Contribution to the certification of fingerprint systems : towards the reproducibility of the evaluation

Abdarahmane Wone

Normandie Université

UNIVERSITÉ CAEN NORMANDIE

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité **INFORMATIQUE**

Préparée au sein de l'**Université de Caen Normandie**

Contribution to the certification of fingerprint systems: towards the reproducibility of the evaluation

Présentée et soutenue par

ABDARAHMANE WONE

Thèse soutenue le 29/11/2023

devant le jury composé de :

M. KIDIYO KPALMA	Professeur des universités - INSA de Rennes	Rapporteur du jury
M. FREDERIC MORAIN-NICOLIER	Professeur des universités - UNIVERSITE REIMS CHAMPAGNE ARDENNE	Rapporteur du jury
M. JOEL DI MANNO	Ingénieur de recherche - FIME EMEA	Membre du jury
MME STÉPHANIE SCHUCKERS	Professeur - Clarkson University	Membre du jury
M. CHAKER LARABI	Professeur des universités - UNIVERSITE POITIERS	Président du jury
M. CHRISTOPHE ROSENBERGER	Professeur des universités - ENSICAEN	Directeur de thèse
M. CHRISTOPHE CHARRIER	Maître de conférences - Université de Caen Normandie	Co-directeur de thèse

Thèse dirigée par **CHRISTOPHE ROSENBERGER** (GREYC ALGORITHMIQUE) et **CHRISTOPHE CHARRIER** (GREYC ALGORITHMIQUE)

MIIS

GREYC
Laboratoire de recherche en sciences du numérique

Acknowledgements

During this PhD thesis, I had the opportunity to make my first steps as a researcher. Thanks to the GREYC laboratory for giving me the opportunity to do my research in a such stimulating and warm place.
Special thanks go to my PhD directors Christophe Rosenberger and Christophe Charrier who supported and guided me all along this project.
I want to thank Joel Di Manno for being my supervisor at Fime during these years. Your advice and bright ideas helped me a lot in directing my research. Thanks to you three for your readings, encouragement, and support contributed a lot to the achievement of the thesis.

I would like to thank the French National Association for Research and Technology (ANRT) and Fime company for funding this project.
I would like to thank my follow-up committee members Amine Nait-Ali and Alexis Lechervy. I want to thank the administrative service of the GREYC laboratory, Sophie, Gaelle, Arielle, Caroline as well as Marie and Sandrine from the doctoral school for their help and support.

I thank all my Jury members, Kidiyo Kpalma, and Frédéric Morain-Nicolier for accepting to evaluate this manuscript, Chaker Larabi and Stephanie Schuckers for evaluating this work.
Thanks to all the members of the GREYC laboratory and SAFE colleagues, especially Tanguy for the discussions we had. Your advice helped me a lot during the writing of my thesis.
I would like to thank all my colleagues from Fime, particularly the Biometric Team for supporting me, and the interest they showed towards my work.
Finally, I would like to thank my family, who supported me and encouraged me despite the distance and the fact of not always understand what my research involved.

Remerciements

Pendant cette thèse, j'ai pu m'initier à la recherche et me familiariser avec la biométrie. Je tiens à remercier le laboratoire GREYC qui m'a accueilli et accompagné tout au long de cette thèse.

Mes premiers remerciements vont évidemment à l'endroit de mes directeurs de thèse Christophe Rosenberger et Christophe Charrier qui m'ont accompagné et guidé pendant ces années. Ils ont su m'insuffler la passion de la recherche. Je les remercie pour leur disponibilité, leurs conseils avisés, et leur relecture malgré des calendriers bien chargés.

Je remercie Joël Di Manno pour m'avoir encadré pendant ces années, tes conseils et tes idées lumineuses m'ont aidé à orienter mes recherches. Vos relectures à tous et votre persévérance ont permis à ce manuscrit d'aboutir. Merci! Je remercie l'ANRT et la société Fime pour avoir financé cette thèse. Je remercie les membres de mon comité de suivi de thèse Amine Nait-Ali et Alexis Lechervy. Je remercie le service administratif du GREYC particulièrement, Sophie, Arielle, Gaëlle et Caroline, ainsi que Marie et Sandrine à l'école doctorale pour leur accompagnement.

Je remercie tous les membres de mon jury, Kidiyo Kpalma et Frédéric Morain-Nicolier pour avoir accepté de rapporter scrupuleusement cette thèse, Chaker Larabi et Stephanie Schuckers pour leur examen. Je remercie les membres du laboratoire GREYC et de l'équipe SAFE, particulièrement Tanguy dont les conseils et les discussions m'ont été utiles lors la rédaction de ce manuscrit.

Je remercie mes collègues de Fime et particulièrement l'équipe biométrique pour m'avoir soutenu et pour l'intérêt qu'ils ont manifesté à l'égard de mes recherches.

Je remercie également mon gai et fidèle compagnon Dr. Brice Wandji (aka la main de la panthère, le conseiller matrimonial, entre autres sobriquets) pour les discussions passionnées et ses conseils qui ont allégé ces trois années.

Enfin, je remercie ma famille qui, malgré la distance et la différence de parcours, m'a accompagné et soutenu toutes années et m'a encouragé à aller au bout de ces

longues études.

À ma famille.

List of abbreviations and acronyms

APCER Attack Presentation Classification Error Rate

AUC Area Under the Curve

BPCER Bonafide Presentation Classification Error Rate

DET Detection Trade-Off

EER Equal Error Rate

FAR False Accept Rate

FMR False Match Rate

FNIR False Negative Identification Rate

FNMR False Non Match Rate

FPIR False Positive Identification Rate

FRR False Reject Rate

FTA Failure To Acquire

FTE Failure To Enrol

GAN Generative Adversarial Network

IAPAR Impostor Attack Presentation Acceptance Rate

IAPIR Impostor Attack Presentation Identification Rate

ISO International Organization for Standardization

MCC Minutia Cylinder-Code

MWGAN Multi-marginal Wasserstein GAN

NFIQ NIST Fingerprint Image Quality

NIST National Institute of Standards and Technology

PAD Presentation Attack Detection

PAI Presentation Attack Instrument

ROC Receiver Operating Characteristic

Contents

Chapter 1

Introduction

Summary : *This short chapter presents the context in which this doctoral thesis was carried out. It presents the industrial framework of the thesis, which is the certification of biometric systems, as well as the problem addressed and its contributions. Finally, we conclude with the prospects.*

1.1 Context

We are living in a digital world where cybersecurity is an essential part. Technology and cyber-solution are ubiquitous. Whether for leisure, home automation, or our interactions with administrative services, etc., computer systems are an integral part of our lives. Their security must be guaranteed to ensure their integrity as well as the protection of personal data and the privacy of users. This is a key element for the confidence one can have in a computer system.

Through the years, many solutions have been proposed to ensure the security and integrity of cyber systems.

The first solutions to emerge were based on *Something you know*. Passwords and passphrases are examples of solutions that still exist today. However, passwords present many risks as people try to set them in a way they can remember them easily. They are often shared or noted somewhere or hints are noted, hints which may contain the password itself. Microsoft reported 1,287 password attacks every second, which represents more than 111 million per day.

Another solution based on physical possession such as a smartcard or a physical key has been proposed. These are based on *Something you have*. It ensures that the person has something that only the right person should possess. The physical object can be lost, forgotten, or replicated, etc.

A third type is *Something you are*. For the last one, biometrics is the best candidate to guarantee security and ensure that people are who they claim they are with the least risks. Biometrics is all the means to authenticate or identify a human being thanks to his/her biometric characteristics, which can be morphological or behavioral. In opposition to passwords, smartcards or tokens that can be reset, our biometric characteristics are unique and cannot be changed. That is why the security of biometric systems is essential and crucial.

1.2 Biometrics

Biometrics is more and more present and used, although at the beginning it was a solution that was used for some specific tasks such as border control or identity checks. Nowadays, biometrics is not only used for administrative or official tasks but is present in our everyday lives. From computers to smartphones or banking cards, biometrics is more and more used for logical access or for some more specific tasks such as second-factor authentication, multifactor authentication, identity, banking, retail and commerce, insurance, ... For example, in 2022, Cisco reported that 81% of smartphones have biometrics enabled [1] while smartphones with fingerprint sensor went from 19% to 60% between 2014 and 2018 [2]. Indeed, as pointed out in the previous section, the solutions based on knowledge or possession are subject to be compromised and therefore the security of the associated application is not guaranteed anymore. Biometric solutions offer the advantage of being set to be user-friendly and not constraining most of the time. Biometrics is a solution that covers up the weaknesses of the two others. It is used to assess the identity of people and give them access to systems they are allowed to.

Therefore, biometric systems need to be secured and people or services using biometrics need to have guarantees regarding the integrity and usability of the biometric systems. Thus, biometric systems need to be evaluated in order to assess their usability and associated security level.

[1] `https://duo.com/resources/ebooks/the-2022-duo-trusted-access-report`

[2] `https://www.statista.com/statistics/804269/global-smartphone-fingerprint-sensor-penetration-rate/`

1.3 PhD Thesis

1.3.1 Context

This thesis is the result of a collaboration between the GREYC laboratory and FIME EMEA. It is co-funded by the French National Association for Research and Technology (ANRT: Association Nationale de la Recherche et de la Technologie) whose objective is to create innovations in partnership with research laboratories and tech companies.

1.3.2 Objectives

The thesis is set in the context of the certification of biometric systems in general and fingerprint-based systems in particular. Indeed, for a certified system, it is important to know the testing scenarios that have been used by the certification laboratory to assess the conformance of a system to a specific test plan. The tests that have led to the approbation of the product under test should be reproducible identically or with a minimum deviation regarding the user experience and the security of the certified product. Many factors have been identified to make difficult the reproducibility of the testing. While some of them depend on the user interaction with the system, the system as well as the environment where it is deployed can significantly impact its recognition capabilities. The context of acquisition is the main factor regarding the non-repeatability of the tests. This concerns both the environmental conditions of acquisition and the capture system itself. As stated earlier, the certification of a biometric (fingerprint) system is assessed on the basis of its ability to recognize the right person and to resist to presentation attacks. The last part is time-consuming and laboratories have no indication regarding the training methods of the tested algorithms. Therefore, it is not fair to rely on the existing datasets, which are often limited to research purposes, to test products. Moreover, the constraints associated with the testing motivate us to go towards the generation of synthetic data to be used for the evaluation of presentation attack detection.

During this thesis, we address the following questions:

- How the acquisition context of a biometric fingerprint system can impact its performance?
- How synthetic fingerprint spoofs can help to evaluate fingerprint systems?
- How can we achieve the reproducibility of biometric testing?

1.4 Contributions

We list hereafter the different contributions of this PhD thesis.

1. **First contribution**: A study to understand the impact of environmental conditions' effects on fingerprint systems is done. Environmental conditions are one of the most overlooked biases for fingerprint systems. Indeed, due to these factors, it is very difficult to predict the behavior of a product once it is deployed. During this PhD thesis, we built a fingerprint dataset under controlled temperature and humidity conditions. We observed the effect of environmental conditions using an objective fingerprint quality metric and different fingerprint-matching algorithms. We also underlined the importance of the acquisition system quality and how it can affect the performance of the final biometric solution.

2. **Second contribution**: We proposed a method to generate fingerprint presentation attack instruments to be used for the evaluation of the presentation attack detection module.
   The generated data is validated from quality and matching points of view. We compare our method with the state-of-the-art solutions.

3. **Third contribution**: We proposed a generic method to validate synthetic biometric data. The proposed method takes into account the usability of the data in a recognition process and their quality using an objective assessment tool for the quality.

1.5 Thesis organization

The thesis is organized as follows:

- Chapter 2 describes the certification of a biometric system. We detail the importance of the certification of biometric systems and how the evaluation is done in the testing laboratories. We also present the main metrics that are normalized for the evaluation of the biometric system and those that we use in the thesis.
- Chapter 3 presents the main datasets that exist in the field of fingerprints. We express the constraints linked to the certification of biometric systems as well as the collection of biometric data and the need to have synthetic data. That motivates us to go towards the generation of synthetic data. We finally propose

a generic method to validate synthetic biometric data using some objective metrics linked to the usability and security of biometric systems.

- Chapter 4 is dedicated to the analysis we did on the acquisition context and the way it can impact the performance of a fingerprint biometric system. This context can be either the environmental conditions of the capture or the capturing device. The control of these contexts is a key to the reproducibility of the testing. Indeed, we want to understand the parameters that the evaluation of biometric systems may be correlated to.
- Chapter 5 presents the method to generate fingerprint biometric datasets for presentation attack detection testing that we propose, and the comparison of the generated datasets with existing datasets. Indeed, this follows the need to have synthetic data expressed in Chapter 1. We present the methodology that we propose in this Ph.D. thesis, as well as the results we obtained using the validation methodology we introduced in Chapter 2.
- Chapter 6 is dedicated to the conclusion of this work, where we present some perspectives.
- Chapter 7 and Chapter 8 give a global overview of the thesis and the conclusions we came up with during this thesis, in French.

Chapter 2

Certification of biometric systems

Contents

Summary : *Biometrics is known to be one of the most secure solutions used for the authentication and identification of people.*
To face the changes across the payment market in particular and take up the technological challenges of tomorrow, this thesis focuses on the biometric evaluation, especially on innovation for future methodologies involving synthesized generated spoof attacks, to evaluate the effectiveness of the biometric system and the effects of biases such as environmental conditions on biometric systems' performance.

2.1 Introduction

Nowadays, biometrics has become a reference tool for security and user experience. We can find it in smartphones, access to secured sites, etc. During the last decade, biometrics became a famous tool thanks to its simplicity and non-intrusiveness.
Originally, biometrics was the study of liveness. A biometric data is a characteristic that is unique and universal. Biometric characteristics are features that anyone has but are different from one person to another. Most of them are physiological: fingerprint, face, iris, retina, voice, etc. Recently, many studies have been done in the field of behaviors: the way of walking (gait), signature, keystroke dynamics, etc. We can encounter biometric usage in everyday applications such as banking, transport, law enforcement [Jain and Kumar, 2012, Walker, 2012] and public security, border, and migration control [Labati et al., 2016], civil identification, healthcare, physical and logical access, commercial applications, identity verification & binding, and so on.

However, due to the multiple sensitive operations relying on biometrics, biometrics systems must be secured, easy to use, and trustworthy. Thus, before deploying a biometric product, the latter has to be tested to assess its behavior and functioning. This step is done in a third-party laboratory that has been accredited by a test authority for the certification process.
An overview of a complete biometric system is given in Figure 2.1.

- The **data capture** module is composed of a sensor that captures a biometric signal. It can be a fingerprint sensor, a camera, a microphone, or any capture device that allows to capture properly the biometric signal,
- The **signal processing unit** is responsible for extracting any useful feature vector,
- A **database** is used for the storage of biometric data to be used as a reference for comparison,

- A **matching unit** performs the similarity measurement between a submitted biometric sample for identification or verification and the one(s) already stored in the database and gives a verdict.

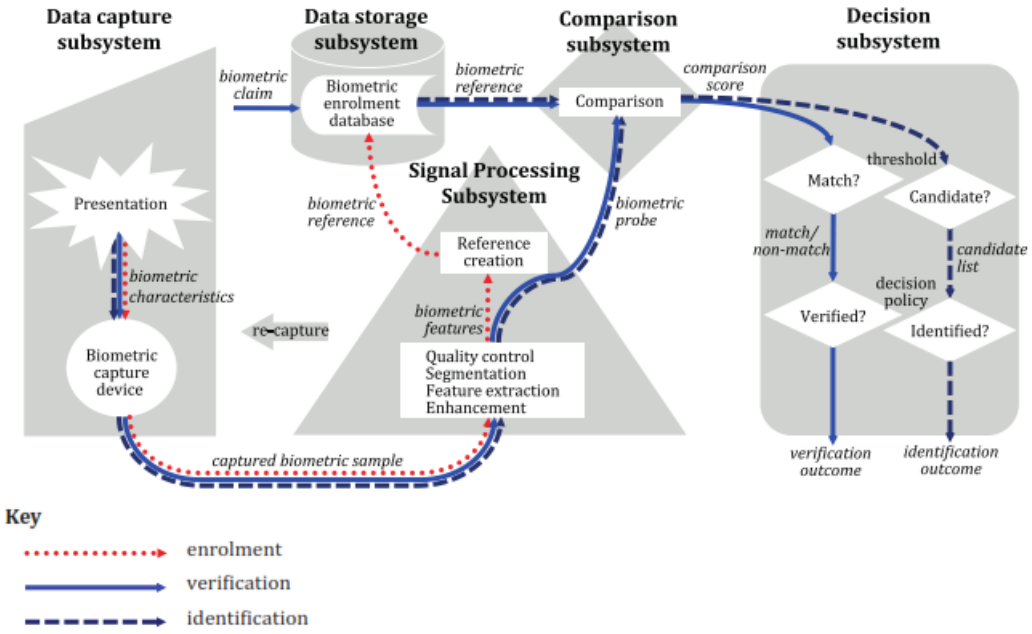


Figure 2.1: Main components of a biometric system from [ISO 19795-1:2021(E), 2021].

In biometrics and in this PhD thesis, we will encounter specific terms that we will first define:

- **enrolment**: phase where a genuine user is registered and his/her biometric data are stored prior to future comparison.
- **Verification**: one or more verification attempts resulting in the resolution of a biometric claim.
- **Authentication**: act of proving or showing to be of undisputed origin or veracity
- **Identification**: process of searching against a biometric enrolment database to find and return the biometric reference identifier(s)

A complete list of biometric-related terms and their definitions can be found in [ISO 2382-37:2022, 2022].

2.2 Biometric modalities

From the Greek 'bios' meaning 'life' and 'metron' meaning 'measurement', "Biometrics" is originally the measure of life. This term represents in modern days science that identifies people using their physiological or behavioral characteristics.

Biometrics is not just something that arose recently. Humans had always the faculty to recognize people by personal traits. Face is one of the most natural ways to identify people. Indeed, the human brain records the faces of people we know and associates an identity with each of them so we can recognize them the next time we meet. Facial recognition is in that way, one of the less invasive due to the history behind that. But we do recognize known people in other different ways such as their voice tone, their handwriting, the way they walk, etc. In modern times, technology offers the computation of parameters that make easier the identification of people. Fingerprint patterns start forming from the friction in the womb with the amniotic fluid.

In antiquity, people were identified by interpersonal recognition. With the evolution of technology, photography became a trusted way to identify people. To improve the proof of the identity, Alphonse Bertillon developed "Bertillonnage", a method to reinforce the identity of criminals and suspects by adding specific information like height, weight, head dimensions, etc. These anthropometric records are the first known official biometric identification. Examples of measured characteristics using that method are visible in Figure 2.2.

In modern days, with the growth of technologies, we are able to calculate and identify millions or billions of users using different types of biometrics. Different biometric modalities have been proposed. We mainly distinguish two types of biometrics: physiological biometrics and behavioral biometrics. Figure 2.3 shows some examples of biometrics by type of modality.

According to Jain [Jain, 2005], to be considered as a biometric characteristic, a physiological or a behavioral signal must respect the following properties:

- **Universality**: each person should have the characteristic;
- **Distinctiveness**: any two persons should be sufficiently different in terms of

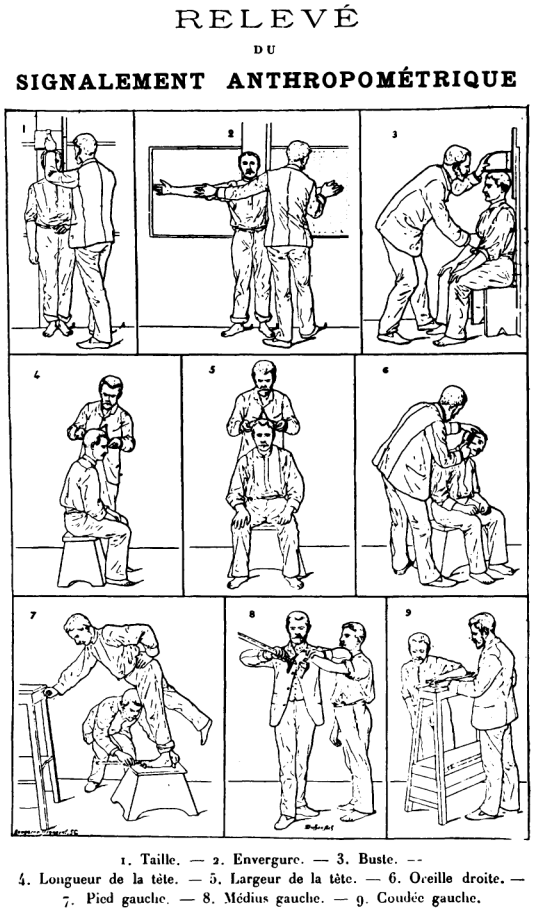


Figure 2.2: Overview of anthropometric record process [Bertillon, 1893].

the characteristic;

- **Permanence**: the characteristic should be sufficiently invariant (with respect to the matching criterion) over a period of time;
- **Collectability**: the characteristic can be measured quantitatively.

Table 2.1 shows the advantages and disadvantages of some of the most used biometrics.

Table 2.1: Comparison of few biometrics from [Jain et al., 1999].

Biometrics	Uniqueness	Collectability	Performance	Acceptability	Permanence
Fingerprint	High	Medium	High	Medium	High
Face	Low	High	Low	High	Medium
Iris	High	Low	High	Low	High
Voice	Low	Medium	Low	High	Low
keystroke	Low	Medium	Low	Medium	Low
Gait	Low	High	Low	High	Low

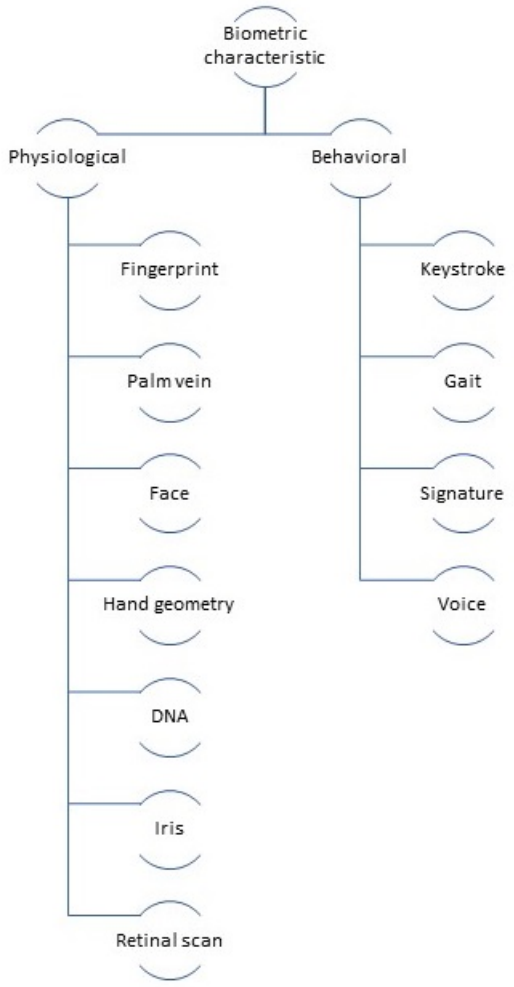


Figure 2.3: Examples of biometric modalities and their category.

2.3 Evaluation of biometric systems

The evaluation of a biometric system consists of a series of tests to establish its recognition capabilities and resistance to fraud. The results of the evaluation provide information on the usability and security level of the tested biometric system. It allows to know its capacity to recognize (authenticate or identify) the right persons and refuse impostors or non-legitimate users. Tested biometric systems can be full systems where a software is associated with hardware parts, as we can see in Figure 2.1. However, it can also be a software-only solution that is aimed to be compatible with a wide range of capture devices of that type of biometric or integrated into

another solution. These types of solutions are more and more present and deployed as the current matching algorithms are mostly based on deep learning solutions, their training is independent of any capture device.
Depending on the nature of the system under test, different types of evaluation are possible. A synthesis of different types of evaluations is given in Figure 2.4.

- **Technology evaluation:** offline evaluation of one or more algorithms for the same biometric modality using an existing or especially-collected corpus of samples. For these tests, the solution consists of an algorithm only. To test it, laboratories have to use existing datasets or collect datasets for this purpose.
- **Scenario evaluation:** evaluation that measures end-to-end system performance in a prototype or simulated application with a test crew. The evaluation is done on a database collected by the acquisition module of the biometric system or an existing dataset that has been processed to match the properties of the system.
- **Operational evaluation:** evaluation that measures the performance of a biometric system in a specific application environment using a specific target population.

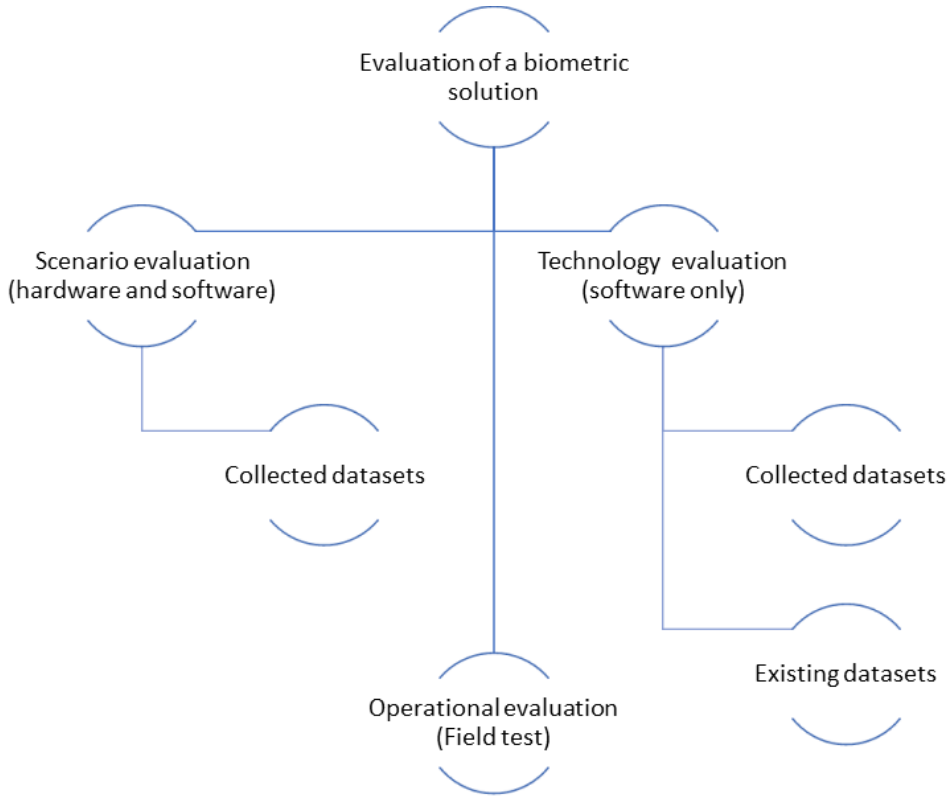


Figure 2.4: Different types of evaluations for a biometric system

The technology and scenario evaluations are the most common ones. The operational evaluation is very rare and is most of the time used to verify if a system is still working as expected and checks if the algorithm's performance has not decreased.

2.4 Databases

The sensitivity of biometric data and their private nature make their storage and retention difficult and subject to numerous regulations. In Europe, the collection, storage, and use of this data are governed by the General Data Protection Regulation (GDPR) [1]. However, there are few biometric datasets available on demand.

One can mention FVC datasets from a Fingerprint Verification Contest that exists since 2000. Few editions have followed since, one in 2002, 2004, and 2006 [Cappelli et al., 2007]. Besides FVC, the Liveness Detection (LivDet) [Ghiani et al., 2017] competition is a contest where academics and industrials are given datasets of fingerprint spoofs of different brands and technologies for training. They submit their trained anti-spoofing algorithms for evaluation on a testing dataset not seen during the training. Other datasets can be found in the literature [Ortega-Garcia et al., 2003]. These are some of the most used fingerprint datasets. More datasets for different modalities are available [2]. Almost all these biometric data are restricted to research use only and can not be used in an industrial context. Moreover, it is impossible to say whether they were used for the training of the solution or not as the tested solution is a black box. So, there is a risk of bias in the test verdict using these data. The lack of publicly available biometric datasets for testing and the biometric solutions being black boxes make the collection of biometric data for every product test a systematic operation.

2.5 Certification of biometric systems

The certification of a biometric system is done by an independent laboratory following instructions given by a certification body which can be specific to a domain, *i.e.* Mastercard or Visa for payment or more generic like Fido Alliance or Android for devices with Android Operating system. The certification of a biometric system gives an overview of the system quality, security, and user experience [El-Abed et al., 2012]. Two generic tests are mainly done during the certification of a biometric system: The

[1] `https://gdpr-info.eu/`

[2] `https://ieee-biometrics.org/index.php/resources/biometric-databases`

assessment of the performance of a biometric system and the presentation attack detection.

2.5.1 Performance of a biometric system

The performance testing of a biometric system gathers metrics that deal with the usability of a biometric system. It measures the accuracy of a biometric system and gives information on the user experience. The performance of a biometric system is expressed as error rates. The rates achieved by the system are compared to the requirements to assess whether the system is ready for deployment or not. The scenarios of evaluation of a biometric system's performance are mostly based on ISO 19795 [ISO 19795-1:2021(E), 2021]. However, tests can be done without certification bodies, but following ISO testing and recommended methods for evaluation.

Error rates

Error rates are percentage figures which indicate how the biometric system wrongly classifies a biometric sample. Some of them are more significant and are present in most of the biometric certification programs and shall be clearly indicated in the test reports.

- Failure-to-enrol rate (FTE): The FTE represents the proportion of users that are unable to enrol in a biometric system.
- Failure to acquire rate (FTA): The FTA represents the proportion of transactions for which a biometric system is not able to capture a sample to perform a verification task.
- False Acceptance rate (FAR): The FAR represents the proportion of zero-effort non-genuine transactions that will be incorrectly accepted.
- False Rejection Rate (FRR): The FRR represents the proportion of genuine verification transactions that are wrongly rejected.
- False Match Rate (FMR): The FMR represents the proportion of zero-effort transactions that wrongly match genuine users.
- False Non-Match Rate (FNMR): The FNMR represents the proportion of genuine verification transactions that do not match their corresponding enrolment.
- False-Positive Identification Rate (FPIR): The FPIR is the proportion of identification transactions by capture subjects not enrolled in the system for which a reference identifier is returned.
- False-Negative Identification Rate (FNIR); The FNIR is the proportion of a specified set of identification transactions by capture subjects enrolled in the

system for which the subject's correct reference identifier is not among those returned.

The FAR and FRR are referred to when talking about a complete biometric system whereas the FMR and FNMR refer to decisions due to the biometric matching algorithm only.
For a given test set of N users, considering that each user has done 10 verification transactions, the FAR or FMR testing consists of doing the comparison of each user with the N-1 others, 10 times. For the FRR or FNMR, each user is compared with himself/herself.

Depending on the target performance and/or level of certification, the size of the dataset may be different. Depending on the destination of the biometric system, the focus may be on the performance of the product or its resistance to attacks. The number of test subjects to be gathered for the performance test is computed according to the rule of 3 of ISO [ISO 19795-1:2021(E), 2021]. It states that the upper bound of the 95% is 3/C with C being the number of combinations. C is computed according to the following formula:

$$C = a^2 \frac{n(n-1)}{2} \tag{2.1}$$

where n is the number of unique test subjects and a is the number of unique samples per test subject (*i.e.* number of fingers, number of eyes, ...). For example, the most advanced level of certification of FIDO Alliance [Schuckers et al., 2023] requires 245 subjects for the performance tests with a minimum of 123 unique test subjects to meet a FAR requirement of 1:10K. This number can be lower if the targeted FAR is higher.

2.5.2 Presentation attack detection

The reputation of biometric systems as being secure leads researchers to work hard on methods of fooling them and succeeding the imposture attempts. This is done to detect vulnerabilities in Presentation Attack Detection (PAD) algorithms and improve them, so imposters or intruders would not achieve their breach attempts. In biometrics testing, different levels of attacks exist depending on the difficulty of creating the spoofs. PAD can be a hardware-based or software-based solution or hybrid. In the past years, many solutions have been proposed. [Marcel et al., 2023] proposed a complete review of anti-spoofing methods for biometric systems.

Evaluation of presentation attack detection

In the literature, multiple key points have been identified as potential access for attackers that are shown in Figure 2.5 but most of the certification schemes focus on the first attack point (*i.e.* Presentation attack). The presentation attack detection test is done to confirm the resistance of a biometric system to attacks.
A presentation attack detection test is done using a presentation attack instrument (PAI), a biometric characteristic or object used to fool a biometric system.

For most of the certification schemes, the testing of the PAD module is mainly based on ISO 30107[ISO 30107-3:2023, 2017].

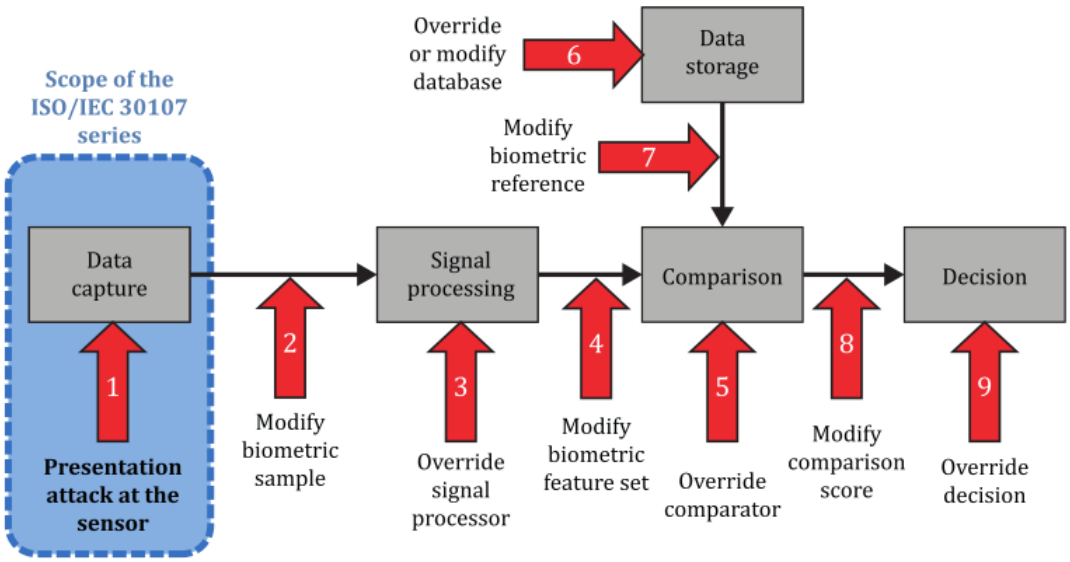


Figure 2.5: Different attack points from [ISO 30107-1:2023, 2023], inspired by [Ratha et al., 2001].

We focus on 2 physiological modalities as illustrations of presentation attack solutions.

- **Fingerprints**: Fingerprint attacks represent a huge challenge for the anti-spoofing system due to the complexity of fingerprint shape. For this reason, impostors do not have another choice than falsify fingerprints. Indeed, by inadvertence, we leave our fingerprints on almost all surfaces without knowing it. The most basic fingerprint attack consists of trying to reactivate latent fingerprints on the surface of the sensor. Due to high humidity, it is possible to reactivate latent fingerprints, especially on capacitive fingerprint sensors.

Biometric (fingerprint) spoofing involves the creation of fake fingerprint samples in order to spoof the identity of a real user. There are several techniques for making fake fingerprints in the literature. Anti-spoofing tests may require up to 14 types of attacks and may involve up to 15 test subjects, according to the latest FIDO Alliance biometric requirements [Schuckers et al., 2023]. In practice, in a testing scenario, there are two groups of sources depending on how the attacker obtains the biometric sources. Figure 2.6 synthesizes the two main types of sources widely used for testing of PAD solutions.

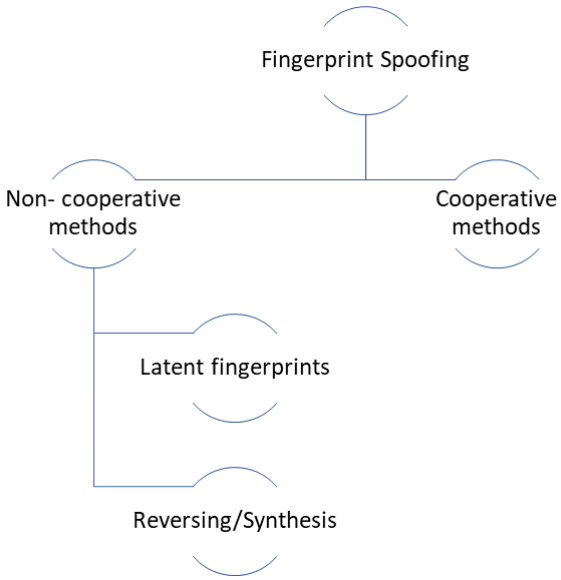


Figure 2.6: Sources and types of attacks

[ISO 30107-3:2023, 2017] gives other types of sources which can be cooperative, recording, regeneration from template, impersonation, and synthetic sample generation.

In **cooperative spoofing**, a genuine user willingly gives his/her biometric sources to the attacker/tester and challenges them to bypass the anti-spoofing system. [Marasco and Ross, 2014] gives a more detailed review of the methods used to make fake fingerprints and existing techniques to buffer against these kinds of attacks. Usually, a fingerprint is cast on some material like dental pasta or clay to make a negative of the fingerprint, a mold. A spoofing material such as gelatin is cast in the obtained mold to create a fake biometric sample that will be used to impersonate the genuine user. Figure 2.7 shows the main steps of presentation attack instruments (spoofs) creation in a cooperative

scenario.

Figure 2.7: Steps of spoof making in a cooperative mode from [Marasco and Ross, 2014], originally from Matsumoto `http://web.mit.edu/6.857/OldStuff/Fall03/ref/gummy-slides.pdf`

In the **non-cooperative methods**, the basic method consists of reactivating a latent fingerprint. The latent fingerprint can be used as a source to make a mold using a 3D printing method in order to create a negative of a finger.

- **Face**: For face recognition, access to the source is easier. Indeed, with the increasing presence of people on social networks, it is easier to get high-quality photos and videos of a person. The most basic spoof for face recognition is a picture of the targeted subject printed on paper or displayed on a screen. A level above is a video of the subject with some movements or facial expressions to fool the liveness detection. The anti-spoofing and spoofing growing mutually, the basic spoofs may not fool the latest technologies. High-quality spoofs have been created to be more realistic [Erdogmus and Marcel, 2013, Korshunov and Marcel, 2018, Rathgeb et al., 2022]. Examples of high-level 3D face masks are given in Figure 2.8.

Depending on the solution under test, there are 3 levels of PAD evaluation.

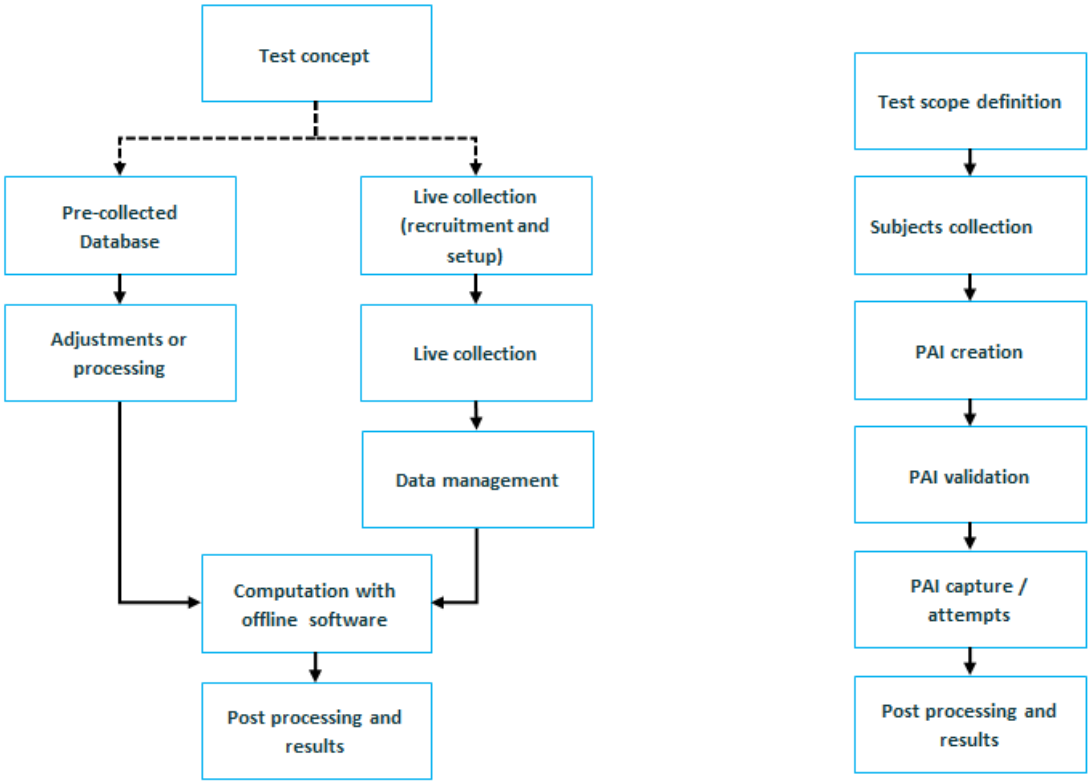
- **Data capture subsystem**: In that case, the reason for failure can be other than the submitted biometric sample failing to pass the liveness checking. The quality check can be the reason for instance.

Figure 2.8: Examples of face masks obtained from ThatsMyFace.com, from [Erdogmus and Marcel, 2013]

- **PAD subsystem**: this evaluation aims to measure the ability of the PAD module to recognize attacks and bona fide presentations and classify them as they should be. This test focuses only on the efficiency of the PAD module.
- **Full system**: This test measures the result of the whole system, integrating the verdict of the PAD subsystem and the matching operation.

The next section presents some of the most used metrics for the PAD evaluation.

PAD metrics



(a) Performance testing (b) PAD testing

Figure 2.9: Major steps to evaluate a biometric solution.

The main metrics used for the evaluation of presentation attack detection are the following:

- **IAPAR**: The Impostor Attack Presentation Acceptance Rate represents the proportion of impostor attack presentations using the same PAI species that result in accept. It can be found under the former name of IAPMR (Impostor Attack Presentation Match Rate) or SAR (Spoof Acceptance Rate).
- **BPCER**: The Bonafide Presentation Classification Error Rate is the proportion of bonafide transactions that are wrongly classified as presentation attacks by the anti-spoofing module.

- **APCER**: The Attack Presentation Classification Error Rate represents the proportion of attacks that successfully fool the anti-spoofing system and managed to be classified as bonafide.
- **IAPIR**: (Impostor Attack Presentation Identification Rate). In a full-system evaluation of an identification system, the IAPIR is the proportion of impostor attack presentations using the same presentation attack instrument (PAI) species in which the targeted reference identifier is among the identifiers returned, or, depending on the intended use case, at least one identifier is returned by the system.

The most advanced level of certification for FIDO Alliance biometric certification program needs 15 people with 14 different PAI recipes (6 being basic attacks and 8 being advanced attacks).

A summary of the testing flows of a biometric system is given in Figure 2.9.

2.6 Conclusion

This chapter presents the need to certify a biometric system. It presents the two main tests done in most of the certification programs, *i.e.* the performance testing and the presentation attack detection testing. However, due to the lack of publicly available data for tests and the non-transparency of the tested biometric solution, data have to be collected for the performance and the anti-spoofing tests. This has a cost. It requires time and skills to gather the data and make spoofs with a high potential for success. On the other hand, the fact of doing a "scenario" test limits the exploration of all possible configurations. Thus, it makes the tests barely repeatable.

During this PhD thesis, we address the problem of reproducibility of testing by mastering the context of dataset acquisition and the use of artificial intelligence in the testing of a biometric system by the simulation of large real/synthetic spoof datasets.

Chapter 3

Biometric data

Contents

Summary :
In the previous chapter, we introduced the concept of certification of biometric systems and the importance of this step prior to the production of a biometric system. As one can conclude, the certification of a biometric system is an important step in the life of a biometric device and can be viewed as the evaluation of its behavior on a collected database. Whether it is a scenario or operational testing, a biometric dataset is used to assess the recognition capabilities of the system under test and another one for the efficiency of its anti-spoofing algorithm. In this chapter, we present the fingerprint dataset that we dealt with during this thesis, the constraints they are associated with, and the protocol we use all along this thesis to validate our results.

3.1 Introduction

As stated in the previous chapter, the certification of a biometric system is based on scenario testing for a complete biometric system (hardware and software) or operational testing for an algorithm-only system. For better trust, the evaluation is done in accredited laboratories in certain conditions. A biometric dataset is used for this purpose. However, depending on the dataset, the results may differ. A good system should handle the variability of existing datasets, then the results should not vary from one dataset to another.

This chapter is organized as follows: Section 2 presents the collection of biometric data for tests as well as the associated constraints. Section 3 gives an overview of the state-of-the-art of fingerprint datasets. In section 4 we propose a generic method to validate synthetic biometric data.

3.2 Biometric data collection

The certification of a biometric system is done by means of a biometric database adapted to the product. The database can be built for this purpose (scenario testing) for a complete biometric system or an existing dataset for an algorithm alone. Depending on the biometric modality and the targeted performance values, the number of implicated users can be high. A FAR of 1:100,000 requires 775 test subject users according to the rule of 3 of ISO [ISO 19795-1:2021(E), 2021]. The collection of data is subject to the laws in force in the place of collection. Most of the existing datasets that can be found in the scientific literature are restricted to

research purposes only and cannot be used in the context of an industry certification of a product that is meant to be commercialized. However, the collection of biometric data comes with both legal and logistical constraints.

3.2.1 Legal constraints

Because of their personal and sensitive nature, biometric data must be protected and not accessible to everyone because raw biometric data are irrevocable. Beyond digital protection methods such as encryption, people need to be protected so that access to their data is restricted and the integrity of their data is maintained. In most countries, the collection of biometric data has a legal framework. In Europe, the General Data Protection Regulation (GDPR)[1] is the regulatory body for the protection of biometric data and personal data in general. GDPR is about the citizenship of the test subject and not the place of the collection. However, there are national bodies that watch personal data usage locally like CNIL (Commission Nationale de l'Informatique et des Libertés) the French data protection authority. Due to these policies, a consent form has to be signed between the collector and the volunteers.
The consent form has to be explicit about what the collected data are intended for, what are the volunteers' rights, how the data will be processed, how long it will be stored, ... An example of a consent form respecting French and European laws is given in Figure 3.1. The form is signed by both parties at the date of the collection. Major rights for the test participant and duties for the collector are reminded as well as different ways to get answers regarding their data. Figure 3.2 summarizes the steps for the data collection before it can be used for the test of a biometric solution.

Plus, the sensitivity of the biometric datasets, and contrarily to passwords that can be changed easily when compromised, biometrics need to be stored securely after transformation and revoked or canceled if the stored template is compromised. Due to legal constraints, labs must ensure everything is in conformity with the regulations which consequently create some operational constraints.

3.2.2 Operational constraints

The collection of datasets for biometric usage is subject to logistical constraints. Indeed, a meticulous organization is needed to carry out the biometric collection.

[1] `https://gdpr-info.eu/`

ID

Formulaire de consentement

Programme de certification biométrique à

Je, soussigné(e), [Nom/Prénom] _____

donne mon accord pour l'utilisation de mes données personnelles selon les conditions décrites dans le programme de certification biométrique à (« Programme Biométrique») qui m'a été présenté aujourd'hui, et ce pour la durée d'une certification et des contre-expertises nécessaires, soit la conservation desdites données jusqu'à 3 ans. J'ai compris que, si je souhaitais être informé de la destruction de mes données personnelles ou recevoir une copie du Programme Biométrique 2.0 d'Octobre 2022, je dois envoyer ma demande à l'adresse mail :

Obligations de

De notre part, s'engage à respecter les dispositions françaises et européennes relatives à la protection des Données Personnelles, à l'occasion de tout traitement de Données Personnelles. Ainsi, s'engage :

- à ne pas communiquer, diffuser et réutiliser les données personnelles, sous quelque forme que ce soit, en tout ou parties, en dehors du cadre définit dans le Programme Biométrique.
- à ne pas conserver les données personnelles au-delà de la durée définie dans le Programme Biométrique.
- à mettre en œuvre tous les moyens nécessaires en vue de garantir la sécurité et la confidentialité des données personnelles traitées, et à prendre toutes les mesures techniques et organisationnelles appropriées afin de restreindre les risques de perte, ou de mauvaise utilisation de celles-ci, ainsi que les risques d'accès aux données personnelles par des tiers non autorisés.
- à assurer que vous pouvez exercer vos droits sous l'Article 18, chapitre III du RGPD, qui sont notamment un droit d'accès, de rectification, de limitation, de portabilité, d'opposition, d'effacement de vos Données Personnelles. Vous pouvez exercer vos droits en envoyant un courrier au délégué à la protection des données de à l'adresse suivante :

Gratification et signature

Je reconnais avoir reçu au titre de gratification le montant de _____ € pour ma participation.

Fait à en 2 exemplaires, le Signature, précédée par la mention « lu et approuvé »

[Date] _____

Fime Participant

Figure 3.1: Example of a consent form for Biometric data collection in French where people have to give their name and surname, mention if they were paid.

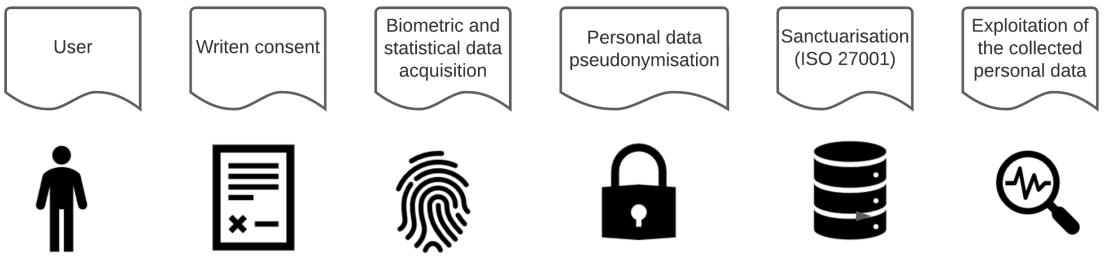


Figure 3.2: Different steps for the collection of Biometric data

Participants have to be recruited and sometimes convinced by a reward. Thus, the reward can very quickly reach a substantial amount which will be reflected in the cost of the test. Volunteers need to be trained in the use of the product or given

sufficient time to familiarise themselves with the product, which can lengthen the collection time.

A typical collection of biometric samples can involve two test operators and last one week or more. It starts with a call for volunteers. In that step, a test crew of the needed number of volunteers for the requested test is hired. The test crew should respect some constraints in terms of composition (age, gender, ethnicity distributions, size, ...) depending on the targeted evaluation and performance figures. Once the needed volunteers are found, the next steps are typically followed:

1. An explication of the volunteers' rights regarding their biometric data is given. This step is really important as it is the main step to explain the duties of the collecting company regarding the personal biometric data of the volunteers. People are given a typical test program explaining all the steps for the exploitation and storage of the to-be-collected data. Data will be stored in encrypted computers and hard drive disks, with access granted to a few people. The data are stored for a reasonable time, typically 3 years to do all the needed tests and expertises that can be required.

2. An explication of the concerned test is given. The test operator explains the type of testing and the concerned type of biometric modality. Usually, the data are collected in one visit but it can also be done in two or more visits. The last case happens when the collection is time-consuming often due to the fact that we want to do the test in real conditions (an in-car collection for example) and recreate the conditions in the lab.

3. Training of the test subjects: Depending on the test methodology, very detailed and full guidance can be provided or just a minimal one for certification schemes like Fido Alliance which suggests leaving the test subject to interact naturally with the product. Whether it is a fully guided collection or not, the test crew is given enough time to be familiar with the product.

4. Anonymity: People get a random number, and only know of them. This number can be written in the volunteers' copy of the consent form. It is not known from the laboratory and can be reminded for every question they may ask to the Data Protection Officer (DPO) regarding their data for exploitation, correction, or deletion.

5. Collection of data: Once people have agreed to give their biometric data, it can be collected using the product under test. Data are stored under the

given number. The format of the data depends on the biometric solution. It can be readable (non-encrypted images) or encrypted (templates) of the biometric samples. The data are stored in a format that can be exploited by the customer's solution for offline usage.

6. Once data are collected, the exploitation can start with the cleaning of the datasets to fix mislabelling or isolate unusable data for example. Then, the computation and the reporting of metrics associated with the test session starts. When the test session is completed and the test report is issued, the data are definitely wiped using a specific tool that makes the retrieval of the deleted data nearly impossible. Volunteers who expressed their will to know when their data are deleted are notified by email.

In addition, a collection of biometric datasets requires a well-defined schedule and dedicated operators. Biometric data collection is therefore a time-consuming and costly task. The next section presents the existing fingerprint datasets.

3.3 State-of-the-art of fingerprint datasets

Most of the existing datasets that can be found in the literature are restricted to research purposes only and cannot be used in the scope of an industrial certification of a product that is meant to be commercialized. Moreover, as mentioned in the previous chapter, most biometric systems are black boxes. As a result, their training method remains unknown from the point of view of a test laboratory, and it is impossible to know whether the semi-public databases that exist were used during this training. Figure 3.3 illustrates the different types of data that can be found in the field of biometrics and how they can be used.

3.3.1 Real data

In this section, we focus on real datasets, acquired from humans. They have been created for the performance and robustness evaluation of fingerprint systems.

Performance evaluation

Depending on the biometrics, some have more public data than others. Fingerprint-based biometrics is represented by a few semi-public university databases available on demand for research purposes only. The most widely used databases in biometric testing of fingerprint systems for performance testing (*i.e.* FAR and FRR) are the

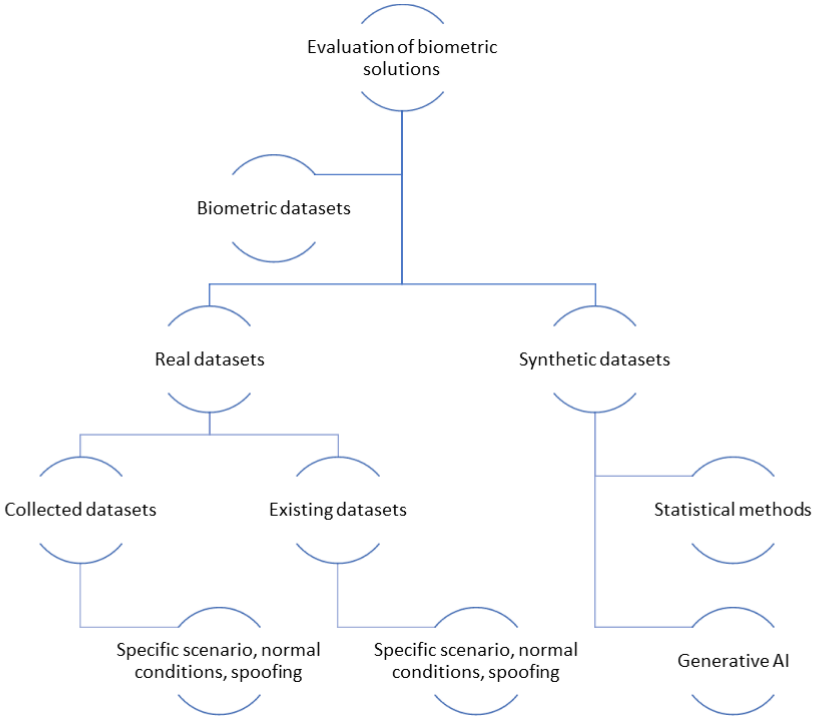


Figure 3.3: Diagram of the types of data that we can find to evaluate a biometric solution.

FVC databases[2],[3],[4],[5].

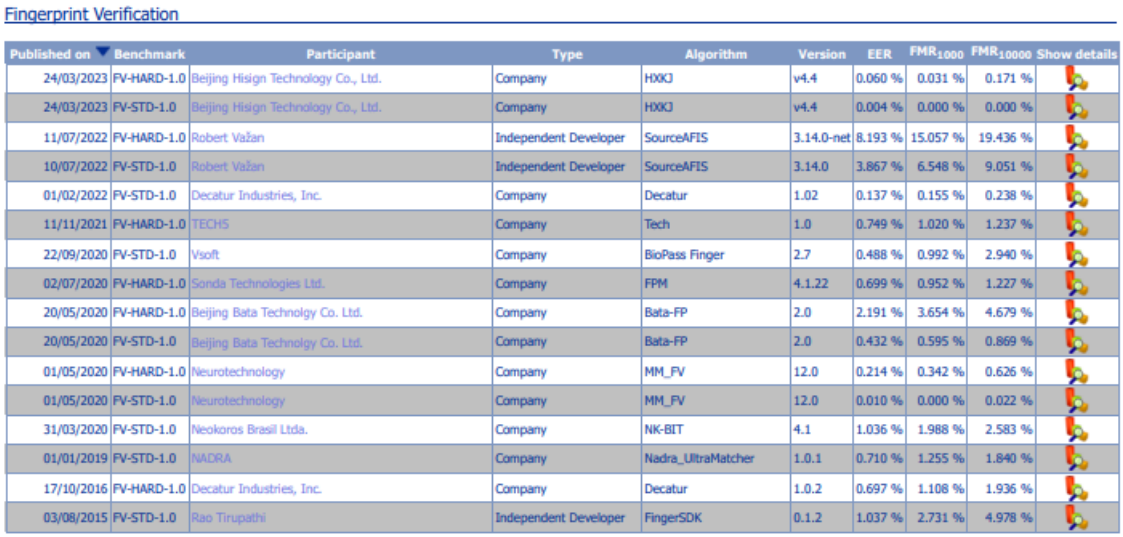
As previously mentioned, FVC is a fingerprint Verification Contest that has existed since 2000 and organized by the Biometric System Laboratory (University of Bologna), the U.S. National Biometric Test Center (San Jose State University), and the Pattern Recognition and Image Processing Laboratory (Michigan State University). Few editions have followed since (2002, 2004, and 2006). They provide fingerprint datasets to test the performance of competitors' algorithms. The datasets are usually made of fingerprint images from capacitive sensors, optical sensors, and synthetic data from SFinGe[Raffaele et al., 2004]. Table 3.1 gives an overview of the 2006 database for illustration.

A continuation of this work is the availability of **FVC Ongoing** platform[6]. It is a web platform where companies, researchers, and individuals can submit their fingerprint-matching algorithms to be evaluated. The results are published privately

[2] http://bias.csr.unibo.it/fvc2000/
[3] http://bias.csr.unibo.it/fvc2002/
[4] http://bias.csr.unibo.it/fvc2004/
[5] http://bias.csr.unibo.it/fvc2006/
[6] https://biolab.csr.unibo.it/fvcongoing/UI/Form/Home.aspx

Table 3.1: Details of FVC 2006 dataset from the competition website

Dataset	Sensor Type	Image Size	Set A size	Set B size	Resolution
DB1	Low-cost Optical Sensor	300x300	100 x 8	10 x 8	500 dpi
DB2	Low-cost Capacitive Sensor	256x364	100 x 8	10 x 8	500 dpi
DB3	Optical Sensor	448x478	100 x 8	10 x 8	500 dpi
DB4	Synthetic Generator	240x320	100 x 8	10 x 8	around 500 dpi



Fingerprint Verification

Published on	Benchmark	Participant	Type	Algorithm	Version	EER	FMR_{1000}	FMR_{10000}	Show details
24/03/2023	FV-HARD-1.0	Beijing Hisign Technology Co., Ltd.	Company	HXKJ	v4.4	0.060 %	0.031 %	0.171 %	
24/03/2023	FV-STD-1.0	Beijing Hisign Technology Co., Ltd.	Company	HXKJ	v4.4	0.004 %	0.000 %	0.000 %	
11/07/2022	FV-HARD-1.0	Robert Važan	Independent Developer	SourceAFIS	3.14.0-net	8.193 %	15.057 %	19.436 %	
10/07/2022	FV-STD-1.0	Robert Važan	Independent Developer	SourceAFIS	3.14.0	3.867 %	6.548 %	9.051 %	
01/02/2022	FV-STD-1.0	Decatur Industries, Inc.	Company	Decatur	1.02	0.137 %	0.155 %	0.238 %	
11/11/2021	FV-HARD-1.0	TECHS	Company	Tech	1.0	0.749 %	1.020 %	1.237 %	
22/09/2020	FV-STD-1.0	Vsoft	Company	BioPass Finger	2.7	0.488 %	0.992 %	2.940 %	
02/07/2020	FV-HARD-1.0	Sonda Technologies Ltd.	Company	FPM	4.1.22	0.699 %	0.952 %	1.227 %	
20/05/2020	FV-HARD-1.0	Beijing Bata Technolgy Co. Ltd.	Company	Bata-FP	2.0	2.191 %	3.654 %	4.679 %	
20/05/2020	FV-STD-1.0	Beijing Bata Technolgy Co. Ltd.	Company	Bata-FP	2.0	0.432 %	0.595 %	0.869 %	
01/05/2020	FV-HARD-1.0	Neurotechnology	Company	MM_FV	12.0	0.214 %	0.342 %	0.626 %	
01/05/2020	FV-STD-1.0	Neurotechnology	Company	MM_FV	12.0	0.010 %	0.000 %	0.022 %	
31/03/2020	FV-STD-1.0	Neokoros Brasil Ltda.	Company	NK-BIT	4.1	1.036 %	1.988 %	2.583 %	
01/01/2019	FV-STD-1.0	NADRA	Company	Nadra_UltraMatcher	1.0.1	0.710 %	1.255 %	1.840 %	
17/10/2016	FV-HARD-1.0	Decatur Industries, Inc.	Company	Decatur	1.0.2	0.697 %	1.108 %	1.936 %	
03/08/2015	FV-STD-1.0	Rao Tirupathi	Independent Developer	FingerSDK	0.1.2	1.037 %	2.731 %	4.978 %	

Figure 3.4: Public results of the FVC ongoing benchmark.

with the report of performance metrics. Submitters have the choice to make public the results of the evaluation. The algorithm benchmark is open to other modalities such as palm vein, face morphing detection, ICAO compliance, etc. An example of the results of the benchmark for fingerprint algorithms is given in Figure 3.4 where EER, FMR100, and FMR10000 are used as metrics of performance. FMR10000 represents the lowest FNMR for FMR0.01%. More details on the test results (ROC and DET curves, distribution of scores) are visible with the "details" button.

CASIA fingerprint datasets [cas, a, cas, b, cas, c] is another widely used dataset. CASIA is composed of 3 large datasets. The first one [cas, a] is the *ATVS-FFp DB* which is made of genuine fingerprint images from 17 users with 4 fingers per user obtained with cooperation on 3 fingerprint sensors. A subset of fake fingers acquired with the same fingerprint sensors exists. The second CASIA dataset [cas, b] is *CASIA Fingerprint Subject Ageing* which contains data from the same test subjects with a capture interval of 4 years (2009 and 2013). The first sub-dataset was captured using one fingerprint device and the second one using 3 sensors (including the 2009

model). The third CASIA dataset [cas, c] is *CASIA-FingerprintV5* which contains 20,000 fingerprint images from 500 unique subjects with 40 samples from 8 fingers each. The dataset was acquired using one fingerprint sensor model during one session.

The NIST (National Institute of Standards and Technology) has freely available datasets. [Fiumara et al., 2019] contains Nail-to-nail fingerprint images that can be used for training and testing tasks. However, it is impossible to predict how long it will remain as the Special Database 4 [SD4,][7] which was widely used has been removed from the website.

Robustness to attacks

Another widely used dataset is LivDet [Ghiani et al., 2013, Ghiani et al., 2017], a liveliness detection competition that has known several editions in the past years. It provides a large dataset from different materials and sensing technologies for anti-spoofing training and testing. The competition is open to academic researchers and industrials. The table presented in Figure 3.5 summarizes the different materials for all editions of the LivDet competition and Table 3.2 gives the accuracy of competing algorithms during the 2017 LivDet contest. The first column shows the submitted algorithms, the last column is the overall TDR (True Detection Rate, number of attacks correctly classified by the presentation attack detection) and the other columns are the sensors used to capture the dataset.

[Grosz et al., 2020b] used Livdet 2015 to train and validate the proposed solution. [Grosz and Jain, 2022] used LivDet 2013 and LivDet 2015 to validate their solution. MSU FPD (Michigan State University Fingerprint Presentation Attack Dataset) [Chugh et al., 2018] is a dataset for spoof training and testing that utilizes two sensors and 4 different spoof materials and the MSU-FPD V2 is made of 12 materials.

Summary

A list of existing and most used datasets for fingerprint recognition-related tasks is given in Table 3.3. These datasets are useful for the evaluation and training of biometric solutions. However, they cannot be used for specific scenario evaluations. Moreover, they are often limited to research purposes only and cannot be used in an industrial evaluation program.

[7] `https://www.nist.gov/srd/nist-special-database-4`

		2009	2011	2013	2015		2017		2019		2021	
Material	**Type**	-	-	-	TR	TS	TR	TS	TR	TS	TR	TS
Body Double	Siliconic		•	•	•	•	•		•			•
Ecoflex	Siliconic		•	•	•	•	•		•			
GLS20	Siliconic											•
Modasil	Siliconic			•								
OOMOO	Siliconic					•						
Play-Doh	Siliconic	•		•	•	•						
RFast30	Siliconic											•
RProFast	Siliconic										•	
RTV	Siliconic					•						
Silicone	Siliconic	•	•									
Latex	Rubber		•	•	•	•		•	•		•	
Gelatine	Hybrid	•	•	•		•		•	•			
Liquid Ecoflex	Hybrid					•		•		•		
Mix1	Hybrid									•		•
Elmers Glue	Glue											•
Mix2	Glue									•		
Woodglue	Glue		•	•	•	•	•		•			

Figure 3.5: From [Ghiani et al., 2017]: Materials characteristics and frequency over the seven LivDet editions. The train and test materials were completely separated from 2017 to examine the PADs' resilience against "never-seen-before" materials. "TR" means Training set and "TS" testing set

3.3.2 Synthetic data

Synthetic data refers to every type of data that has been produced artificially. In opposition to real data collected from living test subjects, these data are digitally created using a synthesis method. Synthetic data allows us to bypass the challenge of building large datasets to test biometrics while resolving the problem of data privacy as the generated data do not belong to living beings. Thus, much work has been done by the community to move towards more and more realistic generation models. There are mainly two methodologies to create synthetic fingerprints: statistical and deep-learning methods. We detail each approach in the following sections.

Statistical methods

The first and oldest methods for synthetic fingerprint generation are based on statistical modeling of fingerprint features. These models try to imitate the key

Table 3.2: Accuracy of the algorithm that participated in the 2017 LivDet competition. The TDR is used as an accuracy metric. The table is from [Mura et al., 2018]

Algorithm	Green Bit	Digital Persona	Orcanthus	Overall
SSLFD	93.58	94.33	93.14	93.68
JLW_A	95.08	94.09	93.52	94.23
JLW_B	96.44	95.59	93.71	95.25
OKIBrB20	84.97	83.31	84.00	84.09
OKIBrB30	92.49	89.33	90.64	90.82
ZYL_1	95.91	95.13	91.66	94.23
ZYL_2	96.26	94.73	93.17	94.72
$SNOTA2017_1$	95.03	91.26	91.58	92.62
$SNOTA2017_2$	94.04	86.72	86.74	89.17
ModuLAB	94.25	90.40	90.21	91.62
ganfp	95.67	93.66	94.16	94.50
PB_LivDet_1	93.85	89.97	91.85	91.89
PB_LivDet_2	92.86	90.43	92.60	91.96
hanulj	97.06	92.34	92.04	93.81
SpoofWit	93.66	88.82	89.97	90.82
LCPD	89.87	88.84	86.87	88.52
PDfV	92.86	93.31	N.A.	N.A.

Table 3.3: Some of the widely used fingerprints datasets and possible usages

Dataset	Unique fingers	number of images	Performance	PAD
FCV2000_A	110	880	x	
FVC2006_A	150	1800	x	
NIST SD302	2,000	2,5093	x	
LivDet			x	x
MSU FPAD				x
MSU FPAD v2				x
CASIA ATVS-FFp	68 + 64	816 + 768	x	x
CASIA Fingerprint Subject Ageing	196	15,680	x	
CASIA-FingerprintV5	2,500	20,000	x	

features of a real fingerprint image sample. Historically, SFinGe [Raffaele et al., 2004] (Synthetic Fingerprint Generator) is one the first and most known fingerprint generator models. It was proposed by researchers from the University of Bologna in 2002 and is based on the mathematical modeling of fingerprint characteristics. The fingerprint generation using SFinGe can be summarized as follows:

- Directional map generation,
- Density map generation,
- Ridge pattern generation,
- Noising and Rendering.

The main shape of the fingerprint is elliptical segments. SFinGe applies a mathematical ridge-flow model from Sherlock and Monro [Sherlock and Monro, 1993] to the positions of the singularities to generate a directional map. Filters similar to the Gabor filters are applied to a white image with random points. The filter's orientation and frequency are locally adjusted according to the directional and density maps which makes appear realistic minutiae. Other effects such as dilatation, erosion, and noise are added to make the generated fingerprint look more realistic. Figure 3.6 gives a visual representation of these steps and Figure 3.7 shows the interface of the software and the different options offered by the software.

(a) Singularities (b) Directional map (c) Ridge pattern (d) Denoising (e) Final

Figure 3.6: Different steps of SFinGe generation in (a) and (b) a Directional map is generated, a ridge pattern is created in (c) and Denoised in (d) to give the final image in (e).

SFinGe is well-known in the field of fingerprints. A set of fingerprint images is always given in the FVC competitions which testifies how realistic fingerprint samples from SFinGe are. Indeed, the validation of the generation methodology was done on the FVC2002 datasets where the results of the dataset generated using SFinGe are similar to those from DB1, DB2, and DB3 made of real data collected from living beings.

Another generator based on statistic modeling is Anguli [Ansari, 2011]. Anguli defines itself as a project motivated and based on SFinGe. The software can be downloaded freely [8]. The authors validate the generation methodology by comparing

[8] `https://dsl.cds.iisc.ac.in/projects/Anguli/`

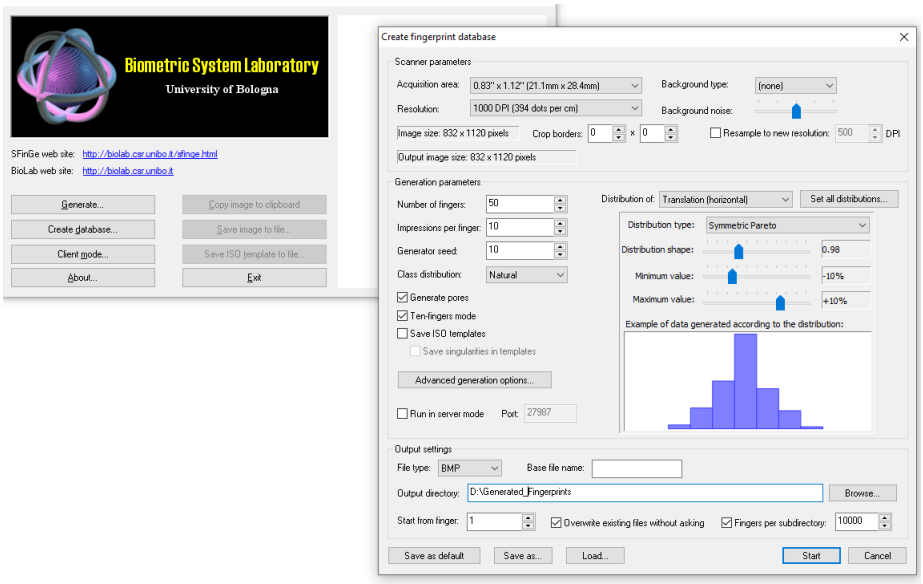


Figure 3.7: A view of the sFinge generator interface.

the performance (distribution of scores and DET curves) of the generated dataset and comparing it to real datasets.

[Zhao et al., 2012] proposed a method of generating fingerprints with multiple impressions which can be summarized in these steps: sampling fingerprint features from statistical models, generating a master fingerprint, generating multiple impressions from the master fingerprint, and rendering fingerprint images. These steps are illustrated in Figure 3.8. The method is composed of four main modules: (a) sampling features (singular points, orientation field, and minutiae) from appropriate statistical feature models; (b) generating a master fingerprint; (c) generating multiple fingerprint impressions from the master fingerprint via distortion (one such impression is shown here); and (d) rendering fingerprint images by simulating finger dryness and adding noise. The authors validate the proposed method by comparing the performance of generated data to real data from the NIST - SD4[SD4,] dataset.

Authors in [Johnson et al., 2013] proposed a method based on texture modeling. The validation of the proposed method is done through a comparison of extracted features from synthesized data and data from existing datasets which shows similarity with features distribution using FVC2004.

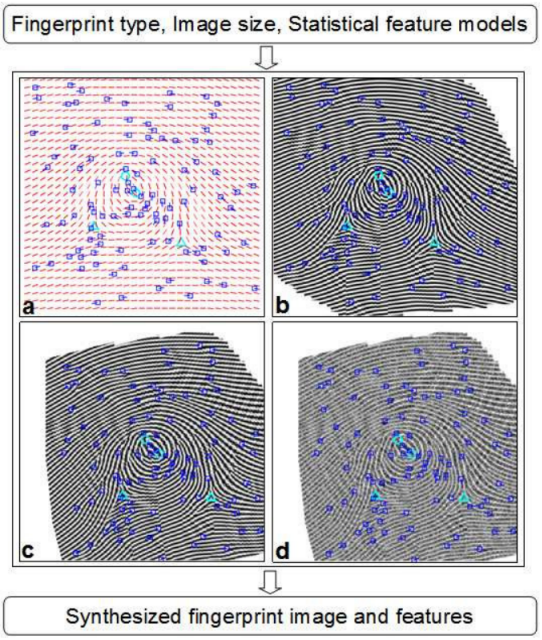


Figure 3.8: Fingerprint image synthesis method from [Zhao et al., 2012].

Deep Learning based methods

F. Chollet [Chollet, 2021] defines Artificial Intelligence (AI) as "the effort to automate intellectual tasks normally performed by humans.". The first versions of AI were designated as symbolic AI because programmers wrote rules and got answers for given data. Indeed, the AI is given data and answers and has to learn the patterns from entry and create its own rules. which can be applied to new data; unseen ones.

Among all possible AI, Generative Adversarial Networks (GANs) are the closest to what we are looking for. GANs first have been introduced in 2014 by a group of researchers from the University of Montréal [Goodfellow et al., 2014]. Many papers and versions of the GANs have followed since with various applications[9].

A global architecture of GANs is given in Figure 3.9. GANs are mostly used in computer vision to generate or translate images, enhance images, etc. but many applications have been found since and searchers are doing great work in this topic making GANs nets more effective and more versatile. GANs are composed of two modules: a generator and a discriminator. The generator G after training generates images from noise while the discriminator D estimates the probability of a sample to be from G. The challenge is to bring D to take generated samples as samples from

[9] `https://github.com/hindupuravinash/the-gan-zoo`

the training set, *i.e.*.which implies generating realistic samples that perfectly look like samples from the training data set. The whole goal is to win the Mimax game which can be summarized by:

$$\min_G \max_D \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[1 - \log D(G(z))] \tag{3.1}$$

G and D modules are built over with perceptron multilayers architecture and the training is done with backpropagation. The work in the field of biometric generation in general and fingerprint generation, in particular, is using GANs or their variants.

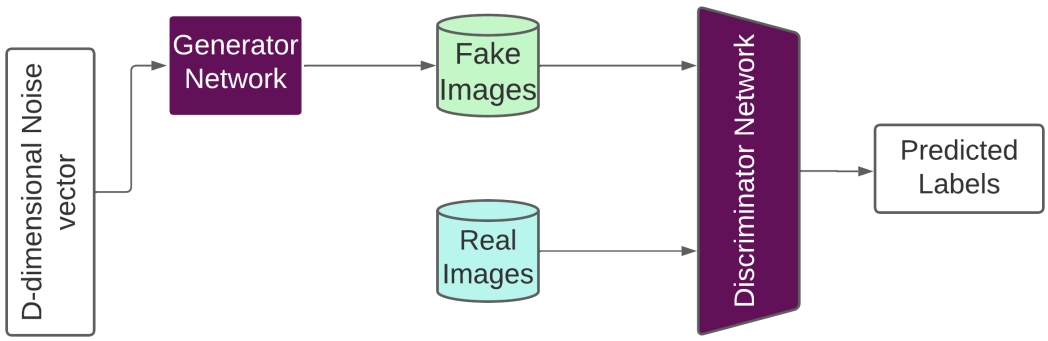


Figure 3.9: A global architecture of a GAN inspired by [Minaee and Abdolrashidi, 2018]

Existing works on fingerprint generation for performance and spoof generation purposes are extensive, so we have chosen only those that seem to offer the best results at the time this chapter is written. In opposition to statistical methods where the models are only focusing on generating realistic "genuine" fingerprints (to be used for performance evaluation only), deep-learning methods offer the possibility not only to create fingerprints to be used to evaluate the matching task but also realistic PAIs for liveness detection.

Model for "genuine" fingerprints generation:

As stated previously, these models aim to generate fingerprint datasets that can be used to train, test, and evaluate fingerprint-matching solutions instead of (or with) datasets from living test subjects.

One of the first studies to use generative Deep-Learning models to create fingerprint datasets was [Minaee and Abdolrashidi, 2018]. The authors used a Deep Convolutional GAN trained on FVC 2006 and PolyU [Zhao et al., 2010] to generate realistic

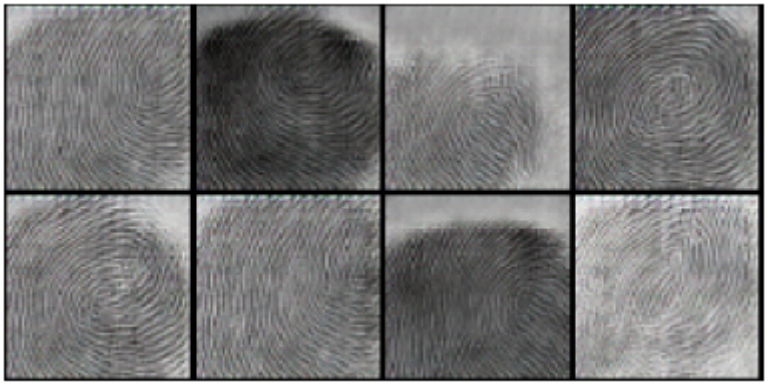


Figure 3.10: Examples of images generated by Finger-GAN[Minaee and Abdolrashidi, 2018].

fingerprint datasets. Figure 3.10 shows examples of images generated by Finger-GAN after training on PolyU datasets. The authors validate their methods by computing the Frechet Inception Distance (FID) which compares statistics of the generated samples to real samples.

[Mistry et al., 2020] proposes a generation method based on I-WGAN [Chen et al., 2022] with identity loss function using fixed-length fingerprint representation. They trained a convolutional auto-encoder (unsupervised CAE) to transform the fingerprint image to latent vector z' (encoder) and again into image (decoder). The trained CAE decoder is used to initialize the Generator of the I-WGAN. The authors used 8 metrics (including matching test and identification accuracy) to validate their methodology of fingerprint generation. They conclude that 7/8 metrics show that the proposed method gives more realistic fingerprints than state-of-the-art works by the time their work was done.

[Wyzykowski et al., 2021] uses Anguli to create multiple instances of fingerprint and cycleGan [Isola et al., 2017] to transform seed images to realistic fingerprints. They validate the proposed method using Bozorth3 [Ko Kenneth, 2007], a pore-based matcher, and human perception using EER as a metric of performance. [Seidlitz et al., 2021] is using 3 GAN (ProgressiveGAN [Karras et al., 2017], StyleGAN [Karras et al., 2019] and StyleGAN2 [Karras et al., 2020]).

Models for Spoof generation:

As for datasets for performance evaluation of fingerprint systems, some works focus on the generation of datasets to be used for the training, testing, or evaluation of PAD or liveness detection. The classical ways of testing biometric systems request the production of physical PAI. Depending on the certification body, the process uses the cooperation of the subject or not. Some of the most used materials are shown in Figure 3.11.

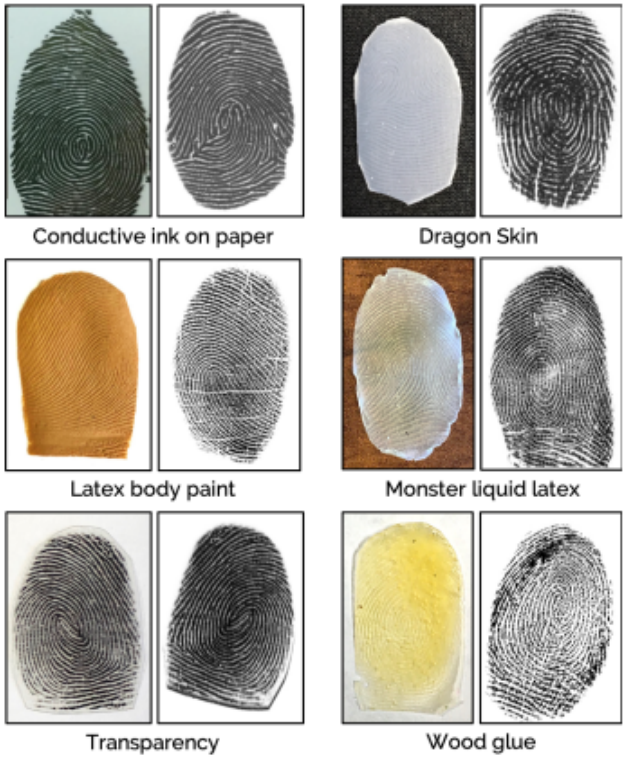


Figure 3.11: Example of physical spoofs or PAI. Images have been captured from that material by a fingerprint sensor (source [Chugh et al., 2018])
.

As stated previously, the existence of legal and operational constraints is not for performance evaluation only. Indeed, for the PAD, the lab is holding sources and presentation attack instruments (PAIs) or spoofs from the volunteers which requests high security for the storage and integrity of the data. Besides the generation of digital biometric data, many researchers have proposed new solutions for generating synthetic PAI samples for training and testing PAD systems (more intelligent data augmentation solution than rotation/translation operations classically used in deep

learning frameworks).

Authors in [Chugh and Jain, 2019] proposed a deep learning method for the generation of synthetic fingerprint PAI samples to deal with unseen materials during the training of a PAD system. This method gives a TDR (True Detection Rate) of 91.78%. The authors of [Bouzaglo and Keller, 2022] proposed recently the SynFing method based on StyleGan2 [Karras et al., 2020] to address the generation of fingerprints and more specifically the reconstruction of latent and rolled fingerprints. The method proposed by [Chugh and Jain, 2019] is a continuity of the universal material generator proposed by authors in [Gajawada et al., 2019].

Authors in [Grosz et al., 2020b] proposed a cross-sensors and cross-materials generation of anti-spoofing. This method achieved a TDR of 87.86%. Engelsma et al. proposed the PrintsGAN method [Engelsma et al., 2022] as a generator of synthetic fingerprints with different impressions. They concluded that their matching algorithm had a true acceptance rate of 87.03% on a real dataset when trained with synthetic images against 73.37% when trained without them. This shows the interest in having synthetic images during the training phase of matching algorithms.

Authors in [Priesnitz et al., 2022] proposed a generation of contactless fingerprint samples using SFinGe images. They applied on these images geometrical transformation, skin characteristics, and environmental influences. They validated the proposed method using quality metrics. A complete state-of-the-art of current work on the generation of synthetic data in the field of biometrics can be found in [Joshi et al., 2022]. More recently, Grosz *et al.* [Grosz and Jain, 2022] proposed SpoofGAN, a model that synthesizes multi-attempt spoofs from spatial modifications (translation, rotation, and non-linear transformations) of a master print.

Table 3.4 gives some of the state-of-the-art works for fingerprint spoof generation. One of the main difficulties in generating synthetic PAI samples is the validation process.

Study	Method	Dataset	Results
Gajawada *et al.* [Gajawada et al., 2019]	Universal Material Translator AdaIn module	LivDet2015	TDR=78.05%
Chugh *et al.* [Chugh and Jain, 2019]	Universal material generator	MSU-PAD	TDR=91.78% @FDR=0.2%
Grotz *et al.* [Grosz et al., 2020b]	Style transfer with a few samples of target sensor + fingerprint images adversarial representation learning	LivDet 2015	TDR=87.86% @ FDR=0.5%
Grotz *et al.* [Grosz and Jain, 2022]	Master print + Deformations and Texture	LivDet 2013,LivDet 2015, GCT 1-5, GCT 6	-

Table 3.4: some of the state-of-the-art works for fingerprints presentation attack instruments generation

3.3.3 Discussion

In the previous sections, we presented the use of real datasets as well as the two major approaches to generating biometric data, particularly fingerprint datasets for the evaluation of a biometric solution. The real data are the most representative solution whether it is a larger dataset for the performance evaluation or a small one for PAD testing. However, as we stated earlier, the collection of biometric samples comes with constraints that do not facilitate the testing of the biometric solution. Plus, it is quite impossible to say if the available dataset was not used for the training of the biometric solution. Indeed, most of the tested biometric solutions are black boxes and it is not fair to use such datasets for conformance testing.

Regarding the generation of biometric samples, statistical approaches which historically are the first ones to bring a huge contribution to the field of biometrics. It allows to train easily algorithms for recognition tasks and quick in-home tests. However, few papers used a quality assessment process and most of them consider the increase in performance of PAD systems when synthetic PAI have been used during the training [Husseis et al., 2019, Boyd et al., 2020]. The statistical methods allow the creation of realistic fingerprint datasets based on modeling of the statistical features of real fingerprint databases whereas the deep learning methods learn to encode the features of real fingerprints and generate new fingerprints. The Deep-Learning methods offer more possibilities regarding the custom of the algorithms (loss function, master of the latent space, etc.). The generation of fingerprint spoofs that we will present in this thesis are based on Deep-Learning solutions.

The biometric data generated or created through a collection process on living beings needs to have a qualification regarding its interest in testing usage. Indeed, a validation protocol is needed to qualify biometric data and interpret the results obtained using such data.

The next section presents the methods used in the literature for the validation of synthetic data generation.

3.4 Evaluation of synthetic data

This section describes the protocol we use throughout this thesis to validate our results on synthetic data generation considering works in the literature. In order to position ourselves in relation to the existing work, we have decided that we must be

able to qualify our results both in terms of the quality of our biometric data and their recognition and results through an anti-spoofing test. We considered the different validation approaches from existing works as detailed in the previous section.

3.4.1 Quality assessment

The metric we use in this work is the NFIQ2 (NIST Fingerprint Image Quality) tool [Bausinger and Tabassi, 2011]. NFIQ2 gives an overall score based on the usability and features of a fingerprint image. Scores go from 0 to 100 (0 bad and 100 good). NFIQ2 score is based on the nature of the image, the fidelity to its source, and the utility of the sample. Indeed, the output score reflects the contribution of the given biometric sample to the overall performance of a biometric system. The tool has been developed thanks to the contribution of multiple algorithm providers and larger fingerprint datasets. This metric is used here to measure how the quality of generated datasets is similar to the quality of real datasets with an objective metric that measures the way that the generated samples would contribute to the performance of a biometric system. Indeed, synthetic data can be considered as valid data if quality of the synthetic data is similar to the quality of real data in the same scenario.

3.4.2 Performance evaluation

The performance evaluation of biometric systems is generally measured using two metrics: the AUC (Area Under the ROC Curve) and the EER (Equal Error Rate). AUC value can be viewed as a ranking measure that is very useful and is based on pairwise comparisons between classifications of two classes. In other words, the AUC value is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. Given two randomly chosen users, one being a legitimate user and the other an impostor, the AUC represents the probability $P(S^{leg} > S^{imp})$ (*i.e...* probability of a good assignment):

$$AUC = \frac{\sum_{p=1}^{n_g} \sum_{q=1}^{n_i} I(S_p^{leg}, S_q^{imp})}{n_g n_i} \tag{3.2}$$

where n_g and n_i are respectively number of legitimate users and impostors and S_p^{leg}, S_q^{imp} the scores of legitimate users and impostors, and the function I is defined by:

$$I(S_p^{leg}, S_q^{imp}) = \begin{cases} 1 \ if \ S_p^{leg} > S_q^{imp} \\ 0 \ otherwise \end{cases} \tag{3.3}$$

That way, AUC can be considered as a global criterion of performance. The higher the AUC is, the better the performance.

The EER value is when the False Acceptance Rate (FAR) is equal to the False Rejection Rate (FRR). It can be viewed as a compromise between usability and security. The goal of a matcher is to minimize this value. A visual representation of AUC and EER is given in Figure 3.12. Both AUC and EER are computed after a bootstrap of 1000 replications and are given within a 95% confidence interval.

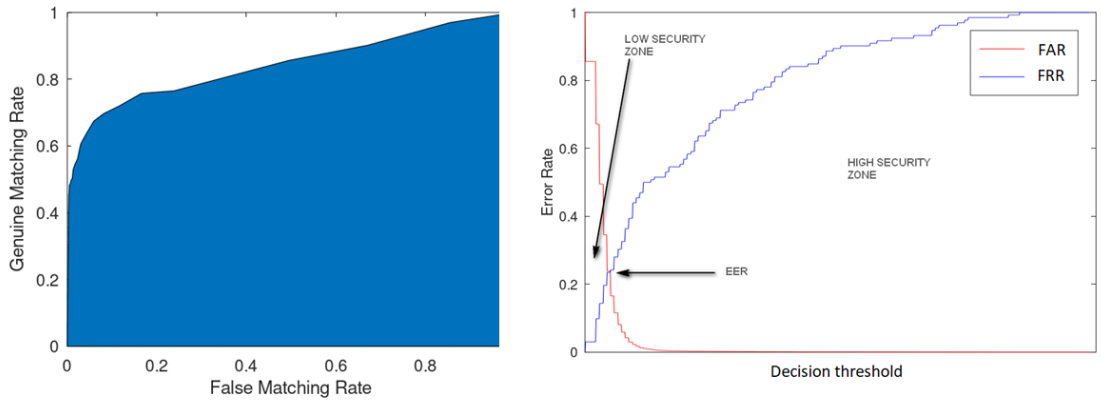


Figure 3.12: Examples of ROC curve which shows the AUC of a biometric system (Left graph) and the link between EER and the FMR and FNMR curves (Right graph)

To compute these metrics, we used two of the most widely used biometric matching algorithms in the literature, Bozoth 3 [Ko Kenneth, 2007] and MCC. Bozorth 3 is a fingerprint matching algorithm from the NIST with minutiae extractor *mindct* [Ko Kenneth, 2007]. The MCC fingerprint matcher is proposed by Raffaele Cappelli [Cappelli et al., 2010]. An illustration of the MCC comparison of 2 fingerprint cylinder-codes is given in Figure 3.13. A commercial matching algorithm is used besides these two to assess the performance of our data.

3.4.3 Presentation attack evaluation

When evaluating a PAD system, the PAIs are validated. The validation method is set to be able to measure the potential of each PAI type. Many metrics do exist to evaluate the resistance to attacks for a PAD system. The metric used in this thesis

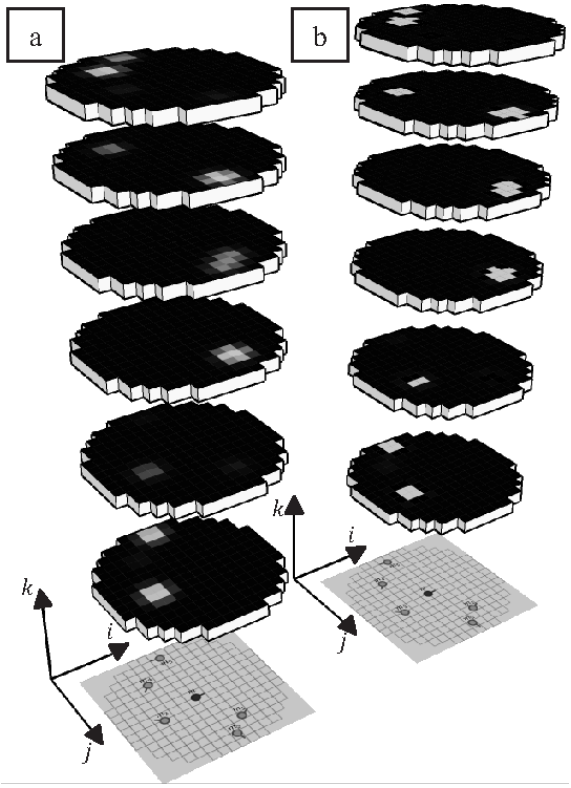


Figure 3.13: Illustration of the fingerprint MCC matcher: comparison of the circular neighborhood of minutiae from two fingerprint samples a and b (source [Cappelli et al., 2010]).

for PAD qualification is the TDR (True Detection Rate), defined as:

$$TDR = 100 \times \frac{\text{number of detected attacks}}{\text{Total number of attacks}}$$

We finally compute the TDR which represents the proportion of attacks detected by the PAD algorithm. We use a commercial PAD system to classify whether given images are seen as real fingerprints or spoofs. The TDR metric is used here to measure if a given PAD module detects similarly synthetic and real PAIs.

3.5 Conclusion

In this chapter, we presented some of the most widely used fingerprint databases that we use for this research. We also presented the methodology that we put in place to validate our work. However, there are many parameters that are difficult to control during the creation of a biometric dataset which can bias the results of a test. That is why we study the impact of parameters that can be controlled in the

next chapter.

We also introduced the concept and most widely used dataset for fingerprint systems. We decided to build datasets under controlled conditions to master each parameter and understand how they can impact the recognition accuracy of a biometric system.

Chapter 4

Analysis of acquisition context impact on the performance of fingerprint systems

Contents

Summary : *This chapter presents the study we realized to understand the effects of acquisition context for the evaluation of a fingerprint system. In the previous chapter, we presented the process of the certification of biometric systems and the need for the associated tests to be reproducible. Among other factors, the acquisition context can have a significant impact and introduce variability in the performance of a biometric system. We study the acquisition environment effects on fingerprint system performance as well as the data capture module.*

4.1 Introduction

Biometrics is more and more employed for user authentication to secure access to digital services or computers/smartphones. The fingerprint modality is a very popular one as we can estimate that 80% of smartphones embed a fingerprint sensor. This biometric solution is very easy to use and is largely used (70% of biometric systems in the USA use digital fingerprints). As this kind of system is designed to avoid attacks, it should meet security and privacy constraints. The certification of systems is a process whose objective is to verify how these constraints are fulfilled. The certification of a biometric system is an important step during the development and use of a biometric system. During this test, a biometric system is subject to many tests in order to establish its conformance to a certain test plan (performance, robustness to attacks, time). Indeed, these tests are done by independent laboratories, the testing scenarios are defined by the testing authority. The testing of biometric systems is mainly based on the ISO 19795 series for the performance tests and ISO 30107 series for the PAD.

In [Wone et al., 2021], it has been demonstrated that environmental conditions (*i.e.* temperature and humidity) can impact significantly the performance of a fingerprint system. This work showed that enrollment and matching in the same conditions improve performance and can reduce the vulnerabilities of a fingerprint system to attacks. The test scenarios that are used by certification schemes and laboratories to evaluate biometric systems recommend the tests to be conducted under monitored but not necessarily controlled conditions. In addition, the used dataset for the testing of biometric systems is mainly heterogeneous, as dataset distribution should target the population that will use the product. So, it is difficult to isolate a particular factor. Generally, particular behaviors of the product are smooth by the rest of the test population. However, some certification bodies want to move towards bias

isolation and bias crossing.
In this chapter, we address three questions that are crucial for the certification of fingerprint systems:

1. How do acquisition conditions influence the performance, the security, the user experience, and finally the trust of fingerprint systems?

2. Are all fingerprint matchers impacted in a similar way?

3. is the performance of a biometric system related to the acquisition system?

4.2 Related works

The operational testing of a fingerprint system is realized by using a dataset collected for the purpose of the test or an existing one. The second case makes sense only if the testing dataset is well known with good confidence it has not been used during the training of the recognition algorithm. However, due to the fact that most of the tested biometric systems are tested as black boxes, it is not easy to say how it is fair and unbiased to use a public existing dataset. For this reason, most of the testing methodologies use a dataset collected for a single product. Thus, taking into account certain parameters would be an excellent way of averaging if not minimizing bias for biometric systems. We list all capture parameters influencing the performance of fingerprint systems:

- **Demography**. It has been demonstrated that demography is a key factor of bias not only for face recognition but also for other biometric modalities. From a study conducted on Malaysian groups, authors in [Heng et al., 2018] conclude that fingerprint patterns could be inherited genetically. They also found that some ethnicities are more likely to have a certain pattern. The same goes for gender and occupation. In [Godbole et al., 2022], the authors conducted a study on groups of Caucasian males, Caucasian women, Black males, and Black women using two fingerprint matching algorithms and quality measurement. Their study allows them to conclude that most observed demographic differentials can be explained by the poor quality of some fingerprint images and the high accuracy of the used matching algorithm makes them less sensitive to demographic bias. Age is also an important factor especially for children and old people as they are more subject to skin transformation [ISO 19795-1:2021(E), 2021]. Therefore, they are big contributors to the False Non-Match rate and

the Failure-to-Acquire rate in performance assessment studies.

- **User**. User's anatomy has always been known as a source of perturbation in a recognition task for a biometric system. Indeed, beards, mustaches, and baldness can lead to bad recognition scores, particularly if the state of that factor was different during the enrollment [Pentland et al., 1994]. This is known as template aging and is applicable to fingerprints. Harvey et *al.* [Harvey et al., 2019] studied this phenomenon over seven years with the same group of people. They propose a methodology that can be used to isolate the effect of biometric template aging. Lanitis [Lanitis, 2010] gives a complete survey of the effects of aging on biometric identity verification with different biometric modalities. Other things such as the subject's motivation, familiarity, behavior [Kukula et al., 2009], appearance (fingernails can impact the positioning), fingerprint condition[ISO 19795-1:2021(E), 2021](depth and spacing ridges, dry, cracked or damp), etc. are known to be a source of quality variation during the capture of fingerprint data.

- **Environment**. The environmental conditions are known to have impacts on the recognition process. In the literature, to the best of our knowledge, very few works have studied the performance of biometric systems through different climatic environments. We can cite a recent work for finger veins [Kirchgasser et al., 2020]. The authors present their work as a preliminary study to get the first results to identify the most challenging factors for finger vein recognition. Tan *et al.* [Tan et al., 2010] have shown that biometric matchers can have different behaviors through different environments. Figure 4.1 from [Tan et al., 2010] illustrates the effects of different conditions on the same subject. The study is more focusing on the PAIs (Presentation Attack Instruments) and the way they are detected with different PAD (Presentation Attack Detection) algorithms. Grosz *et al.* [Grosz et al., 2020a] give a module-by-module certification of a biometric process taking into consideration the moisture of the skin. Krishnasamy *et al.* [Krishnasamy et al., 2011] made a very interesting work for the recognition of fingerprints in wet conditions by building a wet and wrinkled fingerprint database available on demand. The study concludes that the error increases when matching a wet finger against a dry one.

As clearly underlined by Fernandez-Saavedra *et al.* [Fernandez-Saavedra et al.,

2008, Fernandez-Saavedra et al., 2010], the existing certification schemes such as FIDO alliance or Common Criteria are more focusing on the methodology and protocol of testing the performance of a biometric application under lab conditions, and/or ask biometric technology developers to include information on environmental influence and ways of reducing them.

The same observation can be made for ISO 19795-2[ISO 19795-2:2007, 2007] or 19795-6[ISO 19795-6:2012, 2012]. Even if they point out the impact of environmental factors, all these standards and test methods only recommend defining and reporting the conditions of the tests, which are naturally the lab conditions. Fernandez et al. [Fernandez-Saavedra et al., 2008, Fernandez-Saavedra et al., 2010] propose a full test protocol, including the use of a climatic chamber to control the test conditions.

- **Capture system**. The capture system is the object of this study. The sensor quality is the main source of image quality variations. As pointed out by Marasco [Marasco, 2019], the quality variation between different sensors is a big challenge for the operability of biometric systems. There are works that proposed solutions to overcome this issue [Alshehri et al., 2018][Jain and Kumar, 2010]. [Alshehri et al., 2018] deals with the decrease in performance of a fingerprint system when the sensor at the enrollment is not the same used for the recognition operation. They proposed two fingerprint descriptors to encode the fingerprint information and overcome the discriminative characteristics among the fingerprints captured with different fingerprint sensors.

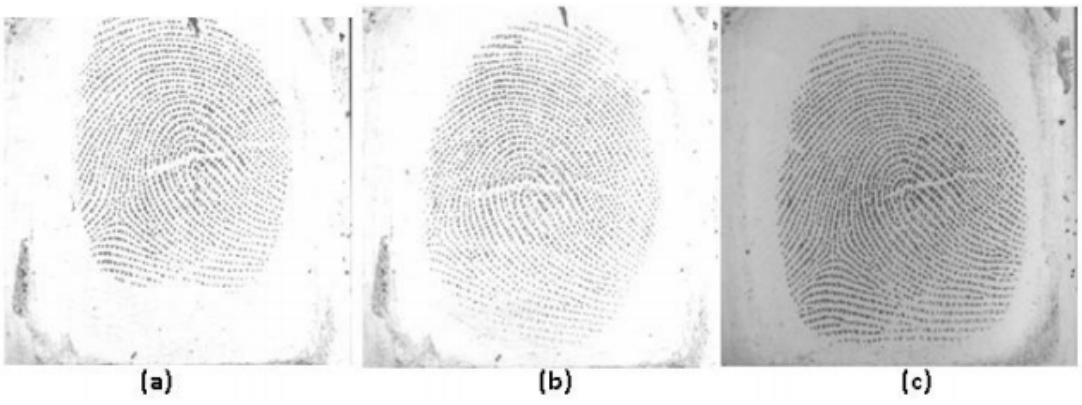


Figure 4.1: Illustration from [Tan et al., 2010] with a) High Temperature / High Humidity, b) High Temperature / Normal Humidity, c) Low Temperature / Normal Humidity

4.3 Effects of environmental conditions

Based on our observations, the texture of fingerprints changes according to the acquisition conditions. Those changes may badly influence the quality of images and matching scores. Thus, we decided to analyze such bias and the effects that environmental conditions could have on the performance of fingerprint systems. In this chapter, we investigate if this assumption can align with experimental observations. This study can help some certification schemes to reduce interoperability risk by including an evaluation of the bias linked to the acquisition environment of biometric products. Despite the good matching performance that we can have inside a testing laboratory using target application conditions, the perception of the end-user and the security can differ if the product is used under unusual environmental conditions. In order to verify our assumption, a database has been created to simulate the uncontrolled environment.

4.3.1 Experimental protocol

The unavailability of a controlled environmental fingerprint dataset enhanced our desire to build our own for the purpose of these experiments by setting up a test protocol to build the database. This database was built with a process respecting data privacy and security: the purpose of the experiment was explicitly provided, and subjects were told about their privacy rights, storage limitations, integrity, and confidentiality of the data. Data are stored in an encrypted hard drive disk using *VeraCrypt* and AES (Advanced Encryption Standard) method. The disk is stored in a safe protected by a password known to only few people.

In a controlled climatic chamber, different environmental conditions with specific values of temperature and humidity have been recreated. 990 fingerprint sample images from 17 unique participants from 22 to 55 y.o. with 65% of males and 35% of females were collected. For comparison, each set of the Fingerprint Verification Competition (FVC) is around 880 images. In each environment, 3 fingerprint scans of the thumb, index, and middle of both hands are performed using the **Digital Persona's EikonTouch 700** sensor, which is a capacitive one. The details of the built database are given in Table 4.1. Finally, six environmental conditions have been simulated. Figure 4.2 shows the **Weiss WK11-180/40** climatic chamber used and the setup of the tests.

Table 4.1: Environments considered in this study.

Environment	Temp (°C)	Humidity (%)	Number of fingers	Total images
#1	15	20	54	162
#2	15	50	42	126
#3	15	80	72	216
#4	25	20	42	126
#5	25	50	48	144
#6	25	80	72	216

(a) A picture of one of the participants during the tests

(b) Inside the chamber

Figure 4.2: Overview of the collection setup

Figure 4.3 shows the steps followed for each sample collection. The climatic chamber has a probe that gives feedback about the current values of the temperature and humidity. People keep the current hand for two minutes for the adaptation, and then we can collect. People were trained to interact with the sensor to avoid misuse and bad data. So, all the data of the same conditions have been collected at the same time before we set new conditions.

Validation protocol

First, the quality of the acquired biometric samples in each environment is computed using the NFIQ2 fingerprint quality assessment metric [Olsen et al., 2016], in order

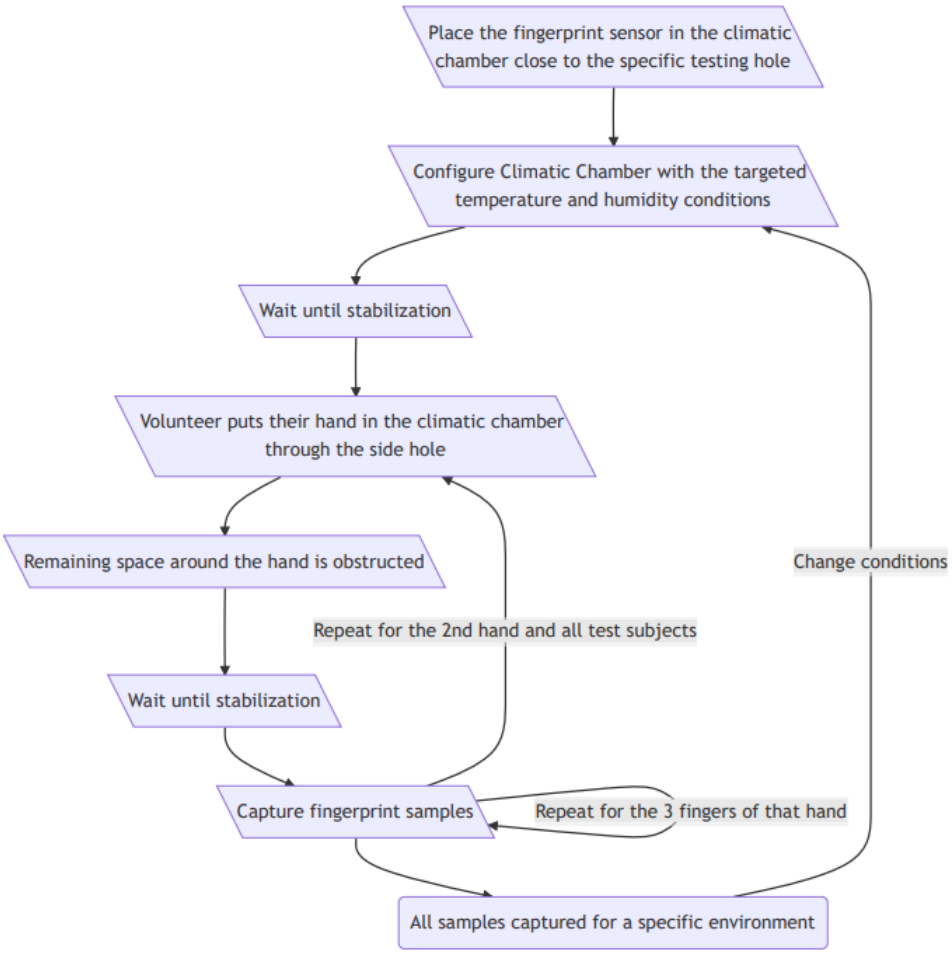


Figure 4.3: Synopsis of the acquisition method

to be compliant with the ISO/IEC 29794-4:2017 recommendation [ISO 29794-4:2017, 2017].
Second, we need to consider the performance of a fingerprint system for each trial environment. Three fingerprint-matching algorithms are used in order to generalize conclusions. The NIST (National Institute of Standards and Technology) matcher Bozorth 3, the Minutia Cylinder-Code (MCC) matching algorithm, and a commercial matcher are used for the fingerprint recognition task. The commercial matcher uses its own minutiae extractor whereas, for Bozorth 3, the *mindtct* extractor is used to extract the biometric features needed for the matching operation.

In this work, we consider the authentication process for the performance evaluation of each trial fingerprint-matching algorithm for the different simulated environments. AUC values 3.2 are computed for each scenario and matcher. As reminder, AUC represents the measure of separability. It indicates how much the model is able to

distinguish between classes. The AUC value can be viewed as a measure ranking, which is very useful and is based on pairwise comparisons between classifications of two classes. We also estimate the confidence intervals (95% confidence) of AUC results, since the used database is not large. It is computed after a bootstrapping of 1000 replications to confine the AUC value in a confidence interval with 95% of certitude. We also compute the EER and the FMR100.

4.3.2 Experimental results

In this section, the statistical analysis of the obtained results is provided in order to evaluate the performance of the trial biometric systems under different environments. First, we achieved a quality assessment of the collected data, then we observed the performance of fingerprint-matching algorithms on the data.

Quality Assessment

The NFIQ2 score is used to assess the quality of the collected fingerprint samples through the different environments we created. The NFIQ2 score for a given fingerprint image expresses the contribution of that fingerprint to the performance of a biometric system. In other words, the NFIQ2 score measures the usability of a fingerprint image in a recognition task. NFIQ2 scores are within a close range between 0 and 100 as referring to [ITU-R BT.500-15,], the NFIQ2 scores can be divided into five intervals:

- **Bad quality**: ranging from [0,20[, mostly samples with almost no readable information. The capture from the source may cause many failures to acquire,
- **Poor quality**: ranging into the interval [20,40[, samples with very few fingerprint ridges which can give bad matching results,
- **Fair quality**: ranging into the interval [40,60[, samples with an average quality,
- **Good quality**: ranging into the interval [60,80[, samples with many fingerprint details that can lead to very good matching results,
- **Excellent quality**: ranging into the interval [80,100], samples of very high quality which lead to extraction of enough features for the recognition task.

Table 4.2 shows the mean NFIQ2 score and its associated standard deviation values of each subset in every environment computed on our acquired database.

We can observe that better average scores are obtained when the environment is dry for the two trial temperatures (15°C and 25°C) (environments #1 and #4). Usually,

Table 4.2: Overview of the NFIQ2 mean values for each environment of our database.

Environment	NFIQ2 $\pm std$
#1	56.2% ±15.72
#2	55.8% ±11.83
#3	53.5% ±15.65
#4	57.2% ±17.05
#5	54.9% ±14.91
#6	55.7% ±15.16

when the humidity rate increases (whatever the considered temperature), we observe that the quality scores decrease. Two remarks can be formulated to explain the obtained results :

1. When increasing the humidity, the sensor struggles to capture the samples because of the high moisture of the finger skin.

2. People tend to press harder their fingers on the sensor, which favors the decrease in the quality of the acquired sample.

These results show that a high degree of humidity deteriorates the quality of acquired fingerprint samples.

The distribution of the mean NFIQ2 score for each environment is provided in Figure 4.4. We can observe that the quality scores vary from very low (values close to 0%) to very high quality (values close to 100%) and cover the whole quality range. Moreover, the environments with average humidity (50%) have distribution with less dispersion around the average value. This may be linked to the NFIQ2 metric itself. Indeed, NFIQ2 is the result of contributions from industrials that provide algorithms. The quality measure is the contribution to the performance of these algorithms. Knowing that most of the existing databases are collected in nominal conditions, average humidity can lead to less spread NFIQ quality distributions. In opposition, the extremes lead to higher standard deviation, as these conditions are less common.

Matching comparison

Since in many testing scenarios and in real-life situations, we perform the enrollment only once, we compare the collected samples in the climatic chamber against the reference samples (captured under nominal conditions: 22°C and 50% of humidity). Two fingerprint samples have been used for the computation of the matching scores.

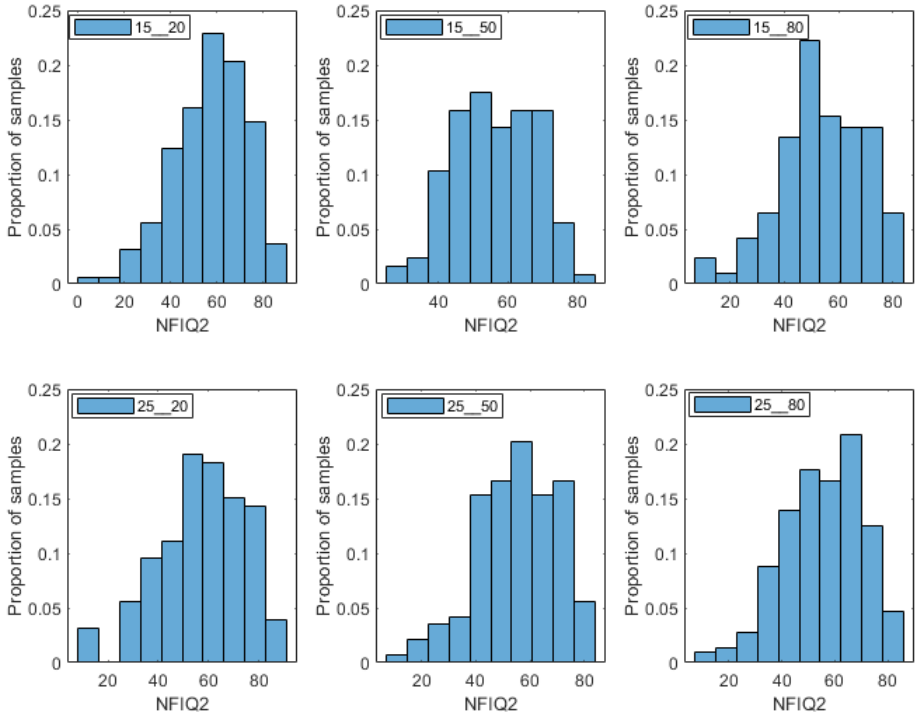


Figure 4.4: Quality comparison of the whole database in different conditions: for the first row, the temperature is set to 15°C and for the second row, the temperature is 20°C, Humidity from 20% to 80% on each row.

We particularly have been interested in the similarity of the AUC values (as a global metric of the biometric system), EER, and FMR100. For each detected minutiae, we have its coordinates, orientation, and quality.

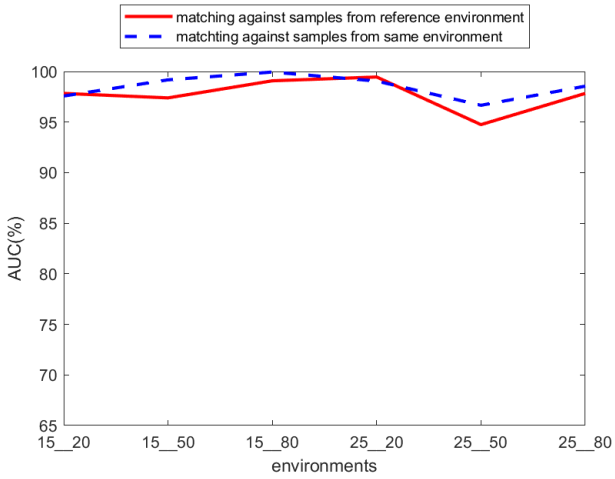
Table 4.3 gives the AUC of the different matchers we used, as well as EER values and FMR100 values.

Table 4.3: Performance of the three matchers when enrollment is done in a "normal" environment (Temperature 22°C, Humidity 50%) and verification in different conditions: AUC is given with a 95% confidence interval.

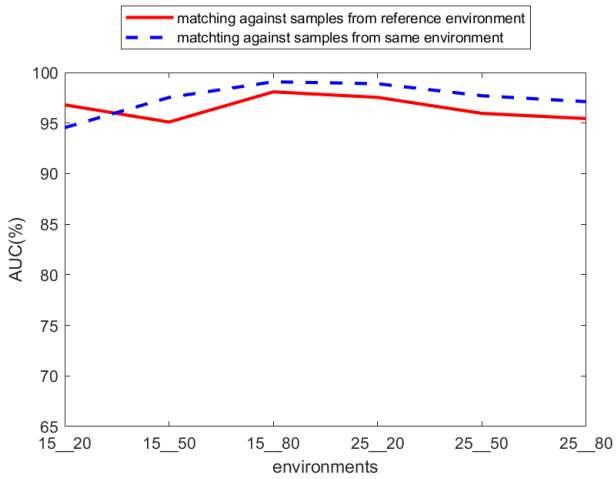
Env.	AUC (%)			EER (%)			FMR100 (%)		
	Commercial	Bozorth3	MCC	Commercial	Bozorth3	MCC	Commercial	Bozorth3	MCC
#1	97.83 ± 0.04	96.80 ± 0.06	74.54 ± 0.19	5.58	7.51	33.15	8.02	13.27	65.51
#2	97.39 ± 0.04	95.10 ± 0.08	65.10 ± 0.20	7.27	9.56	39.42	7.14	18.65	75.39
#3	99.08 ± 0.03	98.08 ± 0.05	68.78 ± 0.21	3.26	5.48	38.24	4.16	10.65	68.52
#4	99.45 ± 0.02	97.53 ± 0.05	80.33 ± 0.16	2.35	6.76	28.60	3.17	11.11	49.20
#5	94.75 ± 0.06	95.96 ± 0.08	77.14 ± 0.18	9.08	9.26	30.94	9.03	13.19	60.42
#6	97.83 ± 0.05	95.44 ± 0.08	69.07 ± 0.22	6.68	8.63	37.89	9.72	14.58	70.73

The behaviors of the different matchers are shown in Figure 4.5 when focusing

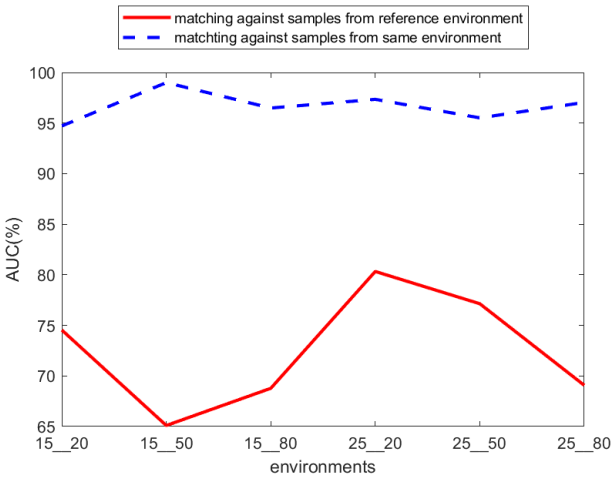
on AUC values. When the matching is done against a template collected in the reference enrolment condition (corresponding to the red lines in the figure 4.5), we observe a variation between environments with different intensities. The two first matchers (Commercial and Bozorth) have a variation <5% AUC, while MCC has a variation >15% AUC. Matchers do not react in the same way while under different environmental conditions.



(a) with the commercial matcher



(b) Bozorth 3



(c) MCC

Figure 4.5: Behaviors of different matchers with enrollment and verification done against reference samples VS when it is done in the same environment.

The MCC matcher seems to be more affected by the difference between the enrollment and the verification conditions. Indeed, it gives the lowest accuracy among the three matchers from the three accuracy metrics.

In order to deeply analyze the previously obtained results, we wanted to investigate

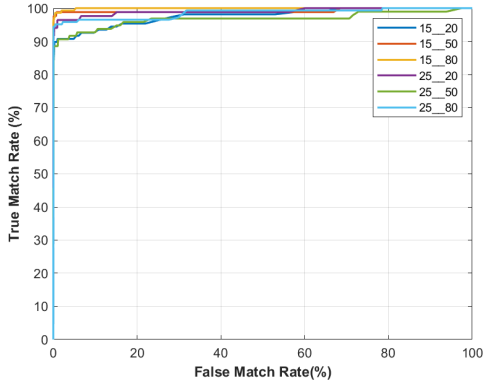
what would happen if the enrollment and the verification are performed within the same conditions. The goal is to evaluate how the performance is less impacted by environmental conditions in this use case. We compute the performance of the trial matchers when the test subject performs the enrollment and the verification in the same environment (dot blue line in Figure 4.5). Even if for all matchers, AUC is increased when the same conditions are used for enrollment and verification compared to when enrollment and verification are not done in the same environments, we notice that the impact of this increase depends on the matcher. The MCC matcher seems to be the most affected. For this matcher, when performing authentication in the same environment that we registered, the global performance increases by 24% in average considering the AUC. The two other matchers are affected by around 1%. Security and User Experience will be increased if enrollment can be redone in the targeted environment when the conditions change.

Table 4.4: Performance of the three matchers when enrollment and tests are done in the same environment: AUC is given with a 95% confidence interval.

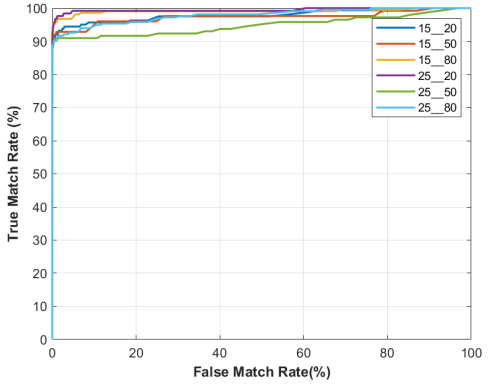
Env.	AUC(%)			EER(%)			FMR100(%)		
	Commercial	Bozorth3	MCC	Commercial	Bozorth3	MCC	Commercial	Bozorth3	MCC
#1	97.57 ± 0.06	94.55 ± 0.10	94.12 ± 0.14	7.37	10.94	8.90	9.26	14.81	25.92
#2	99.17 ± 0.05	97.53 ± 0.09	98.99 ± 0.03	1.19	3.89	6.39	11.91	10.71	14.28
#3	99.95 ± 0.002	99.08 ± 0.02	96.49 ± 0.11	1.39	5.46	6.70	13.89	7.64	15.27
#4	99.05 ± 0.04	98.89 ± 0.03	97.35 ± 0.08	3.62	6.14	9.29	3.57	10.71	16.66
#5	96.66 ± 0.09	97.70 ± 0.05	95.51 ± 0.11	7.16	7.86	10.50	11.46	11.46	18.75
#6	98.55 ± 0.04	97.12 ± 0.06	97.04 ± 0.09	4.10	7.81	7.59	4.86	13.89	20.93

Table 4.4 gives the AUC value, EER, and FMR100 of the different matchers when the enrollment is performed in the same conditions as the verification task. The detailed performance of the three matchers is given in Figure 4.6. It shows the behavior of the three matchers through different environments when the enrollment and the verification attempts are done in the same conditions (Figure 4.6a, 4.6c and 4.6e) and when we perform the verification attempt against the reference environment samples (Figure 4.6b, 4.6d and 4.6f). The behaviors of the three matchers are similar from the Figure 4.6a to Figure 4.6e, the Figure 4.6f is very different and witnesses the effects of the testing environment on MCC matcher

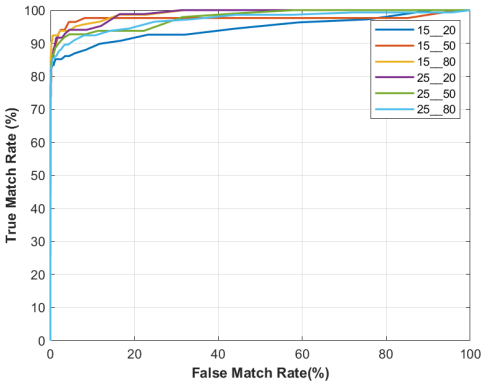
The impact on the MCC matcher is even more visible. Indeed, for an ideal matcher, the ROC curve would be a "step" of 100% (*i.e.* AUC=100%). As one can see from Figure 4.6f, the difference between enrollment and verification conditions moves the ROC curve away from the ideal system. However, changing enrollment conditions makes the performance better, as in Figure the 4.6e.



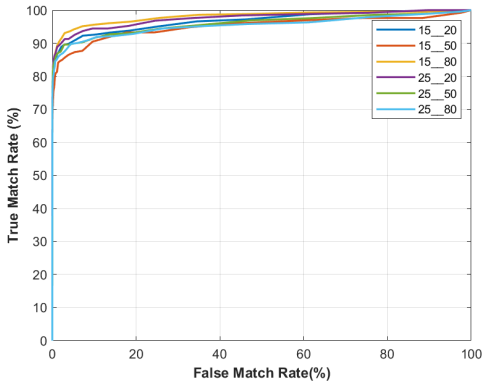
(a) Commercial matcher when enrollment and verification are done in the same conditions



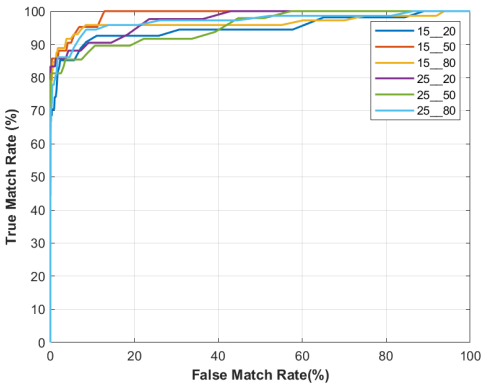
(b) Commercial matcher when verification is done against reference samples



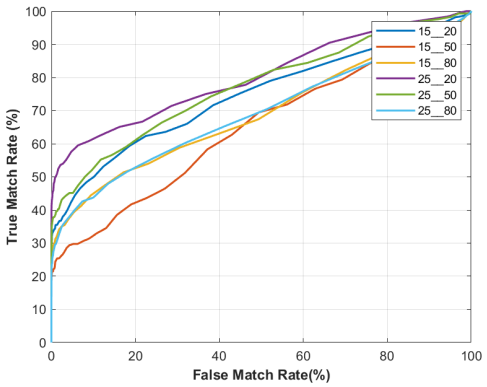
(c) Bozorth3 when enrollment and verification are done in the same conditions



(d) Bozorth3 when verification is done against reference samples



(e) MCC when enrollment and verification are done in the same conditions



(f) MCC when verification is done against reference samples

Figure 4.6: ROC curves of the three matchers through the different environments

To go further, and in order to analyze the influence each parameter can have on the

global AUC value, a statistical hypothesis test is used to measure the dependency of acquisition conditions on the performance of the fingerprint system. A P-value of the AUC for each matcher using the Student's test is computed by observing one parameter (only temperature is varying and humidity is fixed or vice versa) at a time.
For each parameter, we set a value and compute the p-value when the second value varies. So, the two test hypotheses are verified:

- At a given temperature, humidity only does not impact significantly the performance of the fingerprint system.
- At a given humidity, temperature only does not impact significantly the performance of the fingerprint system.

Tables 4.5 and 4.6 give the different p-values through all the possibilities. As an illustration, for the 4th row of Table 4.5, we consider the situation when the humidity is fixed and equal to 50% and compute the p-value of the different AUC supposing it does not depend on the temperature. For Commercial and Bozorth 3 matchers, the p-value is <5% meaning that the temperature has no significant impact. For MCC, the p-value is >5% meaning that temperature has a significant impact on the AUC. This p-value method can highlight Security and User Experience risks for a fingerprint matcher when used under specific environmental conditions. For example, it means that it can represent a high risk of security for the MCC algorithm. The test reveals also that both humidity and temperature greatly influence the AUC of the commercial matcher.

Table 4.5: P-values considering separately the two parameters when enrollment and verification are done in the different conditions where $H_T \in [20, 50, 80]$ with T_E and H_E temperature and humidity at the enrollment and T_T and H_T temperature and humidity during the test (verification)

Parameters	Commercial	Bozorth 3	MCC
$T_E = 22°C, T_T = 15°$C, H_E=50 and $H_T \in \{20, 50, 80\}$	2.66×10^{-5}	7.99×10^{-5}	1.6×10^{-3}
$T_E = 22°C, T_T = 25°C, H_E = 50$ and $H_T \in \{20, 50, 80\}$	2.01×10^{-4}	4.29×10^{-5}	2×10^{-3}
$H_E = 50\%, H_T = 20\%$ and $T_T \in \{15, 25\}°C$	5.2×10^{-3}	2.4×10^{-3}	2.38×10^{-2}
$H_E = H_T = 50\%$ and $T_T \in \{15, 25\}°C$	6.7×10^{-3}	2.9×10^{-3}	5.38×10^{-2}
$H_E = 50\%, H_T = 80\%$ and $T_T \in \{15, 25\}°C$	4×10^{-3}	8.7×10^{-3}	1.4×10^{-3}

4.3.3 Discussion

The work presented here deals with the problem of fingerprint recognition from a matching point of view when environmental conditions change. It brings an eye to

Table 4.6: P-values considering separately the two parameters when enrollment and verification are done in the same condition with T_E and H_E temperature and humidity at the enrollment and T_T and H_T temperature and humidity during the test (verification)

Parameters	commercial matcher	Bozorth 3	MCC
$T_E = T_T = 15°C$, and $(H_E, H_T) \in \{20, 50, 80\} \times \{20, 50, 80\}$	4.98×10^{-5}	1.88×10^{-4}	1.65×10^{-4}
$T_E = T_T = 25°C$, and $(H_E, H_T) \in \{20, 50, 80\} \times \{20, 50, 80\}$	5.51×10^{-5}	2.84×10^{-5}	3.44×10^{-5}
$H_E = H_T = 20\%$ and $(T_E, T_T) \in \{20, 50, 80\} \times \{20, 50, 80\}$	4.8×10^{-3}	1.43×10^{-2}	8.7×10^{-3}
$H_E = H_T = 50\%$ and $(T_E, T_T) \in \{20, 50, 80\} \times \{20, 50, 80\}$	8.2×10^{-3}	4.49×10^{-4}	1.14×10^{-2}
$H_E = H_T = 80\%$ and $(T_E, T_T) \in \{20, 50, 80\} \times \{20, 50, 80\}$	6.4×10^{-3}	4.5×10^{-3}	1.8×10^{-3}

the problem of the sensitivity of biometric systems to environmental conditions that is not well underlined in the existing literature related to fingerprints.

Environmental factors can make biometrics challenging. This study investigates how environmental bias impacts performance results, namely by temperature and humidity. We saw that some worldwide conditions can be reproduced on climatic chambers to obtain an overview of security risk and user experience impact. We have shown that the three trial matchers have their own vulnerability to specific conditions. Knowing this can prevent further challenges when products are deployed over the world and can help to provide more trust in fingerprint biometric systems.

The study is limited to capacitive sensing, as almost all the fingerprint sensors on cards are of this type. The test population is mainly composed of Caucasians at almost 90% which might make the results change if the test is repeated. The study shows the variability of matching errors on genuine verification attempts through the environments and how it changes when the enrollment and the verification are done in the same conditions compared to doing the enrollment in a reference environment once and for all.

4.4 Effects of acquisition quality

As stated earlier, there are plenty of factors that can have an impact on the recognition capacities of a biometric system, thus on user experience and security. In addition, for interoperability, a biometric algorithm should be performing well independently of the hardware part capture device. The quality of a biometric sample, which depends on the capture device, is one parameter that can impact the recognition performance. For each biometric modality, there are some inherent indicators such as resolution, speed, frequency range, etc. which give an indication of the functionality

of an acquisition system. We focus here on the intrinsic quality of a fingerprint system.

In addition to ensure good performance, a good user experience, and a high level of security, a biometric system should be interoperable. Considering the large number of applications of biometric systems and the life of hardware solutions, it is more likely right to assume that a biometric system will be integrated with different acquisition systems. Therefore, the efficiency of a biometric system should not be conditioned by the acquisition module it is associated with. The interoperability of biometric systems is not taken into consideration by most of the certification schemes. The main reason is that most of the tested solutions are end-to-end solutions where a software solution is coupled with a hardware part. However, we observe that there are more and more algorithm-only providers in the market and the need to perform *technology evaluation* is growing.

4.4.1 Experimental protocol

In this section, we present the experimental protocol we followed to analyze the effect of the quality of captured fingerprint samples.

Dataset

For this work, we use a fingerprint dataset generated SFinGe presented in Chapter 3. As a reminder, SFinGe is based on the mathematical modeling of fingerprint characteristics.

Using SFinGe 4.1, we generated a dataset imitating a capacitive and an optical sensor. For each sensing technology, we use the integrated quality indicator to control the quality of the generated data. These different quality sets are simulating the quality of the fingerprint sample. Thus, we generated five different groups of images with different qualities:

- **High quality**: Fingerprints with almost no translation and rotation, most ridge patterns of very high quality, with almost no skin distortion or other perturbations.
- **Medium quality**: Fingerprints with almost no translation and rotation, ridge patterns of medium/high quality, with limited skin distortion and perturbations.
- **Low quality**: Fingerprints with almost no translation and rotation, ridge patterns of low quality, with limited skin distortion and perturbations.

- **Very low quality**: Fingerprints with almost no translation and rotation, ridge patterns of very low quality, with various perturbations.
- **Varying quality**: Fingerprints with varying quality and perturbations: most ridge patterns of medium quality, but some of low or very low quality.

In the proposed methodology, we generated 200 unique fingers for each set, with 5 impressions per finger. The images have a resolution of 1000 dpi and a size of 832x1120 pixels. We chose the highest image resolution and size that SFinGe offers. Scratches are added to the images using the dedicated function in SFinGe. Indeed, the scratches introduce minor degradation to the images to simulate a real-life scenario similar to what we observe when doing a dataset collection from random people.

Examples of the generated data are given in Figure 4.7. Images in the first row are from a capacitive sensor, and second-row images are from an optical one. For a visual comparison, Figure 4.8 gives examples of real images from FVC 2000 contest from the Db1a dataset. The images are from an optical sensor.

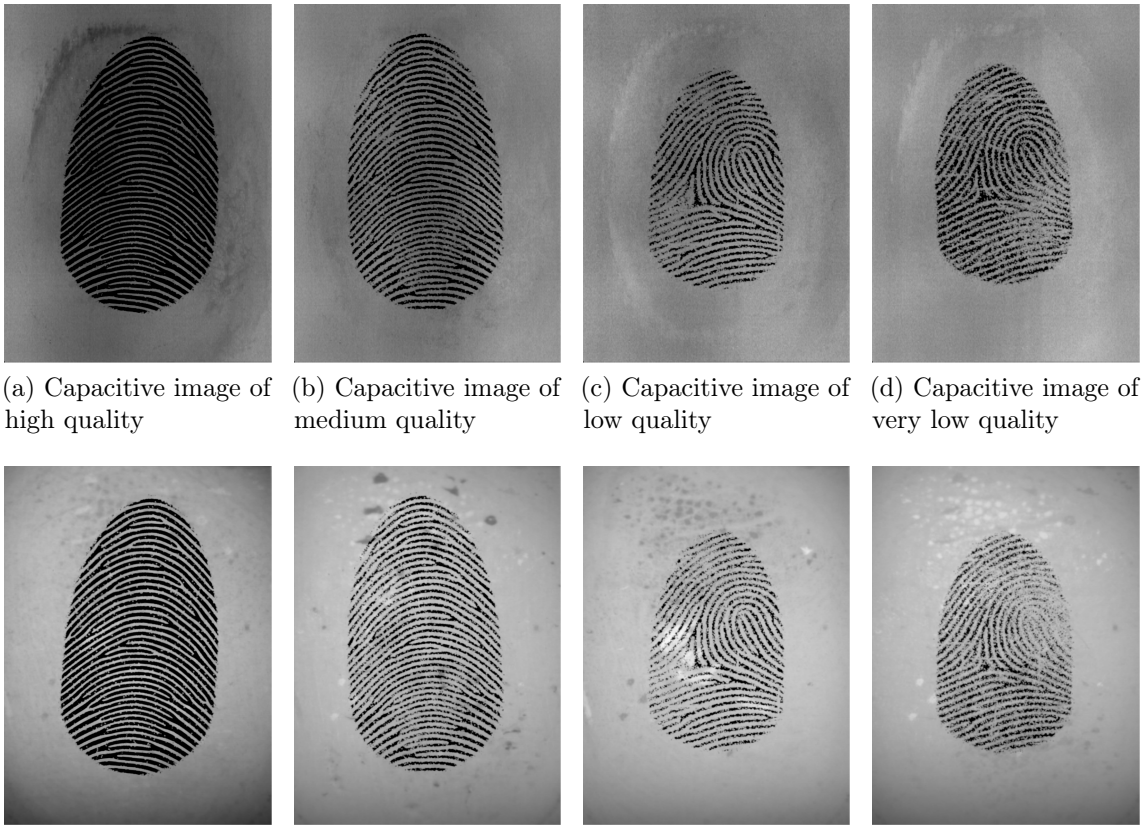
4.4.2 Experimental results

The evaluation methodology is similar to the one used previously in the previous section to assess the impact of acquisition conditions on biometric systems. So, we consider two types of tests for this study: 1) fingerprint quality assessment, and 2) performance evaluation.

Quality assessment

The quality assessment is done using the NIST NFIQ2 metric and the performance testing is done through AUC and EER using Bozorth3 and MCC for the fingerprint-matching operations.

We observe the quality variation over the generation parameters. We compute the average value and standard deviation of the NFIQ2 scores. Tables 4.7 and 4.8



(a) Capacitive image of high quality (b) Capacitive image of medium quality (c) Capacitive image of low quality (d) Capacitive image of very low quality

(e) Optical image of high quality (f) Optical image of medium quality (g) Optical image of low quality (h) Optical image of very low quality

Figure 4.7: Examples of images from the data generated with SFinGe.

show the statistical indicators (mean and standard deviation) of the dataset. Figures 4.9 and 4.10 show the data profiles of the NFIQ2 scores for each set of quality. They also give the range of data points between the 10th and 90th percentile (shaded area), as well as the average value and the standard deviation.

Considering NFIQ2 values, we can understand that the intrinsic quality of the generator may differ from a pure fingerprint quality metric. Indeed, one may expect high-quality generated samples to have higher statistics than other generation settings. This may be due to the fact that if we consider, for example, the set with “varying quality”, it may cover a wider range of quality scores and contain higher scores than the "high-quality" setting which has high-quality samples but is confined to a narrow range.

(a) Sample 1 (b) Sample 2

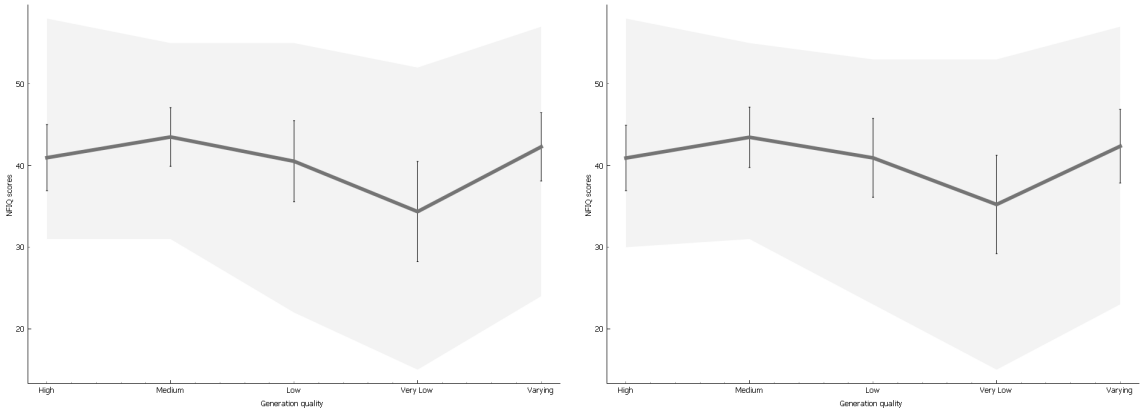
Figure 4.8: Examples of real images from the FVC 2001, Db1 subset.

Table 4.7: NFIQ2 statistical figures of the capacitive dataset

	Capacitive with no option (%)	Scratches (%)
High Quality	40.97 ±4.1	40.92 ±4.0
Medium Quality	43.50 ±3.6	43.47 ±3.7
Low Quality	40.52 ±5.0	40.94 ±4.8
Very Low Quality	34.36 ±6.1	35.23 ±6.0
Varying Quality	42.3 ±4.2	42.37 ±4.5

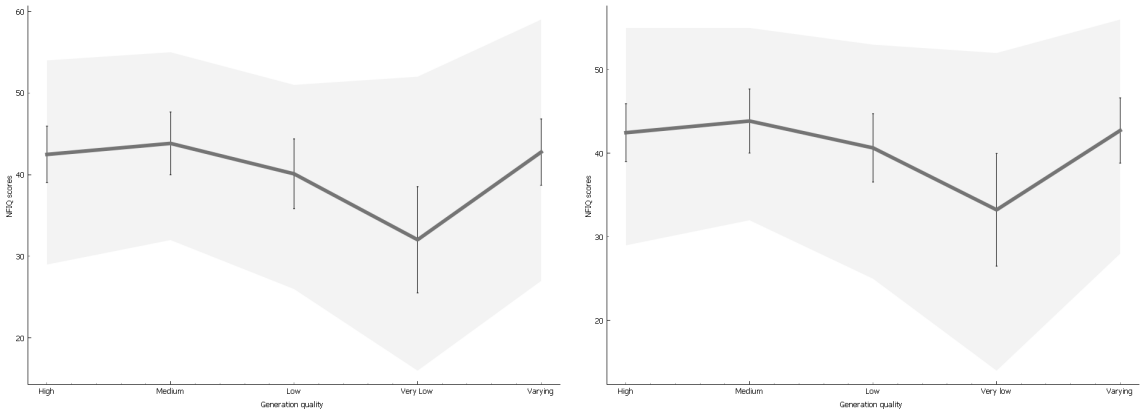
Table 4.8: NFIQ2 statistical figures of the Optical dataset

	Optical with no option (%)	Scratches (%)
High Quality	42.49 ±3.5	42.47 ±3.5
Medium Quality	43.84 ±3.8	43.87 ±3.8
Low Quality	40.11 ±3.6	40.65 ±4.1
Very Low Quality	32.03 ±6.5	33.24 ±6.7
Varying Quality	42.77 ±4.1	42.73 ±3.9



(a) No option (b) Scratches

Figure 4.9: Visualization of data profiles of the NFIQ2 quality scores from the capacitive images.



(a) No option (b) Scratches

Figure 4.10: Visualization of data profiles of the NFIQ2 quality scores from the optical images.

Matching comparison

In this part, we evaluate the performance of the generated dataset considering the AUC and the EER measures using two matching algorithms. Images from SFinGe have been generated with fingerprint templates respecting ISO/IEC 19794:2 format which can be directly used with MCC for comparison tasks. We also extract minutiae from fingerprint samples with NIST *Mindtct* [Ko Kenneth, 2007] for Bozorth3. For each matching algorithm, we compare the AUC and the EER values.

Tables 4.9 and 4.10 show the AUC and EER computed with the NIST Bozorth3 matcher for the capacitive sensor and optical sensor. We can observe that the scratches do not introduce a high decrease in the AUC or EER values. The actual

difference in performance on the generated datasets comes mainly from the acquisition quality. Despite the scratches, the inherent generation quality is the only factor that seems to be very important here when we consider a single capture technology and the same matching algorithm. The same observation applies to the MCC matcher for which the performance is shown in Tables 4.11 and 4.12.

Table 4.9: Performance of Bozorth3 on the capacitive dataset

	Capacitive with no options		Capacitive with Scratches	
Generation quality	AUC (No option)	EER (No option)	AUC (Scratches)	EER (Scratches)
High	81.1314 $\pm 1.4981e-14$	0.27258 $\pm 4.1308e-17$	80.5505 $\pm 2.3793e-14$	0.27908 $\pm 4.8193e-17$
Medium	89.0779 $\pm 3.525e-15$	0.1805 $\pm 6.1962e-17$	89.0546 $\pm 5.2874e-15$	0.17883 $\pm 1.2048e-17$
Low	79.907 $\pm 1.8506e-14$	0.30119 $\pm 1.0327e-17$	79.9247 $\pm 6.1687e-15$	0.27194 $\pm 4.1308e-17$
VeryLow	67.1891 $\pm 7.0499e-15$	0.3942 ± 0	67.4475 $\pm 3.525e-15$	0.38437 $\pm 6.1962e-17$
Varying	86.5812 $\pm 1.3219e-14$	0.21417 $\pm 2.2375e-17$	86.1737 $\pm 2.6437e-14$	0.21369 $\pm 6.3683e-17$

Table 4.10: Performance of Bozorth3 on the optical dataset

	Optical with no options		Optical with Scratches	
Generation quality	AUC (No option)	EER (No option)	AUC (Scratches)	EER (Scratches)
High	84.8549 $\pm 2.115e-14$	0.16339 $\pm 2.4096e-17$	84.6644 $\pm 1.6744e-14$	0.16488 $\pm 8.6058e-18$
Medium	81.9144 $\pm 4.4062e-15$	0.19922 $\pm 6.8847e-17$	82.323 $\pm 2.6437e-15$	0.19497 $\pm 6.8847e-17$
Low	73.6108 $\pm 2.7318e-14$	0.33473 $\pm 8.9501e-17$	73.7676 $\pm 1.7625e-15$	0.31858 $\pm 3.7866e-17$
VeryLow	65.0311 $\pm 4.4062e-15$	0.40782 $\pm 1.3425e-16$	65.7977 ± 0	0.39931 $\pm 7.5731e-17$
Varying	87.2199 $\pm 2.6437e-14$	0.19555 $\pm 4.6472e-17$	86.0072 $\pm 1.0575e-14$	0.17967 $\pm 3.0981e-17$

Table 4.11: Performance of MCC on the capacitive dataset

	Capacitive with no options		Capacitive with Scratches	
Generation quality	AUC (No option)	EER (No option)	AUC (Scratches)	EER (Scratches)
High	100 ± 0	0 ± 0	100 ± 0	0 ± 0
Medium	100 $\pm 1.078e-07$	2.5025e-06 $\pm 2.6951e-07$	100 $\pm 1.057e-07$	2.3618e-06 $\pm 2.6425e-07$
Low	99.2677 ± 0.016163	0.010583 ± 0.00018142	99.2364 ± 0.0168	0.0094371 ± 0.00016929
VeryLow	99.206 ± 0.016697	0.0088748 ± 0.00019145	99.2459 ± 0.016929	0.009172 ± 0.000169
Varying	99.7168 ± 0.010108	0.0046938 ± 0.00012898	99.6998 ± 0.010504	0.0054811 ± 0.00012706

Table 4.12: Performance of MCC on the optical dataset

	Optical with no options		Optical with Scratches	
Generation quality	AUC (No option)	EER (No option)	AUC (Scratches)	EER (Scratches)
High	100 ± 0	0 ± 0	100 ± 0	0 ± 0
Low	99.2654 ± 0.016659	0.010627 ± 0.00018694	99.2217 ± 0.016867	0.0096432 ± 0.0001709
Medium	100 $\pm 1.0852e-07$	2.5528e-06 $\pm 2.7129e-07$	100 $\pm 1.0677e-07$	2.4322e-06 $\pm 2.6693e-07$
Varying	99.7109 ± 0.0099345	0.0047217 ± 0.00012742	99.6982 ± 0.010712	0.0054164 ± 0.00012792
VeryLow	99.1992 ± 0.016722	0.0089747 ± 0.00019379	99.2344 ± 0.017533	0.0092891 ± 0.00017349

4.4.3 Discussion

This study focuses on the bias introduced during the acquisition of fingerprint data. Though there are many uncontrolled impacting factors, we demonstrated in this

chapter the need for an algorithm to be independent of the capture device.

Different sets of fingerprint images have been created, from very low-quality images to high-quality images with a set of varying-quality images, with two capturing technologies with or without scratches. The quality scores reveal no significant variation introduced by the scratches within the same capturing quality group. The "varying" quality generation parameter serves here as a reference as it is the most likely capturing quality one can find in the market and the most representative of what we may observe in a real-life fingerprint collection. So, we compute the relative NFIQ2 score variation to the average value of the "varying" quality set. Results are shown in Figure 4.11. We can see that with respect to our reference, the variation of NFIQ2 scores can be clearly significant. This is an indication of the quality gap we may observe between an average sensor in the market and very distinctive sensors. The quality is an indicator of the usability of a fingerprint as a biometric sample. So, a decrease in quality can lead to a poor recognition capacity and a bad user experience, as a good biometric system should be able to recognize people equally. This variation is more important considering the optical technology.

Figures 4.12 show the relative variation of the AUC of each set against the reference set (*i.e.* "varying") respectively with Bozorth3 and MCC. We can observe that the relative variation of Bozorth is similar to the NFIQ2 score (Figure 4.11) for the capacitive datasets. From an operational perspective, this means that this matching algorithm is highly sensitive to the quality of the fingerprints. Moreover, considering the same type of sensor, we can see a change depending on its quality. For the MCC algorithm, given a type of sensor, the AUC seems to be stable regardless of the quality of the sensor. It is visible with its AUC variable close to 0% for 3 datasets and stays stable for the last set. This can be explained by the good performance it achieves, which makes the algorithm able to handle samples of various qualities.

4.5 Conclusion

In this chapter, we investigated the impact of acquisition conditions for the recognition of fingerprints. Experimental results highlighted there was a dependence between the performance of the matching algorithm and the environmental conditions. The trends of impacts of environmental conditions are shared between matchers, but

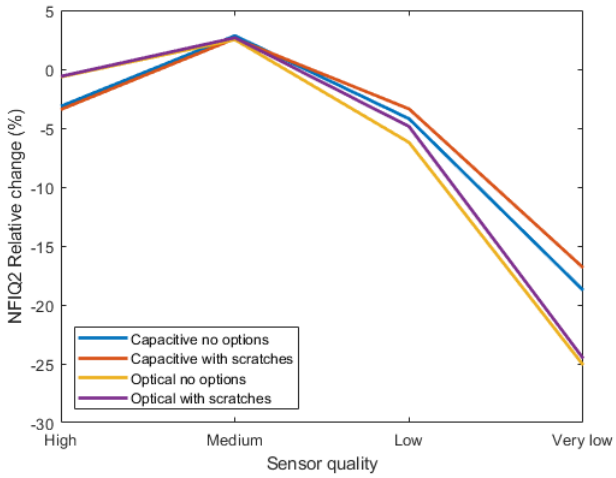
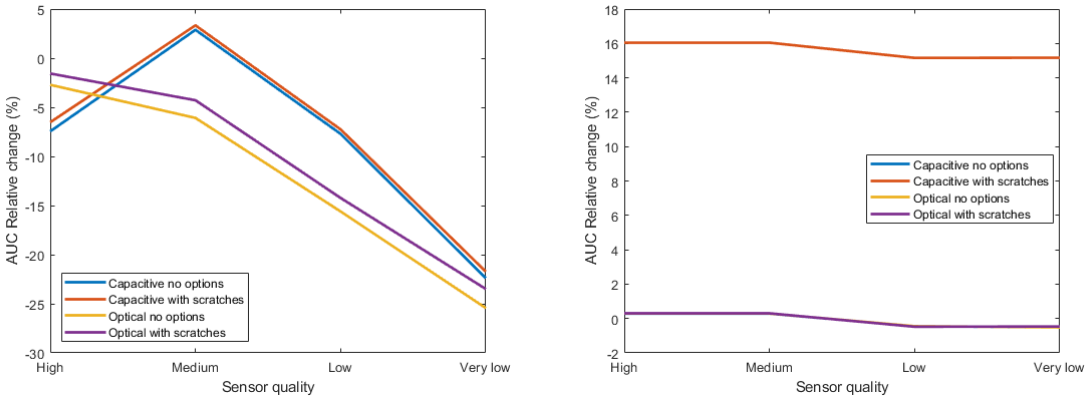


Figure 4.11: Relative change of the NFIQ2: the value of each group is computed against the average score of the "varying" set of that category.



(a) B3 relative change

(b) MCC relative change

Figure 4.12: Relative change of the AUC of the two matching algorithms

the strength of the bias is different. When the enrollment and the verification are done in the same conditions, an increase in performance is observed for all the considered fingerprint-matching algorithms. That means that if we go from location A to location B, there is a major risk for usability and security if we do not redo the enrollment. The biometric fingerprint system may have difficulties to recognize well or be subject to attacks. The p-values are useful to characterize more precisely the factors linked to high-bias impacts.

The MCC gives the worst performance when the test samples are not from the same environment as the enrollment samples whereas it gives the best results on average when we perform a verification in the same conditions that we enrolled the test subject.

Moreover, regarding the quality of the acquisition system, the tested algorithms show that the main challenge for the performance comes from the algorithm. Indeed, even if for Bozorth3 the sensor technology and sensor quality are sources of variations in the performance, the MCC allows us to conclude that recognition algorithms with very high accuracy are less sensitive to sensing technology and more likely to be used with sensors of different quality. Fingerprint algorithms with very high accuracy are more likely to be interoperable.

The next chapter presents a methodology to generate fingerprint spoofs that can be used to evaluate presentation attack detection solutions.

Chapter 5

Generation of synthetic spoofs for the evaluation of fingerprint systems

Contents

Summary : *Nowadays, biometrics is becoming more and more present in our everyday lives. They are used in ID documents, border controls, authentication, and e-payment.... Therefore, ensuring the security of biometric systems has become a major concern. The certification process aims at qualifying the behavior of a biometric system and verifying its conformity to specifications. It involves the evaluation of the system's performance and its robustness to attacks. The anti-spoofing tests imply the creation of physical presentation attack instruments (PAI) and multiple attempts of testing on the device. In this chapter, we propose a new solution based on deep learning to generate synthetic fingerprint spoof images from genuine images. We transform genuine images into what they would look like if they were created from known spoof materials usually involved in fingerprint spoofing tests.*

5.1 Introduction

Biometric systems are now used every day through our devices (face authentication on a smartphone, fingerprint authentication for payment...). Before their deployment, formal evaluations in dedicated laboratories have to be done, in order to assess their conformity to some standards. These tests involve performance and PAD testing. The ISO 30107 series is the basis of PAD testing. The ISO 30107-1 [ISO 30107-1:2023, 2023] identifies 9 different vulnerable points for a given biometric system. The ISO 30107-3 covers only presentation attacks done in the sensor.

The evaluation of PAD requires the creation of physical fingerprint presentation attack instruments (PAI) which can be based on casting some materials on the negative of the fingerprint used as mold [Karampidis et al., 2021]. Testing labs have to spend considerable time to create physically the PAI or spoofs and scan them with the sensor of the device under test. So, testing laboratories need to have skills to create spoofs of good quality to challenge the liveness detection of fingerprint systems. A difficult constraint to achieve within this context is the reproducibility of the evaluation results (that is a mandatory constraint).

We believe that it may be interesting to be able to digitally transform genuine images into known material spoofs. Actually, the security evaluation would be more efficient with a high diversity of PAI species. Digitally synthetic fingerprint spoofs can help to cover a high number of attack attempts with limited time and human resources.

Moreover, it can achieve a much better reproducibility of evaluation results. This work can also help biometric solution providers to train their PAD subsystems in order to improve their resistance to presentation attack instruments. Certification schemes could have ways to diversify their testing by adding testing methods using digitally synthesized images to increase the number of PAI during evaluations.

We investigate how the synthetic PAI look like real ones from the considered material considering 1) their quality, 2) performance from a matching point of view, and 3) detection by an anti-spoofing algorithm.

The contributions of the chapter concern first the definition of a new method for the generation of digitized PAI for fingerprints. Second, we show that we are able to imitate any material of the PAI. Third, we consider the quality of the generated PAI and also their performance compared to real ones. The proposed method could be a useful way for certification bodies to cover a larger scope of attack detection at a low cost with better reproducibility.

5.2 Related works

The classical ways of testing biometric systems request the production of physical PAI. Depending on the certification body, the process uses the cooperation of the subject or not. For instance, Fido Alliance[1] requires that for the creation of attack instruments or spoofs during testing, test subjects have to provide biometric characteristics on which the PAI is based through pressing their finger on a surface creating a latent print, taking a low-resolution photograph, capturing their fingerprint on a fingerprint scanner, or taking a high-resolution photograph (non-cooperative testing). Other certification bodies request that fingerprint molds have to be obtained from an individual by pressing a finger into silicon or other molding material (cooperative testing). Once molds are created, some materials are cast on them to have a presentation attack instrument.

The classical approach is quite tedious for the testing labs as it requests them to hire a large number of people willing to give their personal biometric data for the purpose of a test and commits the lab to ensure data privacy-related regulations such as GDPR are respected. Hence, the use of synthetic biometric data is more

[1] `https://fidoalliance.org/specs/biometric/requirements/`

and more discussed in certification instances. Many papers deal with the creation of synthetic datasets for biometric data generation.

In this chapter, we propose a method for the generation of PAI samples considering it as a style transfer problem, and we validate the process with metrics associated with the evaluation of a biometric solution. The generation of fingerprints or data, in general, using Deep Learning methods can be done in multiple ways. Deep Neural networks have found many applications in computer vision since they have been theorized.

The GANs offer numerous possibilities for the creation of content. As said earlier, many variations of the original GAN have been proposed in the state-of-the-art of generative models for different tasks. Some are specialized in audio [Vasquez and Lewis, 2019], others in image generation. StyleGan and its evolutions [Karras et al., 2019, Karras et al., 2020, Karras et al., 2021] are well known for the generation of realistic face images. These models have been used in the biometric application for fingerprint [Seidlitz et al., 2021, Bouzaglo and Keller, 2022] or face [Colbois et al., 2021] data generation.

However, there is another application of deep neural networks that we are interested in which is the rendering of a content image in different styles from a source domain to another domain is known under the name of ***style transfer***. The image-to-image translation task can be divided into two groups: 1) the two-domain image-to-image translation and 2) multi-domain image-to-image translation.

5.2.1 Two-domain Image-to-Image translation

The two-domain image-to-image translation problem refers to the task of rendering images from domain A to domain B. One of the first to be introduced using conditional adversarial networks is the Pix2Pix [Isola et al., 2017]. Figure 5.1 gives examples of applications and results of Pix2Pix. One of the major criticisms of Pix2Pix is its loss term composed of a combination of an adversarial term with **L1** norm loss, which requires pairs of images from both domains. Pix2Pix gives an answer for two-domain image translation problems, where another study [Zhu et al., 2017] solves the problem of transferring style from A to B and from B to A.

Other solutions have been proposed to improve the results of these architectures: [Liu et al., 2017], [Kim et al., 2017], and the previously mentioned CycleGAN.

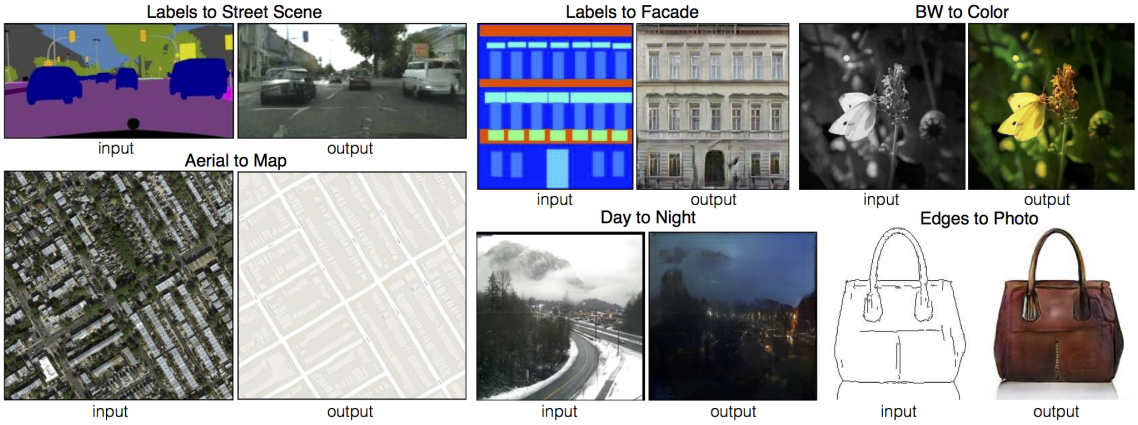


Figure 5.1: Illustration of Pix2Pix results, extracted from [Isola et al., 2017]

Various models have been proposed for the tasks of image-to-image translation across two domains with and without paired data [Zhu et al., 2020, Yin et al., 2020, Lin et al., 2020]. The two-domain image-to-image translation models are restricted to two-domain tasks and if we want to explore more than two domains, we have to train multiple models which can have high training costs without a guarantee that the features from different domains are learned similarly.

For this study, we chose to explore the multi-domain image translation. Indeed, as mentioned in Chapter 3, the testing of PAD involves multiple materials. Some of these networks are limited by the number of domains we can learn simultaneously. So, if we consider multi-domains, we have to train as many networks as styles we want to learn.

5.2.2 Multi-domain Image-to-Image translation

The need for style transfer goes beyond two domain tasks. The basic idea of multi-domain image-to-image translation is to train one generator that learns to map features from a domain to multiple domains.

StarGAN [Choi et al., 2020a] is one of the first to propose the multi-domain translation without having to train cross-domain generators. The method is based on the training of one single generator. AttGAN [He et al., 2019] also uses a single generator training to perform the multi-domain translation. The authors show the efficiency of the method on facial attributes translation. UFDN [Liu et al., 2018] learns domain-invariant representation from data across multiple domains. MWGAN [Cao et al., 2019] is a multi-domain image-to-image translation that proposes to solve the

challenges previous models faced by minimizing the Wasserstein distance across the domains. Other methodologies and methods have been proposed in the research literature [Lee et al., 2020, Choi et al., 2020b, Liu et al., 2020]. A complete list of proposed methodologies for image-to-image translation up to 2021 is proposed by [Pang et al., 2021].

5.3 Proposed method

The proposed method is based on the Wasserstein GAN and particularly MWGAN [Cao et al., 2019] (Multi-marginal Wasserstein GAN). Figure 5.2 shows an illustration of MWGAN on the face for style transfer. Since the introduction of Generative Adversarial Networks (GANs) [Goodfellow et al., 2014] in 2014, many variations of this architecture have been proposed in the literature to generate fake data or to translate content from one domain to another.

We chose MWGAN for its cross-domain performance. In addition, the model is more stable and, as the cross-domain distance is minimized in the loss function, it allows different materials to be learned in the same way without imbalance.

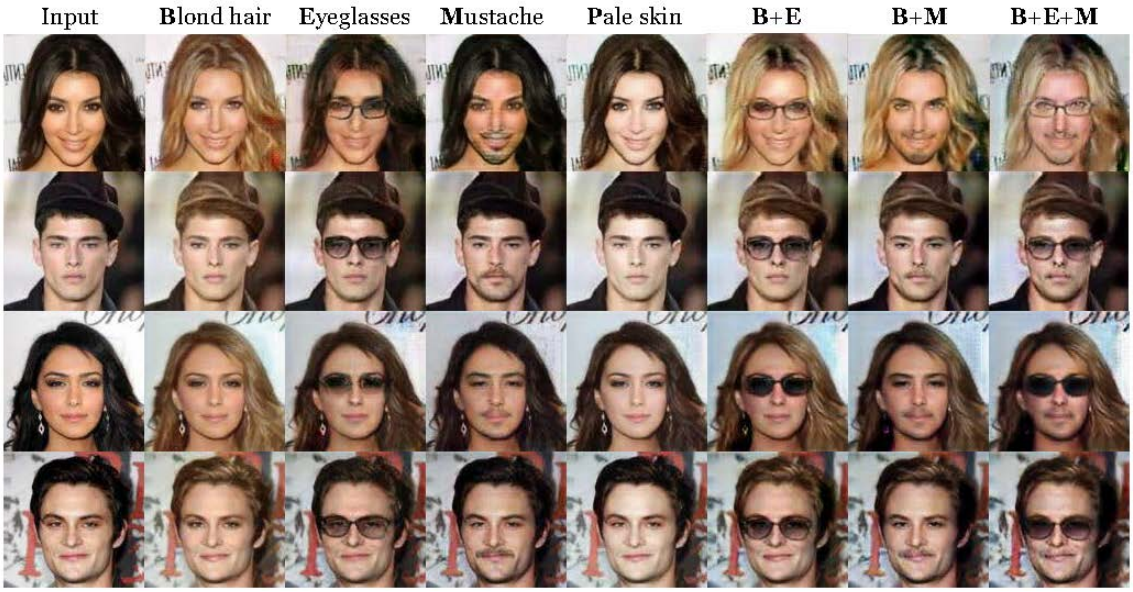


Figure 5.2: Illustrations of the MWGAN method for style transfer on Facial biometric (source [Cao et al., 2019]).

In this work, we use the architecture proposed by Cao et al. [Cao et al., 2019]. They proposed a model of multi-domain image-to-image translation that minimizes

the Wasserstein distance between the learned domains. The model is trained on 30,000 iterations with a learning rate of 0.0001 for the generators and the discriminator, and an Adam Optimizer.

The model as given in the article gives quite good results but we struggle to see visible differences in the generation PAI samples that one may observe for real fingerprint spoofs (PAIs). We decided to add a data-linked term to better differentiate the generated materials from genuine images. Figure 5.3 describes the whole process we proposed.

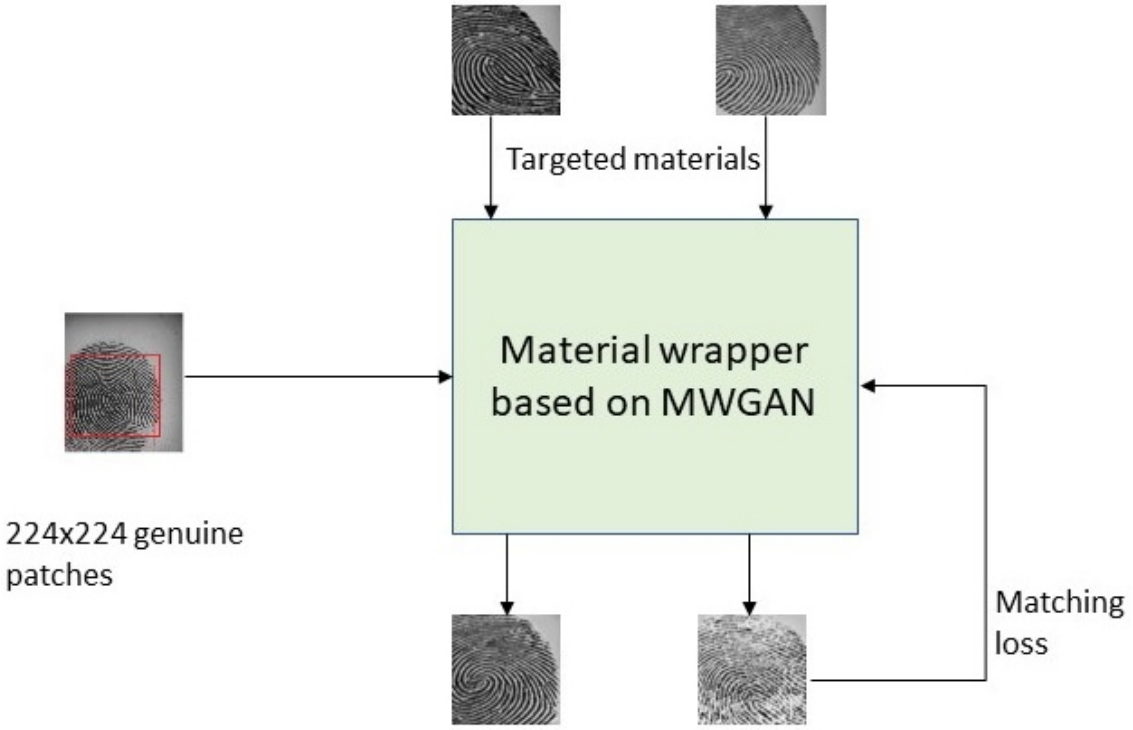


Figure 5.3: Overview of the proposed method (images in the illustration do not necessarily correspond to the same finger).

From [Cao et al., 2019], the domain classification loss of the Multi-Wasserstein GAN is defined as follows. Given an input $x := x^{(0)}$ and generator g_i, the objective is to translate the input x to an output $\hat{x}^{(i)}$ which can be classified to the target domain D_i correctly. To achieve this goal, an auxiliary classifier is introduced $\phi : X \to Y$ to optimize the generators. Specifically, real data $x \sim \hat{\mathbb{P}}_{t_i}$ are labeled as 1, where $\hat{\mathbb{P}}_{t_i}$ is an empirical distribution in the i-th target domain, and generated data $\hat{x}^{(i)} \sim \hat{\mathbb{P}}_{\theta_i}$

are labeled as 0. Then, the domain classification loss w.r.t. ϕ can be defined as:

$$C_\alpha(\phi) = \alpha \cdot \mathbb{E}_{x' \sim \hat{\mathbb{P}}_{t_i} \cup \hat{\mathbb{P}}_{\theta_i}}[l(\Phi(X'), y)] \tag{5.1}$$

where α is a hyperparameter, y is corresponding to $x^{'}$, and $l(\cdot,\ \cdot)$ is a binary classification loss. To add a texture-based term, we define a matching term to the loss to more link the generated data to the real one, defined as:

$$Material\_loss = 1 - match(\hat{x}^{(i)}, Y) \tag{5.2}$$

where $Y = \{y_i\}$ is a batch of real images from the i-th domain.

Then, the classification loss becomes:

$$C_\alpha(\phi) = C_\alpha(\phi) + Material\_loss \tag{5.3}$$

At each epoch, we perform the matching between a generated batch of spoofs and respective real spoofs, for a given material. As the matching score will be higher when comparing an image with itself, maximizing this score will favor the similarity between the synthetically generated spoofs and the real ones. In this study, we use the MCC to compute matching score, as it demonstrates a high efficiency when the two samples are acquired in the same environmental conditions, that is the case here. During the training, we generate a batch of images of each material and compare them to the reference images of the same material. We add the deviation from the maximum reachable value. A detailed implementation of that part is given in Algorithm 1.

Algorithm 1 Attach to material

```
for i < n_epochs do
    for domain in domains do
        translate to domain
        material_loss ← 1 − match(generated images, ref images)   ▷ deviation from the max value
        loss ← loss + material_loss
    end for
end for
```

Figure 5.4 illustrates the application of the proposed method in the context of the certification of biometric systems. The style transfer is applied in order to generate

the PAI samples given a spoofing material. As a reminder, the goal is to generate realistic synthetic fingerprint presentation attack instruments usable to evaluate a fingerprint solution. We propose the use of the synthetic data alone or in combination with an existing dataset to evaluate biometric fingerprint solutions.

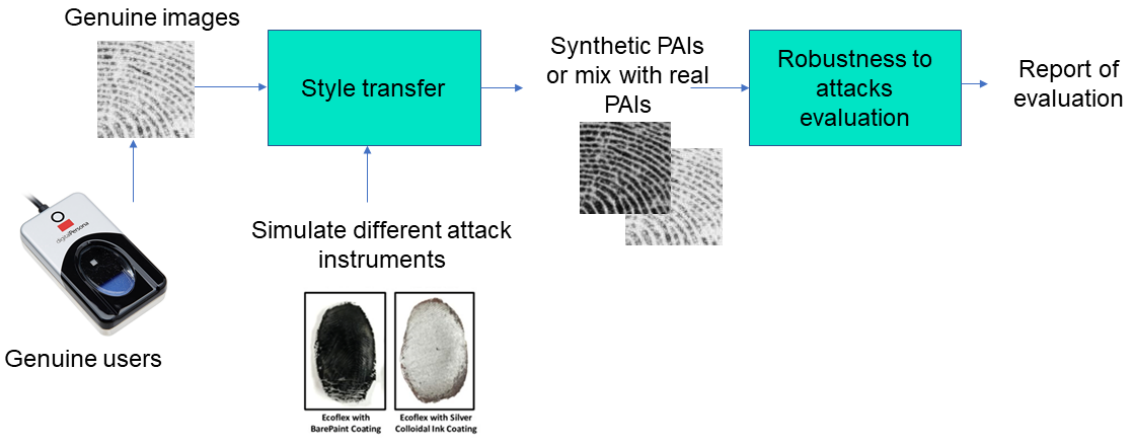


Figure 5.4: Application of the proposed method for the evaluation of biometric fingerprint systems

5.4 Validation of the proposed method

In this section, we present the experimental protocol we set and the experiments we conducted for the validation of the proposed method.

5.4.1 Experimental protocol

Biometric datasets

In this work, we use the PAI datasets LivDet [Ghiani et al., 2013]. They have been created during multiple editions of the international competition of liveness detection on fingerprints. We use data from the 2013 competition from Biometrika sensor, as they were of good quality and from various spoofing materials. This database includes genuine images and spoofs from Ecoflex, Gelatin, Latex, Modasil and Wood Glue. The set is composed of a training set and a testing one of 2,000 images each set (1,000 genuine images and 200 of each spoof material for each set) per used sensor. Usually, train sets are given to the participants to train their algorithms

on that dataset. And the (unseed) test sets are used to validate the trained algorithms.

We train the proposed method on Livdet 2013 fingerprint datasets [Ghiani et al., 2013]. Training is done on the training set of LivDet and validation is done on the Validation set of Livdet 2013. In order to face the few number of available samples in the training set, we applied an augmentation data strategy extracting patches of 224×224. We performed random cropping around the center and extracted 5 patches from each spoof sample to increase the dataset size and facilitate the training of the model. The augmented set is composed of 1,000 images from live fingers and 1,000 images for each material.

To validate the digitally synthesized fingerprint spoofs, different metrics and methods are used to assess the spoofness of generated data. We consider three types: 1) fingerprint quality assessment, 2) performance and 3) presentation attack evaluation.

Presentation attack evaluation

Considering PAD systems is often realized when evaluating the quality of PAI samples. In this work, we use a commercial presentation attack detection solution to analyze how a PAD algorithm could detect such synthetic PAI samples compared to physical ones. The metric used for PAD qualification is the TDR (True Detection Rate), defined in Chapter 3.

5.4.2 Experimental results

We show in this section the obtained experimental results that illustrate the benefit of the proposed method through different evaluation scenarios.

Quality assessment

As a first step, we consider the quality/fidelity of generated PAI samples. Figure 5.5 shows some examples of generated PAI samples from a real fingerprint one (unseen by the model). The first row presents real samples (genuine real samples on the first column) and real PAI with different materials (other columns). Other rows present fingerprint samples for different users (genuine real samples on the 1st column) and other columns correspond to generated PAI samples for each material. The different materials are from left to right: EcoFlex, Gelatin, Latex, Modasil and Wood Glue. We can see visually that generated PAI samples are consistent considering real PAI

samples.

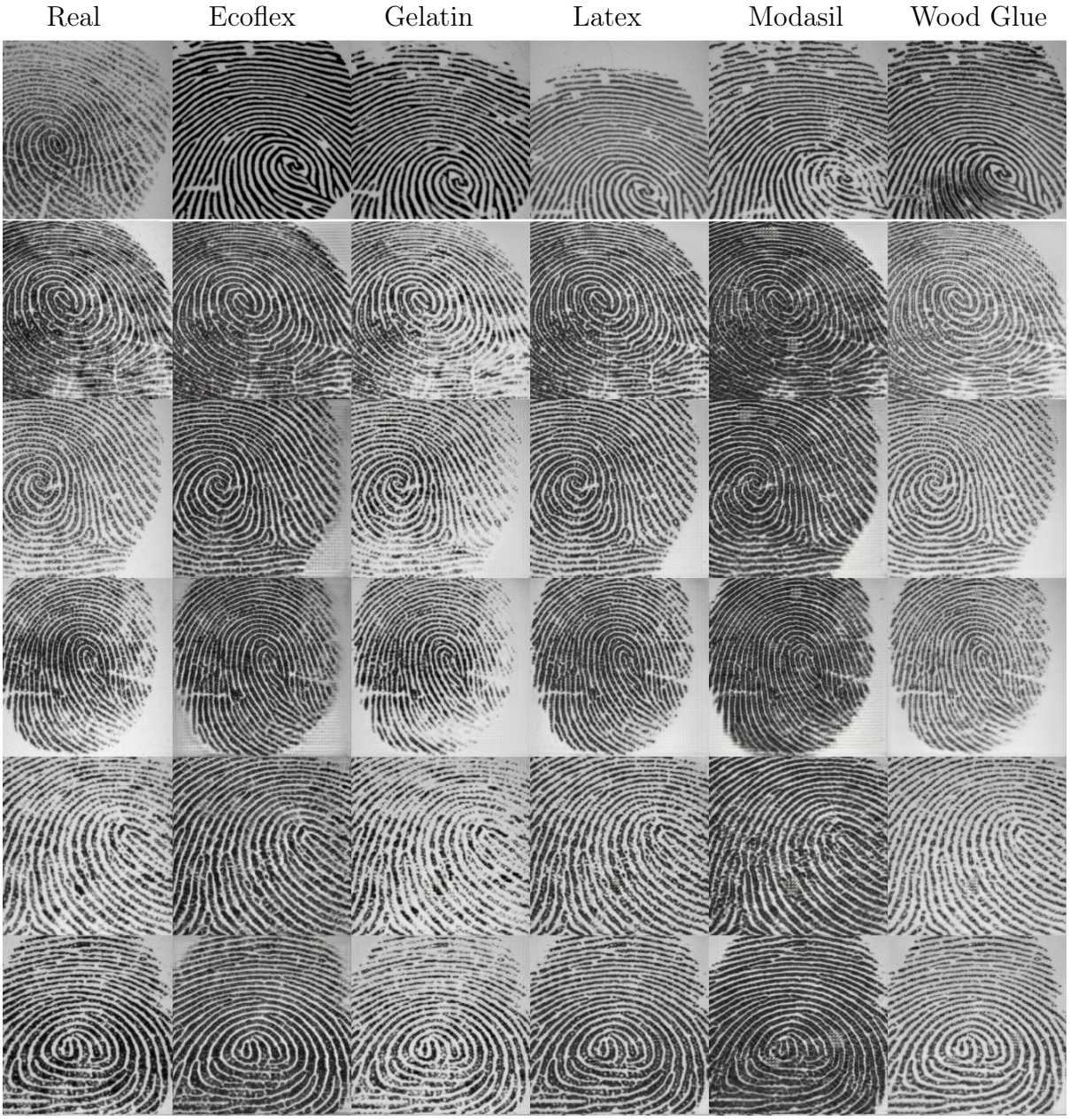
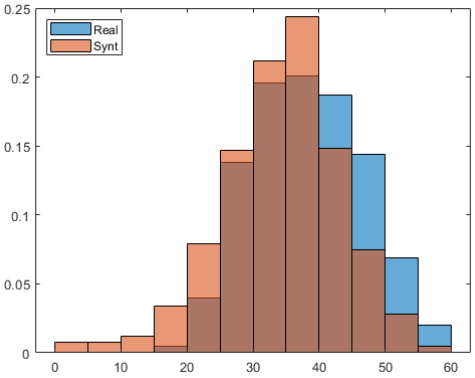


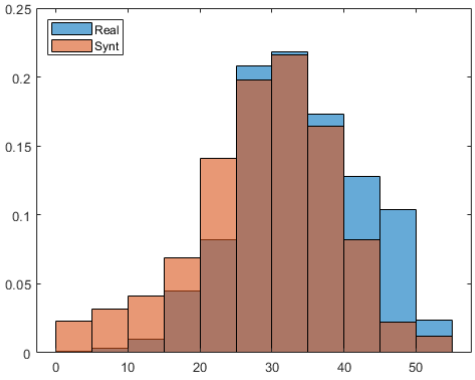
Figure 5.5: Illustration of generated fingerprint PAI samples.

To go beyond the visual aspect of generated PAI images, we consider the fingerprint quality assessment metric using NFIQ2. Figure 5.6 shows the distribution of the NFIQ2 quality scores for synthetic and real PAI samples for each of the 5 materials. We can clearly see there is a high similarity between the two distributions (synthetic

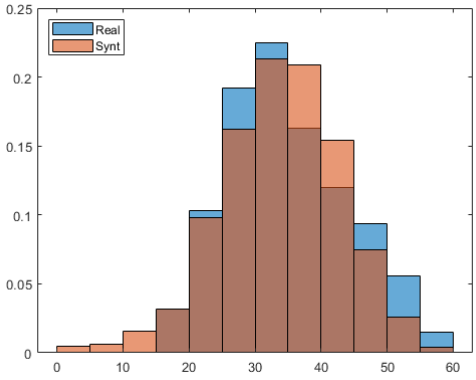
versus real) for each material, and the range of covered quality scores is the same.



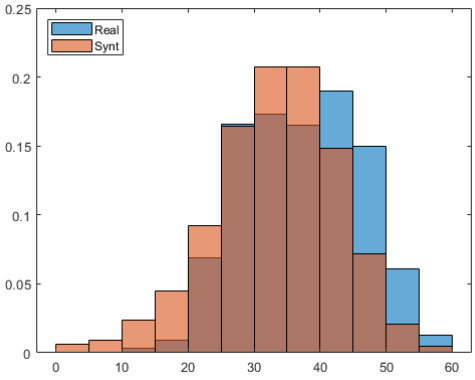
(a) Ecoflex



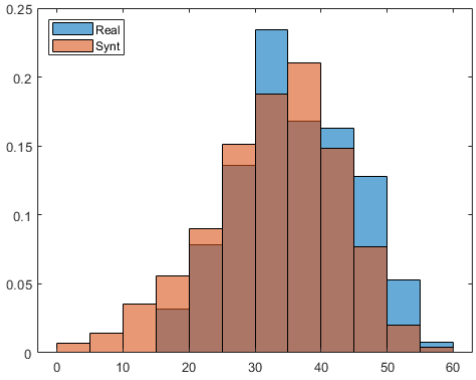
(b) Gelatin



(c) Latex



(d) Modasil



(e) Wood Glue

Figure 5.6: Distribution of NFIQ2 scores for real and our synthetic PAI samples.

We compute for each material the Pearson Linear Correlation Coefficient (PLCC)

between the two distributions of NFIQ2 quality scores to quantify their similarity. The obtained results are presented in Table 5.1. One can observe a decrease of quality scores considering synthetic fingerprint images with respect to real ones. Yet, the computed PLCC is high, since it reaches 97.63% for the lower and 98.78% for the higher. Thus, we can consider synthetic data are of same quality that real ones.

Table 5.1: Statistics of the NFIQ2 scores for real and generated PAI samples for the 5 materials

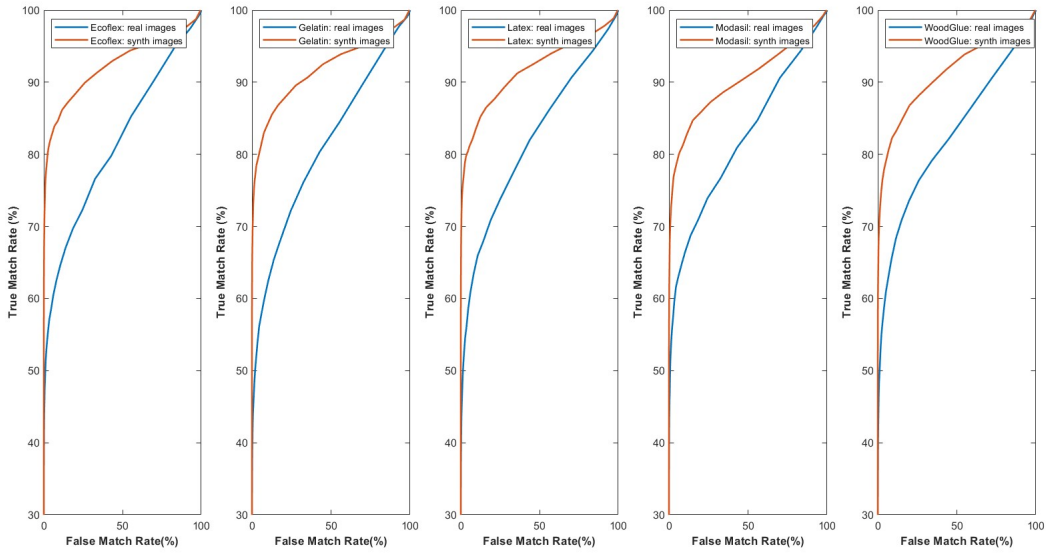
Material	Real PAI avg	Real PAI std	Synthetic PAI avg	Synthetic PAI std	Pearson correlation (%)
EcoFlex	37.879	8.4408	33.842	9.0396	97.63
Gelatin	33.33	8.875	28.549	9.9905	98.6
Latex	34.445	9.0549	33.454	9.1149	98.06
Modasil	36.874	8.8957	32.862	9.4801	98.48
WoodGlue	35.681	8.8119	32.328	10.0967	98.78

Performance evaluation

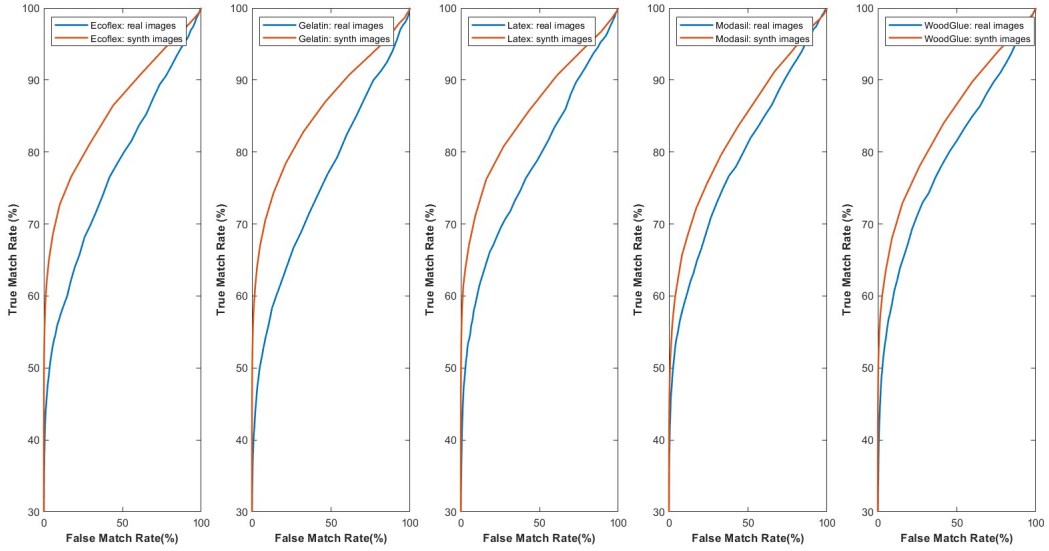
In this part, we evaluate the performance of the proposed approach considering metrics detailed in the previous section. Generated PAI samples should have a similar performance to real PAI ones. First, we extract minutiae from fingerprint samples with NIST *mindtct*. Then we use two fingerprint-matching algorithms from minutiae templates: NIST Bozorth3 and the MCC matching algorithm. For each matching algorithm, we compare the real fingerprint PAIs and the PAIs generated from the genuine fingerprint samples. Figure 5.7 shows the obtained ROC curves. One can remark that with the MCC matching algorithm (which is much more efficient than Borzorth3 when samples are acquired in the same environmental conditions), obtained ROC curves for real and synthetic data are very similar whatever the considered material.

As illustration, Table 5.2 and Table 5.3 give the detailed performance of the MCC matching algorithm. Figure 5.7 shows the detailed performance of the two considered matchers. Even if the work is about the generation of fingerprint presentation attack instruments, it is important for synthetic spoofs to have fingerprint features and be able to be recognized when there is no anti-spoofing. We can see that considering only the "fingerprintness" of generated data, synthetic PAIs sometimes perform better than real spoofs from matching only considering AUC, EER, and even ROC curves.

Indeed, depending on the number of epochs on the network, some artifacts due to the physical characteristics of the material (conductivity, elasticity, etc.) present on the real PAI sample may not be visible on the generated one. Different levels of real



(a) Bozorth3



(b) MCC

Figure 5.7: ROC curves obtained for B3 and MCC considering real spoofs and synthetic spoofs.

PAI quality can be simulated by tuning the epoch number on the network. It is possible to generate PAI samples corresponding to very high spoof quality, which requires a long time and high expertise in physical creation.

Table 5.2: Performance of Bozorth3 on real and synthetic data

Material	real material AUC (%)	synth material AUC(%)	real material EER(%)	Synth material EER(%)
Ecoflex	81.49 ±0.039	92.40±0.029	26.133 ±0.03	13.25 ±0.06
Gelatin	80.81 ±0.04	91.72±0.03	26.18 ±0.03	13.79 ±0.048
Latex	81.89±0.040	91.74±0.03	25.68 ±0.03	14.13 ±0.06
Modasil	82.07 ±0.04	90.23±0.03	25.24 ±0.03	15.18 ±0.03
WoodGlue	82.39 ±0.03	91.25±0.03	24.79 ±0.03	15.28 ±0.03

Table 5.3: Performance of the MCC matcher on real and synthetic PAI samples.

Material	Real PAI AUC (%)	Synthetic PAI AUC(%)	Real PAI EER(%)	Synthetic PAI EER(%)
Ecoflex	77.63 ±0.04	86.04±0.02	29.86 ±0.046	20.83 ±0.037
Gelatin	76.39 ±0.04	85.84±0.02	31.087 ±0.03	21.61 ±0.02
Latex	78.49 ±0.04	85.83±0.02	28.98 ±0.05	21.50 ±0.09
Modasil	79.31 ±0.039	83.63±0.02	28.13 ±0.05	24.32 ±0.03
WoodGlue	79.54 ±0.04	84.57±0.023	27.53 ±0.04	23.98 ±0.05

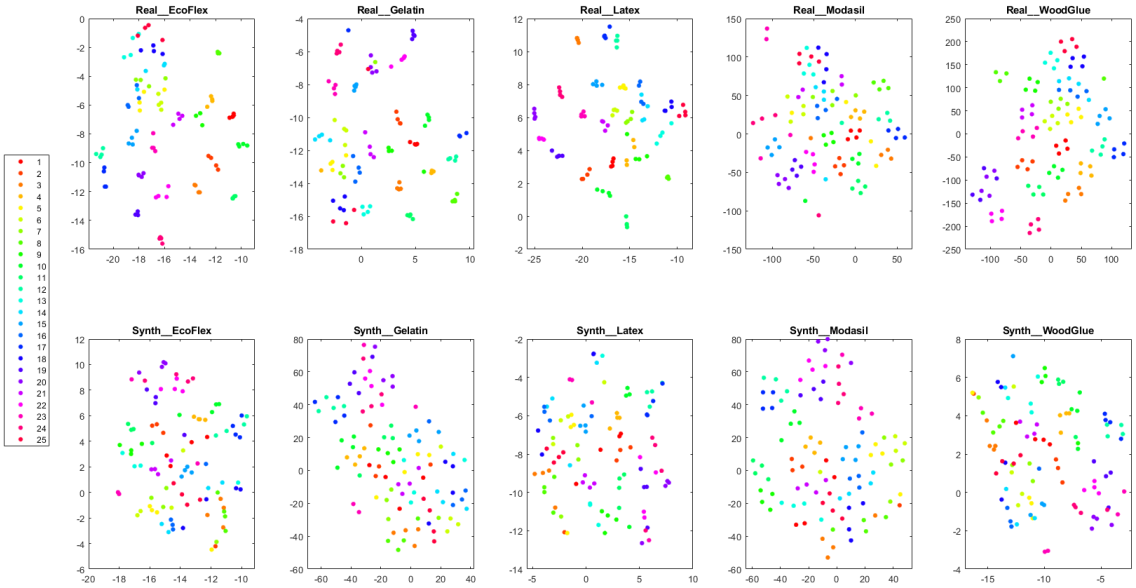


Figure 5.8: Visualization of features of 25 same users for each material.

As a complement, we observe the behavior of deep features for real and synthetic PAIs. We used in this work a pre-trained CNN architecture based on AlexNet [Krizhevsky et al., 2012]. AlexNet outperforms other classification methods in the fingerprint recognition task. In addition, the model has relatively few parameters. We applied transfer learning using synthetic fingerprint samples with the SFinGe software [Raffaele et al., 2004] (60,000 images from 1,200 random users and 50 samples per user) in order to generate reliable deep features describing a fingerprint image. As an example, the obtained EER value with these deep features (cosine

distance for matching) is less than 1.3% on 2006 FVC datasets [2]. We consider this CNN matcher as a generic one and associated deep features are used to better analyze the efficiency of the proposed method.

For each material, we observe the same 25 users from LiveDet. We put an interest in the way this algorithm recognizes these 25 users, considering their real and synthetic PAIs. We extracted the 1,200 features of the AlexNet fully connected layer, then used t-SNE [Van der Maaten and Hinton, 2008] to observe the separation between the classes (here the 25 considered users), features from real images are on the first row and features from synthetic images on the second row. Figure 5.8, we can see that for the synthetic data, most of the time, the CNN matcher succeeds in classifying the user. For the Modasil material, clusters are less visible. These differences may come from the geometrical difference between data. Indeed, during the acquisition of the original data (living fingers and real spoofs), the fingers are not scanned at the same position. The possible translation/rotation may lead to a geometric transformation between the synthetic data and the corresponding real one. This results often in the presence of different parts on the synthetic image and the real one which increases the recognition task. This favors lower scores when the two images are compared which can be observed in Figure 5.9. This figure shows the distribution of the Cosine similarity scores when comparing the 1,200 features vector of real and synthetic PAI of the 25 users for each material.

Presentation attack evaluation

Table 5.4 shows the behavior of a PAD algorithm on synthetic PAI samples. We can see that the PAD algorithm struggles around 50% of the time to distinguish the digital PAI from genuine images, except for the Modasil. This informs us that even if a material does not succeed in fooling a fingerprint system by presentation to the sensor, it might succeed if the counterfeiter manages to intercept the flow of data coming from the sensor in the biometric processing chain. The last case is not covered by presentation attack detection, but is a type of test to evaluate the liveness detection independently of the matching unit or the acquisition system. These "Cyber" attacks are more and more considered. They warn us to increase the efficiency of software anti-spoofing methods.

[2] `https://biolab.csr.unibo.it/FVCOnGoing/UI/Form/Home.aspx`

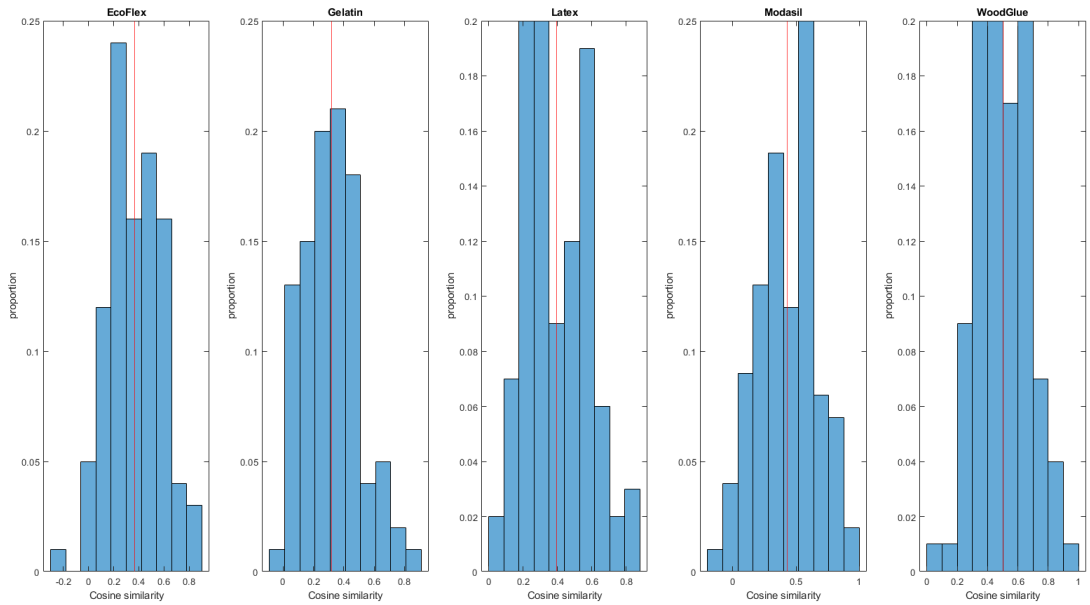


Figure 5.9: Distribution of Cosine distances between real and synthesized data with average Cosine score in red

Table 5.4: TDR of the commercial algorithm PAD module

Material	TDR on real PAI	TDR on synthetic PAI
EcoFlex	99.47%	53.72 %
Gelatin	95.04%	52.56 %
Latex	99.82%	43.02 %
Modasil	97.87%	94.19 %
WoodGlue	99.65%	61.74 %

5.4.3 Discussion

The proposed method for generating synthetic PAI gives very good results that can fool PAD algorithms. A performance comparison with some of the state-of-the-art works is presented in Table 5.5 considering the TDR metric as the metric of performance. The observed differences with the real spoofs from targeted materials can come from different factors. Indeed, the data-linked term is a matching score based on the extraction of minutiae. During the first epochs of the training, the network tries to make the template as similar as possible to the destination material dataset. Indeed, some imperfections coming from the material itself can lead to the occurrence of some features in the images that can be interpreted as minutiae. These imperfections can be unique to some materials but not only, as demonstrated in [Wone et al., 2021].

Table 5.5: Comparative table of performance with some of the state of the art works

Study	Validation set	Performance
Gajawada *et al.* [Gajawada et al., 2019]	LivDet2015	TDR=78.05%
Chugh *et al.* [Chugh and Jain, 2019]	MSU-PAD	TDR=91.78% @FDR=0.2%
Grotz *et al.* [Grosz et al., 2020b]	LivDet 2015	TDR=87.86% @ 0.5%
Grotz *et al.* [Grosz and Jain, 2022]	LivDet 2013,LivDet 2015, GCT 1-5, GCT 6	-
Our study	LivDet 2013 Briometrika Test	TDR=98.37% for real (average) TDR=60.85% for synthetic (average) 98.3% average correlation for NFIQ2

5.5 Conclusion

In this chapter, we present an original work for the generation of fingerprint PAI from different materials given genuine images and a multi-domain style transfer model. Obtained results with generated PAI show that the performance is similar to real PAI from the targeted materials. Moreover, the tested PAD system has its TDR highly dropping when facing these images. This informs us that even if the quality level of some materials did not succeed in fooling a fingerprint system by presentation to the sensor, the generated PAR might succeed if the quality level generated is very high. Our method can be useful in evaluating the robustness of biometric systems to high-level fingerprint PAI. However, the difference between genuine and fake data becomes less important as long as the training goes on. This leads to a decrease in learning performance and visual difference from what we might observe at the beginning of the training, which informs us that matching only is not sufficient to link the loss of the material itself.

Chapter 6

Conclusion

Contents

Summary : *This short chapter presents the conclusion of our work.*
We highlight the contributions of the achieved research.
We also list different perspectives.

6.1 Context

This PhD thesis has been realized in the context of certification of fingerprint biometric systems. Biometrics is more and more used by large public and are not restricted to only administrative tasks anymore. Biometric systems and fingerprint systems in particular are getting more and more popular and deployed as solutions for frictionless and secured solutions for user authentication and identification. Therefore, they need to be certified in order to assess their conformity to some defined test plan. In addition, due to multiple reasons, biometric solutions should not only be compatible or rely on a specific capturing system. So, interoperability is a key to having a fingerprint biometric that works across different sensors. Plus, the tests done in testing laboratories need to be repeatable.

During this Ph.D. study, we have brought three contributions summarized hereafter.

6.2 Contributions

1. **Evaluation of acquisition context impact on the performance of fingerprint systems**
   The first contribution of this thesis is the evaluation of the acquisition conditions' effects on the performance of fingerprint systems. First, we built a fingerprint dataset in a controlled environment and studied the effects of each environment on the quality of the acquired dataset and also on the performance of 3 matching algorithms. This study highlighted an overlooked problem in fingerprint systems which contributes to the non-reproducibility of biometric evaluations. Moreover, we emphasize the major risks of security for a biometric system if it is not able to handle these environmental changes.

   Moreover, we highlighted the necessity to have interoperable biometric systems that are not dependent on the hardware system they are associated with. This study shows how the acquisition quality can impact the performance of a

matching fingerprint. It also highlights that if a fingerprint-matching algorithm is not strong enough, it can be a critical security issue for the final solution.

2. **Generation of presentation attacks instruments for the evaluation of PAD**
   The second contribution is a method to generate fingerprint spoofs that can be used for the evaluation of presentation attack detection. The proposed method is based on an existing deep-learning model (MWGAN) which is a multi-domain image-to-image translation model that has been introduced to overcome learning problems other models had when it comes to more than two domains. We add a matching term to the loss function to give it physical meaning. The results show good similarity when it comes to the quality and the fingerprint features for the recognition, and the PAD test gives good results.

3. **A generic method to validate synthetic biometric data**
   We propose a generic method to validate generated biometric data based on objective metrics to measure. For this work, an objective quality metric (NFIQ2.0) is used to assess the usability of the generated data as biometric samples in a recognition process. We also used the performance contribution of the dataset as well as the presentation attack detection for the synthetic fingerprint spoofs. We applied this methodology to the generation of synthetic spoofs to validate our approach.

6.3 Perspectives

During this PhD, we spotted many ideas to go further.

1. For the acquisition conditions, we highlighted the effects of the environmental conditions on fingerprint and the importance of interoperability for a biometric system for user experience, security, and reproducibility of the fingerprint evaluation. However, for a real evaluation, such tests are very constraining as controlling environmental conditions are quite complicated and requests specific calibrated equipment. A solution could be to use deep-learning solutions to map images acquired in nominal conditions and simulate them in some specific conditions. In that way, it is possible to predict the behavior of the biometric system in a specific targeted condition. Another solution is to go for

a generic methodology to have adaptative processing for data according to the environment where they are acquired.

2. The generation of fingerprint presentation attack instruments has been done using a deep learning model trained on a fingerprint spoof dataset. However, we believe that the results will be better with a larger dataset, and a pattern-linked loss term will be more attached to the texture of the images. Moreover, to be totally free from legal constraints, a generation from scratch or a first step with generation from scratch followed by texture mapping to imitate presentation attack instruments will help to facilitate training and evaluation of biometric fingerprint systems and make them available without restrictions due to privacy.

3. We expressed the need to have interoperable biometric systems. This observation is not only restricted to fingerprint systems but is also meaningful for systems based on other modalities. A perspective of this work is to be able to generate synthetic data that reproduce the physical characteristics of a targeted sensor using a few data captured with this sensor. We want to extend that approach to other biometric modalities.

4. As stated earlier, it is difficult to evaluate the effects of a particular parameter on a biometric system. We wonder if it is possible to use an optimized process (using genetic algorithms, for instance) to find the combination of parameters that would affect the most a biometric system.

Chapter 7

Introduction (Français)

Nous vivons dans un monde où l'informatique occupe une place centrale. Au regard des applications de l'informatique, la sécurité des systèmes informatiques ainsi que les données des utilisateurs doivent être garanties. Elle protège et constitue un élément essentiel ainsi qu'un gage de confiance pour les utilisateurs. Les premières solutions à être proposées reposaient essentiellement sur la **connaissance**. C'est le cas par exemple des mots de passe. Le mot de passe permet de s'assurer que seulement la personne ayant la connaissance d'un secret puisse avoir accès un service ou un système donné. Cependant, les mots de passe restent vulnérables et sujets aux attaques. Selon Microsoft, il y a 1287 attaques de mots de passe toutes les secondes[1]. En plus des attaques, les mots de passe sont souvent partagés, ou notés sur un bout de papier pour mieux les retenir, ce qui participent largement à leur compromission. Une autre solution basée sur la **possession** a été proposée. Il s'agit pour un utilisateur de prouver qu'il est détenteur d'un objet (unique) que lui seul possède. C'est le cas des badges, ou des clés uniques d'accès. Ces objets peuvent être dérobés ou perdus, ou encore attaqués par copie de signature numérique par exemple.
Une autre solution basée sur **qui on est** est également possible. Cela permet de garantir l'accès uniquement à la personne légitime. La biométrie offre ce genre de possibilité.

Selon la CNIL (Commission Nationale de l'Informatique et des Libertés), *la biométrie regroupe l'ensemble des techniques informatiques permettant de reconnaître automatiquement un individu à partir de ses caractéristiques physiques, biologiques, voire comportementales. Les données biométriques sont des données à caractère*

[1] `https://www.microsoft.com/en-us/security/blog/2023/01/09/microsoft-entra-5-identity-priorities-for-2023/`

personnel car elles permettent d'identifier une personne. Elles ont, pour la plupart, la particularité d'être uniques et permanentes (ADN, empreintes digitales, etc.). La biométrie permet de s'assurer qu'une personne est qui elle prétend être en analysant sa donnée biométrique. Elle contribue ainsi, de façon fluide, à garantir la sécurité et l'intégrité d'un système informatique afin de protéger les utilisateurs et leurs données personnelles. Il existe principalement deux familles de biométries: les modalités biométriques physiologiques et comportementales. La première catégorie concerne les biométries les plus "traditionnelles" comme l'empreinte digitale, le visage tandis que la seconde regroupent les façons d'identifier un individu à sa façon de se comporter ou d'interagir avec un appareil (façon de marcher, façon de taper au clavier entre autres). Selon [Jain, 2005], pour être considérée comme donnée biométrique, une caractéristique physiologique ou comportementale doit avoir les propriétés suivantes:

- **Universalité**: Tout individu doit posséder cette caractéristique;
- **Unicité**: La caractéristique doit être suffisamment différente d'une personne à une autre;
- **Permanence**: la caractéristique doit être suffisamment invariante (par rapport au critère de comparaison) sur une certaine période;
- **Collectabilité** : la caractéristique peut être mesurée quantitativement.

La Table 7.1 présente pour quelques modalités biométriques les forces et faiblesses de chacune d'elles.

Table 7.1: Comparaison de quelques modalités biométriques, source [Jain et al., 1999].

Biométrie	Unicité	Collectabilité	Performance	Acceptabilité	Permanence
Empreinte digitale	Haute	Moyenne	Haute	Moyenne	Haute
Visage	Faible	Haute	Faible	Haute	Moyenne
Iris	Haute	Faible	Haute	Faible	Haute
Voix	Faible	Moyenne	Faible	Haute	Faible
Dynamique de frappe au clavier	Faible	Moyenne	Faible	Moyenne	Faible
Démarche	Faible	Haute	Faible	Haute	Faible

Un système biométrique complet est principalement composé d'un module de capture, d'une partie traitement du signal, d'une base de données, d'un module de comparaison et d'un système de décision. Le module de capture sert à acquérir

une donnée biométrique. Il peut être composé d'un capteur (lecteur d'empreintes digitales, caméra,...). Il capture la donnée biométrique qui est traitée par le module de traitement du signal. Celui-ci contrôle la qualité de la donnée, la segmente et extrait les paramètres utiles dans le cadre d'une comparaison biométrique. Lors de la phase d'enregistrement, les paramètres biométriques de la personne sont stockés dans une base de données et serviront de référence pour cette personne. Lors d'une opération d'identification ou d'authentification, les paramètres biométriques extraits de la donnée biométrique soumise sont comparés aux paramètres de références, et en fonction du score de similarité et du degré de confiance fixé, un verdict d'authentification ou d'identification est retourné. Un schéma montrant les principaux composants d'un système biométrique est visible à la Figure 7.1.

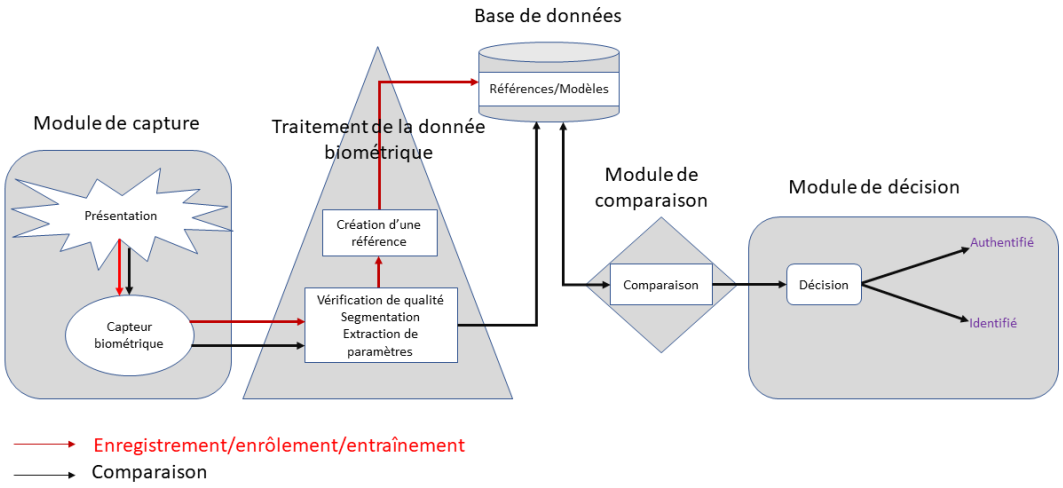


Figure 7.1: Principaux composants d'un système biométrique (schéma inspiré de [ISO 19795-1:2021(E), 2021])

La biométrie est de plus en plus utilisée aujourd'hui dans notre vie quotidienne. D'une action la plus répétitive comme déverrouiller son téléphone à des actions plus délicates comme valider une transaction bancaire, la biométrie est utilisée pour prouver son identité.

L'empreinte digitale est la plus ancienne des données biométriques utilisées. Des traces d'utilisations de l'empreinte digitale comme signature chez les babyloniens dès -5000 et chez les chinois dès -1900 ont été retrouvées. Les empreintes ou

dermatoglyphes se forment dès les premières semaines de vie du fœtus et résultent des frottements des doigts encore en formation avec le liquide amniotique et les structures utérines [2].De ce fait, les empreintes digitales ne sont ni héréditaires, ni partagés entre jumeaux et les chances d'avoir exactement la empreinte digitales sont de 1 sur 64 milliards. L'identification par empreinte digitale se fait principalement par comparaison de motif d'image ou comparaison de gabarits qui contiennent souvent les coordonnées des minuties. Les minuties correspondent principalement aux points de terminaison de ligne ou de bifurcation.
De nos jours, les empreintes sont largement utilisés que ce soit dans un cadre légal ou embarqué dans des appareils portables. Selon un rapport fourni par CISCO, en 2022, 81% des téléphones "intelligents" embarquent une fonctionnalité biométrique, et d'autres études montrent que lecteurs d'empreintes digitales embarqués sur ces téléphones ont bondi de 19% à 60% entre 2014 et 2018. La biométrie et les empreintes digitales en particulier sont déployées pour simplifier certaines tâches tout en garantissant un degré de confiance et une sécurité élevés. Elles sont employées dans les documents d'identité, les transactions bancaires, les accès physiques et logiques, la santé, dans le domaine militaire, etc.
Au regard des services concernés et des enjeux que cela implique, la sécurité des systèmes biométriques doit elle aussi être certaine. Ainsi, les systèmes biométriques doivent être certifiés conformes à un plan de test proposé par une autorité de test qui témoigne de l'utilisabilité et de la sécurité des systèmes biométriques. La certification des systèmes biométriques est une étape cruciale de la production qui précède le déploiement du produit. Elle est réalisée par un laboratoire, sous l'autorité d'un organisme de test, afin de vérifier la conformité d'un système biométrique. Un bon système biométrique doit pouvoir reconnaître une personne légitime, rejeter un imposteur et résister aux attaques notamment par présentation.

Objectifs de la thèse

Cette thèse est le résultat d'une collaboration entre le laboratoire GREYC et la société FIME SAS. Elle est co-financée par l'Association Nationale de la Recherche et de la Technologie (ANRT) dont l'objectif est de promouvoir des innovations via des partenariats entre les laboratoires de recherches et les entreprises du numérique. La thèse s'inscrit dans le contexte de la certification des systèmes biométriques en général et des systèmes à base d'empreintes digitales en particulier. En effet, pour un système certifié, il est important de connaître les scénarios d'essai qui ont été

[2] https://www.police-scientifique.com/vrais-jumeaux-mÃłmes-empreintes/

utilisés par le laboratoire de certification pour évaluer la conformité d'un système à un plan de test spécifique. Les tests qui ont conduit à l'approbation du produit testé doivent pouvoir être reproduits à l'identique ou avec un écart minimal en ce qui concerne l'expérience de l'utilisateur et la sécurité du produit certifié. De nombreux facteurs ont été identifiés comme rendant difficile la reproductibilité des tests. Si certains d'entre eux dépendent de l'interaction entre l'utilisateur et le système, le système et l'environnement dans lequel il est déployé peuvent avoir un impact significatif sur ses capacités de reconnaissance du système biométrique. Le contexte d'acquisition est le principal facteur de non-reproductibilité des tests. Cela concerne à la fois les conditions environnementales de l'acquisition et le système de capture lui-même. Comme indiqué précédemment, la certification d'un système biométrique (empreintes digitales) est évaluée sur la base de sa capacité à reconnaître la bonne personne et à résister aux attaques par présentation. Cette dernière partie prend du temps et les laboratoires n'ont aucune indication sur les méthodes d'apprentissage des algorithmes testés car la plupart des produits testés sont des boîtes noires. Il serait donc incertain d'utiliser des bases de données existantes, qui sont souvent limitées à des fins de recherche, pour ce genre de test. En outre, les contraintes associées aux tests nous incitent à nous orienter vers la génération de données synthétiques réalistes, pouvant être utilisées pour l'évaluation de la détection des attaques par présentation d'un système biométrique à empreintes digitales.

Tout au long de cette thèse, nous adressons les questions suivantes:

- Comment le contexte d'acquisition d'un système biométrique à empreintes digitales peut avoir un impact sur ses performances ?
- Les données synthétiques peuvent-elles aider dans l'évaluation de solutions biométriques à base d'empreintes digitales ?
- Comment peut-on assurer la reproductibilité des évaluations de solutions biométriques ?

Contributions

Dans cette section, nous présentons les principales contributions de cette thèse:

1. **Première contribution**: Nous avons mené une étude pour comprendre l'impact des conditions environnementales sur les systèmes d'empreintes digitales. Les conditions environnementales sont l'un des biais les plus négligés pour les systèmes d'empreintes digitales. En effet, à cause de ces facteurs, il est très difficile de prédire le comportement d'un produit une fois qu'il est déployé

dans une région d'une monde. Au cours de cette thèse, nous avons construit une base de données d'empreintes digitales acquises dans des conditions de température et d'humidité contrôlées. Nous avons observé l'effet des conditions environnementales en utilisant une mesure objective de la qualité des empreintes digitales et différents algorithmes de comparaison d'empreintes digitales. Nous avons également souligné l'importance de la qualité du système d'acquisition et la manière dont elle peut affecter la performance de la solution biométrique finale.

2. **Seconde contribution**: Nous avons proposé une méthode de génération d'instruments d'attaque par présentation utilisable dans le cadre de l'évaluation du module de détection d'attaques par présentation pour un système biométrique à base d'empreintes digitales. Nous avons proposé une méthode générique pour valider les données biométriques synthétiques. Les données générées sont validées du point de vue de la qualité et de la reconnaissance. Nous comparons notre méthode avec les solutions qui existent dans l'état de l'art.

Organisation du manuscrit

Le manuscrit est organisé comme suit:

- Le chapitre 2 présente la certification des systèmes biométriques. On y présente les contraintes auxquelles les laboratoires de certification font face lors de cette étape. On exprime les motivations qui nous poussent à partir vers des données synthétiques. On y détaille le processus de certification, sa logistique, ses métriques et méthodes ainsi que la nécessité d'avoir des données synthétiques. On propose une méthode générique pour valider les données biométriques synthétiques avec des méthodes objectives liées à l'utilisabilité d'un système biométrique ainsi que sa sécurité.

- Le chapitre 3 est dédié à l'analyse du contexte d'acquisition et à la manière dont il peut influer sur les performances d'un système biométrique à base d'empreintes digitales. Ce contexte peut être soit les conditions environnementales dans lesquelles la donnée biométrique est acquise, soit le dispositif de capture. Le contrôle de ces contextes est la clé de la reproductibilité des tests biométriques et la minimisation de ses effets permettrait d'aller vers des systèmes biométriques interopérables et des évaluations répétables. En effet, on souhaite comprendre les paramètres

auxquels l'évaluation des systèmes biométriques peut être corrélée.

- Le chapitre 4 présente la méthode que nous proposons pour générer des bases de données d'empreintes digitales synthétiques afin de tester la détection des attaques par présentation, ainsi que la comparaison des données générées avec des données existantes. En effet, cela répond au besoin de disposer de données synthétiques exprimé au chapitre 1. Nous présentons la méthodologie que nous proposons dans cette thèse ainsi que les résultats que nous avons obtenus en utilisant la méthodologie de validation que nous avons introduite dans le chapitre 2.

- Le chapitre 5 est dédié à la conclusion de ce travail. Nous y présentons également des perspectives à ce travail.

Chapter 8

Conclusion (Français)

Contexte

Cette thèse doctorale a été réalisée dans le cadre de la certification des systèmes biométriques à base d'empreintes digitales. La biométrie est de plus en plus utilisée par le grand public et ne se limite plus à des tâches administratives. Les systèmes biométriques, et en particulier les systèmes d'empreintes digitales, sont de plus en plus populaires et déployés en tant que solutions d'authentification et d'identification sans contraintes et sécurisées pour les utilisateurs. Ils doivent donc être certifiés afin d'évaluer leur conformité à un plan de test défini par une autorité. En outre, pour de multiples raisons, les solutions biométriques ne doivent pas seulement être compatibles ou dépendre d'un système de capture spécifique. Les matériaux (partie physique d'un système biométrique) ayant une durée de vie, ou une disponibilité limitée, l'interopérabilité est donc essentielle pour disposer d'une solution biométrique à base d'empreintes digitales qui fonctionne avec différents capteurs et ce quelque soit l'endroit dans lequel il opère. Les tests effectués dans les laboratoires d'évaluation doivent pouvoir être répétés afin de garantir le bon fonctionnement d'un produit.

Contributions

1. **Evaluation de l'impact des conditions d'acquisition sur les systèmes à base d'empreintes digitales**
   La première contribution de cette thèse est l'évaluation des effets des conditions d'acquisition sur les performances des systèmes d'empreintes digitales. Dans un premier temps, nous avons construit une base de données d'empreintes digitales dans des environnements contrôlés et étudié les effets de chaque

environnement sur la qualité de l'ensemble de données acquises ainsi que sur la performance de 3 algorithmes de reconnaissance d'empreintes digitales. Cette étude a mis en évidence un problème négligé dans les systèmes d'empreintes digitales qui contribue à la non-reproductibilité des évaluations biométriques. De plus, nous soulignons les risques majeurs de sécurité pour un système biométrique s'il n'est pas capable de gérer ces changements environnementaux. En effet, nous avons démontré qu'un système sensible à ce genre de paramètre était susceptible d'être attaqué car vulnérable. De plus, il est susceptible de posséder des taux de faux positifs à l'identification élevés.

D'un autre côté, nous avons mis l'accent la nécessité d'avoir des systèmes biométriques interopérables qui ne dépendent pas du système matériel auquel ils sont associés. Cette étude montre comment la qualité du module d'acquisition peut influer sur les capacités de reconnaissance des empreintes digitales. Elle souligne également que si un algorithme de reconnaissances d'empreintes digitales n'est pas suffisamment efficace, cela peut constituer un problème de sécurité critique pour la solution biométrique qui l'intègre.

2. **Génération d'instruments d'attaque synthétiques**
   Nous avons présenté une méthode pour générer des instruments d'attaque qui peuvent être utilisés pour l'évaluation de la détection des attaques par présentation dans le cadre des empreintes digitales. La méthode proposée est basée sur un modèle d'apprentissage profond existant (MWGAN) qui est un modèle de transfert de style multi-domaine qui a été introduit pour surmonter les problèmes d'apprentissage que d'autres modèles avaient lorsqu'il s'agissait de plus de deux domaines. Nous ajoutons un terme d'attache aux données à la fonction de coût pour lui donner une signification physique. Les résultats montrent une bonne similarité en ce qui concerne la qualité et les caractéristiques des empreintes digitales synthétiques et des données réelles pour leur capacité à être reconnues; de même, le test d'attaques par présentation donne de bons résultats.

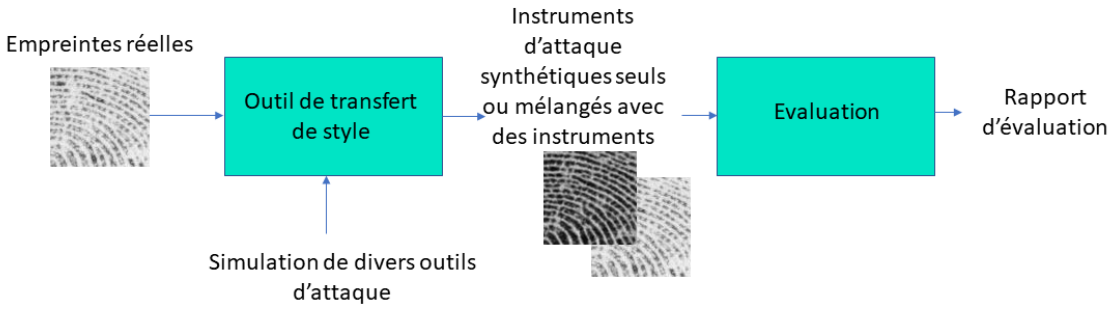


Figure 8.1: Possible utilisation d'instruments d'attaques synthétiques dans le cadre d'une évaluation.

3. **Une méthode générique pour valider les données biométriques synthétiques**

   Nous avons proposé une méthode générique pour valider les données biométriques synthétiques, basée sur des mesures objectives. Pour ce travail, une métrique de qualité objective (NFIQ2.0) est utilisée pour évaluer l'utilisabilité des données générées en tant qu'échantillons biométriques dans un processus de reconnaissance d'empreintes digitales. Nous avons également utilisé la capacité de ces données à être reconnues comme des empreintes digitales ainsi que leur capacité à tromper un algorithme de détection d'attaques par présentation. Nous avons appliqué cette méthodologie à la génération de fausses empreintes digitales pour valider notre approche.

Perspectives

Durant cette thèse, nous avons identifié plusieurs idées à explorer.

1. Pour les conditions d'acquisition, nous avons mis en évidence les effets des conditions environnementales sur les empreintes digitales et l'importance de l'interopérabilité d'un système biométrique pour l'expérience de l'utilisateur, la sécurité et la reproductibilité de l'évaluation de celui-ci. Cependant, pour une évaluation réelle, de tels tests sont très contraignants car le contrôle des conditions environnementales est assez compliqué et nécessite un équipement calibré spécifique. Une solution pourrait consister à utiliser des solutions d'apprentissage profond pour transformer les images acquises dans des conditions nominales et les simuler dans certaines conditions spécifiques.

De cette manière, il serait possible de prédire le comportement du système biométrique dans une condition spécifique considérée. Une autre solution consiste à opter pour une méthodologie générique afin d'adapter le traitement des données biométriques en fonction de l'environnement dans lequel elles sont acquises.

2. La génération d'instruments d'attaque par présentation d'empreintes digitales a été réalisée à l'aide d'un modèle d'apprentissage profond entraîné sur un jeu de données. Cependant, nous pensons que les résultats peuvent être encore meilleurs avec une base de données plus conséquente, et un terme d'attache aux données lié à la texture des images. De plus, pour s'affranchir totalement des contraintes juridiques, une génération à partir rien ou une première étape de génération à partir de rien suivie d'un transfert de style des textures pour imiter les instruments d'attaque par présentation facilitera l'entraînement et l'évaluation des systèmes d'empreintes digitales biométriques et rendra disponibles ces données, sans les restrictions liées à la protection de la vie privée.

3. On a exprimé la nécessité de disposer de systèmes biométriques interopérables. Cette observation n'est pas seulement limitée aux systèmes à empreintes digitales, mais est également applicable aux systèmes basés sur d'autres modalités. Une perspective de ce travail est de pouvoir générer des données synthétiques qui reproduisent les caractéristiques physiques d'un capteur ciblé en utilisant quelques données capturées issues de ce capteur. Nous souhaitons étendre cette approche à d'autres modalités biométriques.

4. Comme indiqué précédemment, il est difficile d'évaluer les effets d'un paramètre particulier sur un système biométrique. On s'interroge sur la possibilité d'utiliser d'une méthode optimisée (à l'aide d'algorithmes génétiques par exemple) pour trouver la combinaison de conditions qui affecterait le plus un système biométrique donné.

Publications

International conferences

- Abdarahmane Wone, Joël Di Manno, Christophe Charrier, Christophe Rosenberger. Impact Of Environmental Conditions On Fingerprint Systems Performance. 18th Annual International Conference on Privacy, Security, and Trust (rank CORE B) 2021, Dec 2021, Auckland (virtual), New Zealand. ⟨hal-03436063v2⟩
- Abdarahmane Wone, Joël di Manno, Christophe Rosenberger, Christophe Charrier. Digitally synthetized fingerprint spoofs: a threat for anti-spoofing systems?. 2022 International Conference on Cyberworlds (rank CORE B), Sep 2022, Kanazawa, Japan. ⟨hal-03748579⟩
- Abdarahmane Wone, Joël Di Manno, Christophe Rosenberger, Christophe Charrier. Capture Biases in Fingerprint Systems. 2023 International Conference on Cyberworlds (rank CORE B), Oct 2023, Sousse (Tunisie), ⟨hal-04176199⟩

Summer School

- Abdarahmane Wone, Joël di Manno, Christophe Rosenberger, Christophe Charrier. Artificial Intelligence in certification of biometric systems, 2nd Inria-DFKI European Summer School on AI (IDESSAI 2022)

Bibliography

[cas, a] ATVS-FFp DB, howpublished = http://biometrics.idealtest.org/findtotaldbbymode.do?mode=fingerprint#/datasetdetail/11, note = Accessed: 2023-10-05. 30

[cas, b] CASIA Fingerprint Subject Ageing, howpublished = http://biometrics.idealtest.org/findtotaldbbymode.do?mode=fingerprint#/datasetdetail/15, note = Accessed: 2023-10-05. 30

[cas, c] CASIA-FingerprintV5, howpublished = http://biometrics.idealtest.org/findtotaldbbymode.do?mode=fingerprint#/datasetdetail/7, note = Accessed: 2023-10-05. 30, 31

[SD4,] NIST Special Database 4, NIST 8-Bit Gray Scale Images of Fingerprint Image Groups (FIGS), webpage = https://www.nist.gov/srd/nist-special-database-4, note = Accessed: 2023-10-05. 31, 35

[Alshehri et al., 2018] Alshehri, H., Hussain, M., Aboalsamh, H. A., and Al Zuair, M. A. (2018). Cross-sensor fingerprint matching method based on orientation, gradient, and gabor-hog descriptors with score level fusion. *IEEE Access*, 6:28951–28968. 51

[Ansari, 2011] Ansari, A. H. (2011). Generation and storage of large synthetic fingerprint database. *ME Thesis, Jul.* 34

[Bausinger and Tabassi, 2011] Bausinger, O. and Tabassi, E. (2011). Fingerprint sample quality metric nfiq 2.0. In *BIOSIG*. 42

[Bertillon, 1893] Bertillon, A. (1893). *Identification anthropométrique. Instructions signalétiques*. 11, 120

[Bouzaglo and Keller, 2022] Bouzaglo, R. and Keller, Y. (2022). Synthesis and reconstruction of fingerprints using generative adversarial networks. *arXiv preprint arXiv:2201.06164*. 40, 76

[Boyd et al., 2020] Boyd, A., Fang, Z., Czajka, A., and Bowyer, K. W. (2020). Iris presentation attack detection: Where are we now? *Pattern Recognition Letters*, 138:483–489. 41

[Cao et al., 2019] Cao, J., Mo, L., Zhang, Y., Jia, K., Shen, C., and Tan, M. (2019). Multi-marginal wasserstein gan. In *Advances in Neural Information Processing Systems*. 77, 78, 79, 122

[Cappelli et al., 2007] Cappelli, R., Ferrara, M., Franco, A., and Maltoni, D. (2007). Fingerprint verification competition 2006. *Biometric Technology Today*, 15(7):7–9. 14

[Cappelli et al., 2010] Cappelli, R., Ferrara, M., Maltoni, D., and Tistarelli, M. (2010). Mcc: A baseline algorithm for fingerprint verification in fvc-ongoing. In *2010 11th International Conference on Control Automation Robotics & Vision*, pages 19–23. IEEE. 43, 44, 121

[Chen et al., 2022] Chen, Y., Gao, Q., and Wang, X. (2022). Inferential wasserstein generative adversarial networks. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(1):83–113. 38

[Choi et al., 2020a] Choi, Y., Uh, Y., Yoo, J., and Ha, J.-W. (2020a). Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8188–8197. 77

[Choi et al., 2020b] Choi, Y., Uh, Y., Yoo, J., and Ha, J.-W. (2020b). Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8188–8197. 78

[Chollet, 2021] Chollet, F. (2021). *Deep learning with Python*. Simon and Schuster. 36

[Chugh et al., 2018] Chugh, T., Cao, K., and Jain, A. (2018). Fingerprint spoof buster: Use of minutiae-centered patches. *IEEE Transactions on Information Forensics and Security*, PP:1–1. 31, 39, 121

[Chugh and Jain, 2019] Chugh, T. and Jain, A. K. (2019). Fingerprint spoof generalization. *arXiv preprint arXiv:1912.02710*. 40, 90

[Colbois et al., 2021] Colbois, L., de Freitas Pereira, T., and Marcel, S. (2021). On the use of automatically generated synthetic image datasets for benchmarking face recognition. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE. 76

[El-Abed et al., 2012] El-Abed, M., Charrier, C., and Rosenberger, C. (2012). Evaluation of biometric systems. 14

[Engelsma et al., 2022] Engelsma, J. J., Grosz, S. A., and Jain, A. K. (2022). Printsgan: synthetic fingerprint generator. *arXiv preprint arXiv:2201.03674*. 40

[Erdogmus and Marcel, 2013] Erdogmus, N. and Marcel, S. (2013). Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect. In *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6. 19, 20, 120

[Fernandez-Saavedra et al., 2008] Fernandez-Saavedra, B., Sanchez-Reillo, R., Alonso-Moreno, R., and Mueller, R. (2008). Evaluation methodology for analyzing environment influence in biometrics. In *2008 10th International Conference on Control, Automation, Robotics and Vision*, pages 1342–1346. 51

[Fernandez-Saavedra et al., 2010] Fernandez-Saavedra, B., Sanchez-Reillo, R., Moreno, R., and Miguel-Hurtado, O. (2010). Environmental testing methodology in biometrics. 51

[Fiumara et al., 2019] Fiumara, G. P., Flanagan, P. A., Grantham, J. D., Ko, K., Marshall, K., Schwarz, M., Tabassi, E., Woodgate, B., and Boehnen, C. (2019). Nist special database 302: Nail to nail fingerprint challenge. 31

[Gajawada et al., 2019] Gajawada, R., Popli, A., Chugh, T., Namboodiri, A., and Jain, A. K. (2019). Universal material translator: Towards spoof fingerprint generalization. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE. 40, 90

[Ghiani et al., 2013] Ghiani, L., Yambay, D., Mura, V., Tocco, S., Marcialis, G. L., Roli, F., and Schuckcrs, S. (2013). Livdet 2013 fingerprint liveness detection competition 2013. In *2013 International Conference on Biometrics (ICB)*, pages 1–6. 31, 81, 82

[Ghiani et al., 2017] Ghiani, L., Yambay, D. A., Mura, V., Marcialis, G. L., Roli, F., and Schuckers, S. A. (2017). Review of the fingerprint liveness detection (livdet)

competition series: 2009 to 2015. *Image and Vision Computing*, 58:110–128. 14, 31, 32, 120

[Godbole et al., 2022] Godbole, A., Grosz, S. A., Nandakumar, K., and Jain, A. K. (2022). On demographic bias in fingerprint recognition. In *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. 49

[Goodfellow et al., 2014] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc. 36, 78

[Grosz et al., 2020a] Grosz, S., Engelsma, J., and Jain, A. (2020a). White-box evaluation of fingerprint recognition systems. 50

[Grosz et al., 2020b] Grosz, S. A., Chugh, T., and Jain, A. K. (2020b). Fingerprint presentation attack detection: A sensor and material agnostic approach. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. 31, 40, 90

[Grosz and Jain, 2022] Grosz, S. A. and Jain, A. K. (2022). Spoofgan: Synthetic fingerprint spoof images. *arXiv preprint arXiv:2204.06498*. 31, 40, 90

[Harvey et al., 2019] Harvey, J., Campbell, J., and Adler, A. (2019). Characterization of biometric template aging in a multiyear, multivendor longitudinal fingerprint matching study. *IEEE Transactions on Instrumentation and Measurement*, 68(4):1071–1079. 50

[He et al., 2019] He, Z., Zuo, W., Kan, M., Shan, S., and Chen, X. (2019). Attgan: Facial attribute editing by only changing what you want. *IEEE Transactions on Image Processing*, 28(11):5464–5478. 77

[Heng et al., 2018] Heng, G. S., Ismail, N. A., Rahman, Z. A. A., and Anan, A. (2018). Distribution of fingerprint patterns among young adults and siblings in malaysia. *Int. J. Med. Sci*, 3(1):11–7. 49

[Husseis et al., 2019] Husseis, A., Liu-Jimenez, J., Goicoechea-Telleria, I., and Sanchez-Reillo, R. (2019). A survey in presentation attack and presentation attack detection. In *2019 International Carnahan Conference on Security Technology (ICCST)*, pages 1–13. IEEE. 41

[ISO 19795-1:2021(E), 2021] ISO 19795-1:2021(E) (2021). ISO/IEC 19795-1: 2021 Information technology — Biometric performance testing and reporting — Part 1: Principles and framework. Standard, International Organization for Standardization. 9, 15, 16, 24, 49, 50, 97, 120, 122

[ISO 19795-2:2007, 2007] ISO 19795-2:2007 (2007). ISO/IEC 19795-2:2007 Information technology — Biometric performance testing and reporting — Part 2: Testing methodologies for technology and scenario evaluation. Standard, International Organization for Standardization. 51

[ISO 19795-6:2012, 2012] ISO 19795-6:2012 (2012). ISO/IEC 19795-6:2012 Information technology — Biometric performance testing and reporting — Part 6: Testing methodologies for operational evaluation. Standard, International Organization for Standardization. 51

[ISO 2382-37:2022, 2022] ISO 2382-37:2022 (2022). ISO/IEC 2382-37:2022 Information technology — Vocabulary — Part 37: Biometrics. Standard, International Organization for Standardization. 9

[ISO 29794-4:2017, 2017] ISO 29794-4:2017 (2017). ISO/IEC 29794-4:2017 Information technology — Biometric sample quality — Part 4: Finger image data. Standard. 54

[ISO 30107-1:2023, 2023] ISO 30107-1:2023 (2023). ISO/IEC 30107-1: 2023 Information technology — Biometric presentation attack detection — Part 1: Framework. Standard, International Organization for Standardization. 17, 74, 120

[ISO 30107-3:2023, 2017] ISO 30107-3:2023 (2017). ISO/IEC 30107-3: 2023 Information technology — Biometric presentation attack detection — Part 3: Testing and reporting. Standard, International Organization for Standardization. 17, 18

[Isola et al., 2017] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 38, 76, 77, 122

[ITU-R BT.500-15,] ITU-R BT.500-15. Methodologies for the subjective assessment of the quality of television images. Standard Recommendation ITU-R BT.500-15, year=2023, month=may. 55

[Jain et al., 1999] Jain, A., Bolle, R., and Pankanti, S. (1999). *Biometrics: personal identification in networked society*, volume 479. Springer Science & Business Media. 11, 96, 123, 124

[Jain, 2005] Jain, A. K. (2005). Biometric recognition: how do i know who you are? In *Image Analysis and Processing–ICIAP 2005: 13th International Conference, Cagliari, Italy, September 6-8, 2005. Proceedings 13*, pages 19–26. Springer. 10, 96

[Jain and Kumar, 2010] Jain, A. K. and Kumar, A. (2010). Biometrics of next generation: An overview. *Second generation biometrics*, 12(1):2–3. 51

[Jain and Kumar, 2012] Jain, A. K. and Kumar, A. (2012). Biometric recognition: an overview. *Second generation biometrics: The ethical, legal and social context*, pages 49–79. 8

[Johnson et al., 2013] Johnson, P., Hua, F., and Schuckers, S. (2013). Texture modeling for synthetic fingerprint generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 154–159. 35

[Joshi et al., 2022] Joshi, I., Grimmer, M., Rathgeb, C., Busch, C., Bremond, F., and Dantcheva, A. (2022). Synthetic data in human analysis: A survey. *arXiv preprint arXiv:2208.09191*. 40

[Karampidis et al., 2021] Karampidis, K., Rousouliotis, M., Linardos, E., and Kavallieratou, E. (2021). A comprehensive survey of fingerprint presentation attack detection. *Journal of Surveillance, Security and Safety*, 2(4):117–161. 74

[Karras et al., 2017] Karras, T., Aila, T., Laine, S., and Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*. 38

[Karras et al., 2021] Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., and Aila, T. (2021). Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems*, 34:852–863. 76

[Karras et al., 2019] Karras, T., Laine, S., and Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410. 38, 76

[Karras et al., 2020] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119. 38, 40, 76

[Kim et al., 2017] Kim, T., Cha, M., Kim, H., Lee, J. K., and Kim, J. (2017). Learning to discover cross-domain relations with generative adversarial networks. In *International conference on machine learning*, pages 1857–1865. PMLR. 76

[Kirchgasser et al., 2020] Kirchgasser, S., Kauba, C., and Uhl, A. (2020). *Towards Understanding Acquisition Conditions Influencing Finger Vein Recognition*. Springer, Cham. 50

[Ko Kenneth, 2007] Ko Kenneth, W. J. S. (2007). User's guide to nist biometric image software (nbis). Technical report. 38, 43, 67

[Korshunov and Marcel, 2018] Korshunov, P. and Marcel, S. (2018). Deepfakes: a new threat to face recognition? assessment and detection. *arXiv preprint arXiv:1812.08685*. 19

[Krishnasamy et al., 2011] Krishnasamy, P., Belongie, S., and Kriegman, D. (2011). In *International Joint Conference on Biometrics (IJCB)*, Washington, DC. 50

[Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25. 87

[Kukula et al., 2009] Kukula, E. P., Blomeke, C. R., Modi, S. K., and Elliott, S. J. (2009). Effect of human-biometric sensor interaction on fingerprint matching performance, image quality and minutiae count. *International Journal of Computer Applications in Technology*, 34(4):270–277. 50

[Labati et al., 2016] Labati, R. D., Genovese, A., Muñoz, E., Piuri, V., Scotti, F., and Sforza, G. (2016). Biometric recognition in automated border control: a survey. *ACM Computing Surveys (CSUR)*, 49(2):1–39. 8

[Lanitis, 2010] Lanitis, A. (2010). A survey of the effects of aging on biometric identity verification. *International Journal of Biometrics*, 2(1):34–52. 50

[Lee et al., 2020] Lee, H.-Y., Tseng, H.-Y., Mao, Q., Huang, J.-B., Lu, Y.-D., Singh, M., and Yang, M.-H. (2020). Drit++: Diverse image-to-image translation via disentangled representations. *International Journal of Computer Vision*, 128:2402–2417. 78

[Lin et al., 2020] Lin, J., Pang, Y., Xia, Y., Chen, Z., and Luo, J. (2020). Tuigan: Learning versatile image-to-image translation with two unpaired images. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*, pages 18–35. Springer. 77

[Liu et al., 2018] Liu, A. H., Liu, Y.-C., Yeh, Y.-Y., and Wang, Y.-C. F. (2018). A unified feature disentangler for multi-domain image translation and manipulation. *Advances in neural information processing systems*, 31. 77

[Liu et al., 2017] Liu, M.-Y., Breuel, T., and Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, 30. 76

[Liu et al., 2020] Liu, Y., De Nadai, M., Yao, J., Sebe, N., Lepri, B., and Alameda-Pineda, X. (2020). Gmm-unit: Unsupervised multi-domain and multi-modal image-to-image translation via attribute gaussian mixture modeling. *arXiv preprint arXiv:2003.06788*. 78

[Marasco, 2019] Marasco, E. (2019). Biases in fingerprint recognition systems: Where are we at? In *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–5. 51

[Marasco and Ross, 2014] Marasco, E. and Ross, A. (2014). A survey on antispoofing schemes for fingerprint recognition systems. *ACM Computing Surveys (CSUR)*, 47(2):1–36. 18, 19, 120

[Marcel et al., 2023] Marcel, S., Fierrez, J., and Evans, N. (2023). *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*. Springer Nature. 16

[Minaee and Abdolrashidi, 2018] Minaee, S. and Abdolrashidi, A. (2018). Finger-gan: Generating realistic fingerprint images using connectivity imposed gan. 37, 38, 121

[Mistry et al., 2020] Mistry, V., Engelsma, J. J., and Jain, A. K. (2020). Fingerprint synthesis: Search with 100 million prints. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE. 38

[Mura et al., 2018] Mura, V., Orrù, G., Casula, R., Sibiriu, A., Loi, G., Tuveri, P., Ghiani, L., and Marcialis, G. L. (2018). Livdet 2017 fingerprint liveness detection competition 2017. In *2018 international conference on biometrics (ICB)*, pages 297–302. IEEE. 33, 123

[Olsen et al., 2016] Olsen, M. A., Šmida, V., and Busch, C. (2016). Finger image quality assessment features–definitions and evaluation. volume 5, pages 47–64. IET. 53

[Ortega-Garcia et al., 2003] Ortega-Garcia, J., Fierrez-Aguilar, J., Simon, D., Gonzalez, J., Faundez-Zanuy, M., Espinosa, V., Satue, A., Hernaez, I., Igarza, J.-J., Vivaracho, C., et al. (2003). Mcyt baseline corpus: a bimodal biometric database. *IEE Proceedings-Vision, Image and Signal Processing*, 150(6):395–401. 14

[Pang et al., 2021] Pang, Y., Lin, J., Qin, T., and Chen, Z. (2021). Image-to-image translation: Methods and applications. *IEEE Transactions on Multimedia*, 24:3859–3881. 78

[Pentland et al., 1994] Pentland, A., Moghaddam, B., Starner, T., et al. (1994). View-based and modular eigenspaces for face recognition. 50

[Priesnitz et al., 2022] Priesnitz, J., Rathgeb, C., Buchmann, N., and Busch, C. (2022). Syncolfinger: Synthetic contactless fingerprint generator. *Pattern Recognition Letters*, 157:127–134. 40

[Raffaele et al., 2004] Raffaele, C., Dario, M., Davide, M., et al. (2004). Sfinge (synthetic fingerprint generator). 29, 33, 87

[Ratha et al., 2001] Ratha, N. K., Connell, J. H., and Bolle, R. M. (2001). Enhancing security and privacy in biometrics-based authentication systems. *IBM systems Journal*, 40(3):614–634. 17, 120

[Rathgeb et al., 2022] Rathgeb, C., Tolosana, R., Vera-Rodriguez, R., and Busch, C. (2022). *Handbook of digital face manipulation and detection: From deepfakes to morphing attacks*. Springer Nature. 19

[Schuckers et al., 2023] Schuckers, S., Cannon, G., and Tekampe, N. (2023). Fido biometrics requirements. Technical report. 16, 18

[Seidlitz et al., 2021] Seidlitz, S., Jürgens, K., Makrushin, A., Kraetzer, C., and Dittmann, J. (2021). Generation of privacy-friendly datasets of latent fingerprint images using generative adversarial networks. In *VISIGRAPP (4: VISAPP)*, pages 345–352. 38, 76

[Sherlock and Monro, 1993] Sherlock, B. G. and Monro, D. M. (1993). A model for interpreting fingerprint topology. *Pattern recognition*, 26(7):1047–1055. 34

[Tan et al., 2010] Tan, B., Lewicke, A., Yambay, D., and Schuckers, S. (2010). The effect of environmental conditions and novel spoofing methods on fingerprint anti-spoofing algorithms. In *2010 IEEE International Workshop on Information Forensics and Security*, pages 1–6. 50, 51, 121

[Van der Maaten and Hinton, 2008] Van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(11). 88

[Vasquez and Lewis, 2019] Vasquez, S. and Lewis, M. (2019). Melnet: A generative model for audio in the frequency domain. *arXiv preprint arXiv:1906.01083*. 76

[Walker, 2012] Walker, D. R. (2012). Biometric technology in law enforcement. *Neurosurgery*, 71(2):197–200. 8

[Wone et al., 2021] Wone, A., Di Manno, J., Charrier, C., and Rosenberger, C. (2021). Impact of environmental conditions on fingerprint systems performance. In *2021 18th International Conference on Privacy, Security and Trust (PST)*, pages 1–5. IEEE. 48, 89

[Wyzykowski et al., 2021] Wyzykowski, A. B. V., Segundo, M. P., and de Paula Lemes, R. (2021). Level three synthetic fingerprint generation. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 9250–9257. IEEE. 38

[Yin et al., 2020] Yin, X., Li, Y., and Shin, B.-S. (2020). Dagan: A domain-aware method for image-to-image translations. *Complexity*, 2020:1–15. 77

[Zhao et al., 2012] Zhao, Q., Jain, A. K., Paulter, N. G., and Taylor, M. (2012). Fingerprint image synthesis based on statistical feature models. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 23–30. IEEE. 35, 36, 121

[Zhao et al., 2010] Zhao, Q., Zhang, D., Zhang, L., and Luo, N. (2010). High resolution partial fingerprint alignment using pore–valley descriptors. *Pattern Recognition*, 43(3):1050–1061. 37

[Zhu et al., 2017] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*. 76

[Zhu et al., 2020] Zhu, P., Abdal, R., Qin, Y., and Wonka, P. (2020). Sean: Image synthesis with semantic region-adaptive normalization. In *Proceedings of the*

IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5104–5113. 77

List of Figures

List of Tables

List of Algorithms

Contribution to the certification of fingerprint systems: towards the reproducibility of the evaluation

Computer science is more and more present in our daily lives for multiple tasks. Considering its multiple services and applications, its security is an essential guarantee of the trust that a final user can have. Biometrics is a solution that is more and more used not only to secure some solutions but also to simplify our lives.
We contribute to the understanding of the interoperability and reproducibility problems. We explore the factors impacting the behavior of a fingerprint system in its recognition capacities. Moreover, the testing of resistance to attack has legal and operational constraints that make the testing difficult. We explored the generation of synthetic attack instruments using a deep-leaning solution for the evaluation of biometric fingerprint solutions. We validated our studies with an objective method that we propose in this dissertation.

Keywords : Biometrics, Fingerprint Systems, Certification, Evaluation, Biases, Synthetic Biometric Data, Deep learning.

Contribution à la reproductibilité de l'évaluation des systèmes à empreinte digitale

Les systèmes informatiques sont de plus en plus utilisés au quotidien pour diverses tâches. Au regard des services et des applications concernés, leur sécurité est une garantie essentielle pour leur fonctionnement, ainsi qu'un gage de confiance pour l'utilisateur final. La biométrie est une solution pour cela. Cependant, la sécurité des systèmes biométriques doit être assurée. Un bon système biométrique doit à la fois pouvoir reconnaître la bonne personne et résister aux attaques. Cette thèse s'inscrit dans le cadre de la certification des systèmes biométriques qui est une étape qui qualifie la conformité d'un système biométrique à un plan de test proposé par une autorité. Nous avons contribué sur la compréhension de l'interopérabilité et le problème de la répétabilité de ces tests. Pour cela, nous explorons les impacts qu'ont les conditions environnementales d'acquisition tout comme la technologie d'acquisition pour les systèmes basés sur les empreintes digitales. D'un autre coté, nous avons exploré la création, basée sur les réseaux de neurones profonds d'instruments d'attaques utilisables dans le cadre d'une évaluation. Nous avons validé ces études par une méthodologie objective qui a un sens physique.

Mots-clés : Biométrie, empreintes digitales, systèmes biométriques, certification, évaluation de systèmes biométriques, apprentissage profond, biais biométriques.

Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France