic recognition of people is a challenging issue present in many fields, from authentication in your own laptop to on-line bank transactions. In the last decades, biometrics have emerged as an alternative to traditional PIN- or token-based security applications, offering a reliable way of verifying the identity of a given subject. However, as with any other new technology, a great effort, shared by all the biometrics community (i.e., researchers, vendors, etc.), has to be done in order to improve the performance of the systems, reduce the impact of external attacks, and increase the trust of the final user. This work tries to shed some light into the area, and encourage future studies that may help the introduction of biometrics into everyday life.

In this Chapter biometrics are introduced in Sect. 1.1, and several aspects, closely related to the topic at hand, are described: in Sect. 1.2, multibiometrics are presented and Sect 1.3 describes the complex problem of security evaluation in biometrics. Finally, the objectives of this work are presented in Sect. 1.4 and an outline of the Dissertation is included in Sect. 1.5.

1.1. Biometric Systems

"A biometric system measures one or more physical or behavioural characteristics [...] of an individual to determine or verify his identity" [Jain *et al.*, 2011]. Biometric systems can be therefore regarded as an specific application of pattern recognition algorithms. Instead of asking *what you posses* or *what you know*, these systems base their decisions on *who you are or what you produce*. On the one hand, who you are refers to *physiological* characteristics such as fingerprints, iris, or face. On the other hand, what you produce refers to *behavioral* patterns which entail a learning process and that characterise your identity such as the voice or the written signature.

This new paradigm proposed by biometrics not only provides enhanced security but also avoids, in authentication applications, the need to remember multiple passwords and maintain multiple authentication tokens.

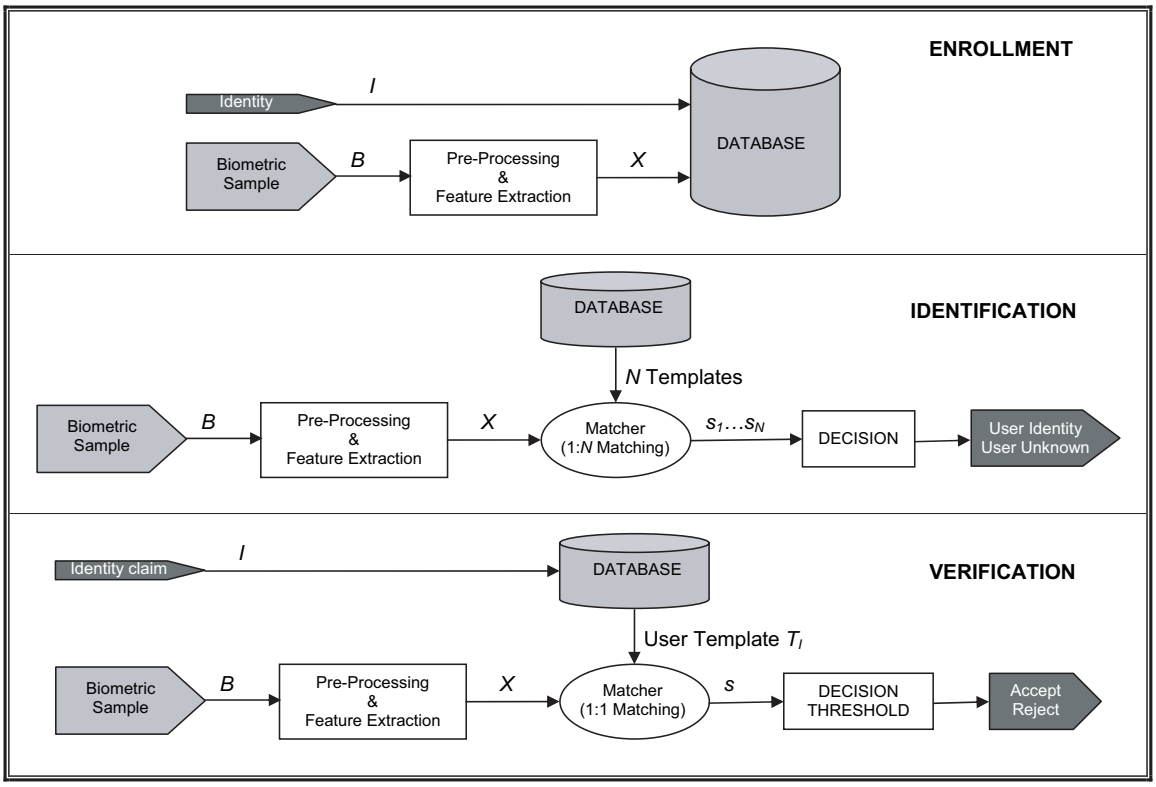

Figure 1.1: *Diagrams of the typical modes of operation in a biometric system.*

In order to process the information comprised in biometric traits, their characteristics or features have to be extracted and transformed into the digital domain. This digital representation is known as *template*. In Fig. 1.1 (top), the *enrolment* or *training* process by which templates are stored in the system is depicted. Once the users have been enrolled to the system, the recognition process can be performed in two modes [Jain *et al.*, 2006]:

- **Identification**. In this mode, a one-to-many comparison must be performed in order to answer the question "is this person in the database?". The answer might thus be Â'NoÂ' (the person is unknown to the system), or any of the registered identities in the database (Fig. 1.1, center).

- **Verification**. In this case, a one-to-one comparison must be carried out: we want to know is if a person is really who she claims to be (i.e., is this person truly John Doe?). The system compares the biometric trait to the enrolled pattern associated with the claimed identity (Fig. 1.1, bottom), in order to determine if the subject is a *client* (the identity claim is *accepted*), or an *impostor* (the identity claim is *rejected*). Typical verification applications include network logon, ATMs, physical access control, credit-card purchases, etc.

This Dissertation is focused on the security evaluation of biometric systems working under the verification mode (also known as *authentication*). In this mode, the *clients* or *targets* are

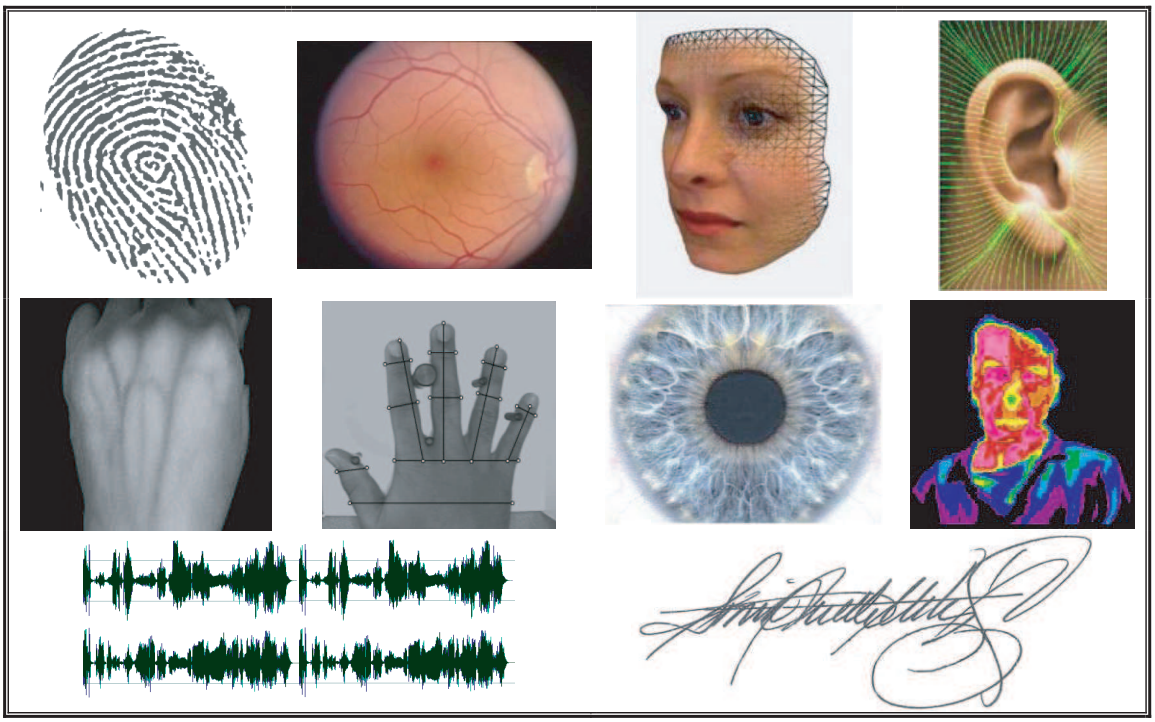

Figure 1.2: *Examples of common biometrics.*

known to the system (through the enrolment process), whereas the *impostors* can potentially be the world population. The result of the comparison between the feature vector X (extracted from the biometric sample B provided by the user) and the template T_I corresponding to his/her claimed identity I is a similarity score s which is compared to a decision threshold. If the score is higher than the decision threshold, then the claim is accepted (client), otherwise the claim is rejected (impostor).

In order to perform the user authentication, a number of different biometrics have been proposed and are used in various applications [Jain *et al.*, 2006]. The above mentioned classification into physical and psychological traits is just indicative, as some of them are not easy to categorize in any of the groups. The voice, for instance, is commonly accepted to be a behavioural biometric (as the voice is something that we *learn* to *produce*), however its distinctiveness largely depends on physiological characteristics (e.g., vocal tracts, mouth, nasal cavities, or lips). On the other hand, other very distinctive human feature, the DNA, is usually not considered a biometric modality as recognition systems based on it still require manual operation and cannot be used in (pseudo) real-time. Example images from various biometrics are given in Fig. 1.2.

In theory, any human characteristic can be used as a biometric identifier as long as it satisfies the following requirements:

- **Universality**, which indicates to what extent a biometric is present in the world population.
- **Distinctiveness**, which means that two persons should have sufficiently different biomet-

rics.

- **Permanence**, which indicates that the biometric should have a compact representation invariant over a sufficiently large period of time.
- **Collectability**, which refers to the easiness of the acquisition process and to the ability to measure the biometric quantitatively.

Other criteria required for practical applications include:

- **Performance**, which refers to the efficiency, accuracy, speed, robustness and resource requirements of particular implementations based on the biometric.
- **Acceptability**, which refers to which people are willing to use the biometric and in which terms.
- **Circumvention**, which reflects the difficulty to fool a system based on a given biometric by fraudulent methods.
- **Exception handling**, which has to do with the possibility to complete a manual matching process for those people that cannot interact in a normal way with the system (e.g., impossibility to perform the feature extraction process due to an excessive degradation of the trait).
- **Cost**, which refers to all the costs that would be necessary to introduce the system in a real-world scenario.

1.2. Multibiometrics

An ideal biometric system should meet all the aforementioned requirements. Unfortunately, no single biometric trait satisfies all of them: while some biometrics have a very high distinctiveness (e.g., fingerprint or iris), they are relatively easy to circumvent (e.g., using a gummy finger, or an iris printed photograph); similarly, other biometrics such as the face thermogram or the vein pattern of the retina are very difficult to circumvent, but their distinctiveness is low and are not easy to acquire. Multibiometrics, or the combination of several biometric sources, have been considered as a possible solution to this problem: several sources usually compensate for the inherent limitations of one another Jain *et al.* [2011].

Combining information from several biometric traits also increases the accuracy of the recognition system, an essential requirement for many practical applications such as the US-VISIT program and the Unique Identification (UID) system in India, where the identification of very large numbers of subjects must be resolved with negligible error rates.

However, multibiometric systems not only include those based on several traits. According to the multiple sources used, multibiometric systems can be broadly classified into: *i*) multi-sensor, that is, several sensors are used in the acquisition stage; *ii*) multi-algorithm, that is, several

representations of the trait are considered and processed; *iii*) multi-instance, for example, both the right and the left eyes of the subject are considered; *iv*) multi-sample, that is, one sensor is used to acquire multiple samples of the same trait; and *v*) multimodal systems, which combine different biometric traits. Therefore, in the first four scenarios, only one biometric trait is used.

The information offered by several unimodal biometric systems can be fused at five different levels, namely: *i*) sensor-level, *ii*) feature-level, *iii*) score-level, *iv*) rank-level, and *v*) decision-level. While the first two options fuse the information prior to matching, the remaining fusion approaches use the information available after matching: either a numeric score or a binary decision (accept/reject). Even though fusion at an earlier processing stage should perform better (the information available is richer), specific and more complex algorithms have to be applied. The development of such algorithms is not straightforward, and therefore most multibiometric systems use score- or decision-level approaches.

As well as increasing the performance of biometric recognition systems, multibiometrics have usually been considered to increase the security of the systems for the final users. They have also been regarded as more robust to external attacks: the vulnerabilities shown by one trait can be compensated by the rest of the traits used.

However, in the last decade, some studies have questioned this common belief for spoofing attacks [Akhtar and Alfarid, 2011; Akhtar *et al.*, 2011; Chetty and Wagner, 2005; Johnson *et al.*, 2010; Rodrigues *et al.*, 2010, 2009]. In most cases, attacking the most robust trait leads to success in the global attack. The question of whether the same behaviour may be observed for indirect attacks remains unanswered.

In the present Dissertation, we study for the first time the vulnerabilities of multimodal biometric systems with score-level fusion to indirect software attacks.

1.3. Security Evaluation in Biometrics

A great deal of attention has been given recently to the key area where this work is focused: evaluating the security of biometric systems. Many different researchers have addressed the vulnerabilities of biometric systems to spoofing attacks (those carried out at the sensor level using, for instance, a gummy finger or a printed iris image) [Galbally *et al.*, 2011; Matsumoto, 2004; Rodrigues *et al.*, 2009; Thalheim and Krissler, 2002], and to software-based attacks (carried out against some of the internal modules of the system) [Galbally *et al.*, 2010; Martinez-Diaz *et al.*, 2011; Uludag and Jain, 2004]. Furthermore, the interest in the analysis of system vulnerabilities has permeated the scientific field and different standardization initiatives at the international level have emerged in order to deal with the problem of security evaluation in biometric systems, such as the Common Criteria [CC, 2006] and its associated Biometric Evaluation Methodology [BEM, 2002] or the ISO/IEC-19792:2009 for biometric security evaluation [ISO/IEC 19792, 2009].

In addition to research works which address the vulnerabilities of multimodal systems to spoofing attacks [Akhtar and Alfarid, 2011; Akhtar *et al.*, 2011; Chetty and Wagner, 2005;

Hämmerle-Uhl *et al.*, 2011; Johnson *et al.*, 2010; Marasco, 2010; Rodrigues *et al.*, 2010, 2009], different studies may be found in the literature regarding the analysis of indirect attacks to unimodal systems [Galbally *et al.*, 2010; Martinez-Diaz *et al.*, 2011; Uludag and Jain, 2004]. However, the problem of whether multimodal approaches are vulnerable or not to software-based attacking methodologies, and if their level of protection against these threats is really increased compared to unimodal schemes, still remains unexplored.

1.3.1. Transparency vs Obscurity

When addressing the problem of providing a security service, two main approaches may be adopted to guarantee that the level of security offered to the user is not compromised: *security through obscurity* (also *security by obscurity*) or *security through transparency* (also known as *security by design*).

The security through obscurity principle relies on secrecy (of design, implementation, formats and protocols used, etc.) to provide security. A system using this approach may have theoretical or practical security vulnerabilities, but its designers believe that attackers are unlikely to find or to exploit them. Developers supporting this methodology argue that if details of countermeasures employed in biometric systems are publicized, it may help attackers to avoid or defeat them. Similarly, if attackers know what countermeasures are not employed, this will help them to identify potential weaknesses in the system, enabling the attacks towards those weak areas. Furthermore, an attacker's first step is usually information gathering; this step is delayed by security through obscurity.

In opposition, the security through transparency scheme follows the Kerckhoffs' principle (stated by Auguste Kerckhoffs in the 19th century) [Kerckhoffs, 1883]: a cryptosystem should be secure even if everything about the system, except the key, is public knowledge. Although it was first thought for cryptography, the principle was later reformulated to be applied to any security system as "the enemy knows the system". Undoubtedly, any security system depends on keeping *some* things secret, the question is, *what* things?. The Kerckhoffs' principle points out that the things which are kept secret ought to be those which are least costly to change if inadvertently disclosed. In other words, the fewer and simpler the things one needs to keep secret in order to ensure the security of the system, the easier it is to maintain that security. Quoting B. Schneier, one of the world's leading security technologists, "*Kerckhoffs' principle applies beyond codes and ciphers to security systems in general: every secret creates a potential failure point. Secrecy, in other words, is a prime cause of brittleness and therefore something likely to make a system prone to catastrophic collapse. Conversely, openness provides ductility*" [Schneier, 2000].

Applying security through transparency to biometrics would mean, in words of the Biometric Working Group [BWG, 2009]: "*make public exposure of countermeasures and vulnerabilities which will lead to a more mature and responsible attitude from the biometrics community and promote the development of more secure systems in the future*" [BWG, 2003].

Our view on biometric security, based on which this Dissertation has been developed, is

aligned with the security through transparency principle. This way, throughout the Dissertation different threats that may affect biometric systems are pointed out, systematically evaluated, and new countermeasures that can guarantee the final level of security offered to the user are proposed.

1.4. Motivation and Objectives

Certainly, biometrics are very attractive and useful for the final user: forget about PINs and passwords, you are your own key Jain *et al.* [2006]; Wayman *et al.* [2005]. However, we cannot forget that as any technology aimed to provide a security service, biometric systems are exposed to external attacks which could compromise their integrity [Schneier, 1999]. Thus, it is of special relevance to understand the threats to which they are subjected and to analyse their vulnerabilities in order to prevent possible attacks and increase their benefits for the users.

Although several works have already studied different specific vulnerabilities of biometric systems [Hennebert *et al.*, 2007; Hill, 2001; Thalheim and Krissler, 2002], the problem has been addressed on most cases from a *yes-or-no* perspective (i.e., the question being answered is, *can a biometric system be bypassed using this attacking method?*). However, in most of those valuable research contributions, a far more complex question remains unanswered: ***how*** *vulnerable is the biometric system to the attack?*. Identifying the threats is the first stage in a vulnerability evaluation, however quantifying the danger is just as important in order to assess the security level provided by the application.

The second observation is strongly related to the first one. In these existing publications, experimental results are obtained and reported without following any general or systematic protocol, and thus, even in the case of performing an statistical analysis of a given vulnerability, results cannot be compared, losing this way part of their utility.

This new concern which has arisen in the biometric community regarding the security of biometric systems has led to the appearance of several international projects, like the European Tabula Rasa [2010] and BEAT [2012], which base their research on the security through transparency principle, which was already introduced in Sect. 1.3.1 [Kerckhoffs, 1883; Schneier, 2000]: in order to make biometric systems more secure and reliable, their vulnerabilities need to be analysed and useful countermeasures need to be developed.

As well as those international projects, different initiatives are currently trying to develop standard security evaluation protocols [BEM, 2002; CC, 2006; ISO/IEC 19792, 2009]. These standards are in general directed to the very wide range of Information Technology security products, which means that additional documents are required in order to apply the general guidelines given in the norms to the particular specificities of a given technology. This is specially important in the biometric field due to the large amount of different existing biometric modalities and the multiple areas of knowledge that it covers (pattern recognition, computer vision, electronics, etc.)

Thus, following the same transparency principle, the main objectives and contributions of

the present work are: *i*) proposal of a fully novel software-based attacking methodology against multimodal systems, *ii*) study of the vulnerabilities of a realistic multimodal system to the previous attack under a replicable scenario, *iii*) comparison of the performance of the multimodal attack to that obtained against the unimodal modules in order to determine if the multimodal approach increases the security of the system against this type of threat, and *iv*) study of some biometric-based countermeasures which may prevent such attacks.

1.5. Outline of the Dissertation

The Dissertation is structured as follows:

- Chapter 1 introduces the topic of biometrics and gives the motivation, outline and objectives of this MSc. Thesis.
- Chapter 2 summarizes the state of the art in biometric security, presenting related works.
- Chapter 3 describes the methods proposed in the Dissertation.
- Chapter 4 presents the experimental protocol followed in the experiments.
- Chapter 5 describes the experiments carried out and analyses the results.
- Chapter 6 concludes the Dissertation summarizing the main results obtained and outlining future research lines.

Chapter 2

Related Works and State of the Art

Biometrics is a wide field of research, where, as with any other new technology, many aspects must be analysed. Many studies have been carried out in order to improve the recognition performance or analyse the social acceptability, but in order to increase the trust of the final user, other aspects should be carefully looked into. The level of security and privacy offered by these systems are a key to the final expansion of this thriving technology.

Only the most closely related topics to the Dissertation within the biometrics field of research are summarized in this Chapter. The main goal of the present work is to evaluate the vulnerabilities of a multimodal biometric system to an indirect attack. The general method proposed, which may be applied to any multimodal system, is described in Chapter 3, and experiments are carried out on a face and iris-based biometric system in Chapter 5.

This Chapter thus focuses on vulnerability evaluations to indirect attacks. In Sect. 2.1, attacks to biometric recognition systems are presented and categorized, emphasizing indirect attacks, more closely related to our work. Finally, in Sect. 2.2 and 2.3, brief summaries of face and iris recognition, respectively, are presented.

2.1. Attacks to Biometric Recognition Systems

In 2001, Ratha *et al.* identified and classified in a biometric recognition system eight possible points of attack [Ratha *et al.*, 2001]. These vulnerable points, depicted in Fig. 2.1, can be broadly divided into direct (first point) and indirect attacks (seven remaining points).

Direct attacks. Also known as spoofing-attacks, these are attacks at the sensor level, carried out with synthetic biometric traits, such as gummy fingers or high quality printed iris images, and thus requiring no knowledge for the attacker of the inner parts of the system (matching algorithm used, feature extraction method, template format, etc.)

Some research regarding the vulnerabilities of multimodal systems to these attacks has been carried out over the last recent years: in 2005, Chetty and Wagner [2005] tested the performance of spoofing attacks against a novel multimodal system based on face and voice; in 2009, Tan [2009] investigated methods for increasing the security of multimodal systems based on face

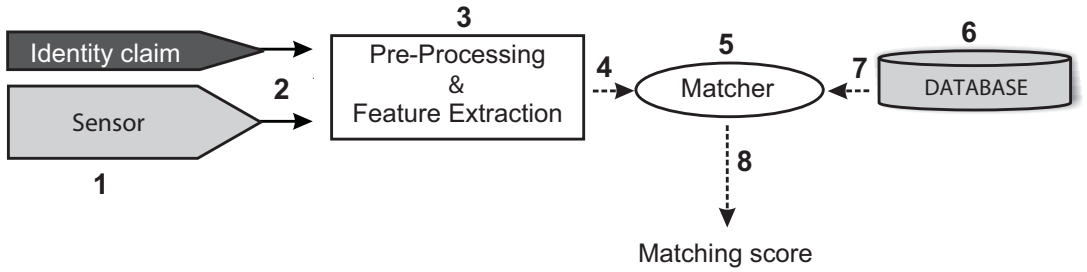

Figure 2.1: *Architecture of an automated biometric verification system. Possible attack points are numbered from 1 to 8.*

and voice against spoofing attacks; in 2010 [Rodrigues *et al.*, 2010] and 2011 [Rodrigues *et al.*, 2009], Rodrigues *et al.* evaluated the vulnerabilities of a multimodal system based on face and fingerprint, using different fusion techniques and proposing new ones more robust to the attacks; in 2010, Johnson *et al.* [2010] analysed the effect of spoofing attacks against a multimodal system based on face and iris, proposing a method for the vulnerabilities assessment of these systems; later in 2010, Marasco [2010] analysed the security risk in multimodal biometric systems based on face and fingerprint coming from spoofing attacks; in 2011, Akhar *et al.* [Akhtar and Alfarid, 2011; Akhtar *et al.*, 2011] used real rather than simulated spoof samples for the evaluation of the vulnerabilities of a multimodal system based on fingerprint, face and iris, proposing a new learning algorithm able to improve the security offered by the system against spoofing attacks. All these works have proven that combining several traits in one system for person authentication does not necessarily increment the security offered against spoofing attacks, since the system can be bypassed by breaking only one of the unimodal traits.

Indirect attacks. These attacks can be further divided in three groups, namely: *i*) attacks 2, 4, 7 and 8 lie in the communication channels of the system, extracting, adding or changing information; *ii*) attacks 3 and 5 may be carried out using a Trojan Horse that bypasses the feature extractor and the matcher, respectively; and *iii*) attack 6, in which the system database may be manipulated in order to gain access to the application, by changing, adding or deleting a template. While for direct attacks the intruder needed no knowledge about the inner modules of the system, this knowledge is a main requisite here, together with access to some of the system components (database, feature extractor, matcher, etc.).

Most of these indirect attacks are based on some variation of a hill-climbing algorithm. A general hill-climbing approach starts by synthetically generating an initial pool of templates (see Fig. 2.2 left). Each of them is used in an attempt to access the system. If none of them succeeds, the templates are iteratively changed, according to a particular modification scheme, so that at each iteration a higher similarity score between the synthetic and the attacked template is obtained, until eventually access to the system is granted. A limit on the number of iterations is usually established in order to avoid infinite loops. The main difference between hill-climbing attacks is therefore the modification scheme.

In Fig. 2.2 (right), an example of the evolution of the scores across the iterations of a hill-

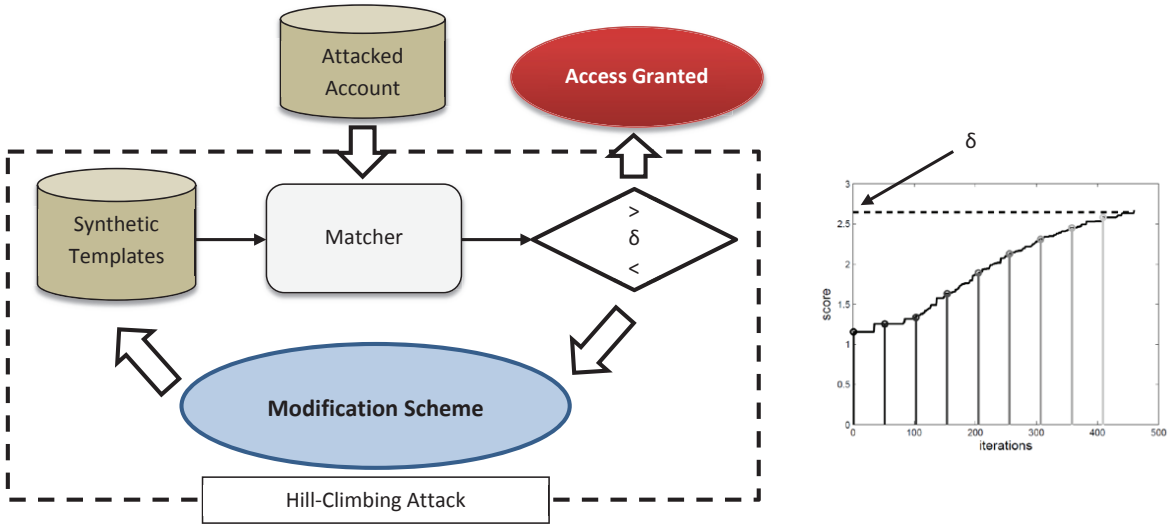

Figure 2.2: *Diagram of a generic hill-climbing attack (left) with an example of the evolution of the scores (right). The threshold δ for which access to the system is granted is depicted by an horizontal dashed line.*

climbing attack is depicted. As may be observed, with each iteration the score increases, meaning that we are closer to our objective: reaching an score higher than the threshold δ, depicted with an horizontal dashed line.

Even though some research has been done in the field of indirect attacks to unimodal systems, based on different traits, to the best of our knowledge there is no previous analysis of the vulnerabilities of multimodal biometric systems against this kind of attacks. Some of those works include the following ones. In [Adler, 2004], the authors successfully recover face images from their templates, as a linear combination of eigenfaces. The hill-climbing algorithm works on randomly selected quadrants of the eigenfaces, adding noise and modifying its pixels. This attack also overcomes the score quantization countermeasure recommended by the BioAPI Consortium [2001]. Later that year, in [Uludag and Jain, 2004], a hill-climbing attack that synthesizes the target minutia templates is proposed. Starting with a random set of minutiae, individual minutia are perturbed, added, replaced or deleted, until access to the system is granted. More recently, Galbally *et al.* [2010] presented a hill-climbing attack algorithm based on Bayesian adaptation to test the vulnerability of two face recognition systems, an eigenface- and a part-based system, to indirect attacks. This algorithm models the faces subspace with an initial Gaussian distribution, and adapts it to the specificities of the user account under attack. Finally, Martinez-Diaz *et al.* [2011] thoroughly analyse the performance of several variations of hill-climbing attacks to minutia-based fingerprint recognition systems, taking or not Regions Of Interest into account. All the variations studied successfully generate fingerprint images able to break into two different state-of-the-art verification systems. Score quantization is also tested as a possible countermeasure, showing that it is an effective measure in order to prevent the studied hill climbing attack, as the performance of the attacking algorithm drops drastically.

2.2. Summary of Face Recognition

Many and quite different methods for face recognition have been proposed during the past 30 years. Face recognition is such a challenging problem that it has attracted researchers who have different backgrounds (e.g., psychology, pattern recognition, neural networks, computer vision, and computer graphics). Literature on face recognition is therefore vast and diverse. A single system often involves techniques motivated by different principles, which leads to a usage of a mixture of techniques that makes it difficult to classify these systems based purely on what types of techniques they use for feature representation or classification. Following psychological criteria on how humans use holistic and local features, face recognition systems can be broadly divided into:

- Holistic matching methods. These methods use the whole face region as the raw input to a recognition system. One of the most widely used representations of the face region is eigenpictures, which are based on principal component analysis (PCA) Turk and Pentland [1991]. Another example includes the use of Fisherfaces, based on Fisher's Linear Discriminant (FLD) instead of PCA Belhumeur *et al.* [1997]. PCA is the method considered in the present work.

- Feature-based (structural) matching methods. Typically, in these methods, local features such as the eyes, nose and mouth are first extracted and their locations and local statistics (geometric and/or appearance) are fed into a structural classifier. Some examples include the use of Hidden Markov Models Samaria and Fallside [1993] or Convolutional Neural Networks Lawrence *et al.* [1997].

- Hybrid methods. Just as the human perception system uses both local features and the whole face region to recognize a face, an automatic recognition system can also take advantage of both techniques. For example, the algorithm proposed in [Huang *et al.*, 2004], combines holistic and feature analysis-based approaches using a Markov Random Field (MRF) method.

Other techniques have been recently applied to face recognition, such as sparse representation [Wright *et al.*, 2009] and Local Binary Patterns (LBP) [Ahonen *et al.*, 2006]. More detailed surveys on face recognition methods may be found in [Tolba *et al.*, 2006; Zhao *et al.*, 2003].

2.3. Summary of Iris Recognition

The objective of this section is to briefly summarize those aspects of an iris recognition system which are directly related to the present study and which are essential for the correct understanding of the work. For a more comprehensive, descriptive and self-contained review on automatic iris recognition the reader is referred to [Bowyer *et al.*, 2008; Daugman, 2003, 2004, 2006, 2007, 2008].

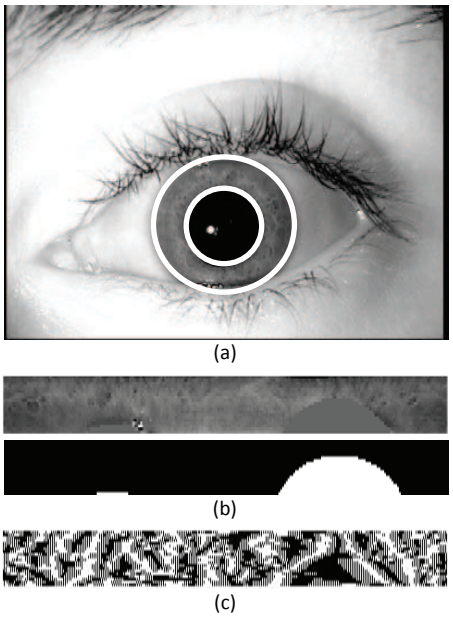

Figure 2.3: *Example illustrating the segmentation (a), the normalization and occlusion mask (b), and the encoding (c) stages used by most iris recognition systems.*

Common iris recognition systems comprise five different stages: image acquisition, iris location and segmentation, normalization, encoding and matching. As has been mentioned before, the main objective of this work is to analyse the vulnerabilities of an iris-based recognition system to the proposed indirect attack. Thus, although the acquisition and segmentation tasks may be very challenging under certain scenarios (e.g., long distance acquisition, uncontrolled lighting conditions, eye deviation, etc.) they are not relevant to this study and will not be treated here.

- **Normalization**. Once the iris has been segmented, the vast majority of iris recognition systems transform the annular-like iris pattern in cartesian coordinates to a normalized rectangular image of fixed dimensions in pseudo-polar coordinates. These are the type of images that will be reconstructed using the algorithm described in this work. The normalization process may be reversed and the normalized iris patterns can be incorporated again into the original eye images (of the same or of a different user).

- **Encoding**. Although a number of methods have been reported in this stage, most of them use some type of filtering strategy (typically based on Gabor Wavelet filters) prior to quantizing the phasor response of the filtered output resulting in a binary representation of the iris image (i.e., the iriscode).

Finally, two iriscodes are compared using a bitwise operator such as the Hamming distance. In most cases, in the segmentation stage, a mask showing the occluded areas of the iris (e.g., due to the eyelids or eyelashes) is also generated. Thus, the matching score is only computed using the "non-masked" bits of the iriscode.

In Fig. 2.3 an example of the normalization and encoding stages is shown. The original iris image appears on top (a) with the two white circles denoting the outer and inner boundaries of the segmented iris. The corresponding normalized image along with the mask indicating the occluded areas (b) and the final iriscode (c) are also shown.

Chapter 3

Proposed Methods

In this chapter we introduce the methods proposed for the security evaluation of unimodal and multimodal recognition systems described in Sect 4.2. Experimental results using these methods are detailed in Chapter 5.

Until now, only the vulnerabilities of unimodal systems to indirect attacks have been analysed. In this section we present the first algorithm for the evaluation of the vulnerabilities of multimodal systems to this type of threat. As can be observed in Fig. 3.1 (top), the input to the algorithm are the scores given by the matcher, and the output the templates to be compared to the client account.

For simplicity, the attacking methodology is described here for the particular case of a multimodal system based on the score fusion of a real valued (e.g. face) and a binary (e.g. iris) matcher. However, the proposed approach is general and may be applied with very small modifications to attack multimodal systems working on: *i*) more than two traits represented with real-valued or binary templates (by adding new blocks after the switch in Fig. 3.1), or *ii*) feature-based fusion strategies (by rearranging the template disposition).

In order to attack a multimodal biometric system where one of the biometric traits is represented with real values and the other is binary (most iris recognition systems work on binary templates), the algorithm here presented combines two sub-algorithms. Each of them attacks one segment of the template: the real-valued or the binary segment. In the following subsections, each of the individual sub-algorithms is described (Sect. 3.1 and Sect. 3.2). Finally, the multimodal attacking algorithm based on the previous two is presented (Sect. 3.3).

In Chapter 5, Sub-Algorithm 1 is used to attack a face verification system and the vulnerabilities of an iris recognition system are evaluated with Sub-Algorithm 2. Finally, the combined algorithm is used to attack a multimodal biometric system based on face and iris, whose unimodal subsystems are the previous two unimodal systems. The same protocol, described in Chapter 4, is followed in all evaluations, so that the vulnerabilities of these three systems are compared in a fair fashion.

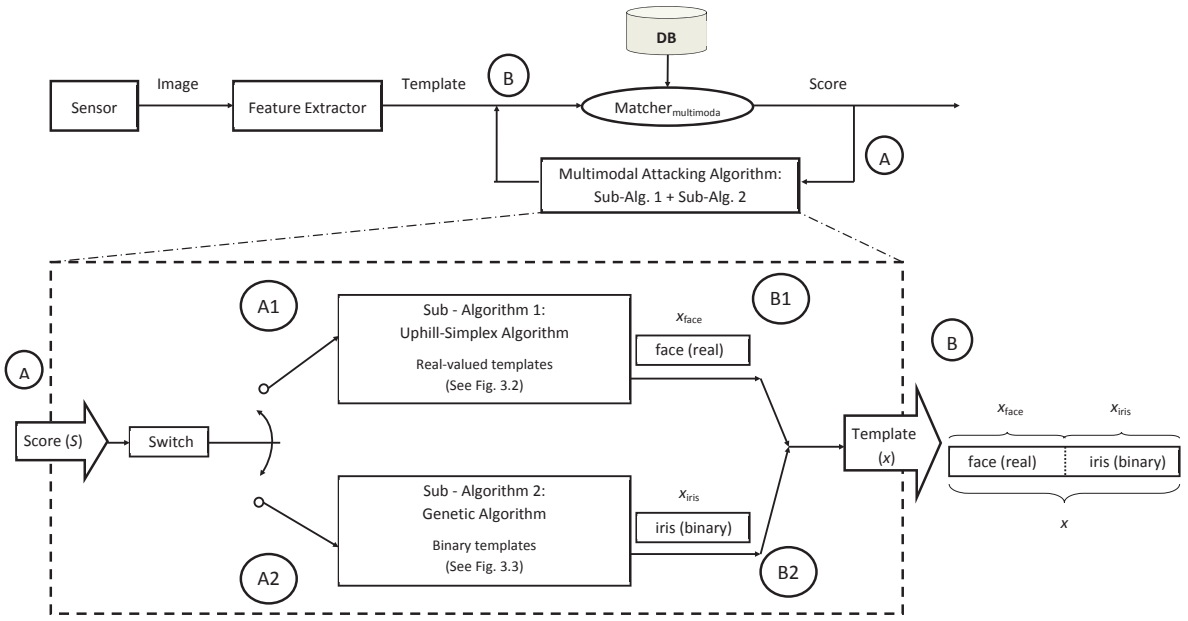

Figure 3.1: *Diagram of a general hill-climbing attack (top), with the specific modification scheme for the combined algorithm (bottom).*

3.1. Sub-Algorithm 1: Hill-Climbing based on the Uphill Simplex Algorithm

Problem statement. Consider the problem of finding a K-dimensional vector of real values x_{face} which, compared to an unknown template $\mathcal{C}_{\text{face}}$ (in our case related to a specific client), produces a similarity score bigger than a certain threshold δ_{face}, according to some matching function J_{face}, i.e., $J_{\text{face}}(\mathcal{C}_{\text{face}}, x_{\text{face}}) > \delta_{\text{face}}$. The template can be another K-dimensional vector or a generative model of K-dimensional vectors.

Assumptions. Let us assume:

- That there exists a statistical model G (K-variate Gaussian with mean μ_G and a diagonal covariance matrix $\mathbf{\Sigma}_G$, with $\sigma_G^2 = \text{diag}(\mathbf{\Sigma}_G)$), in our case related to a background set of users, overlapping to some extent with $\mathcal{C}_{\text{face}}$.

- That we have access to the evaluation of the matching function $J_{\text{face}}(\mathcal{C}_{\text{face}}, x_{\text{face}})$ for several trials of x_{face}.

Algorithm. The problem stated above can be solved by adapting the Downhill Simplex algorithm first presented in [Nelder and Mead, 1965] to maximize instead of minimize the function J_{face}. We iteratively form new simplices by reflecting one point, x_{face}^l, in the hyperplane of the remaining points, until we are close enough to the maximum of the function. The point to be reflected will always be the one with the lowest value given by the matching function, since

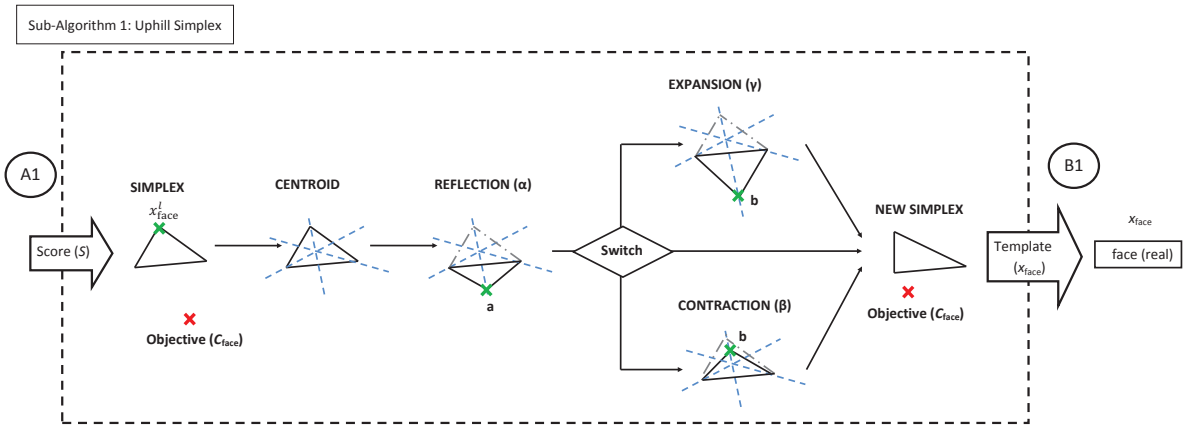

Figure 3.2: *Diagram of the modification scheme for the Sub-Algorithm 1, based on the Uphill Simplex Algorithm.*

it is in principle the one furthest from our objective. Thus, as can be observed in Fig. 3.2, the different steps followed by the sub-algorithm 1 are:

1. Compute the statistical model $G(\mu_G, \sigma_G)$ from a development pool of users.

2. Take $K+1$ samples (x^i_{face}) defining the initial simplex from the statistical model G and compute the similarity scores $J_{\text{face}}(\mathcal{C}_{\text{face}}, x^i_{\text{face}}) = s^i_{\text{face}}$, with $i = 1, \ldots, K+1$.

3. Compute the centroid $\bar{x}_{\text{face}}$ of the simplex as the average of x^i_{face}: $\bar{x}_{\text{face}} = \frac{1}{K+1} \sum_i x^i_{\text{face}}$.

4. Reflect the point x^l_{face} according to the next steps, adapted from the Downhill Simplex algorithm [Nelder and Mead, 1965]. In the following, the indices l and h are defined as $h = \arg\max_i(s^i_{\text{face}})$, $l = \arg\min_i(s^i_{\text{face}})$.

   a) **Reflection**: Given a constant $\alpha > 0$, the *reflection coefficient*, we compute:

   $$a = (1+\alpha)\bar{x}_{\text{face}} - \alpha x^l_{\text{face}}.$$

   Thus, a is on the line between x^l_{face} and $\bar{x}_{\text{face}}$ being α the ratio between the distances $[a\bar{x}_{\text{face}}]$ and $[x^l_{\text{face}}\bar{x}_{\text{face}}]$. If $s^l_{\text{face}} < s^a_{\text{face}} < s^h_{\text{face}}$ we replace x^l_{face} by a. Otherwise, we go on to step 4b.

   b) **Expansion or contraction**.

   1) **Expansion**: If $s^a_{\text{face}} > s^h_{\text{face}}$ (i.e., we have a new maximum) we expand a to b as follows:

   $$b = \gamma a + (1-\gamma)\bar{x}_{\text{face}},$$

   where $\gamma > 1$ is another constant called *expansion coefficient*, which represents the ratio between the distances $[b\bar{x}_{\text{face}}]$ and $[s\bar{x}_{\text{face}}]$. If $s^b_{\text{face}} > s^h_{\text{face}}$, we replace x^l_{face} by b. Otherwise, we have a failed expansion and replace x^l_{face} by a.

2) **Contraction**: If we have reached this step, then $s^a_{\text{face}} \leq s^l_{\text{face}}$ (i.e. replacing x^l_{face} by **a** would leave s^a_{face} as the new minimum). Afterwards we compute

$$b = \beta x^l_{\text{face}} + (1-\beta)\bar{x}_{\text{face}},$$

where $0 < \beta < 1$ is the *contraction coefficient*, defined as the ratio between the distances $[b\bar{x}_{\text{face}}]$ and $[x^l_{\text{face}}\bar{x}_{\text{face}}]$. If $s^b_{\text{face}} > \max(s^l_{\text{face}}, s^a_{\text{face}})$, then we replace x^l_{face} by b; otherwise, the contracted point is worse than x^l_{face}, and for such a failed contraction we replace all the x^i_{face}'s by $(x^i_{\text{face}} + x^h_{\text{face}})/2$.

5. With the new x^l_{face} value, update the simplex and return to step 3.

Stopping criteria. The algorithm stops when: *i*) the maximum similarity score of the simplex vertices is higher than the threshold δ_{face} (i.e., the account is broken), *ii*) the variation of the similarity scores obtained in a number of iterations is lower than a certain threshold or *iii*) a maximum number of iterations is reached.

It is important to notice for the computation of the Efficiency (defined in Sect. 4.4) of this sub-algorithm that at each iteration (except for the initial one) a maximum of 2 matchings will be performed (i.e., $s^a_{\text{face}} + s^b_{\text{face}}$). On average, the number of matchings computed per iteration will be lower than 2 and greater than 1.

The hill-climbing based on the Uphill Simplex algorithm was first presented in [Gomez-Barrero *et al.*, 2011], where it was used to successfully attack a signature verification system.

3.2. Sub-Algorithm 2: Indirect Attack based on a Genetic Algorithm

Problem statement. Consider the problem of finding an L-dimensional binary vector x_{iris} which, compared to an unknown template $\mathcal{C}_{\text{iris}}$ (in our case related to a specific client), produces a similarity score bigger than a certain threshold δ_{iris}, according to some matching function J_{iris}, i.e., $J_{\text{iris}}(\mathcal{C}_{\text{iris}}, x_{\text{iris}}) > \delta_{\text{iris}}$. The template can be another L-dimensional vector or a generative model of L-dimensional vectors.

Assumptions. Let us assume:

- That we have access to the evaluation of the matching function $J_{\text{iris}}(\mathcal{C}_{\text{iris}}, x_{\text{iris}})$ for several trials of x_{iris}.

Algorithm. The problem stated above may be solved by using a genetic algorithm, which has shown a remarkable performance in binary optimization problems [Brindle, 1981], to optimize the similarity score given by the matcher, that is, the fitness value for an individual is $s_{\text{iris}} = \mathcal{J}_{\text{iris}}(x_{\text{iris}}, \mathcal{C}_{\text{iris}})$. As can be seen in Fig. 3.3 the steps followed by the sub-algorithm 2 are:

1. Generate an initial population P_i with N individuals of length L, being L the length of the iris code.

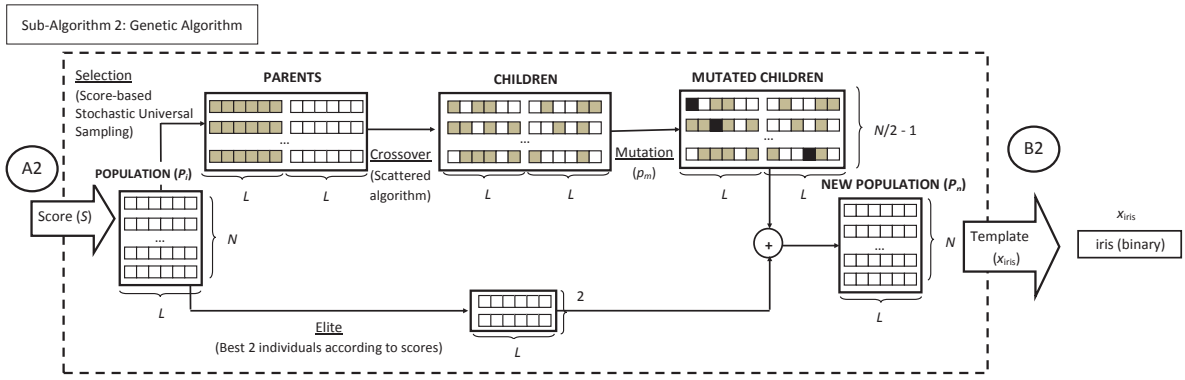

Figure 3.3: *Diagram of the modification scheme for the Sub-Algorithm 2, based on a genetic algorithm.*

2. Compute the similarity scores s^i of the individuals (x^i_{iris}) of the population P_i, $s^i_{\text{iris}} = J(x^i_{\text{iris}}, \mathcal{C}_{\text{iris}})$ with $i = 1, \ldots, N$.

3. Four rules are used at each iteration to create the next generation P_n of individuals from the current population:

   a) **Elite**: the two individuals with the maximum similarity scores are kept unaltered for the next generation.

   b) **Selection**: certain individuals, the *parents*, are chosen by stochastic universal sampling [Baker, 1987]. This way, the individuals with the highest fitness values (similarity scores) are more likely to be chosen as parents for the next generation: one subject can be selected 0 or many times. From the original N individuals, $N/2 - 1$ *fathers* and $N/2 - 1$ *mothers* are chosen.

   c) **Crossover**: parents are combined to form the $N - 2$ *children* of the next generation following a scattered crossover method. A random binary vector is created and the genes (bits) of the child are selected from the first parent where the value of the random vector is 1, and from the second when it is 0 (vice versa for the second child).

   d) **Mutation**: random changes are applied to the bit values of the new children with a mutation probability p_m.

4. Redefine $P_i = P_n$ and return to step 2.

Stopping criteria. The algorithm stops when: *i*) the best fitness score is higher than the threshold δ_{iris} (i.e., the account is broken), *ii*) the variation of the similarity scores obtained in a number of generations is lower than a previously fixed value, or *iii*) when the maximum number of generations is reached.

It is important to notice for the computation of the Efficiency (defined in Sect. 4.4) of this sub-algorithm that at each iteration (i.e., generation) N matchings are performed (one for each of the members of the population).

3.3. Multimodal Attack: Combination of Sub-Algorithms 1 (Uphill-Simplex) and 2 (Genetic-Algorithm)

Problem statement. Consider the problem of finding a $(K + L)$-dimensional vector x of real and binary values which, compared to an unknown template $\mathcal{C}$ (in our case related to a specific client), produces a similarity score bigger than a certain threshold δ, according to some matching function J, i.e., $J(\mathcal{C}, x) > \delta$. The template can be another $(K+L)$-dimensional vector or a generative model of $(K + L)$-dimensional vectors.

Assumptions. Let us assume:

- That we know the distribution of the two subtemplates (real-valued x_{face} and binary x_{iris}) within the multimodal template x.
- That we have access to the evaluation of the matching function $J(\mathcal{C}, x)$ for several trials of x.

Algorithm. The problem stated above may be solved by dividing the template x into its real-valued (x_{face}) and binary parts (x_{iris}) and alternately optimize each of them as can be seen in Fig. 3.1. In order to optimize each of the parts, the algorithms described in the previous subsections are used: the Sub-Algorithm 1 for the real-valued segment (face) and the Sub-Algorithm 2 for the binary segment (iris). Thus, the steps followed are:

1. Generate a synthetic template (x) randomly initializing the real-valued (x_{face}) and binary (x_{iris}) segments, and compute the similarity score $S = J(\mathcal{C}, x)$, which will be used as optimization criterion.
2. Leaving one of the segments unaltered, optimize the other segment of the template using the appropriate sub-algorithm until one of the stopping criteria of the sub-algorithm is fulfilled.
3. Change the optimization target to the segment which was previously left unaltered and go back to step 2.

Stopping criteria. The algorithm stops when: *i*) the verification threshold is reached (i.e., access to the system is granted) or *ii*) the total number of iterations (i.e., changes between the optimized segments) exceeds a previously fixed value (i.e., the attack has failed).

As will be analysed in the experimental section this algorithm may present different results depending on whether it starts attacking the real-valued or binary part of the template.

It is very important to notice that the multimodal attacking algorithm does not have access at any point to the partial scores of the unimodal modules (s_{face} and s_{iris}) but only uses the final fused score given by the system (S). This way, in the description of the previous two sub-algorithms s_{face} ad s_{iris} should be changed by S when they are used as part of the multimodal attack and not individually.

Chapter 4

Experimental Framework

The experimental framework followed throughout the present work is described in this Chapter: firstly, the database used in the experiments is presented (Sect. 4.1); the recognition systems attacked are then described (Sect. 4.2); a performance evaluation of those systems is afterwards carried out (Sect. 4.3); and finally the vulnerability evaluation metrics are presented (Sect. 4.4).

The results of these experiments are latter presented in Chapter 5.

The standard experimental protocol applied in other vulnerabilities evaluations and followed in the present work comprises two different steps, namely: *i*) performance evaluation (Sect. 4.3), and *ii*) security evaluation (Sect. 4.4). The first step allows us to establish the operating point at which the system works in terms of the False Acceptance Rate, FAR, or percentage of impostors accepted as genuine users, and the False Rejection Rate, FRR, or percentage of genuine users rejected by the system. Once the operating point is fixed, in the second step we are able to evaluate the attacking algorithm.

The reason why the two aforementioned stages of the experimental protocol must be carried out is the dependence of the performance of the attacking schemes on the operating point of the system considered (FAR-FRR). In general, the higher the FAR of the system under attack, the higher the success chances of a given attack: a higher FAR means a higher probability of accepting an impostor. Therefore, several operating points should be tested and specified in the experimental framework, in order to compare these results with future studies.

4.1. Database

The experiments are carried out on the face and iris subcorpora included in the Desktop Dataset of the multimodal BioSecure database [Ortega-Garcia *et al.*, 2010], which comprises voice, fingerprints, face, iris, signature and hand of 210 users, captured in two time-spaced acquisition sessions. This database was acquired thanks to the joint effort of 11 European institutions and has become one of the standard benchmarks for biometric performance and security

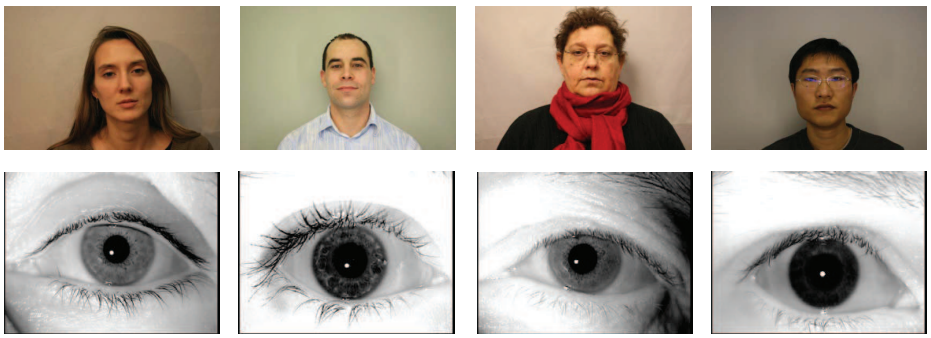

Figure 4.1: *Typical samples of the face and iris images available in the Desktop Dataset of the multimodal BioSecure database.*

evaluations [Mayoue *et al.*, 2009]. It is publicly available through the BioSecure Foundation[1].

The database comprises three datasets captured under different acquisition scenarios, namely: *i*) Internet Dataset (DS1, captured through the Internet in an unsupervised setup), *ii*) Desktop Dataset (DS2, captured in an office-like environment with human supervision), and *iii*) the Mobile Dataset (DS3, acquired on mobile devices with uncontrolled conditions). The face subset used in this work includes four frontal images (two per session) with an homogeneous grey background, and captured with a reflex digital camera without flash ($210 \times 4 = 840$ face samples), while the iris subset includes four grey-scale images (two per session as well) per eye, all captured with the Iris Access EOU3000 sensor from LG. In the experiments only the right eye of each user has been considered, leading this way as in the face case to $210 \times 4 = 840$ iris samples.

Some samples of the iris and face images contained in the database are shown in Fig 4.1.

4.2. Recognition Systems Attacked

4.2.1. Face Recognition System

The hill-climbing attack based on the uphill-simplex algorithm described in Sect. 3.1 is used to evaluate the security of an Eigenface-based face verification system [Turk and Pentland, 1991]. This technique uses Principal Component Analysis (PCA) to derive a vector which represents the face images in a lower dimensional space, and it was used to present initial face verification results for the recent Face Recognition Grand Challenge [Phillips *et al.*, 2005].

The evaluated system uses cropped face images of size 64×80 to train a PCA vector space where 80% of the variance is retained. This leads to a system where the original image space of 5120 dimensions is reduced to 100 dimensions (or eigenvectors). Similarity scores are computed in this PCA vector space using the Euclidean distance, as it showed a very competitive performance compared to the rest of the similarity measures tested.

[1] http://biosecure.it-sudparis.eu/AB

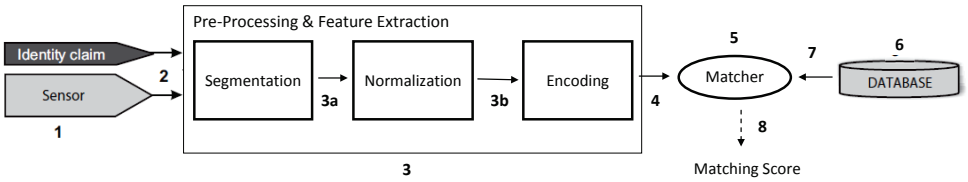

Figure 4.2: *Architecture of an automated iris verification system. Possible attack points are numbered from 1 to 8.*

The experiments are carried out on the face subcorpus included in the Desktop Dataset of the BioSecure multimodal database [Ortega-Garcia *et al.*, 2010]. The performance of the evaluated system is computed using the experimental protocol shown in Fig. 4.4. The database is divided into: *i*) a training set comprising the first three samples of 170 clients (used to compute both the PCA transformation matrix and the enrolment templates), and *ii*) an evaluation set formed by the fourth image of the previous 170 users (used to compute the genuine scores), and all the 4 images of the remaining 40 users with which the impostor scores are calculated. As a result of using the same subjects for PCA training and client enrolment, the system performance is optimistically biased, and therefore harder to attack than in a practical situation (in which the enrolled clients may not have been used for PCA training). This means that the results presented in this work are a conservative estimate of the attack's success rate.

The final score given by the system is the average of the scores obtained after matching the input vector to the three templates of the attacked client model $\mathcal{C}$.

4.2.2. Iris Recognition System

In our experiments, we have used a modified version of the iris recognition system developed by Masek and Kovesi [2003] [1], which is widely used in many iris related publications. As depicted in Fig. 4.2, the system comprises four different steps:

- **Segmentation**: the method proposed in [Ruiz-Albacete *et al.*, 2008] is followed: the system uses a circular Hough transform in order to detect the iris and pupil boundaries, which are modelled as two circles.
- **Normalization**: a technique based on Daugman's rubber sheet model [Daugman, 2004] is used, mapping the segmented iris region into a 2D array.
- **Feature encoding**: the normalized iris pattern is convolved with 1D Log-Gabor wavelets. The encoding process produces a binary template of $20 \times 480 = 9,600$ bits and a corresponding noise mask that represents the eyelids areas.
- **Matching**: the inverse of the Hamming distance is used for matching. It is modified so that it incorporates the noise mask, using only the significant bits. This modified Hamming

[1] The source can be freely downloaded from www.csse.uwa.edu.au/ pk/studentprojects/libor/sourcecode.html

distance is given by the formula

$$HD = \frac{\sum_{j=1}^{L} X_j(XOR)Y_j(AND)\bar{X}n_j(AND)\bar{Y}n_j}{L - \sum_{k=1}^{L} Xn_k(OR)Yn_k}$$

where X_j and Y_j are the two bitwise templates to compare, Xn_j and Yn_j are the corresponding noise masks for X_j and Y_j, and L is the number of bits comprised by each template. $\bar{X}n_j$ denotes the logical not operation applied to Xn_j. A number of Hamming distance values are calculated from successive shifts [Daugman, 2004], correcting this way for misalignments in the normalized iris pattern caused by rotational differences during imaging, being the lowest value finally taken.

For the experiments, the images that were not successfully segmented by the recognition system (3.04% of the 1,680 images available) were segmented manually, allowing us this way to use all of the available dataset. Furthermore, by doing this manual aided segmentation the system performance is optimistically biased and therefore harder to attack than in a practical situation (where the segmentation would be fully automatic).

The iris subset of the BioSecure DS2 [Ortega-Garcia *et al.*, 2010] used in this work includes four grey-scale images (two per session) per eye, all captured with the Iris Access EOU3000 sensor from LG. As before, the system performance evaluation protocol followed is the one depicted in Fig. 4.4.

The final score given by the system is the average of the scores obtained after matching the input binary vector to the three templates (i.e., iris codes) of the attacked client model $\mathcal{C}$. Furthermore, we run two different sets of experiments: *i)* we consider the left and right eyes of one person as different clients, thus having twice as many clients (340) and impostors (80); *ii)* we take the mean of the scores given by the matcher for each eye of the client as final score (multimodal system considering both eyes for verification).

4.2.3. Multimodal Recognition System based on Face and Iris

The multimodal verification system evaluated in this work is the fusion of the two unimodal systems previously analysed, namely: *i*) a modified version of the iris recognition system developed by L. Masek[1] [Masek and Kovesi, 2003], which is widely used in many iris related publications; and *ii*) an Eigenface-based face verification system [Turk and Pentland, 1991], used to present initial face verification results for the recent Face Recognition Grand Challenge [Phillips *et al.*, 2005].

Given an input vector x, the system performs the following tasks in order to obtain the final score, S, as can be seen in Fig. 4.3:

1. Compute the similarity scores obtained by the face (s_{face}) and iris (s_{iris}) traits, as given by the matchers described in Sect. 4.2.1 and Sect. 4.2.2.

[1]The source can be freely downloaded from www.csse.uwa.edu.au/ pk/studentprojects/libor/sourcecode.html

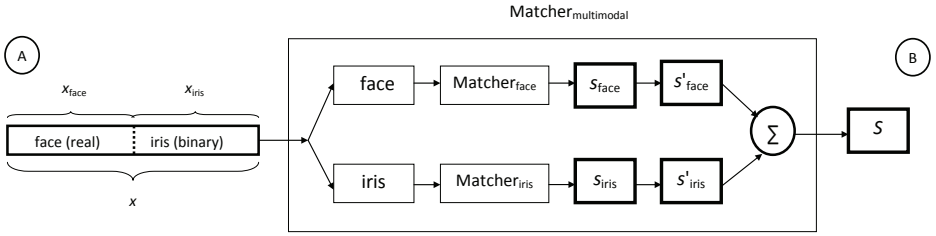

Figure 4.3: *Similarity score obtained from one multimodal template (x) consisting of two different segments, containing: face features (x_{face}, real values) and the iris code (x_{iris}, binary). The unimodal verification subsystems give the corresponding scores (s_{face}, s_{iris}), which are then normalised (s'_{face}, s'_{iris}) and fused to obtain the final output of the global system: S.*

		BioSecure DS2 DB (210 Users)	
Session	Sample	170 Users	40 Users
1	1	Training	Test (Impostors)
	2		
2	1		
	2	Test (Clients)	

Figure 4.4: *Partition of the BioSecure DS2 DB according to the performance evaluation protocol defined.*

2. Normalize the scores s_k, with $k = \{\text{face}, \text{iris}\}$, using hyperbolic tangent estimators (its robustness and high efficiency are proven in [Jain *et al.*, 2005]):

$$s'_k = \frac{1}{2}\left\{\tanh\left(0.01\frac{s_k - \mu}{\sigma}\right) + 1\right\}$$

   where s_k is the original similarity score obtained by the iris (respectively face) section of the template, μ and σ the mean and standard deviation of the scores distribution of the iris (respectively face), and s'_k the normalised score. This way, both partial scores (face and iris) lie in the interval $[0, 1]$.

3. Finally, both normalised scores are fused with a sum, given the very good results that this fusion rule has presented even when compared with more sophisticated methods like decision trees [Ross and Jain, 2003] or neural networks [Wang *et al.*, 2003]:

$$S = s'_{\text{iris}} + s'_{\text{face}}$$

4.3. Performance Evaluation

The performance of the evaluated system is computed using the experimental protocol shown in Fig. 4.4. The database is divided into: *i*) a training set comprising the first three samples of 170 clients, used as enrolment templates; and *ii*) an evaluation set formed by the fourth image

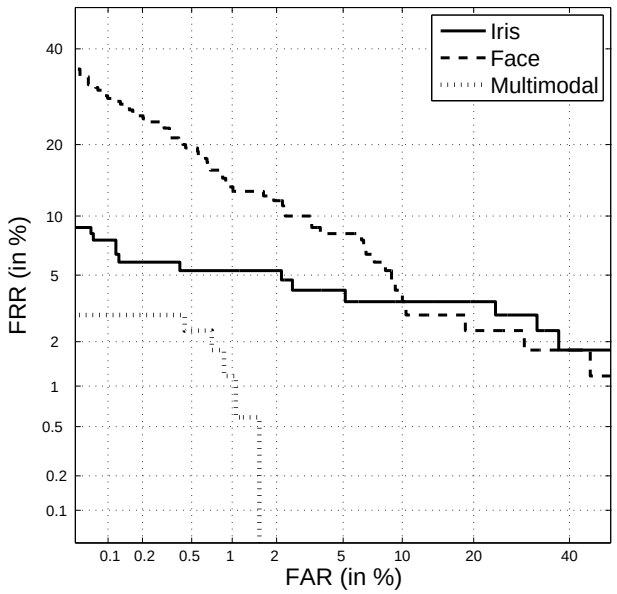

Figure 4.5: *DET curves of the three systems .*

	EER (%)		
	Face	Iris	Multimodal
Before Norm.	6.55	4.11	-
After Norm.	6.61	4.04	0.83

Table 4.1: *EER of the unimodal and multimodal systems, based on face and iris, before and after the normalization of the scores.*

of the previous 170 users (used to compute the genuine scores), and all the 4 images of the remaining 40 users (used to compute the impostor scores). The three operating points where the hill-climbing algorithms are evaluated (corresponding to FAR = 0.1%, FAR = 0.05%, and FAR = 0.01%) correspond to a low, medium, and high security application according to [ANSI, 2001].

In Fig. 4.5 the Detection Error Trade-off (DET) curves of the three systems considered using the described protocol are depicted. The genuine scores are computed using the fourth sample of each of the 170 users enrolled (170 genuine scores). In order to compute the impostor scores, the first sample of the last 40 users is compared to each user model ($40 \times 170 = 6,800$ impostor scores).

Finally, the performance of each system is measured with the Equal Error Rate, or operating point at which FAR = FRR, in Table 4.1. As may be observed, the EER of the multimodal system (0.83%) is considerably lower that those of the unimodal systems fused (and, therefore,

the system performance is higher): while the EERs of the unimodal systems were higher than 4% and 6%, respectively, the multimodal system shows an EER lower than 1%.

4.4. Vulnerabilities Evaluation

In order to generate the user accounts to be attacked with the proposed hill-climbing algorithms, we used the train set defined in the performance evaluation protocol (i.e., three first samples of 170 users as shown in Fig. 4.4). The performance of the attack will be evaluated in terms of the Success Rate and Efficiency, defined as:

- **Success Rate (SR)**: it is the expected probability that the attack breaks a given account. It is computed as the ratio between the number of broken accounts (A_B) and the total number of accounts attacked ($A_T = 170$): SR $= A_B/A_T$. This parameter indicates how dangerous the attack is: the higher the SR, the bigger the threat.
- **Efficiency (Eff)**: it is computed as the inverse of the average number of matchings needed by the attack to break an account. It is defined as Eff $= 1/\left(\sum_{i=1}^{A_B} n_i/A_B\right)$, where n_i is the number of matchings computed to bypass each of the broken accounts. This parameter gives an estimation of how easy it is for the attack to break into the system in terms of speed: the higher the Eff, the faster the attack.

Chapter 5

Experimental Results

The experimental results of this thesis are covered in this chapter. Here, we have carried out the security evaluations of two unimodal systems based on face (Sect. 5.1) and iris (Sect. 5.2), and the fusion of both systems (Sect. 5.3). To that end, the attacking schemes described in Chapter 3 are used, following the protocol presented in Chapter 4, so that a fair comparison may be established.

5.1. Case Study I: Vulnerability Assessment of a Face Recognition System

The goal of these experiments is twofold, namely: *i*) on the one hand, study the vulnerability of an automatic face recognition system to the proposed hill-climbing algorithm, and *ii*) on the other hand, prove the efficiency of the attack against a biometric system based on a different biometric trait (it was already used to successfully attack an on-line signature verification system in [Gomez-Barrero *et al.*, 2011]).

The performance of the attack at different operating point is studied in the first set of experiments. Score Quantization is afterwards analysed as a possible countermeasure: its impact on the SR and the Eff of the attack is studied.

5.1.1. Vulnerability Evaluation of Different Operating Points

The performance of the attack is tested at three different operating points, namely: *i)* FAR = 0.10%, *ii)* FAR = 0.05%, *iii)* FAR = 0.01%, which correspond to a low, medium and high security application according to [ANSI, 2001]. Furthermore, these are the same operating points at which a very similar face verification system was evaluated in [Galbally *et al.*, 2010]. Therefore, the results obtained in both works may be compared in a fair fashion. The results of the experiments are detailed in Table 5.1. As can be observed, the algorithm presented here successfully breaks all the attacked accounts, contrary to the Bayesian hill-climbing algorithm described in [Galbally *et al.*, 2010], existing a significant SR difference between the two approaches at the last

FAR	Uphill-simplex HC		Bayesian HC	
	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)
0.10%	100%	22.124	99.0%	11.905
0.05%	100%	22.472	98.5%	9.364
0.01%	100%	21.930	86.0%	2.226

Table 5.1: *Eff and SR at the operating points tested, compared to those obtained by the Bayesian hill-climbing attack in [Galbally* et al.*, 2010].*

FAR	Face Verif. System		Signature Verif. Syst.	
	SR	Eff ($\times 10^{-3}$)	SR	Eff ($\times 10^{-3}$)
0.05%	100%	2.2472	91.32%	0.8489
0.01%	100%	2.1930	88.43%	0.7391

Table 5.2: *Eff and SR at the operating points tested, compared to those obtained with the on-line signature verification system tested in [Gomez-Barrero* et al.*, 2011].*

operating point (i.e., 100% vs 86%). Moreover, while the efficiency decreases substantially along the three operating points for the previous algorithm, for the attack proposed in the present work it remains almost invariant, regardless of the operating point considered. This leads to an efficiency which is ten times faster at the last operating point between the two attacks (i.e., 456 scores needed to break an account against almost 4,500).

It is also worth noting that the parameters of the algorithm are the same ones which were optimized to attack an on-line signature verification system in [Gomez-Barrero *et al.*, 2011]: $\alpha = 1.1, \beta = 0.8, \gamma = 1.1$. In that work, we performed three successive steps fixing in each of them two of the parameters and sweeping the other in a given range. According to the original Downhill Simplex algorithm [Nelder and Mead, 1965], the best values for the parameters are $\alpha = 1$, $\gamma = 2$ and $\beta = 0.5$. Thus, the selected ranges were centred on those values, taking always into account the constraints explained in Sect. 3.1, namely: $\alpha > 0$, $\gamma > 1$ and $0 < \beta < 1$. Finally, the parameters values (that will be used in the experiments in the present research work) were set to $[\alpha, \gamma, \beta] = [1.1, 1.1, 0.8]$.

In Table 5.2 we show the Eff as well as the SR of the two common operating points (in terms of FAR) attacked in both works. The fact that the SR and the Eff improve for the present case study proves the robustness of the algorithm: it is able to break totally heterogeneous systems working on different biometric traits, even improving its performance, without specifically adjusting its parameters.

It should also be emphasized that in the present work the hill-climbing attack is initialized from a normal distribution of zero mean and unit variance, that is, the first simplex is generated without needing any training faces.

In Fig. 5.1 an example of the execution of the attack is shown. The evolution of the score

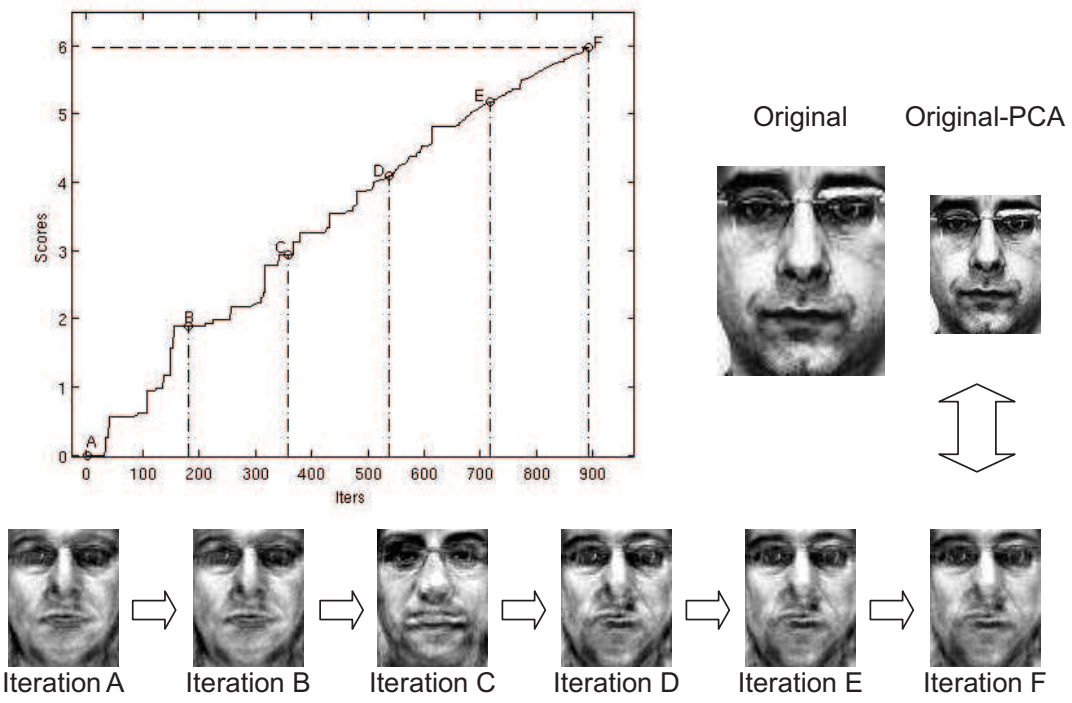

Figure 5.1: *Example of the evolution of the score and the synthetic eigenfaces through the iterations of the attack for a broken account. The dashed line represents the objective threshold.*

through the iterations of the algorithm is depicted together with six points (including the first and the last one) of the iterative process (marked with letters A to F). The dashed line represents the objective value to be reached (i.e, the threshold δ). The two upper faces correspond to one of the original images of the attacked user and its representation in the PCA space. The sequence of the six faces below correspond to the feature vector that produced each of the six scores marked with A to F. We can observe how the hill-climbing algorithm is able to evolve from a totally random face (A) to an image (F) which is very similar to the target user (labelled as "Original-PCA").

This figure shows that the algorithm not only is able to break the system, but it is also capable of reconstructing in a fairly precise way the face of the user, arising this way privacy concerns.

5.1.2. Countermeasure: Score Quantization

The results achieved by the hill-climbing attack based on the uphill-simplex algorithm to the face recognition system considered in the experiments have shown its high vulnerability to this type of attacking approach and the need to incorporate some attack protection method that increases its robustness to this threat.

When a countermeasure is introduced in a biometric system to reduce the risk of a particular attack, it should be statistically evaluated considering two main parameters:

QS	10^{-6}	10^{-4}	10^{-2}	10^{-1}	1	2.5	5
PI (%)	9.83	9.79	7.62	2.22	0.25	0.01	0.01
EER (%)	4.85	4.85	4.85	4.87	4.90	5.85	9.12

Table 5.3: *Percentage of the iterations of the hill-climbing attack with a positive score increase (PI), and EER of the system for different quantization steps (QS) of the matching score.*

QS	10^{-6}	10^{-1}	2.5
SR	97.65%	78.24%	61.76%
Eff ($\times 10^{-3}$)	2.0828	2.020	1.3158

Table 5.4: *Performance (in terms of SR and Eff) of the hill-climbing attack against the system for different quantization steps (QS).*

- Impact of the countermeasure in the system performance. The inclusion of a particular protection scheme might change the FAR and FRR of a system, and these changes should be evaluated and reported (other performance indicators such as speed or computational efficiency might also change, but are not considered here).
- Performance of the countermeasure, i.e. impact of the countermeasure in the SR and Eff of the attack.

It is often argued that a simple account lockout policy (i.e., blocking the user accounts after a number of consecutive unsuccessful access attempts) would be enough to prevent an attack such as the one proposed in the present work. However, such countermeasures still leave the system vulnerable to a spyware-based attack that interlaces its false attempts with the attempts by genuine users (successful attempts) and collects information over a period of time (i.e. piggyback attack). Furthermore, it may be used by the attacker to perform an account lockout attack (i.e., the intruder tries to illegally access a great amount of accounts blocking all of them and collapsing the system).

In this scenario, a specific design of the matching algorithm can also be implemented in order to reduce the effects of this type of threats, providing this way an additional level of security through a biometric-based protection scheme complementary to other possible non-biometric countermeasures.

Among the biometric-based approaches to reduce the effects of hill-climbing attacks, score quantization has been proposed as an effective countermeasure Adler [2004]. In fact, the BioApi Consortium BioAPI [2009] recommends that biometric algorithms emit only quantized matching scores. Such quantization means that small changes in the randomly generated templates will normally not result in a modification of the matching score, so that the attack does not have the necessary feedback from the system to be carried out successfully.

With this precedents, in this section we analyse the performance of score quantization as a possible countermeasure against the proposed attack. We will consider the Eigenface-based system operating at medium security operating point (FAR = 0.05%). For the hill-climbing attack we will assume the same configuration used in the vulnerability assessment experiments, $[\alpha, \beta, \gamma] = [1.1, 0.8, 1.1]$ (which, as mentioned before, is taken from [Gomez-Barrero *et al.*, 2011]).

In order to select the appropriate quantization step according to the trade-off that should be met in terms of its impact on the system performance (ideally as small as possible) and on the attack performance (as big as possible), several Quantization Steps (QS) are tested in terms of their corresponding Positive Increment, PI (i.e., percentage of iterations that produced an increase in the similarity score higher than the quantization step considered). The Equal Error Rate (EER) is the point at which FAR = FRR, and gives a measure of the system performance: the lower the EER, the better the performance.

From results shown in Table 5.3 we can see that for the last QS considered (5) the EER suffers a big increase (QS is too big), while for the previous QS values the system performance is not affected. Therefore, the hill-climbing attack is repeated considering the three QS values QS $= 10^{-6}$, QS $= 10^{-1}$, and QS = 2.5 . Results are presented in Table 5.4, where we can see that score quantization reduces the success chances of the attack (for bigger QS, the SR decreases). However, it can also be noticed that the attacking algorithm is quite robust to this type of countermeasure, as even for the biggest value of QS (increasing it would imply a deterioration of the system EER as shown in Table 5.3), the SR of the attack is still over 60%.

5.2. Case Study II: Vulnerability Assessment of an Iris Recognition System

The goal of these experiments is twofold, namely: *i*) study the vulnerability of an automatic iris recognition system against the proposed attack, and *ii*) find the most consistent bits in the iris code and analyse whether that information can increase the robustness of the system.

The performance of the attack at different operating point is studied in the first set of experiments. The impact of using only the most consistent bits of the iriscode for verification on the SR and the Eff of the attack is then studied.

5.2.1. Vulnerability Evaluation of Different Operating Points

The performance of the attack is measured at four different operating points, namely: *i)* FAR = 0.10%, *ii)* FAR = 0.05%, *iii)* FAR = 0.01%, and *iv)* FAR $\ll$ 0.01%, representing a very high security point. As can be observed in Table 5.5, where the results of the experiments are detailed, the attacking algorithm proposed in this work successfully breaks most of the attacked accounts (around 80% SR on average for all the scenarios considered). Moreover, the number of comparisons needed increases only about 25% between the operating points FAR = 0.1% and FAR = 0.01% (while a brute force attack using randomly chosen real irises to access the system

FAR	One iris		Both irises	
	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)
0.10%	91.18%	1.400	98.24%	0.736
0.05%	80.89%	1.255	95.29%	0.686
0.01%	62.36%	1.102	83.53%	0.603
$\ll$0.01%	52.06%	1.051	59.41%	0.506

Table 5.5: *Eff and SR of the attack at the operating points tested, for the system employing only one eye and both eyes of the user for the verification of the identity claim.*

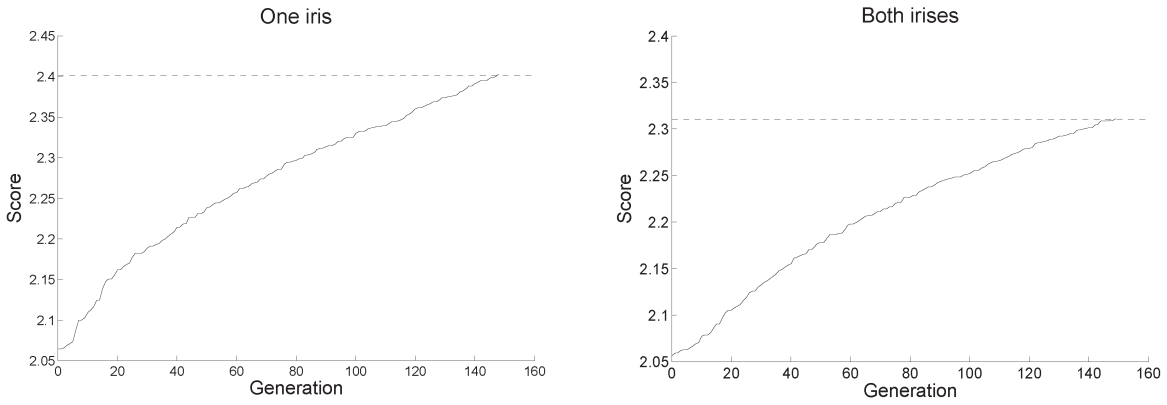

Figure 5.2: *Evolution of the maximum score (s_{max}) reached for every generation of the genetic algorithm for two different broken accounts in the two scenarios studied: one (left) and both irises (right). The verification threshold is marked with a horizontal dashed line.*

would need about 1/FAR $\simeq$ ten times as many matchings).

On the other hand, when comparing the system for the two scenarios considering one and both eyes of a client for verification, an improvement in terms of SR can be observed. The main reason for this is that with larger individuals (the number of bits is multiplied by two in the scenario which takes into account both eyes) and a population twice as big, more diversity can be achieved. Therefore, the algorithm converges to a global rather than to a local minimum of the fitness function (i.e., the score given by the matcher). The difference in terms of efficiency between the two scenarios is due to the population sizes: for one eye, only 50 individuals are computed, while for both eyes, 100 individuals are created in order to have enough diversity in the population. Having twice as many individuals implies that the number of matchings computed per generation is double, and hence the efficiency is divided by two.

Also worth noting the fact that the efficiency does not depend greatly on the operating point attacked. This is confirmed by the experiment run at an operating point of FAR $\ll$ 0.01% (Table 5.5), where the efficiency has only decreased in about 30% compared to the point with FAR = 0.1% (an eventual brute force attack would need on average over 100 times more matchings).

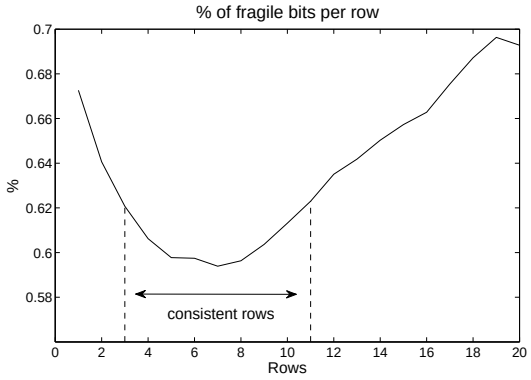

Figure 5.3: *Percentage of fragile bits (the ones flipping at least once across the images of an iris) in a row.*

Finally, in Fig. 5.2 the evolution of the maximum score obtained by the best individual of each generation for two successful attacks, against the scenarios considering one and both irises for verification, are depicted. The verification threshold, where access is granted to the system, is marked with a horizontal dashed line. As we can see in both cases, the matching score steadily increases through the generations (i.e., in each generation the best individual of the population is more similar to the client being attacked) until the positive recognition value is reached.

5.2.2. Countermeasure: Most Consistent Bits in the Iris Code

According to the analysis made by Hollingsworth *et al.* in [Hollingsworth *et al.*, 2009], there are some bits more fragile than others in an iris code, that is, bits that flip between 0 and 1 in different images of the same iris with a high probability. Here we consider that a bit is consistent, (i.e., not fragile), when it does not flip in any of the four images available for each user. In order to determine the most consistent rows of bits in the iris code, we follow the method described in [Hollingsworth *et al.*, 2009]: we compute the frequency (that must lie between 0% and 50%) that each unmasked bit flips, and take the average frequency across all bits in a row for each subject. All the codes of each user are previously aligned, keeping the rotation that gives the minimum Hamming distance to the first code of that user. In Fig. 5.3, the mean percentage of bits considered fragile in each row across all users is depicted. As can be observed, rows 3 to 11 are the more consistent ones, having the lower percentages of fragile bits.

Based on these results and the ones in [Hollingsworth *et al.*, 2009], where the more consistent rows were found to be 5 to 12, we run some experiments testing the impact of reducing the number of rows of the iris codes: from using all rows (1 - 20) to only the best ones found in [Hollingsworth *et al.*, 2009] (5 - 12). The results, all obtained at an operating point of FAR = 0.05%, can be observed in Fig. 5.4. While the Hamming distance between iris codes is lower for a lower number of rows used in order to bypass the system, the Equal Error Rate (EER) does not vary greatly (see Table. 5.6) and as can be observed the attack improves its performance, both in terms of Eff and SR. The main reason for this is that, by decreasing the

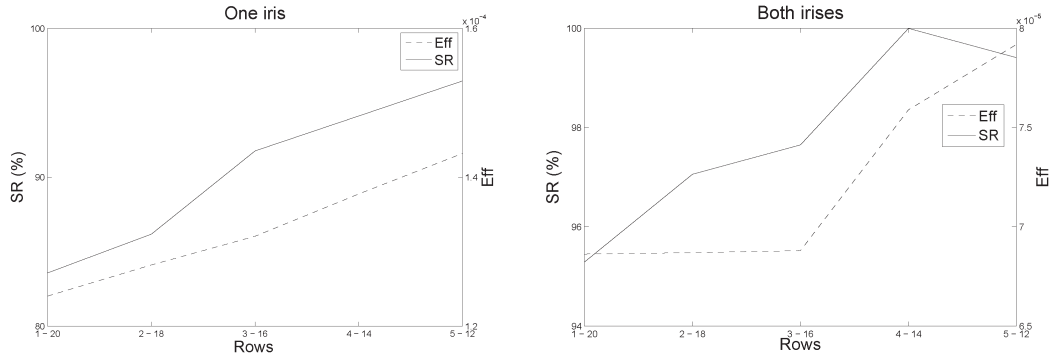

Figure 5.4: *SR and Eff of the attack varying the number of rows used by the matcher, for a system comprising only one eye per client (left) and both eyes (right).*

	EER				
Rows	1 - 20	2 - 18	3 - 16	4 - 14	5 - 12
One iris	3.82%	3.23%	3.56%	4.41%	5.00%
Both irises	0.58%	0.59%	0.59%	0.60%	1.17%

Table 5.6: *EER for the system tested under the two scenarios considered (one and both eyes), and for a decreasing number of rows of the iris code (from the least to the most consistent according to [Hollingsworth* et al.*, 2009]).*

number of rows compared, the number of bits drops drastically while the number of individuals in the population remains the same, thus increasing the diversity of the population and thereby enabling the genetic algorithm to find a maximum faster.

Finally, in Table 5.7 the performance of the attack against the system using only the most consistent bits found in the present work (rows 3 to 11) is shown and compared to the performance using all the bits. The EER for the system is 5.00% for only one eye and 0.69% for both eyes, again similar to the EER of the original system. However, the SR is as high as 98% and 100%, needing also less matchings on average. Thus, we may conclude that using only the most consistent bits in the iris code does not improve the robustness of the system to the proposed attacking algorithm.

Rows	One eye		Both eyes	
	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)
1 - 20	80.89%	1.255	95.29%	0.586
3 - 11	98.53%	1.684	100%	0.690

Table 5.7: *Eff and SR at the operating point FAR = 0.05%, for the systems employing only one eye of the user and both eyes for the verification of the identity claim, and all rows of the iris code or only the most consistent ones.*

FAR	Starts face		Starts iris	
	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)
0.10%	100%	1.9372	100%	1.4180
0.05%	100%	1.8218	100%	1.3585
0.01%	100%	1.3702	100%	1.0532

Table 5.8: *Eff and SR of the attack at the operating points tested, for attack starting with either the face or the iris part of the template.*

FAR	Unimodal Attacks				Multimodal Attack			
	Sub-Alg. 1 vs Face		Sub-Alg. 2 vs Iris		Starts Face		Starts Iris	
	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)
0.10%	100%	22.472	91.18%	1.400	100%	1.9372	100%	1.4180
0.05%	100%	22.124	80.89%	1.255	100%	1.8218	100%	1.3585
0.01%	100%	21.930	62.36%	1.102	100%	1.3702	100%	1.1112

Table 5.9: *Eff and SR for the Sub-Algorithm 1 (Uphill-Simplex) and Sub-Algorithm 2 (Genetic Algorithm) attacks carried out against the corresponding unimodal systems, and for the Multimodal Attack against the multimodal system.*

5.3. Case Study III: Vulnerability Assessment of a Multimodal Recognition System based on Face and Iris

The main goal of this first study of the vulnerabilities of a multimodal system to an indirect attack is to prove whether the combination of two different and independent biometric traits increases the security offered by the system as well as its performance.

The first set of experiments evaluates the performance of the attack on different operating points (security evaluation). Afterwards, score quantization is analysed as a possible countermeasure, studying its impact on the SR and the Eff of the attacking scheme.

5.3.1. Vulnerability Evaluation of Different Operating Points

As it was previously stated, the multimodal system is evaluated at three different operating points. Two set of results are shown in Table 5.8: the attack can start working on the part of the template corresponding to the face (it starts with the proposed hill-climbing, randomly initializing the iris part) or on the iris part of the template (it starts with the genetic algorithm, randomly initializing the face part). As it may be observed, the SR is in both cases as high as 100% at all operating points. However, the Eff of the algorithm decreases about 25% when it starts with the genetic algorithm. This is due to the fact that the hill-climbing needs much less comparisons that the GA (see Tables 5.1 and 5.5). Additionally, when the attack starts

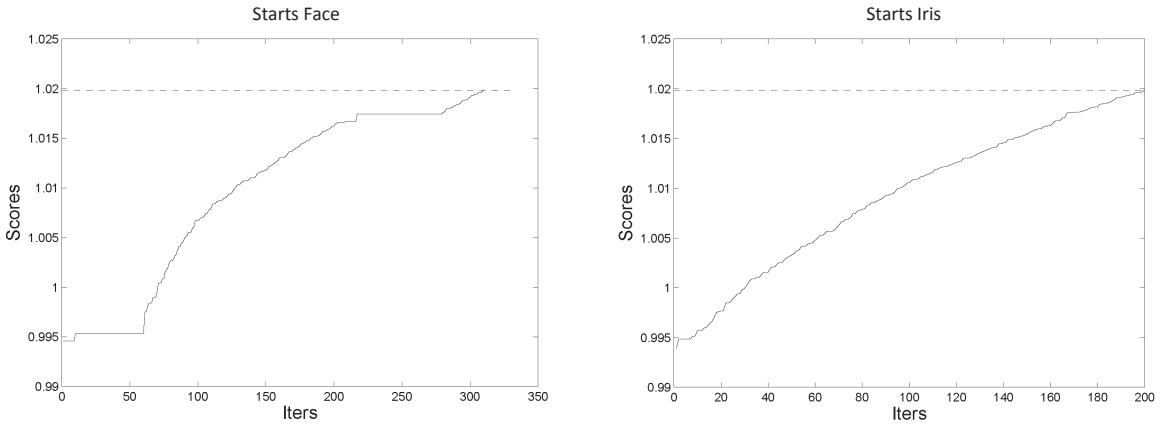

Figure 5.5: *Evolution of the maximum score (s_{max}) obtained in each generation by the genetic algorithm or in each iteration by the hill-climbing under the studied scenarios for two different broken accounts: starting with the face part of the template (left) or with the iris part (right). The verification threshold is represented by a dashed horizontal line.*

with the iris part of the template, between a 40% (at the operating point corresponding to FAR = 0.01%) and 60% (at the operating point corresponding to FAR = 0.1%) of the accounts are broken by modifying only the iris part of the template and leaving unaltered the initial random face part. On the other hand, when the attack starts with the face part, both segments of the template have to be modified for all the attacked accounts.

In Table 5.9, the results achieved when starting with the face part of the template are compared to those achieved by the unimodal systems. We can observe that the most robust system both in term of SR and Eff is the unimodal iris-based. The weakest one is the face-based unimodal system: even though the SR is 100% for the multimodal system as well, the Eff decreases over 10 times at all operating points.

Finally, the evolution of the maximum scores obtained in each generation (by the genetic algorithm) or in each iteration (by the hill-climbing) is depicted in Fig. 5.5. On the leftmost figure, the attack starts working on the face part of the template, and we can observe segments with different slopes, depending on which part of the template is being modified. Furthermore, in the first and third segments, after the score remains constant for 50 iterations, the algorithm automatically switches to the other part of the template. However, On the rightmost figure, the curve comprises only one segment: the attack started with the genetic algorithm and never changed to the hill-climbing.

Taking into account the results here presented, we may conclude that security is not improved by the fusion of the two unimodal system as it occurs with the performance of the system (EER decreases from 5% to 0.8%); the proposed attack seriously compromises the security offered by the multimodal system.

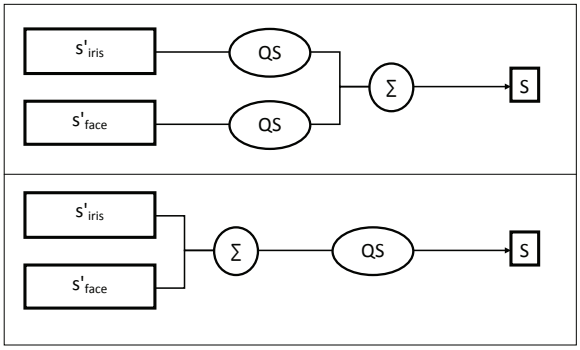

Figure 5.6: *Two possible quantization modes: before the fusion of the scores (top) or after (bottom), where s'_{face} and s'_{iris} are the normalised scores given by the face and iris matchers, and S the final score given by the multimodal system.*

5.3.2. Countermeasure: Score Quantization

The efficiency of quantizing the scores, already evaluated for the face-based unimodal system, is here analysed against the proposed multimodal attack.

Since the global score in this multimodal system is obtained from two previous partial (face and iris) scores that are normalised and then fused, the quantization can take place either before or after this sum or fusion. Both schemes are presented in Fig. 5.6, and will be evaluated and compared in this section. The nomenclature of Sect. 5.1.2 will be used.

In order to select the appropriate quantization step according to the trade-off that should be met in terms of its impact on the system performance (ideally as small as possible) and on the attack performance (as big as possible), we try several Quantization Steps (QS) and consider the corresponding PI. The EER of the system with the different QS is computed when the quantization is applied before and after the score fusion. The QS considered range from 10^{-8} and 10^{-1}. Results are shown in Table. 5.10. For the last QS (10^{-1}), the EER increases considerably (i.e., the QS is too big), while for the remaining values the performance of the system is not significantly affected. The multimodal attack is therefore repeated applying the other four QS values, namely: *i*) QS = 10^{-4}, *ii*) QS = 10^{-3}, *iii*) QS = 10^{-2}, and *iv*) QS = 10^{-1}.

In Table 5.11 the results of these experiments are shown. As it can be seen, the quantization of the scores is effective as a countermeasure against the combined attacking algorithm presented in this work when it is applied:

- Before the fusion with a QS = 10^{-2}. Since the rounding effect of quantizing the scores and then summing them is bigger than that obtained when fusing the scores before applying the quantization, the performance of the attack decreases more when applying the quantization before the fusion. This leads to a SR = 0% for the QS = 10^{-2} when the partial scores are quantized before fusing them.

QS	10^{-8}	10^{-4}	10^{-3}	10^{-2}	10^{-1}
PI (%)	33.69	21.68	0.58	0.01	0.01
EER (%) Before	0.8272	0.8363	0.9990	1.3676	32.659
EER (%) After	0.8272	0.8364	0.9706	1.6912	15.2941

Table 5.10: *Percentage of the iterations of the combined attack with a positive increment (PI), and EER of the system for different quantization steps (QS) for the similarity scores, applying the quantization before and after the fusion of the scores.*

QS		10^{-4}	10^{-3}	10^{-2}	10^{-1}
Before	SR	100%	100%	0%	0%
	Eff ($\times 10^{-4}$)	1.8932	1.6113	-	-
After	SR	100%	100%	100%	0%
	Eff ($\times 10^{-4}$)	1.7806	1.7921	1.7470	-

Table 5.11: *Performance (in terms of SR and Eff) of the combined attack against the system considering different quantization steps (QS), applied before and after the fusion of the scores.*

- Before or after the fusion with a QS = 10^{-1}. With this QS, the system is able to stop the attack regardless of the point where the scores are quantized. As in the previous case, the attack does not receive the necessary feedback from the system on whether it has managed to increase or not the similarity score, and thus fails to achieve its objective.

In both cases listed above, no account is broken, while for the remaining trials the SR of the attack is still 100%, only decreasing its Eff (i.e., more comparisons are needed to break an account). However, while the performance of the system is not considerably affected in the first case (EER = 1.37%), it is barely acceptable with a QS = 10^{-1}: the EER is as high as 32.66%.

Chapter 6

Conclusions and Future Work

This MSc Thesis has considered the problem of evaluating the security offered by biometric systems. Three different hill-climbing attacks were considered throughout the experiments, based on hill-climbing and genetic algorithms, aiming at supplanting the identity of the user. After a summary of the state of the art in face and iris recognition systems, the vulnerabilities evaluation methodologies followed in the MSc Thesis have been presented. These procedural guidelines for the systematic and objective vulnerabilities evaluation of biometric systems have been applied in the experimental studies described in the last chapters of the Dissertation to several publicly available or commercial systems. Besides, in the experimental chapters of the Dissertation, the efficiency of these attacking schemes has been explored, and possible countermeasures tested.

6.1. Conclusions

As case study, we have tested the proposed algorithms on a system based on face and iris (multimodal attack), a trait combination regarded as user-friendly: the features of both traits may be extracted from images the can be captured at the same time, being the acquisition process transparent to the user. The attacking algorithm shows a remarkable performance, thus proving the vulnerabilities of multimodal systems to this type of attacks. Furthermore, the multimodal system has not presented an improvement in the security level against this kind of attack compared to the face and iris modules on their own. This fact confirms what previous studies on spoofing attacks pointed out: even though multimodal systems recognition performance is higher, they do not necessarily increase the robustness of unimodal approaches to external attacks.

The quantization of the scores given by the matcher is analysed as a possible countermeasure. Two different approaches are studied and compared: the partial scores can be quantized before fusing them, or the final score can be quantized after the fusion. The first scenario leads to a null success rate without affecting the verification performance of the system, being thus a suitable countermeasure for the proposed attack. The second case also protects the system against the

attack but at the cost of drastically reducing its verification performance.

In summary, the main results and contributions obtained from this MSc Thesis are:

- The vulnerabilities evaluation protocol followed throughout the first set of experiments.
- The three different algorithmic methods developed and used for evaluating the security of biometric systems, based on optimization algorithms.
- The experimental evidence of the vulnerabilities of biometric systems to this attacks.
- The results of applying score quantization or the use of the most consistent bits of the iriscode as possible countermeasures against hill-climbing based attacks.

6.2. Future Work

Studies as the one performed in the present work are necessary for the improvement and development of any technology. That is specially the case with new fields of research, such as biometrics.

In order to prevent future attacks, we have to know what we can expect from the eventual attackers. That is the base of the security through transparency principle described in Chapter 1. However, knowing what the attacker can do is not enough: we should develop appropriate countermeasures so that eventual impostors are rejected by the recognition systems.

A number of research lines arise from the work carried out in this MSc Thesis. We consider of special interest the following ones:

- Inverse biometrics: this reverse engineering process tries to obtain the initial biometric trait starting from the template. In the past, it has been a common belief that templates do not comprise enough information in order to reconstruct the original sample from them [International Biometric Group, 2002]. However, recent studies have arisen several concerns regarding the soundness of this widely spread belief for traits such as the fingerprint [Cappelli *et al.*, 2007], the iris [Venugopalan and Savvides, 2011] or the face [Galbally *et al.*, 2010]. Applying the hill-climbing algorithms here proposed to the appropriate input may lead to the reconstruction of biometric traits [Gomez-Barrero *et al.*, 2012], that may be regarded as a new kind of attack (once we have the reconstructed trait, we can present it to gain access as a different user), or used to create synthetic databases or even synthetically enlarge existing ones for development purposes.
- Countermeasures: liveness detection. One of the most common countermeasures against spoofing attacks is liveness detection: if the trait presented to the sensor does not fulfil certain requirements, such as heat variation or a minimal trembling, the user is automatically rejected. Even though some research has been conducted on this area, this kind of countermeasures still needs to be further explored.

- Template Protection: a new topic of interest is the protection of biometric templates using techniques imported from the cryptography experts [Rathgeb and Uhl, 2011]. This is a difficult problem, since the input for cryptographic systems must remain invariant in order to obtain the correct key. However, biometric templates suffer from noise and the acquisition conditions often vary (pose, illumination, etc.), and those changes lead to similar but not identical templates. Protecting biometric templates would prevent inverse biometric attacks and at the same time protect the privacy of the user.

Appendix A

Publications

Part of the work of this thesis has been presented in the following articles:

- **Journal articles:**
  - **M. Gomez-Barrero**, J. Galbally and J. Fierrez, "Efficient software attack to multimodal biometric systems and its application to face and iris fusion", Pattern Recognition Letters (JCR 2011 = 1.034), 2013.
  - **M. Gomez-Barrero**, J. Galbally, A. Morales, M. A. Ferrer, J. Fierrez and J. Ortega-Garcia, "A Novel Hand Reconstruction Approach and its Application to Vulnerability Assessment", in Information Sciences (JCR 2011 = 2.833), 2013.
  - J. Galbally, A. Ross, **M. Gomez-Barrero**, J. Fierrez and J. Ortega-Garcia, "Iris Image Reconstruction from Binary Templates: An Efficient Probabilistic Approach Based on Genetic Algorithms", in Computer Vision and Image Understanding (JCR 2011 = 1.340), 2013.
- **Refereed conference papers:**
  - **M. Gomez-Barrero**, J. Galbally, J. Fierrez and J. Ortega-Garcia, "Face verification put to test: a hill-climbing attack based on the uphill-simplex algorithm", in Proc. Intl. Conf. on Biometrics, ICB, pp. 40-45, New Delhi, India, March 2012.
  - **M. Gomez-Barrero**, J. Galbally, P. Tome and J. Fierrez, "On the Vulnerability of Iris-based Systems to a Software Attack based on a Genetic Algorithm", in Proc. Iberoamerican Conference on Pattern Recognition, CIARP, Springer LNCS 7441, pp. 114-121, Buenos Aires, Argentina, September 2012.
  - **M. Gomez-Barrero**, J. Galbally, J. Fierrez and J. Ortega-Garcia, "Hill-Climbing Attack Based on the Uphill Simplex Algorithm and its Application to Signature Verification", in Proc. European Workshop on Biometrics and Identity Management, BioID, Springer LNCS-6583, pp. 83-94, Brandenburg, Germany, March 2011

- **M. Gomez-Barrero**, J. Galbally, A. Morales, M. A. Ferrer, J. Fierrez and J. Ortega-Garcia, "Inverse Biometrics: A Case Study in Hand Geometry Authentication", in Proc. IAPR Int. Conf. on Pattern Recognition, ICPR, Tsukuba, Japan, November 2012.

Pattern Recognition Letters xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Efficient software attack to multimodal biometric systems and its application to face and iris fusion

Marta Gomez-Barrero *[,1], Javier Galbally [1], Julian Fierrez [1]

Biometric Recognition Group – ATVS, EPS, Universidad Autonoma de Madrid, C/ Francisco Tomas y Valiente 11, 28049 Madrid, Spain

ARTICLE INFO

Article history:
Available online xxxx
Communicated by Luis Gomez Deniz

Keywords:
Multimodal system
Security
Vulnerabilities
Hill-climbing
Countermeasures

ABSTRACT

In certain applications based on multimodal interaction it may be crucial to determine not only *what* the user is doing (commands), but *who* is doing it, in order to prevent fraudulent use of the system. The biometric technology, and particularly the multimodal biometric systems, represent a highly efficient automatic recognition solution for this type of applications.

Although multimodal biometric systems have been traditionally regarded as more secure than unimodal systems, their vulnerabilities to spoofing attacks have been recently shown. New fusion techniques have been proposed and their performance thoroughly analysed in an attempt to increase the robustness of multimodal systems to these spoofing attacks. However, the vulnerabilities of multimodal approaches to software-based attacks still remain unexplored. In this work we present the first software attack against multimodal biometric systems. Its performance is tested against a multimodal system based on face and iris, showing the vulnerabilities of the system to this new type of threat. Score quantization is afterwards studied as a possible countermeasure, managing to cancel the effects of the proposed attacking methodology under certain scenarios.

1. Introduction

Multimodal systems represent a new direction for computing that embraces users' natural behaviour as the center of human–computer interaction (Oviatt and Cohen, 2000). As with any other novel discipline, the research community is just beginning to understand how to design robust and well integrated multimodal systems. But only trough multidisciplinary cooperation among those with expertise in individual component technologies can multimodal systems reach its final aim: building more general and robust systems that will reshape daily computing tasks and have significant commercial impact (Oviatt, 1999).

One of the main areas of research in multimodal interaction, where specific expertise is needed, is *recognition*, generally regarded as a form of processing users' commands. However, for certain applications based on multimodal interaction, a second form of recognition is crucial: it is not only necessary to distinguish *what* the user is doing, but *who* is doing it, so that non-authorized individuals cannot use the system. For these cases, a robust personal automatic recognition solution such as the one provided by *biometrics* is required. Although being relatively young compared to other mature and long-used security technologies, biometrics have emerged in the last decade as a pushing alternative for applications where automatic recognition of people is needed. Certainly, biometrics are very attractive and useful for the final user: forget about PINs and passwords, you are your own key (Jain et al., 2006; Wayman et al., 2005). However, we cannot forget that as any technology aimed to provide a security service, biometric systems are exposed to external attacks which could compromise their integrity (Schneier, 1999). Thus, it is of special relevance to understand the threats to which they are subjected and to analyse their vulnerabilities in order to prevent possible attacks and increase their benefits for the users.

External attacks to biometric systems are commonly divided into: *direct attacks* (also known as *spoofing attacks*), carried out against the sensor, and *indirect attacks*, directed to some of the inner modules of the system. In the last recent years important research efforts have been conducted to study the vulnerabilities of biometric systems to both direct and indirect attacks (Galbally et al., 2010, 2011; Matsumoto, 2004; Uludag and Jain, 2004).

This new concern which has arisen in the biometric community regarding the security of biometric systems has led to the appearance of several international projects, like the European Tabula Rasa (2010), which base their research on the security through transparency principle (Schneier, 2000; Kerckhoffs, 1883): in order to make biometric systems more secure and reliable, their vulnerabilities need to be analysed and useful countermeasures need to be developed.

* Corresponding author. Tel.: +34 91 497 33 63.
E-mail addresses: marta.barrero@uam.es (M. Gomez-Barrero), javier.galbally@uam.es (J. Galbally), julian.fierrez@uam.es (J. Fierrez).

[1] This work has been partially supported by projects Contexts (S2009/TIC-1485) from CAM, Bio-Challenge (TEC2009-11186) and Bio-Shield (TEC2012-34881) from Spanish MINECO, TABULA RASA (FP7-ICT-257289) and BEAT (FP7-SEC-284989) from EU, and *Cátedra UAM-Telefónica*.

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.04.029

A Novel Hand Reconstruction Approach and its Application to Vulnerability Assessment

Marta Gomez-Barrero[a], Javier Galbally[a], Aythami Morales[b], Miguel A. Ferrer[b], Julian Fierrez[a], Javier Ortega-Garcia[a]

[a]*Biometric Recognition Group - ATVS, EPS, Universidad Autonoma de Madrid*
C/ Francisco Tomas y Valiente 11, 28049 Madrid, Spain
[b]*Instituto Universitario para el Desarrollo Tecnológico y la Innovación en Comunicaciones*
(IDeTIC) Universidad de Las Palmas de Gran Canaria,
Campus de Tafira s/n, E35017 Las Palmas de Gran Canaria, Spain

Abstract

The present work proposes a novel probabilistic method to reconstruct a hand shape image from its template. We analyse the degree of similarity between the reconstructed images and the original samples in order to determine whether the synthetic hands are able to deceive hand recognition systems. This analysis is made through the estimation of the success chances of an attack carried out with the synthetic samples against three different state-of-the-art independent systems. The experimental results show that there is a high probability of breaking a hand recognition system using this approach. Furthermore, since it is a probabilistic method, several synthetic images can be generated from each original sample, which increases the success chances of the attack.

Keywords: Biometric systems, Hand recognition, Hand reconstruction, Security, Vulnerabilities

1. Introduction

Biometrics are nowadays being introduced into many applications as an alternative to traditional security mechanisms [42, 83]. The main advantage of bio-

Email addresses: marta.barrero@uam.es (Marta Gomez-Barrero), javier.galbally@uam.es (Javier Galbally), amorales@gi.ulpgc.es (Aythami Morales), mferrer@dsc.ulpgc.es (Miguel A. Ferrer), julian.fierrez@uam.es (Julian Fierrez), javier.ortega@uam.es (Javier Ortega-Garcia)

Iris Image Reconstruction from Binary Templates: An Efficient Probabilistic Approach Based on Genetic Algorithms

Javier Galbally[a], Arun Ross[b], Marta Gomez-Barrero[a], Julian Fierrez[a], Javier Ortega-Garcia[a]

[a]*Biometric Recognition Group - ATVS, EPS, Universidad Autonoma de Madrid. C/ Francisco Tomas y Valiente 11, 28049 Madrid. Spain.*
[b]*Integrated Pattern Recognition and Biometrics Lab (i-PRoBe), Michigan State University. East Lansing, MI 48824. USA.*

Abstract

A binary iriscode is a very compact representation of an iris image. For a long time it was assumed that the iriscode did not contain enough information to allow for the reconstruction of the original iris. The present work proposes a novel probabilistic approach based on genetic algorithms to reconstruct iris images from binary templates and analyzes the similarity between the reconstructed synthetic iris image and the original one. The performance of the reconstruction technique is assessed by empirically estimating the probability of successfully matching the synthesized iris image against its true counterpart using a commercial matcher. The experimental results indicate that the reconstructed images look reasonably realistic. While a human expert may not be easily deceived by them, they can successfully deceive a commercial matcher. Furthermore, since the proposed methodology is able to synthesize multiple iris images from a single iriscode, it has other potential applications including privacy enhancement of iris-based systems.

Keywords: Image reconstruction, Biometric systems, Iris recognition, Binary iriscodes, Security, Privacy

Email addresses: javier.galbally@uam.es (Javier Galbally), arun.ross@mail.wvu.edu (Arun Ross), marta.barrero@uam.es (Marta Gomez-Barrero), julian.fierrez@uam.es (Julian Fierrez), javier.ortega@uam.es (Javier Ortega-Garcia)

Face Verification Put to Test: A Hill-Climbing Attack Based on the Uphill-Simplex Algorithm

Marta Gomez-Barrero, Javier Galbally, Julian Fierrez, Javier Ortega-Garcia
Biometric Recognition Group -ATVS, EPS, Universidad Autonoma de Madrid
C/ Francisco Tomas y Valiente 11, 28049 Madrid, Spain
marta.barrero@uam.es, javier.galbally@uam.es, julian.fierrez@uam.es, javier.ortega@uam.es

Abstract

The vulnerabilities of a PCA-based face verification system against a hill-climbing attack using the uphill-simplex algorithm are studied. Experiments are carried out on the face subcorpus of the publicly available BioSecure DB, where the attack has shown a remarkable performance proving the lack of robustness of the tested system to this type of threat. Furthermore, the proposed attacking scheme is not only able to bypass the security of the recognition system, but it is also capable of reconstructing the users face image, with the privacy concerns that this entails. As a possible countermeasure to minimize the effect of the attack, score quantization is applied. This protection method is able to reduce both the success rate and the efficiency of the attack, however it does not completely succeed in preventing a possible intruder from accessing the system. The study also highlights the high adaptation capabilities of the proposed attack which had already been used to break a signature-based verification system.

1. Introduction

Due to the fact that biometrics [10], as an automatic means of human recognition, constitutes a relatively novel field of research, most efforts undertaken by the different parties involved in the development of this technology (researchers, industry, evaluators, etc.) have been mainly (but not exclusively) directed to the improvement of its performance [11]. This has left partially uncovered other important aspects involved in the complex biometric recognition problem.

In particular, it has not been until recently when biometric security assessment has emerged in the biometric community as a primary field of research, as a consequence of the concern arisen after the classification of the vulnerability points presented in [16] (shown in Fig. 1), and the different efficient attacking algorithms developed in order to compromise the security level given by biometric applications [6, 19].

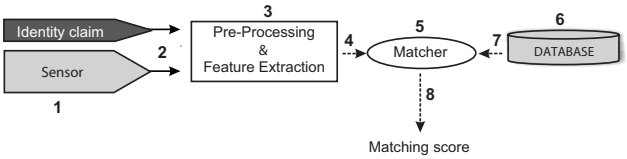

Figure 1. Architecture of an automated biometric verification system. Possible attack points given in [16] are numbered from 1 to 8.

These vulnerability studies have helped to improve the biometric technology by making public certain flaws and by encouraging the industry and researchers to look for solutions to the different threats [9, 22, 18]. This way, the level of security and the convenience offered to the final user are increased.

Most of the existing works studying the vulnerabilities of biometric systems to attacks against the inner modules of the system (those labeled from 2 to 8 in Fig. 1), use some type of variant of the hill-climbing algorithm presented in [17]. Some examples include an attack to a face-based system in [2], and to standard and Match-on-Card minutiae-based fingerprint verification systems in [21] and [12] respectively. These types of attacks take advantage of the score given by the matcher to iteratively change a synthetically generated template until the similarity score exceeds a fixed decision threshold and thereby access to the system is granted. Except for the algorithm proposed in [7], all of these hill-climbing approaches are highly dependent on the technology used, only being usable for very specific types of matchers.

However, recently, a general hill-climbing algorithm based on the uphill-simplex algorithm was presented and tested using a signature verification system [8]. In the present contribution this general method is successfully applied to attack an automatic face recognition system based on eigenfaces, proving this way its biometric independency

On the Vulnerability of Iris-based Systems to a Software Attack based on a Genetic Algorithm

Marta Gomez-Barrero, Javier Galbally, Pedro Tome, and Julian Fierrez

Biometric Recognition Group-ATVS, EPS, Universidad Autonoma de Madrid
C/ Francisco Tomas y Valiente 11, 28049 Madrid, Spain

Abstract. The vulnerabilities of a standard iris verification system to a novel indirect attack based on a binary genetic algorithm are studied. The experiments are carried out on the iris subcorpus of the publicly available BioSecure DB. The attack has shown a remarkable performance, thus proving the lack of robustness of the tested system to this type of threat. Furthermore, the consistency of the bits of the iris code is analysed, and a second working scenario discarding the fragile bits is then tested as a possible countermeasure against the proposed attack.

Keywords: Security, vulnerabilities, iris recognition, genetic algorithm, countermeasures.

1 Introduction

Due to their advantages over traditional security approaches, biometric security systems are nowadays being introduced into many applications where a correct identity assessment is a crucial issue, such as access control or sensitive data protection [1]. These systems perform automatic recognition of individuals based on anatomical (e.g., fingerprint, face, iris, etc.) or behavioural characteristics (e.g., signature, gait, keystroke dynamics). Among these traits, the iris has been traditionally regarded as one of the most reliable and accurate [1].

However, biometric systems are vulnerable to external attacks, that can be divided into two different groups, namely: *i) direct attacks*, carried out against the sensor using synthetic traits [2]; and *ii) indirect attacks*, carried out against one of the inner modules of the system [3], and thus requiring some knowledge about the inner working of the system. Several works have already studied the robustness of iris-based biometric systems against direct attacks, including attackers wearing contact lenses with artificial textures printed onto them [4] and fake iris images [5].

In the present paper, a novel indirect attack based on a genetic algorithm is presented. The point of attack are binary templates, as depicted in Fig. 1 (top), where a general hill-climbing attack is shown. Although other hill-climbing attacks have been proposed [6, 3, 7], none of them work on binary templates, but on real-valued feature vectors or directly on the sample images.

Although in commercial systems the number of consecutive unsuccessful access attempts is usually restricted, this countermeasure has been circumvented in different occasions or may even be used to compromise the system by performing an *account*

Hill-Climbing Attack based on the Uphill Simplex Algorithm and its Application to Signature Verification

Marta Gomez-Barrero, Javier Galbally, Julian Fierrez, Javier Ortega-Garcia

Biometric Recognition Group–ATVS, EPS, Universidad Autonoma de Madrid,
C/ Francisco Tomas y Valiente 11, 28049 Madrid, Spain
{marta.barrero, javier.galbally, julian.fierrez, javier.ortega}@uam.es

Abstract. A general hill-climbing attack to biometric systems based on a modification of the downhill simplex algorithm is presented. The scores provided by the matcher are used in this approach to adapt iteratively an initial estimate of the attacked template to the specificities of the client being attacked. The proposed attack is evaluated on a competitive feature-based signature verification system over both the MCYT and the BiosecurID databases (comprising 330 and 400 users, respectively). The results show a very high efficiency of the hill-climbing algorithm, which successfully bypassed the system for over 90% of the attacks with a remarkably low number of scores needed.

1 Introduction

Biometric security systems are nowadays being introduced in many applications, such as access control, sensitive data protection, on-line tracking systems, etc., due to their advantages over traditional security approaches [1]. Nevertheless, they are also susceptible to external attacks that can decrease their security level. Therefore, it is of the utmost importance to analyse the vulnerabilities of biometric systems so that their weaknesses can be found and useful countermeasures against foreseeable attacks can be developed.

There are two main types of attacks that may put at risk the security offered by a biometric system: (*i*) *direct attacks*, carried out against the sensor using synthetic traits, such as printed iris images or gummy fingers [2]; and (*ii*) *indirect attacks*, carried out against some of the inner modules of the system [3, 4], and thus requiring for the attacker to have some knowledge about the system (e.g., storage format or matcher used). A more detailed analysis of the vulnerable points of biometric systems is made by Ratha *et al.* in [5]. In this work 8 possible points of attack are identified, the first corresponding to direct ones and the remaining seven to indirect attacks.

Several works have already studied the robustness of biometric systems against direct attacks, specially fingerprint- and iris-based, including [2, 3, 6]. In the case of indirect attacks, most of the studies use some kind of variant of the hill-climbing algorithm [4]. Some examples include an indirect attack to a face-based

Inverse Biometrics: A Case Study in Hand Geometry Authentication

Marta Gomez-Barrero, Javier Galbally, Julian Fierrez and Javier Ortega-Garcia
Biometric Recognition Group - ATVS, EPS, Universidad Autonoma de Madrid
{*marta.barrero, javier.galbally, julian.fierrez, javier.ortega*}*@uam.es*

Aythami Morales and Miguel A. Ferrer
Instituto Universitario IDeTIC, Universidad de las Palmas de Gran Canaria
amorales@gi.ulpgc.es, mferrer@dsc.ulpgc.es

Abstract

Recently, a considerable amount of research has been focused on inverse biometrics, that is, regenerating the original biometric sample from its template. In this work, the first reconstruction approach to recover hand geometry samples from their feature vectors is proposed. Experiments are carried out on the publicly available GPDS Hand DB, where the method has shown a remarkable performance, after reconstructing a very high percentage of the hands included in the dataset. Furthermore, the proposed technique is general, being able to successfully reproduce the original hand shape sample regardless of the information and format of the template used.

1. Introduction

Automatic access of persons to services is becoming increasingly important in the information era. This has resulted in the establishment of a new technological area known as biometric recognition, or simply *biometrics* [7]. A biometric system is essentially a pattern recognition application that makes use of biometric traits (e.g., fingerprints, iris, face or hand images) to automatically recognize individuals.

After acquiring the biometric sample, these systems extract the intrinsic information which is discriminant and exclusive of that particular user. This information, or feature vector, is stored in a *biometric template* which is later used as a representation of the subject identity and deployed for authentication purposes. The biometric template should be a compact and precise representation of the user identity (e.g., iriscode or minutiae points of a fingerprint). Therefore, one of the key aspects of the biometric recognition process to which great efforts have been dedicated is the efficient "translation" of the raw biometric sample into a stable, simple and highly discriminant template.

However, over the last recent years important attention has also been paid to the opposite, and very challenging, reverse engineering problem: recovering the original biometric sample from its parameterized template. Such a reconstruction process has been generally referred to as *inverse biometrics* [9], and, for a long period of time it was believed to be unfeasible for certain traits such as the fingerprints or the iris.

In this context, several methods have been proposed for the reconstruction of face images [1], fingerprint impressions [2], or irides [8]. The reconstructed samples have been used for multiple applications such as vulnerability assessment, synthetically increasing the amount of data available for a certain user, or as a possible solution for privacy-related issues.

However, in spite of its many potential uses, the reverse engineering of biometric templates still remains unexplored for largely deployed traits such as the handwritten signature or the hand. In the present contribution we address the inverse biometrics problem for the hand geometry trait and propose the first reconstruction approach to recover hand geometry samples from their templates. The technique, which is based on a combination of the uphill-simplex algorithm and a synthetic hand generator, clearly shows the feasibility of such an inverse biometrics methodology. Furthermore, the proposed approach is general in the sense that it does not need to know the format of the template (i.e., what information is stored in it, or how it is stored).

Appendix B

Short Biography

Marta Gomez-Barrero was born in 1988. She started her Computer Science Engineering and Mathematics studies in 2006 and received the MSc degree in Computer Science Engineering and the MSc degree in Mathematics in 2011 from Universidad Autonoma de Madrid, Spain. Since September of 2010, she is with the Biometric Recognition Group - ATVS at the Universidad Autonoma de Madrid, where she is currently collaborating as an assistant researcher pursuing the PhD degree. Her research interests include signal and image processing, pattern recognition, computer vision and biometrics. Her current research focuses on template protection and cryptobiometrics.

References

A. Adler. Images can be regenerated from quantized biometric match score data. In *Proc. Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 469–472, 2004. 11, 32

T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006. 12

Z. Akhtar and N. Alfarid. Secure learning algorithm for multimodal biometric systems against spoof attacks. In *Proc. International Conference on Information and Network Technology (IPCSIT)*, volume 4, pages 52–57. IACSIT Press, 2011. 5, 10

Z. Akhtar, S. Kale, and N. Alfarid. Spoof attacks in mutimodal biometric systems. In *Proc. International Conference on Information and Network Technology (IPCSIT)*, volume 4, pages 46–51. IACSIT Press, 2011. 5, 10

ANSI, 2001. ANSI X9.84-2001, Biometric Information Management and Security. 26, 29

J. E. Baker. Reducing bias and inefficiency in the selection algorithm. In *Proc. International Conference on Genetic Algorithms and their Application (ICGAA)*, pages 14 – 21. L. Erlbaum Associates Inc., 1987. 19

BEAT. Biometrics evaluation and testing (beat), 2012. URL `http://www.beat-eu.org/`. 7

P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):711 – 720, 1997. 12

BEM. Biometric Evaluation Methodology. v1.0, 2002. 5, 7

BioAPI. The BioAPI consortium, 2009. http://www.bioapi.org. 32

BioAPI Consortium. BioAPI specification (version 1.1), March 2001. www.bioapi.org/Downloads/BioAPI11

K. Bowyer, K. Hollingsworth, and P. Flynn. Image understanding for iris biometrics: a survey. *Computer Vision and Image Understanding*, 110:281–307, 2008. 12

A. Brindle. *Genetic Algorithms for Function Optimization.* PhD thesis, University of Alberta, Edmounton, 1981. 18

BWG. Biometric security concerns, v1.0. Technical report, CESG, UK Government, 2003. 6

BWG. Communications-electronics security group (CESG) – biometric working group (BWG) (UK government), 2009. http://www.cesg.gov.uk/policy_technologies/biometrics/index.shtml. 6

R. Cappelli, D. Maio, A. Lumini, and D. Maltoni. Fingerprint image reconstruction from standard templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29:1489–1503, September 2007. 42

CC. Common Criteria for Information Technology Security Evaluation. v3.1, 2006. Available on-line at http://www.commoncriteriaportal.org/. 5, 7

G. Chetty and M. Wagner. Audio-visual multimodal fusion for biometric person authentication and liveness verification. In *Proc. NICTA-HCSNet Multimodal User Interaction Workshop (MMUI)*, 2005. 5, 9

J. Daugman. The importance of being random: Statistical principles of iris recognition. *Pattern Recognition*, 36: 279–291, 2003. 12

J. Daugman. How iris recognition works. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14 (1):21–30, 2004. 12, 23, 24

J. Daugman. Probing the uniqueness and randomness of iris codes: Results from 200 billion iris pair comparisons. *Proceedings of the IEEE*, 94:1927–1935, 2006. 12

J. Daugman. New methods in iris recognition. *IEEE Trans. on Systems Man and Cybernetics - Part B: Cybernetics*, 37:1167–1175, 2007. 12

J. Daugman. *Iris Recognition*, chapter 4, pages 71–90. Springer, 2008. 12

J. Galbally, J. Fierrez, F. Alonso-Fernandez, and M. Martinez-Diaz. Evaluation of direct attacks to fingerprint verification systems. *Telecommunication Systems, Special Issue on Biometrics*, 47:243–254, 2011. 5

J. Galbally, C. McCool, J. Fierrez, and S. Marcel. On the vulnerability of face verification systems to hill-climbing attacks. *Pattern Recognition*, 43:1027–1038, 2010. xiii, 5, 6, 11, 29, 30, 42

M. Gomez-Barrero, J. Galbally, J. Fierrez, and J. Ortega-Garcia. Hill-climbing attack based on the uphill simplex algorithm and its application to signature verification. In *Proc. European Workshop on Biometrics and Identity Management (BioID), 2011*, pages 83–94. LNCS-6583, 2011. xiii, 18, 29, 30, 33

M. Gomez-Barrero, J. Galbally, A. Morales, M. A. Ferrer, J. Fierrez, and J. Ortega-Garcia. Inverse biometrics: A case study in hand geometry authentication. In *Proc. Int. Conf. on Patter Recognition (ICPR), 2012*, 2012. 42

J. Hämmerle-Uhl, K. Raab, and A. Uhl. Attack against robust watermarking-based multimodal biometric recognition systems. In *Proc. of the COST 2101 European conference on Biometrics and ID management (BioID)*, pages 25–36. LNCS-6583, 2011. 6

J. Hennebert, R. Loeffel, A. Humm, and R. Ingold. A new forgery scenario based on regaining dynamics of signature. In *Proc. IAPR Int. Conf. on Biometrics (ICB)*, pages 366–375. Springer LNCS-4642, 2007. 7

C. J. Hill. Risk of masquerade arising from the storage of biometrics. Master's thesis, Australian National University, 2001. 7

K. P. Hollingsworth, K. W. Bowyer, and P. J. Flynn. The best bits in an iris code. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(6):964–973, 2009. xiii, 35, 36

R. Huang, V. Pavlovic, and D. N. Metaxas. A hybrid face recognition method using markov random fields. In *Proc. Int. Conf. on Pattern Recognition, ICPR*, pages 157–160, 2004. 12

International Biometric Group. Generating images from templates. White paper, 2002. 42

ISO/IEC 19792. ISO/IEC 19792:2009, information technology - security techniques - security evaluation of biometrics., 2009. 5, 7

A. Jain, A. Ross, and K. Nandakumar. *Introduction to Biometrics*. Springer, 2011. 1, 4

A. K. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38:2270–2285, 2005. 25

A. K. Jain, A. Ross, and S. Pankanti. Biometrics: a tool for information security. *IEEE Trans. on Information Forensics and Security*, 1(2):125–143, 2006. 2, 3, 7

P. Johnson, B. Tan, and S. Schuckers. Multimodal fusion vulnerability to non-zero effort (spoof) attacks. In *Proc. Workshop on Information Forensics and Security (WIFS)*, 2010. 5, 6, 10

A. Kerckhoffs. La cryptographie militaire. *Journal des Sciences Militaires*, 9:5–83, 1883. Available on-line at http://www.petitcolas.net/fabien/kerckhoffs. 6, 7

S. Lawrence, C. Giles, A. Tsoi, and A. Back. Face recognition: A convolutional neural-network approach. *IEEE Trans. Neural Networks*, 8:98–113, 1997. 12

E. Marasco. *Secure Biometric Systems*. PhD thesis, University of Naples Federico II, 2010. 6, 10

M. Martinez-Diaz, J. Fierrez, *et al.* An evaluation of indirect attacks and countermeasures in fingerprint verification systems. *Pattern Recognition Letters*, 2011. 5, 6, 11

L. Masek and P. Kovesi. Matlab source code for a biometric identification system based on iris patterns. Master's thesis, School of Computer Science and Software Engineering, University of Western Australia, 2003. 23, 24

T. Matsumoto. Artificial irises: importance of vulnerability analysis. In *Proc. 2nd Asian Biometrics Workshop*, 2004. 5

A. Mayoue, B. Dorizzi, L. Allano, G. Chollet, J. Hennebert, D. Petrovska-Delacretaz, and F. Verdet. *Guide to biometric reference systems and performance evaluation*, chapter BioSecure multimodal evaluation campaign 2007 (BMEC 2007), pages 327–372. Springer, 2009. 22

J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:368 – 313, 1965. 16, 17, 30

J. Ortega-Garcia, J. Fierrez, *et al.* The multi-scenario multi-environment BioSecure multimodal database (BMDB). *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32:1097–1111, 2010. 21, 23, 24

J. Phillips, P. Flynn, *et al.* Overview of the face recognition grand challenge. In *Proc. IEEE CCVPR*, pages 947–954, 2005. 22, 24

N. Ratha, J. H. Connell, and R. M. Bolle. An analysis of minutiae matching strength. In *Proc. IAPR AVBPA*, pages 223–228. Springer LNCS-2091, 2001. 9

C. Rathgeb and A. Uhl. A survey on biometric cryptosystems and cancelable biometrics. *EURASIP Journal on Information Security*, 3, 2011. 43

R. Rodrigues, N. Kamat, and V. Govindaraju. Evaluation of biometric spoofing in a multimodal system. In *Proc. IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, sept. 2010. 5, 6, 10

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. 5, 6, 10

A. Ross and A. K. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24(13):2115 – 2125, 2003. 25

V. Ruiz-Albacete, P. Tome-Gonzalez, *et al.* Direct attacks using fake images in iris verification. In S. LNCS-5372, editor, *Proc. BioID*, pages 181–190, 2008. 23

F. Samaria and F. Fallside. *Face identification and feature extraction using hidden markov models*. Citeseer, 1993. 12

B. Schneier. The uses and abuses of biometrics. *Communications of the ACM*, 48:136, 1999. 7

B. Schneier. *Secrets and lies*. Wiley, 2000. 6, 7

Tabula Rasa. Trusted biometrics under spoofing attacks (tabula rasa), 2010. URL `http://www.tabularasa-euproject.org/`. 7

B. Tan. *Assessing and recducing spoofing vulnerability for multimodal and fingerprint biometrics*. PhD thesis, Clarkson University, 2009. 9

L. Thalheim and J. Krissler. Body check: biometric access protection devices and their programs put to the test. *ct magazine*, pages 114–121, November 2002. 5, 7

A. Tolba, A. El-Baz, and A. El-Harby. Face recognition: A literature review. *International Journal of Signal Processing*, 2(2):88–103, 2006. 12

M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proc. IEEE CCVPR*, pages 586–591, 1991. 12, 22, 24

U. Uludag and A. Jain. Attacks on biometric systems: a case study in fingerprints. In *Proc. SPIE Seganography and Watermarking of Multimedia Contents VI*, volume 5306, pages 622–633, 2004. 5, 6, 11

S. Venugopalan and M. Savvides. How to generate spoofed irises from an iris code template. *IEEE Trans. on Information Forensics and Security*, 6:385–394, 2011. 42

Y. Wang, T. Tan, and A. K. Jain. Combining face and iris biometrics for identity verification. In *Proc. of Fourth International Conference on Audio- and Video-Based Person Authentication (AVBPA)*, pages 805 – 813, 2003. 25

J. Wayman, A. Jain, D. Maltoni, and D. Maio. *Biometric systems. Technology, design and performance evaluation*. Springer, 2005. 7

J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009. 12

W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003. 12