UNIVERSITÀ DEGLI STUDI ROMA TRE
ENGINEERING DEPARTMENT
DOCTORAL PROGRAM IN APPLIED ELECTRONICS

UNIVERSITÀ DEGLI STUDI ROMA TRE
DIPARTIMENTO DI INGEGNERIA
DOTTORATO DI RICERCA IN ELETTRONICA APPLICATA

INVISIBLE BIOMETRICS

Doctoral Dissertation of:
Emanuela Piciucco

Supervisor:
Prof. Patrizio Campisi

XXXII Cycle

"Everything we see hides another thing, we always want to see what is hidden by what we see. There is an interest in that which is hidden and which the visible does not show us. This interest can take the form of a quite intense feeling, a sort of conflict, one might say, between the visible that is hidden and the visible that is present."

René Magritte

Abstract

We live in a world where a wide variety of systems require automatic, efficient and reliable personal recognition schemes to either confirm or determine the identity of an individual requesting their services. Financial, military, national security services and industry IT experts – all want to make sure that the right person has access to the right account, sensitive data or technological processes and they are looking for the best access control methods. Token-based and knowledge-based automatic personal identification approaches have been the two traditional techniques widely used. Token-based approaches use something you have to make a personal identification, such as a passport, driver's license, ID card, credit card, or keys. Knowledge-based approaches use something you know to make a personal identification, such as a password or a personal identification number (PIN).

The traditional approaches suffer from some disadvantages. Passwords, the ubiquitous user login method, is widely known to be the weakest link in cybersecurity today: every year millions of accounts are stolen and millions or domains are violated. Tokens may be lost, stolen, forgotten, or misplaced, and a PIN may be forgotten by a valid user or guessed by an impostor, entailing the system to be compromised. Besides, knowledge- and token-based approaches suffer of the further disadvantage to be unable to differentiate between an authorised person and an impostor who fraudulently acquires the token or knowledge of the authorised person, resulting in being unsatisfactory means of achieving the security requirements of our electronically interconnected information society. The vulnerabilities of the

aforementioned solutions generated efforts to search for more secure user recognition techniques.

In this scenario, improvements in technology have led to the emergence of new increasingly advanced security and authentication systems. Between them, we can find biometric systems, or simply biometrics, that is automatic recognition of individuals based on their physiological and/or behavioural characteristics. Because many physiological or behavioural characteristics are distinctive to each person, biometric identifiers are inherently more reliable and more capable than knowledge-based and token-based techniques in differentiating between an authorised person and a fraudulent impostor. Besides, biometrics offers greater security and convenience compared to traditional identification methods and in some cases can be integrated or totally replace existing technology, opening new business opportunities in different markets.

All systems have their shortcomings and biometric technology is not an exception. Although biometrics holds enormous promises and has in fact been used for decades by government agencies across the world to safeguard data, enhance border security and identify those who may have hostile intents, there are some risks a biometric system can incur into. Some kind of biometric data are inherently public and can be stolen. If a person has ever entered a place where there are public cameras, chances are that their face is on record somewhere; when they make a phone their voice may be recorded; and fingerprints are left everywhere a person touches. So also biometric identifiers are not immune to the problem of stolen data and once a hacker has a person's biometric identifier, it is easy for him to gain access to an account that requires biometric recognition. Besides, since many biometric characteristics does not change over time, if the aforesaid biometric data is stolen or compromised, is unfortunately compromised forever. Eventually, since every human being has a limited number of biometric traits, when the biometric data are compromised, the "reset" of a biometric identifier is not as straightforward as the case of user ID and password and revoking and replacing a biometric characteristic is not as easy as in the token case. In the event of biometric record leak or theft, users will have permanent and most private personal data in the hands of bad actors. The user has permanently lost control of that form of identification.

The aforementioned vulnerabilities and disadvantages of the biometric technology can be faced and reduced simply choosing the right biometric trait. Some biometric identifiers are inherently more difficult to steal and replicate. These biometrics traits, that in this thesis we will refer at with the name *invisible biometrics*, share the properties to be hidden in the human

body and not easy to be captured at a distance with traditional acquisition devices and methodologies. The aforementioned properties entail the difficulty in performing a spoofing attack, that consists in defeating the biometric system's security by stealing, copying and replicating synthetically a biometric trait to gain unauthorised access, on biometric systems based on invisible biometrics. Eventually, an invisible biometric trait inherently provide liveness detection.

In this thesis, two invisible biometric identifiers, namely *vein pattern* and *electroencephalogram*, are presented. Despite their differences in terms of physiological origins, these biometric traits share the advantage of being hidden inside the human body, thus resulting in being difficult to be stolen and replicated. In fact, both the network of blood vessels and brain waves cannot be acquired at a distance and with conventional devices, making a spoofing attack almost impossible to be implemented. On the other hand, the acquisition procedure is generally affected by the environment conditions and the "hiddeness" of the physiological signals makes the capturing step very sensitive to noise. In this thesis, some possible countermeasures to the aforementioned limitations are proposed. Specifically, biometric fusion approaches and deep learning methods are exploited in order to compensate the errors of the system linked to the noise in the acquired data and, consequently, to improve the performance of invisible-biometrics-based biometric system. Besides, biometric systems exploiting invisible identifiers don't overcome the issue related to the security of the system and the problem linked to the revocation of compromised biometrics. In this regard, a template protection approach based on cancelable biometrics is proposed as a solution to deal with security issues in vein-based biometric systems.

Eventually, being biometrics behavioural or physiological characteristics, the process of biometric recognition leaves behind trails of private information. The invisible biometric identifiers considered in this thesis are not immune from the aforesaid shortcoming. Some information are *visible* although the biometric identifier belongs to the class of invisible biometrics, disclosing some aspects of the privacy of system's users and threatening it. Specifically, health and diseases can be revealed by both the hand vascular systems and brain waves, while mental and emotional states as well as the processing cost associated to a specific task can be disclosed by electroencephalographic signals. In this thesis, an example of visible information able to be revealed by brain signals recorded as responses to sentences presented as auditory stimuli is studied. More in detail, given an electroencephalographic signal, the brain cost associated to the processing of the discourse is computed and it is demonstrated that, given specific

features related to the brain cost, it is possible to determine whether the presented sentence contains linguistic misalignment or not.

Sommario

VIVIAMO IN UN MONDO dove una quantità sempre maggiore di sistemi richiedono schemi di riconoscimento automatico, efficienti ed affidabili, il cui scopo è quello di confermare o determinare l'identità dell'individuo richiedente specifici servizi. Servizi finanziari, nazionali, di sicurezza e industrie, sono esempi di applicazioni che condividono la stessa necessità: assicurarsi che solo le persone autorizzate abbiano accesso all'account, dati sensibili o processi tecnologici e che il metodo di controllo degli accessi sia il più sicuro possibile. A tal fine, gli approcci generalmente più utilizzati sono quelli basati su "ciò che si possiede" (*token-based*), come un documento di riconoscimento o una carta, o su "ciò che si conosce" (*knowledge-based*), per esempio PIN o password.

I suddetti approcci soffrono di alcune limitazioni. Le password, il metodo onnipresente quando viene richiesto un login, è noto per essere ad oggi il punto più debole nell'ambito della cybersecurity: ogni anno si assiste ad oltre un milione di account rubati o domini violati. I token possono essere persi, rubati o dimenticati e i PIN possono essere scordati dell'utente legittimo o indovinati da un impostore, compromettendo la sicurezza del sistema. Inoltre, gli approcci *token-based* e *knowledge-based* soffrono dell'ulteriore svantaggio di non essere in grado di distinguere una persona autorizzata da un utente che viene in possesso in maniera fraudolenta della password o del token dell'utente legittimo, risultando pertanto mezzi insoddisfacenti per ottenere i requisiti di sicurezza adeguati. Le vulnerabilità delle suddette soluzioni hanno comportato grandi investimenti nella ricerca di metodi di autenticazione più sicuri.

All'interno di questo scenario, il miglioramento della tecnologia ha implicato la nascita di misure di sicurezza e sistemi di autenticazione sempre più avanzati. Tra essi si distinguono i sistemi biometrici, ovvero sistemi di riconoscimento automatico basato sulle caratteristiche fisiologiche e/o comportamentali dell'utente. Dal momento che molte caratteristiche biometriche sono distintive per ciascuna persona, gli identificatori biometrici sono intrinsecamente più affidabili ed appropriati delle tecniche *token-based* e *knowledge-based* nel compito di distinguere un utente autorizzato da uno non autorizzato. Inoltre, il riconoscimento biometrico offre una maggiore sicurezza e praticità rispetto ai metodi di riconoscimento tradizionali, e, in alcuni casi, può essere integrato o sostituito completamente alle tecnologie esistenti.

Tutti i sistemi hanno i propri limiti e di certo i sistemi biometrici non fanno eccezione. Nonostante il riconoscimento biometrico sia un campo molto promettente ed è, difatti, stato utilizzato per decadi dalle agenzie governative al fine di salvaguardare i dati, migliorare i controlli al confine e identificare i soggetti con possibili intenzioni ostili, vi sono alcuni rischi in cui è possibile imbattersi se si adotta come soluzione una tecnologia basata sulla biometria. Alcuni dati biometrici sono per loro natura pubblici e possono essere rubati. Si pensi ad esempio ad una persona che entra in un luogo pubblico dove vi sono delle videocamere di sorveglianza, le probabilità che il proprio volto sia stato acquisito e memorizzato sono molto alte. Nel momento in cui avviene una chiamata, la voce degli interlocutori potrebbe essere registrata. Tracce di impronte digitali di una persona vengono lasciate su qualsiasi superficie essa entri in contatto ed il processo di acquisizione e duplicazione dell'impronta a partire dalle sue tracce non è di difficile implementazione. Inoltre, dal momento che gli identificatori biometrici generalmente non cambiano nel tempo, se un tratto biometrico viene rubato o compromesso, risulta difficile revocarlo e sostituirlo, operazioni d'altro canto estremamente semplici nel caso di password, PIN o token.

Le suddette vulnerabilità e svantaggi delle tecnologie biometriche possono essere fronteggiate e ridotte semplicemente scegliendo opportunamente il tratto biometrico su cui basare l'identificazione degli utenti. In maggior dettaglio, alcuni tratti biometrici sono per loro natura più difficili da rubare e replicare. Questi identificatori biometrici, che nel seguito della tesi verranno indicati come *tratti biometrici invisibili*, condividono le proprietà di essere nascosti all'interno del corpo umano e non facili da acquisire a distanza con metodi e dispositivi di acquisizione tradizionali. Le suddette proprietà implicano la difficoltà di realizzazione di un attacco di

tipo *spoofing*, che consiste nel rubare, copiare, replicare sinteticamente e ripresentare al sistema un tratto biometrico falsificato per contrastare le misure di sicurezza del sistema. Infine, un tratto biometrico invisibile fornisce intrinsecamente informazioni riguardo la vivezza dell'utente.

In questa tesi verranno presentati due identificatori che fanno parte della categoria dei tratti biometrici invisibili, ovvero le *strutture venose della mano* e l'*elettroencefalogramma*. Nonostante essi risultino essere molto diversi dal punto di vista fisiologico, i suddetti tratti biometrici condividono la caratteristica vantaggiosa di essere difficili da rubare e replicare. Infatti, sia la rete di vasi sanguigni che le onde celebrali non possono essere acquisiti a distanza e la procedura di acquisizione non può essere effettuata con dispositivi tradizionali, rendendo un attacco di tipo *spoofing* quasi impossibile da implementare. D'altra parte anche questi identificatori presentano degli svantaggi, principalmente legati alla procedura di acquisizione stessa: essendo per loro natura "nascosti", i segnati acquisiti sono generalmente sensibili al rumore. In questa tesi verranno presentate alcune soluzioni atte a migliorare le performance di riconoscimento di un sistema basato sui tratti biometrici invisibili, andando a compensare gli errori legati al rumore presente nei dati acquisiti tramite tecniche di fusione biometrica e *deep learning*. Inoltre, come tutti i sistemi biometrici, anche i sistemi di riconoscimento basati su tratti biometrici invisibili soffrono di alcuni limiti legati alla sicurezza del sistema e alla revoca di tratti biometrici compromessi. Nell'ambito di questa tesi viene presentata una soluzione atta a migliorare la sicurezza e superare le vulnerabilità di sistemi basati sulle strutture venose della mano. Nello specifico, all'interno della tesi, vengono studiate tecniche di protezione del *template* basate su trasformazioni irreversibili e viene analizzato l'impatto dei suddetti algoritmi sulle performance e la sicurezza del sistema.

Infine, essendo i tratti biometrici caratteristiche fisiologiche e/o comportamentali di un soggetto, il processo di riconoscimento biometrico rivela informazioni riguardanti la privacy degli utenti. I tratti biometrici analizzati in questa tesi non sono immuni dal suddetto problema. Alcune informazioni sono *visibili* a partire dagli identificatori biometrici invisibili, rivelando alcune informazioni personali degli utenti e minacciando la loro privacy. Nello specifico, informazioni riguardo la salute e malattie di un soggetto possono essere propalate sia dal sistema venoso della mano che dalle onde celebrali, mentre stati mentali ed emotivi così come il costo di elaborazione dell'informazione possono essere estratte processando segnali elettroencefalografici. In questa tesi, viene presentato un esempio di informazione visibile ottenibile a partire dalle onde celebrali. In particolare, viene stu-

diato il costo di elaborazione da parte del cervello di frasi, presentate sotto forma di stimolo uditivo, e viene dimostrato che è possibile comprendere se lo stimolo presenta discordanze linguistiche, sovraccaricando l'utilizzo della *working memory* durante l'elaborazione del discorso stesso.

Acknowledgements

I wish to thank my PhD advisor, Prof Patrizio Campisi, who always trusted in my capabilities, even when I did not trust them anymore. I would also like to thank my co-advisor and friend, Prof. Emanuele Maiorana, always there for precious suggestions.

I would sincerely thank all the people that share special moments with me everyday at the university, between them Gabriel, colleague but above all important friend, always there to support me and be judged by me, and Federico, without him every break would not be so hilarious.

I would like to sincerely thank my parents, for their irreplaceable love and support. I would like to thank my sister, Valeria, for all the strength she is able to give me everyday and for all the constructive talks. I would like to thank my uncle Piero, always there to give me useful suggestions and have constructive chats about always new and interesting topics, and Diana, a person unique in her kindness.

A special thanks goes to my beloved Simone, able to see in me the light even in the darkest moments and to make every day of my ordinary life extraordinary, just with sights, smiles, kindness and love.

I would like to sincerely thank my best friends, to be always there for me, in the happiest moments but above all in the darkest ones: Liliana, we will never stop to be dreamers, Elisa, lifelong friend, Giulia e Elena, irreplaceable source of friendship, strength and support. Without all my friends I would not have been able to learn how to get up after falling down, and getting up again, as many times as necessary.

Last but not least, I would like to dedicate my successes but above all the

strength in dealing with failures to my grandparents, my angels, examples of life and wisdom.

Contents

List of Figures

List of Tables

CHAPTER 1

An Introduction to Biometric Recognition

EVERY DAY organisations in financial services, health care, e-commerce, telecommunication, and government require reliable personal recognition schemes to either confirm or determine the identity of an individual requesting their services; millions of time every day applications like secure access to buildings, computer systems, laptops, cellular phones, and ATMs need to know if a person who wants to access their system is who he or she claims to be and if the applicant should be given the access to the system itself. It's then necessary for these systems to be provided with a scheme which purpose is to ensure that the rendered services are accessed only by a legitimate user and no one else [124]. In fact, the lack of a robust personal recognition schemes makes these systems vulnerable to the attacks of an impostor entailing frauds in welfare disbursements, credit card transactions, cellular phone calls, and ATM withdrawals that total billion of dollars every year [122].

More and more organisations are looking to highly accurate automated personal identification systems for the purpose of increasing the level of security of their applications and saving critical resources as well as improving customers' satisfaction and systems' operating efficiency. Personal

identification refers to the process of associating a particular individual with an identity. The traditional and widely used automatic personal identification approaches are knowledge-based and token-based techniques. Token-based approaches use something that a person has to perform a personal identification, such as a passport, ID card, or keys. Knowledge-based approaches use something people know to carry out a personal identification, such as a password or a personal identification number (PIN). The token-based approaches suffer from the disadvantages that tokens may be lost, stolen, forgotten, or misplaced while the disadvantages of knowledge based approaches are that PIN may be forgotten by a valid user or guessed by an impostor. Because knowledge-based and token-based approaches are unable to differentiate between an authorised person and an impostor who fraudulently acquires the token or knowledge of the authorised person, they cannot be used as proper means for achieving the satisfactory security requirements of most of the systems in our society [201].

Biometric recognition, or simply *biometrics*, refers to the process of automatic recognition of individuals based on their physiological and/or behavioural characteristics, termed as biometric identifier [124]. Strictly speaking, a biometric characteristic is the measurable trait that can be used for automated recognition. Because many physiological or behavioural characteristics are distinctive to each person, compared to knowledge-based and token-based techniques, biometric identifiers are inherently more reliable and more able to differentiate between an authorised person and a fraudulent impostor.

Not every biological measurement qualify to be a biometric, but a biological and/or or behavioural characteristic can be used as a biometric characteristic as long as it satisfies the following requirements [124]:

- **universality:** every individual should possess the trait;
- **uniqueness:** given a biometrics trait, it should be sufficiently different across individuals comprising the user population;
- **permanence:** the biometric trait should be sufficiently invariant, with respect to the matching criterion, over time;
- **measurability:** it should be possible to acquire and quantitatively measure the biometric characteristic by a sensing device that do not cause undue inconvenience to the individual.

However, in a practical biometric system applications, in addition to the previous requirements that a biometric characteristic need to satisfy,

there are a many other issues that a designer of the system must consider, including:

- **performance:** the level of accuracy and speed of recognition that the system is able to achieve and the the resources required to reach the desired recognition accuracy and speed, as well as the operational and environmental factors that affect the accuracy and speed;
- **acceptability:** the user population of the system and people in general should have no problem to accept the use of a particular biometric identifier in their daily lives and they should have no strong objections to the measurement and collection of the biometric characteristic.
- **circumvention:** the ease with which the trait of an individual can be imitated using artefacts, in the case of physical trait, and mimicry, in the case of behavioural traits. It also refers to the process of obfuscation, where a user deliberately alters his biometric trait to evade recognition.

A practical system should reach the specified accuracy, speed, and resource requirements, and, at the same time, it should be accepted by the intended population and be sufficiently robust to various fraudulent methods and attacks.

This Chapter is structured as follows. Section 1.1 depicts some real-life scenarios where biometric recognition is employed. Section 1.2 gives short overviews of the most researched and industrially exploited biometrics and defines invisible biometrics. In Section 1.3 the mode of operation of a biometric system is presented, while the errors that can occur in a biometric recognition process are illustrated in Section 1.4. Section 1.5 introduces multibiometrics systems. Eventually, in Section 1.6 an overview of the attacks a biometric systems is exposed and possible countermeasures are presented.

1.1 Use of Biometric Technologies

Biometric technology can be used for a great number of applications, that can be divided into the following three main groups [124]:

- **commercial applications:** computer network login, ATM access, credit card access, Internet access, electronic data security, e-commerce, account access over phone, physical access control, medical records management, hospital record security and consumer electronics;

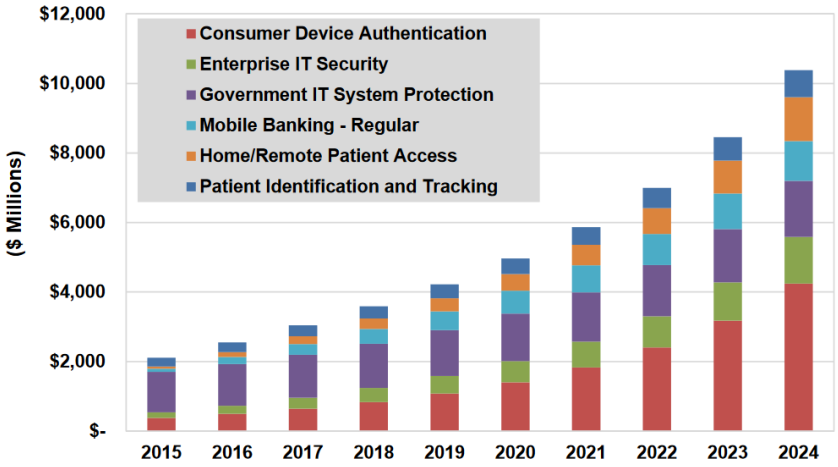


Figure 1.1: *Biometrics revenue by key use cases, World Markets: 2015-2024 (Source: Tractica)*

- **government applications:** low enforcement, government administration, aerospace and defence, border control, travel document, national ID card, driver's license, social security, and welfare disbursement;
- **forensic applications:** corpse identification, criminal investigation, terrorist identification, parenthood determination, and missing children.

Traditionally, commercial applications have used knowledge-based systems (e.g., PINs and passwords), government applications have used token-based systems (e.g., ID cards and badges), and forensic applications have been based on human experts to match biometric features. However, nowadays the use of biometric systems is largely increasing also in civilian and government applications and in forensic applications automatic biometric recognition systems are used.

Fig. 1.1 shows biometrics markets by identifying specific use cases, more in detail the six highest revenue uses cases globally, from Tractica's report, *Biometrics Market Forecasts*[1]. These six use cases account for just about 70% of all biometrics revenue throughout the 10-year forecast period. The remaining 35 use cases contribute the other 30% of revenue. As it can be noticed, consumer device authentication generates the most revenue of all use cases.

[1] `https://www.tractica.com/download-proxy?report_id=4018&type=White%20Paper`

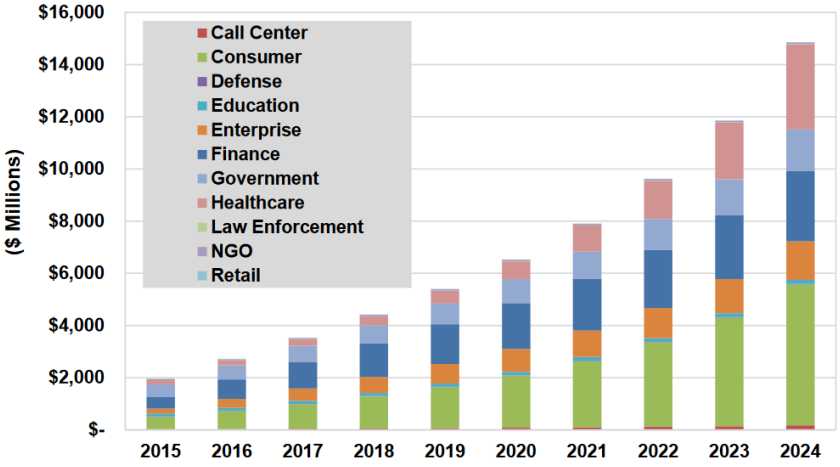


Figure 1.2: *Biometrics revenue by industry, World Markets: 2015-2024 (Source: Tractica)*

Fig. 1.2 depicts Tractica's biometrics market forecast by industry. The horizontal enterprise use cases are represented by the orange data series approximately in the middle of each bar on the chart. Visually, it is clear that use cases that sweep across industries are not the largest for biometrics and that orange enterprise data series accounts for about 10% of total revenues throughout the 10-year forecast period. Again, it can be noticed that consumers generate and are supposed to generate the the most revenue, followed by finance and government.

1.2 Biometric Identifiers and Invisible Biometrics

The existing biometric characteristic are copious and are used in many different practical applications for the task of user recognition; each biometric characteristic has its strengths and weaknesses, and no single biometric trait is ideal, that is it does not exist a biometric trait meeting all the aforementioned requirements [125]. The choice of which biometric characteristic to use depends on the application and the nature and constraints of the application lead the decision to which requirements to favour.

Biometrics are commonly classified into two categories: *physiological* and *behavioural* biometrics. A physiological biometric is based on a characteristic that a person owns, such as fingerprints or iris, while behavioural characteristics rely on something that the user do, that is some aspect of behaviour, e,g, voice, gait or keystroke.

In this thesis, a novel classification of biometric characteristics is pro-

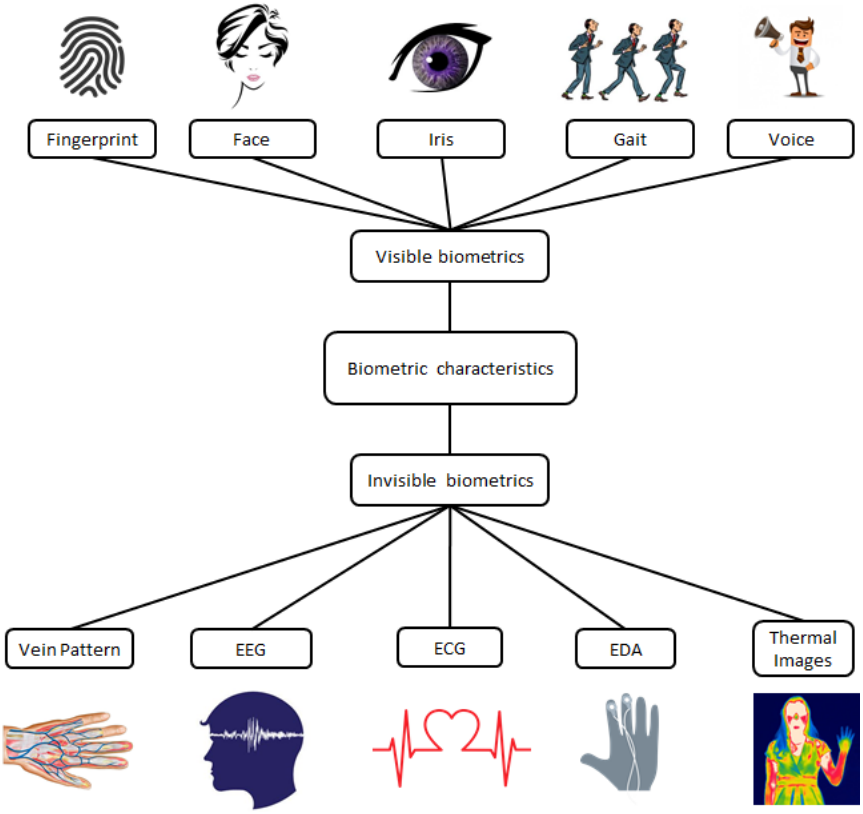


Figure 1.3: *Classification of biometric characteristics: visible and invisible biometrics.*

posed; specifically biometric traits can be categorised into *invisible* and *visible* biometrics, as shown in Figure 1.3. Invisible biometrics are physiological or behavioural characteristics hidden inside the human body. This property ensure a low level of circumvention of the biometric system, that is the system cannot be easily fooled using fraudulent methods when an invisible identifier is exploited. This is mainly due to the "hiddeness" of invisible biometrics, entailing the difficulty in acquiring the aforementioned trait at a distance and/or with conventional capturing devices, and implying the system to be secure and resistant to spoofing attacks. Besides, invisible biometrics are able to provide liveness detection, since they rely on physiological signal linked to the liveness of the user. Eventually, invisible biometrics are able to provide continuous authentication on a secure application. On the other hand, visible biometrics, being inherently visible and easy to acquire, are easy to be stolen and replicated by a fraudulent user and do not inherently provide liveness detection and continuous authentication.

Table 1.1: *Comparison of various biometric technologies. High, Medium, and Low are denoted by H, M, and L, respectively.*

Biometrics	Universality	Uniqueness	Permanence	Measurability	Performance	Acceptability	Circumvention
Fingerprint	M	H	H	M	H	M	M
Face	H	L	M	H	L	H	H
Iris	H	H	H	M	H	L	L
Voice	M	L	L	M	L	H	H
Signature	L	L	L	H	L	H	H
Gait	M	L	L	H	L	H	M
Keystroke	L	L	L	M	L	M	M
Vein pattern	M	M	M	M	M	M	L
EEG	H	H	M	M	M	L	L
ECG	H	M	M	H	M	H	L
EDA	M	M	M	H	M	H	L
Thermal images	H	M	M	M	M	H	L

Presented next are short overviews of the most researched and industrially exploited biometrics, following the classification proposed in this thesis. A brief overview of visible biometric characteristics will be given first, followed by a summary of invisible biometrics. A summary of the features of the aforesaid traits is given in Table 1.1.

Visible Biometrics

- **Fingerprint**. A fingerprint is the pattern of ridges and valley located on the surface of the tip of each finger and it can be acquired through compact sensors that provide digital images of these patterns. The formation of a fingerprint is determined during the first seven months of fetal development and fingerprints of identical twins as well as prints on each finger of the same person are different. One of the problem with the fingerprint recognition systems is that the most of fingerprint sensors are too sensitive to finger surface conditions and humidity in the atmosphere. Besides, fingertips physical conditions and injuries,

such as cuts and bruises, affect the performance of the system. Eventually stealing and replicating a fingerprint is extremely easy, entailing few difficulties in the circumvention of the system.

- **Face**. Face is the most common biometric characteristic used by humans to make a personal recognition; face recognition is a non-intrusive methods that allows to recognise people from facial images. The applications of facial recognition range from static ("mug shots") to dynamic, uncontrolled face identification in a cluttered background (e.g., subway, airport) [41]. The disadvantage of the face recognition systems commercially available is that they impose a number of restrictions on how the facial images are obtained, sometimes requiring a fixed and simple background or special illumination. Besides, faces can be captured with simple cameras, making the privacy and security of this biometric trait extremely low.

- **Iris**. The iris is the annular region of the eye bounded by the pupil and the sclera on either side. The visual texture of the iris begins to form in the third month of gestation, the structures creating its pattern are largely complete by the eight month and the pattern stabilises during the first two years of life. The complex iris texture carries very distinctive information useful for personal recognition and it is being used in various applications including national ID projects and border security. Each iris is distinctive and the irises of identical twins are different. It is extremely difficult to surgically tamper the texture of the iris and it's possible to detect artificial irises (e.g., designer contact lenses) [64]. Anyway, some iris recognition systems are still confused by textured contact lenses and printed image of an iris, showing also several limitations in the task of liveness detection.

- **Voice**. Voice is a combination of physiological and behavioural biometrics. The features of an individual's voice are based on the shape and size of the appendages (e.g., vocal tracts, mouth, nasal cavities, and lips) that are used in the synthesis of the sound. These physiological characteristics of human speech are invariant for an individual, but the behavioural part of the speech of a person changes over time due to age, medical conditions, and emotional state. Voice is also not very distinctive and may not be appropriate for large-scale identification; besides, a disadvantage of voice-based recognition is that speech features are sensitive to a number of factors such as background noise. Eventually, voice is very easy to record, copy and replicate to fool a voice-based recognition system.

- **Signature**. The way a person signs his or her name is known to be a characteristic of that individual; signatures are a behavioural biometric that change over a period of time and are influenced by physical and emotional conditions of the person. Although signatures require contact with the writing instrument and an effort on the part of the user, they have been accepted in government, legal, and commercial transactions as a method of verification.

- **Gait**. Gait is the peculiar way one walks and it is a complex spatio-temporal biometrics. It is not supposed to be very distinctive, but it allows to recognise people at a distance and can be used in some low-security applications. Gait is a behavioural biometric and may not remain the same over a long period of time, due to change in body weight or serious brain damage. To acquire the gait a video sequence is acquired; for this reason the acceptability of the gait as a biometric is high but since a video sequence is used this method is computationally expensive. Besides, mimicking, or imitation, of the human gait is a possible way to fool a gait-based biometric system.

- **Keystroke**. It is hypothesised that each person types on a keyboard in a characteristic way and keystroke dynamics is a behavioural biometric. This biometric trait is not unique to each individual but it may be expected to offer sufficient discriminatory information to permit identity verification. Large intra-class variations in a person's typing pattern can be observed due to changes in emotional state, position of the user with respect to the keyboard, type of keyboard used, etc; besides, the keystroke of a person can be monitored unobtrusively as that person is keying information. The advantage of the use of keystroke as biometric is that it can be used to perform a continuous verification of an individual over a session, after the fist authentication done with a stronger biometric such as fingerprint or iris.

Invisible Biometrics

- **Vein pattern**. Recently many researches showed that the pattern of vein underneath the skin is unique to each individual and so it can be used as a biometric characteristic for user recognition. The vein pattern can be acquired using an near infra-red light, and that property makes this biometric trait secure and resistant to falsification. Furthermore, the vein information represents liveness and for that reason liveness detection and authentication can be provided concurrently.

On the other hand, disadvantages of this biometric trait are that the acquisition devices are often large and that the acquired images have generally worst quality compared to images of other biometric traits.

- **Electroencephalogram (EEG)**. EEG signals are the result of the electrical field generated in the brain by the synchronous firing of specific spatially-aligned neurons of the cortex, i.e. pyramidal neurons. The brain electrical activity can be measured by sensing the differences of electrical voltage between specific positions on the scalp surface, by means of proper electrodes [188]. EEG signals can be captured in response to a presented stimulus or while performing a given task. Biometric recognition based on EEG signals is an emerging research topic. Several recent results have shown its feasibility and potential for personal identification. Without regard to the somewhat cumbersome data recording process, biometric recognition based on EEG signals shows several advantages compered to the traditional biometric identifiers. It is confidential and hard to imitate, since EEG signals are a reflection of individual-dependent inner mental tasks. Furthermore, EEG signals are difficult to imitate, as similar mental tasks are person dependent. In addition, one can not force a person to give ideal EEG signals as those recorded in normal situations, as brain activity is easy to be influenced by the stress and mood of a person [192].

- **Electrocardiogram (ECG)**. An ECG is a recording of the electrical activity of the heart. Electrodes placed on the surface of the body are used to measure the electrical signals originating from the myocardium, the heart muscle. The ECG signal is an emerging novel behavioural biometric for human identification. Individual differences in the heart structure, such as chest geometry, position, and size, manifest unique characteristics in their ECG signals which can be used as a biometric trait. The main advantages of ECG signals is the difficulty in counterfeiting them. Furthermore, the ECG signal, being present in all living individuals, can provide real-time liveness feedback. Besides, ECG provides additional information related to psychological states and physiological status, which may be of interest for applications such as health monitoring or body area sensor network systems. The main disadvantage of the ECG biometric trait is that the ECG signal inherently varies at different heartbeats of the same subject due to variations in fitness, physical and emotional states as well as variabilities caused by sensor position changes and long term baseline shifts [174, 214].

- **Electrodermal activity (EDA).** The electrodermal activity (EDA), also known as galvanic skin response (GSR), is the measure of electrical resistance between two points across the skin. In its most basic form, human skin is used as an electrical resistor whose value changes when a small quantity of sweat is secreted. So EDA refers to the variation of the electrical properties of the skin in response to sweat secretion. The sweat gland activity are reflective of the intensity of our emotional state, otherwise known as emotional arousal. Our level of emotional arousal changes in response to the environment we're in, e.g. if something is scary, threatening, joyful, or otherwise emotionally relevant, then the subsequent change in emotional response that we experience also increases eccrine sweat gland activity. EDA depends on physical conditions and thus is quite subjective. It has been successfully employed for biometric human recognition in the last decade [20, 193].

- **Thermal images.** Infrared thermography (IRT), thermal imaging, and thermal video are examples of infrared imaging science. Thermographic cameras usually detect radiation in the long-infrared range of the electromagnetic spectrum (roughly 9,000–14,000 nm or 9–14 μm) and produce images of that radiation, called thermograms. The amount of radiation emitted by an object increases with temperature; therefore, thermography allows one to see variations in temperature. In the field of biometric recognition, thermal imaging has been applied to hand vein pattern and face recognition.

The invisible biometrics this thesis focuses on are vein pattern and electroencephalogram. More in detail, an overview of the aforementioned identifiers is given in Chapter 2 and Chapter 5 respectively. Chapters 3 and 6 propose different algorithms aiming at improving performance in vein- and EEG-based biometric systems respectively, while Chapter 4 deals with security issues and countermeasures of finger-vein biometrics. Eventually, Chapter 7 presents some drawbacks of invisible biometric identifiers, paying particular attention at the issues of privacy and personal information leakage.

1.3 Biometric System

A biometric system measure one or more physiological and/or behavioural characteristics of the user and allows to automatically determine or verify the identity of the subject. The aforementioned task is performed by extracting discriminative features from the captured template and by comparing the obtained feature set against the template set stored in the database. A generic biometric system is composed by various blocks, listed and detailed hereafter:

- **sensor module:** an user interface incorporating the biometric sensor records the raw biometric data of the user;
- **feature extraction module:** the aim of this module is to generate the template, a compact but expressive representation of the identifier containing the distinctive features of the biometric data. The template is expected to contain only the salient discriminatory information that is essential for recognising the person;
- **database module:** it act as the repository of biometric information;
- **matching module:** the purpose of a biometric matcher is to compare the query features against the stored templates. It generates a matching score, that is a measure of similarity between the template and the query. The matcher module also encapsulates a decision module, which makes the identity decision, that is confirms an user's claimed identity or establishing a user's identity according to the matching score.

The process of user biometric recognition consists of two main phases, namely enrolment and recognition. During the *enrolment* phase the system register users into the biometric system database and associate an identity with its biometric characteristics. More in detail, the biometric trait is first acquired by the sensor and its digital representation is produced. After a quality check, the digital representation is processed in order to generate the template, later stored in the central database of the biometric system or on the device issued to the individual. The block diagrams of user enrolment is graphically illustrated in Fig. 1.4a. During the *recognition* phase, the biometric reader acquires the characteristic of the individual to be identified and its digital format is further processed to produce the feature with the same representation as the template. The resulting representation is fed to the matcher that compares it against the template to establish if the user is allowed to access to the system or not.

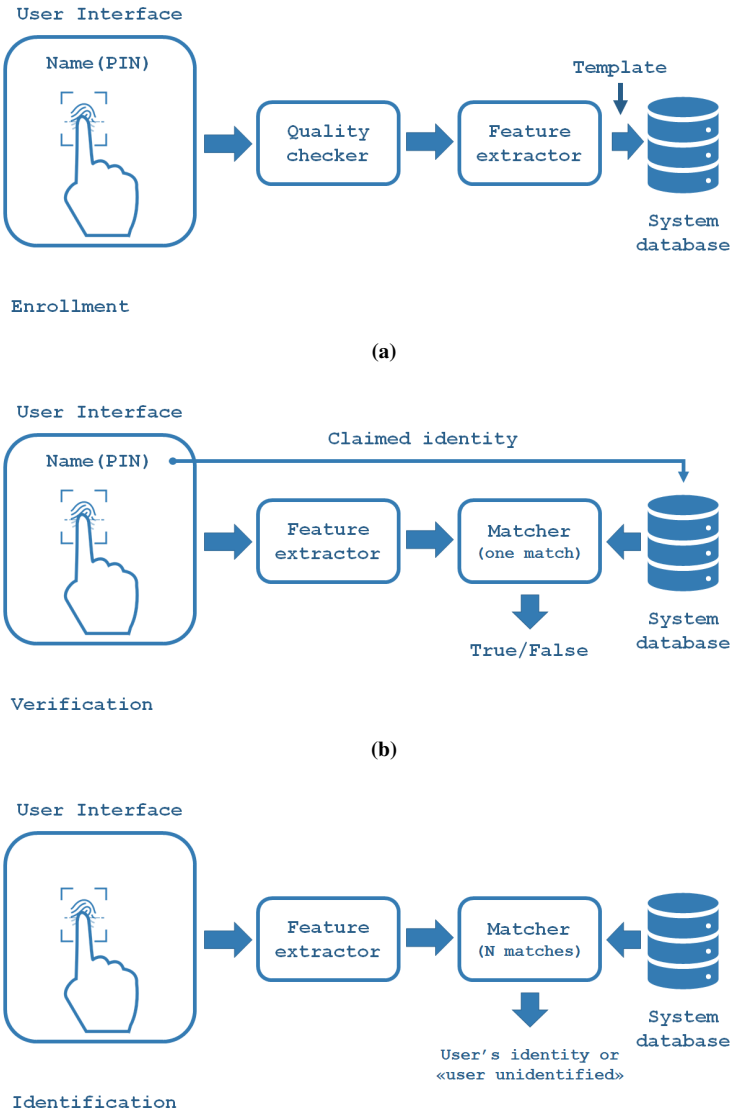


Figure 1.4: *Block diagram of (a) enrolment, (b) verification and (c) identification tasks.*

Depending on the application context, a biometric system may operate either in verification mode or identification mode. In the *verification mode*, the system authenticates a person's identity by comparing the captured biometric characteristic with the person's biometric template stored in the system database in a previous moment. In such a system, an individual who desires to be recognized claims an identity, usually via a personal

identification number (PIN), a user name, or a smart card, and the system either rejects or accepts the submitted claim of identity conducting a one-to-one comparison. Identity verification is typically used for positive recognition, where the aim is to prevent multiple people from using the same identity [315]. The block diagram of a verification system is depicted in Fig. 1.4b.

The verification problem can be formally posted as follow: given the input feature extracted from the acquired biometric data X_Q and the claimed identity I, determine if (I, X_Q) belongs to the class w_1 or w_2, where w_1 indicates that the claim is true and the user is then a genuine user, and w_2 indicates that the claim is false and then the user is an impostor. In order to determine its category, X_Q is matched against X_I, the biometric template corresponding to the user I and stored in the database. Thus

$$(I, X_Q) \in \begin{cases} w_1 & \text{if } S(X_Q, X_I) \geq t \\ w_2 & \text{otherwise} \end{cases} \tag{1.1}$$

where S is the the function that measures the similarity between feature vectors X_Q and X_I, and t is a predefined threshold. The value $S(X_Q, X_I)$ is generally termed as similarity or matching score between the biometric measurements of the user and the claimed identity. Therefore, every claimed identity is classified into w_1 or w_2 based on the variables X_Q, I, X_I and t and the function S.

In the *identification mode*, the system establishes a subject's identity or fails if the subject is not enrolled in the system database by searching the templates of all the users in the database for a match, without the subject having to claim an identity. In a recognition system operating in the identification mode the system conducts a one-to-many comparison to establish the users' identity or returning a failure in case of the subject is not enrolled in the system database. Identification is a critical component in negative recognition applications where the system establishes whether the person is who he or she denies to be. The purpose of negative recognition is to prevent a single person from using multiple identities [315]. While traditional methods of personal recognition such as passwords, PINs, keys, and tokens may work for positive recognition, negative recognition can only be established through biometrics. The block diagram of an identification system is depicted in Figure 1.4c.

The identification problem is formulated as follows. Given an input feature vector X_Q, determine the identity I_k, with $k \in \{1, 2, ..., N, N+1\}$. In this case $I_1, I_2, ..., I_N$ are the identities enrolled and I_{N+1} indicate the

case where non suitable identity can be determined for the user, that is the case in which the user is rejected. Hence

$$(I, X_Q) \in \begin{cases} I_k & \text{if } \max_k \{S(X_Q, X_{Ik})\} \geq t, \quad k = 1, 2, \ldots N \\ I_{N+1} & \text{otherwise} \end{cases} \tag{1.2}$$

where X_{Ik} is the biometric template corresponding to identity I_k, and t is a predefined threshold.

1.4 Biometric System Errors

Because of imperfect imaging conditions, changes in the user's physiological or behavioural characteristics, ambient conditions, and user's interaction with the sensor, two samples of the same biometric characteristic from the same person are not exactly the same [124]. For this reason, the response of the biometric system is typically a single number, the matching score $S(X_Q, X_I)$, that quantifies the similarity between the template extracted during the recognition phase (X_Q) and the template stored in the database (X_I). The higher the score, the higher is the probability that the two biometric measurements come from the same person. In order to decide if the user who wants to access the system is allowed to, a comparison with the threshold (t) is performed: if the matching score is higher than or equal to the threshold, the pairs of biometric samples are inferred as *mate pairs*, that is traits belonging to the same person; on the contrary if the matching score is lower compared to the threshold value, the pairs of biometric samples are meant as *nonmate pairs*, that is belonging to different people.

Matching all the possible pairs of biometric samples from the same person in the system, the distribution of genuine scores, called *genuine distribution*, will be generated; the score distribution obtained by matching pair of biometric characteristics from different persons is called *impostor distribution*. An example of genuine score and impostor score distributions is shown in Fig. 1.5.

A *biometric verification system* makes two types of errors:

- **false match** (or false accept): the biometric measurements are from two different persons but the decision module state that they belong to the same person;
- **false nonmatch** (or false reject): two biometric measurements from

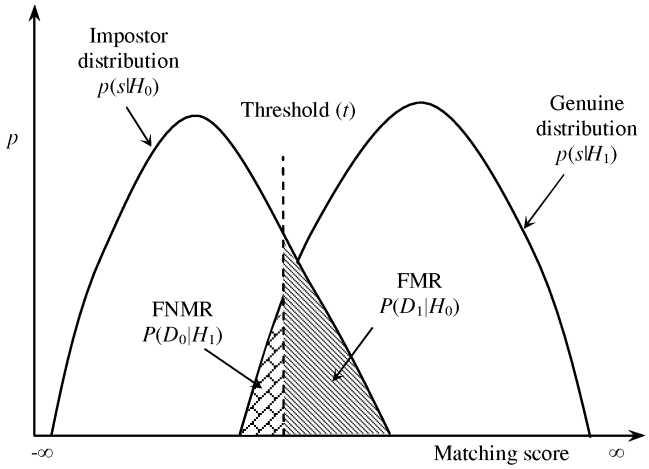


Figure 1.5: *Genuine score and impostor score distribution. FMR and FNMR for a given threshold t are also displayed over the genuine and impostor score distributions.*

the same person are not considered belonging to the same person by the system.

Mathematically, the errors in a verification system can be formulated as a binary classification problem. If X_I represents the stored biometric template of the user I and X_Q is the acquired input during the recognition stage, then there are two possible hypotheses:

- H_0: the input X_Q and the template X_I don't come from the same person;
- H_1: the input X_Q and the template X_I come from the same person.

The associated decisions are as follows:

- D_0: the person is not who she claims to be;
- D_1: the person is who she claims to be.

The decision rule is the following: if the matching score $S(X_q, X_I)$ is less than the system threshold t, then decide D_0, else decide D_1. Assuming such a hypothesis testing formulation, two kind of error could occur:

1. *false match*: the system decide that X_I and X_Q belong to the same user, while they belong to different users, that is H_0 is true but D_1 is decided;

2. *false non-match*: the system decide that X_I and X_Q belong to different users, while they belong to the same person, that is H_1 is true but D_1 is decided;

The probability that the system incorrectly authorises a non-authorised person, due to incorrectly matching the biometric input with a template, that is the probability of type-I, is referred as *false match rate (FMR)*; the probability that the system incorrectly rejects access to an authorised person, due to failing to match the biometric input with a template, that is the probability of type-II, is called *false nonmatch rate (FNMR)*. These two probabilities are defined as:

$$\begin{aligned} FMR &= P(D_1/H_0) \\ FNMR &= P(D_0/H_1). \end{aligned} \tag{1.3}$$

To evaluate the accuracy of a biometric system, it's necessary to collect scores generated comparing multiple images from the same biometric trait, that is the distribution $p(S(X_Q, X_I)/H_1)$, and scores generated comparing images from biometric data belonging to different users, that is the distribution $p(S(X_Q, X_I)/H_0)$. The performance indicator can be then calculated as follows:

$$\begin{aligned} FMR &= \int_t^{\infty} p(S(X_Q, X_I)/H_0)dS \\ FNMR &= \int_{-\infty}^{t} p(S(X_Q, X_I)/H_1)dS. \end{aligned} \tag{1.4}$$

The computation of FMR and FNMR over genuine and impostor distributions, for a given threshold, is graphically illustrated in Fig. 1.5. FMR and FNMR can be represented together in the same curve called *detection error trade-off (DET)*; the DET curve plots the FNMR against the FMR at different thresholds; an example of a DET curve is shown in Fig. 1.6. A variant of the DET curve to represent the performance of a biometric system is the *receiver operating characteristic (ROC)* curve. A ROC curve is similar to the DET curve except that the the y-axes represents the probability of the genuine accept of the system, also known as *genuine accept rate (GAR)* and defined as $GAR = 1 - FNMR$. The DET and ROC curves are often non-linearly scaled in order to highlight the region of error rates of interest; commonly used scales include normal deviate scale and logarithmic scale.

As told before, the values of FMR and FNMR depends on the chosen value of the threshold; the choice of the value of the threshold is linked

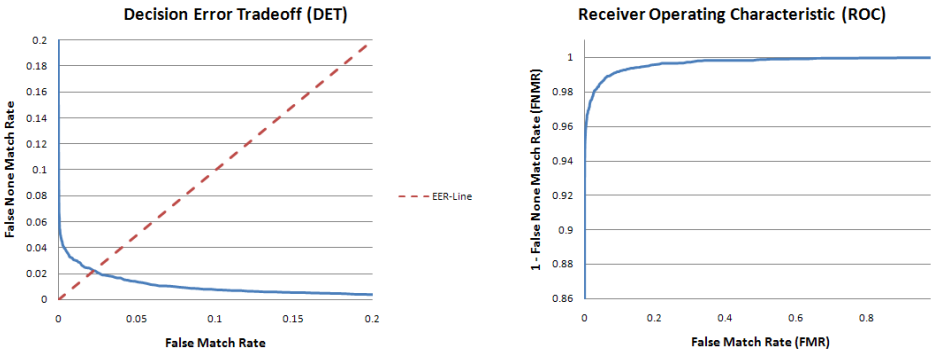


Figure 1.6: *Choosing different threshold values, different FMR and FNMR values are obtained. The curves relating FMR to FNMR at different thresholds are known as DET (left) and ROC (right) curves.*

to the application of the biometric system and there is a trade-off between FMR and FNMR in every biometric system: the system is more tolerant to input variations and noise decreasing the threshold but the FMR increases; on the other hand, if the threshold is increased to make the system more secure, the FNMR increases accordingly. Sometimes it's useful to have a performance measurement independent of the operating point; to express the performance of the system using a metric independent of the operating point, one of the possible performance metric used is the *equal error rate (EER)*, defined as the rate where FMR and FNMR are equal; in particular the EER of a matcher can be also seen as the operating point at the intersection of the line $FMR = FNMR$ with the ROC curve.

The accuracy of a biometric system in the *identification mode* can be inferred using the system accuracy in the verification mode under simplifying assumptions. Let us denote the identification false nonmatch and false match rates with $FNMR_N$ and FMR_N, respectively, where N represents the number of identities in the system database. If the assumption that and only a single identification attempt is made per subject, a single biometric template is used for each enrolled user, and the impostor scores between different users are uncorrelated is valid, then:

$$FNMR_N \cong FNMR$$

$$FMR_N = 1 - (1 - FNMR)^N \cong N \cdot FMR. \tag{1.5}$$

This approximations hold good only when $N \cdot FMR < 0.1$ [124].

1.5 Multimodal Biometric Systems

Even though biometric system have been successfully installed in various civilian and governmental applications, biometric recognition is not still a fully solved problem; issues related to reducing error rate, enhancing the usability of biometric systems and reducing the biometric system limitations are tasks that the research community still aims to address. More in detail, *unimodal biometric systems*, that is biometric application operating using a single biometric trait, are characterised by the following limitations [124]:

- **noise in acquired data:** the captured data might be noisy or distorted; this could be linked to the physical conditions of the biometric trait or to the defective or improperly maintenance of the sensors or also to unfavourable ambient conditions. The use of noisy biometric data could entail errors in the matching process resulting in a user being incorrectly rejected;

- **intra-class variations:** the biometric data acquired during the enrolment stage and the one acquired in the authentication step and stored in the database should be very similar in order to limit the errors in the matching stage. In real-life applications, the data used to obtain the enrolled template and the ones captured for authentication may be very different, due to the fact that the user incorrectly interacts with the sensor or linked to changing in characteristics of the sensor itself; this causes a large intra-class variation that entails problems during the matching phase;

- **distinctiveness:** while it is expected a large inter-class variation, that means that samples of biometrics traits of different individuals should be sufficiently different, there may be large inter-class similarities in the feature sets used to represent these identifiers. This limitation restricts the discriminability provided by the biometric trait and every biometric trait has some theoretical upper bound in terms of its discrimination capability;

- **non-universality:** although it is expected that a biometric identifier used for a specific recognition system is possessed by all the users of the system, it is possible that the biometric trait is not possessed by a subset of them. This could be due to the fact that the physical conditions of the biometric trait make the feature extraction process unable to extract features from the biometric characteristic, entailing

a failure to enrol (FTE) error associated with the use of the single biometric trait;

- **spoof attacks:** one of the possible threat for a biometric system is the spoofing attack, that is an impostor may attempt to spoof the biometric trait of a legitimate enrolled user in order to circumvent the system. Spoofing attacks are possible to be performed both in the case of physiological- and behavioural-characteristic-based biometric systems.

Some of the limitations imposed by unimodal biometric systems can be overcome by using *multimodal biometric systems*, that is biometric systems that use multiple biometric modalities instead of a single biometric characteristic. These systems are expected to be more reliable due to the presence of multiple pieces of evidence and they are also able to improve the performance of the system. Multimodal biometric systems address the problem of non-universality, since several traits of an user are taken into account, the system can reach a certain degree of flexibility and ensure a sufficient population coverage. Besides, a *hierarchical approach* can be used in the recognition stage: some samples can be used to determine a subset of candidate from the database an the remaining samples can be employed to determine the final identity, chosen between the ones of the small pre-selected set. Furthermore, multimodal biometric systems increase security and provide anti-spoofing measures by making it difficult for an intruder to simultaneously spoof the multiple biometric traits of a legitimate user. Eventually, the noisy data problem can be mitigated, since multiple traits are acquired and there are more changes to get some traits less noisy than others.

Multimodal biometrics can be designed to operate in different scenarios, as shown in Fig. 1.7; more in detail, it is possible to classify the multibiometric systems based on the information that they integrate as follows:

- **multi-sensors systems:** multiple sensors are used to acquire the same biometric trait of an individual;

- **multi-instance systems:** different instances of the same biometric trait are captured and used for the enrolment and/or recognition stages;

- **multi-sample systems:** a single sensor can be used to acquire multiple samples of the same biometric trait under different conditions to obtain a more complete representation of the trait;

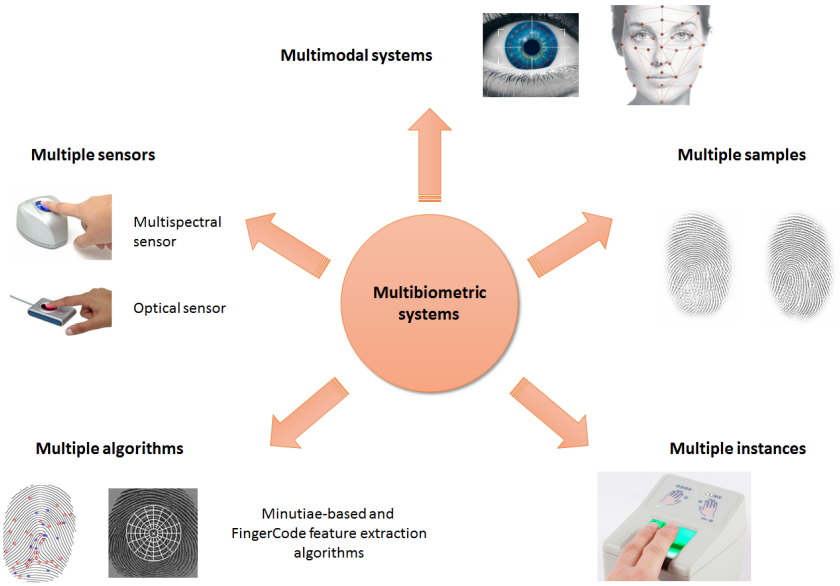


Figure 1.7: *Various scenarios in a multimodal biometric system.*

- **multi-biometrics systems:** multiple biometric traits are used and combined in order to later establish the identity of the user;
- **multi-algorithm systems:** the same biometric data is processed using multiple algorithms. Multiple feature extraction methods or multiple matching algorithms are involved and combined for the task of biometric recognition;
- **hybrid systems:** some of the scenarios described above can be integrated and combined for the aim of user recognition.

The information obtained by the multiple sensors scenario are moderately dependent, so it's expected to result in a smaller improvement in recognition accuracy compared to the multiple biometrics and multiple instance scenarios in which the information combined are independent. However, on the other hand, there is the inconvenience for the user to provide multiple cues and it entails a longer acquisition time. Besides, the cost of the systems where only one sensor is used is less compared to the cost of the system where more sensors are required. Finally, the multi-algorithm scenario, combining different representation and matching algorithms to improve the recognition accuracy, is more cost effective and convenient than scenarios where multiple biometric traits or sensors are used; on the other

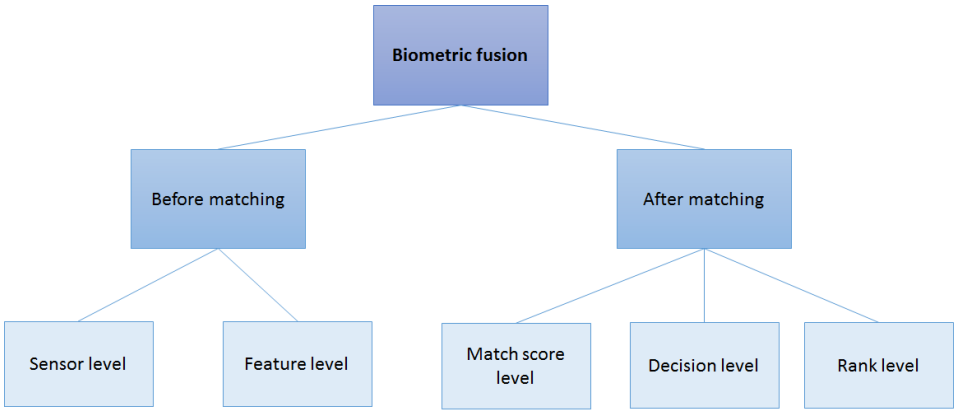


Figure 1.8: *Classification of fusion strategies in multibiometric systems.*

hand, in these systems, as well as systems that use different instances of the same biometric trait, the computational complexity is increased and as a result, the response time of the system is bigger. Even though multi-algorithm and multi-instance require more computational and storage resources than a unimodal biometric system, in principle, different feature extractors and matchers can work in parallel and the overall response time of the system is limited by the slowest individual feature extractor and/or matcher.

The multiple information obtained in multibiometric systems can be integrated and combined at various levels. In particular, the level of fusion can be categorised into two broad categories, as illustrated in Fig. 1.8: *fusion before matching* and *fusion after matching*.

In the case of the fusion before matching, the fusion can be performed at two different levels:

- **sensor-level:** data belonging to the same biometric trait are acquired from multiple sensors and then combined;
- **feature-level:** a combination of different feature vectors is performed. The different features can be obtained by either data acquired from multiple sensors or by applying multiple feature extraction algorithms on the same biometric data. If the features extracted from one biometric indicator are independent of those extracted from the others, feature reduction techniques may be employed to extract a small number of salient features from the larger set of features.

In the case of the fusion after matching, the fusion can be performed at three different levels:

- **match score-level:** a similarity score indicating the closeness between two compared features is provided by each biometric matcher. These matching scores can be combined to assert the genuineness of the claimed identity of the user;
- **decision-level:** the output of multiple classifiers relying on different feature vectors are combined in order to take the final decision;
- **rank-level:** when a biometric system operates in the identification mode, the output can be seen as a ranking of the enrolled entities; the rank-level fusion can be performed in order to consolidate the rank output.

Performing one or a combination of the different fusions methodologies listed before, the biometric system is expected to be more reliable and the performance achieved by the system are generally higher compared to what reached in unimodal biometric recognition systems.

In this thesis, we exploited the advantages of multimodal biometric system in the field of invisible biometrics. More in detail, in Section 3.1 and Section 3.3 the advantages of biometric fusion techniques in the area of vein-based biometric recognition are studied. Section 6.1, instead, shows the application of fusion techniques in the framework of EEG-based biometric identification.

1.6 Vulnerable Points of a Biometric Systems and Countermeasures

The advantages related to biometric-based recognition, when compared to token- or knowledge-based approaches, are remarkable and, for that reason, there has been a significant surge in the use of biometrics for user recognition in recent years. It's important for a biometric system to be designed to withstand attacks when used in security-critical application, above all in unsupervised remote applications such as e-commerce [252]. Even if a considered biometric system is characterised by excellent recognition performance, it could be rendered useless by an attack to the system itself.

In a biometric system different weak links vulnerable to a variety of attacks aimed at mining the integrity of the recognition process can be identified. In particular, the performed attacks are intended to either circumvent the security afforded by the system or to deter the normal functioning of the system. Given a biometric system, there are different attacks that can be perpetrated at different levels of the system, that are shown in Fig. 1.9 and that can be summarised as follows:

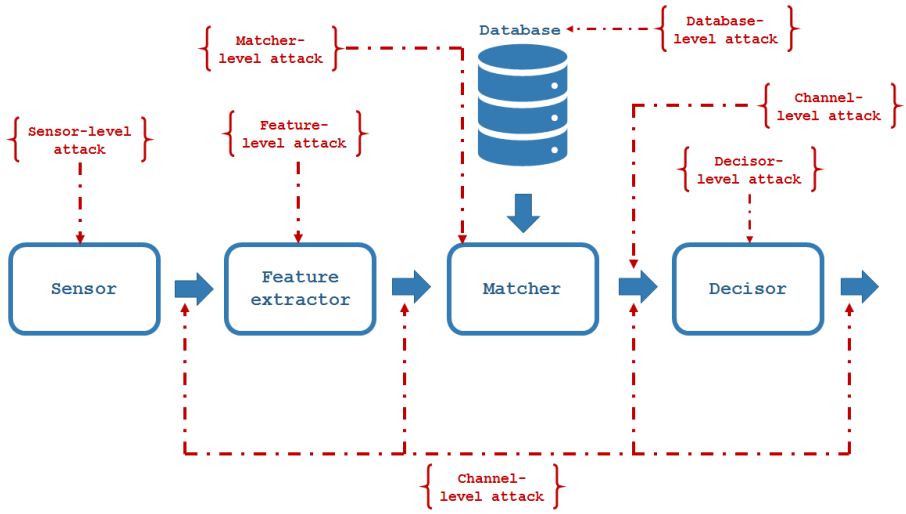


Figure 1.9: *Possible attacks point in a biometric system.*

- **sensor level:** a reproduction of the biometric trait is presented at the system by an impostor (*spoofing attack*) or the true biometric is presented but by an unauthorised manner, e.g. forcing the legitimate user to grant an impostor access to the system;
- **feature-extraction level:** the feature extraction module can be forced in order to overriding the feature extraction process and forcing the feature extractor to produce feature sets pre-selected by an attacker;
- **matcher level:** the matcher might be attacked and corrupted in order to produce pre-selected match scores;
- **decisor level:** the final decision is overrode, making the authentication system disabled;
- **Channel level:** channels interconnect the different modules of the system; they can be intercepted and controlled by an attacker. Among the possible channel-level attacks the following ones can be mentioned:
  - *eavesdropping attack*: biometric data transmission is surreptitiously listened;
  - *man in the middle attack*: the attacker manipulates the information exchanged between two modules, without the two parties knowing that the link has been compromised;

  - *brute force attack*: exhaustive presentation of a large set of biometrics input to the recognition system to find the one that works;
  - *replay attack*: previously stored digitised biometric signals, features or scores are resubmitted to the system, bypassing the modules;
  - *manipulation attack*: matching scores or the decision is captured and their values are changed.

- **database level:** the records in the database are read and modified; the templates stored in the database are modified or removed or a new template is introduced in the database itself.

One of the most potentially damaging attack on biometric system is against the biometric template stored in the system database [123]. There exist several security techniques to thwart attacks at these various points and the wide spread use of biometric authentication imposes serious threats to the security and privacy of its users: the system should be able to guarantee the impossibility of leakage of information to unauthorised individuals. Between the possible solutions, template protection systems ensure the secure and private handling of the personal biometric templates or biometric data during the authentication process inside a biometric recognition systems in order to protect the users' privacy and security. According to the ISO/IEC 24745 standard on biometric information protection [121], in biometric template protection (BTP) systems the following three properties are required to protect the privacy of the users:

- **irreversibility:** given a protected template, it should leak no biometric information, i.e., it should not be possible to go back from the template to the biometric sample that originated it;

- **unlinkability:** given two templates protected with different keys, it should not be feasible to decide whether they conceal the same biometric instance;

- **renewability:** if one template is lost or stolen, it should be possible to issue a new one that does not match with the old template.

Besides, an ideal template protection scheme should satisfy the following properties [123]:

- **performance:** the introduction of a BTP scheme should not significantly degrade the recognition performance (FAR and FRR) of the unprotected biometric system;

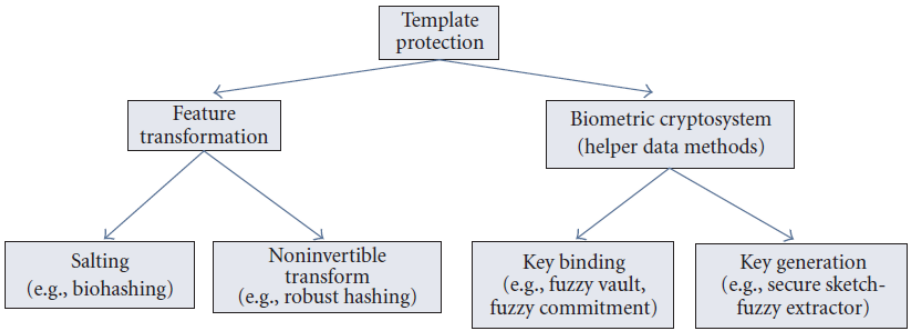


Figure 1.10: *Classification of template protection schemes [123].*

- **security:** it must be computationally hard to obtain the original biometric template from the secured template. This property prevents an attacker from creating a fake copy of the original biometric trait from a stolen template;
- **diversity:** the protected template should not match with the other previously generated template from the same data, that is cross-matching has to be avoided. In this way, the user's privacy is ensured;
- **revocability:** it should be straightforward to revoke a compromised template and reissue a new one based on the same biometric data.

In the recent years, many different solutions have been already proposed for the generation of secure and protected templates. Among the possible classifications of BTP algorithms, BTPs can be categorised into two main categories [123], as shown in Fig. 1.10: *biometric cryptosystems* and *feature transformation* approaches. Some template protection techniques make use of more than one basic approach, that is *hybrid schemes* are employed. A brief summary of the various template protection approaches is depicted in the following and presented in Table 1.2.

Biometric Cryptosystems. Biometric cryptosystems provide the means to adapt cryptographic protocols to biometric data which are inherently noisy data. In a biometric cryptosystem, some public information about the biometric template, generally referred to as helper data, is stored; the helper data doesn't reveal any significant information about the original biometric template and it is needed during the matching stage to extract a cryptographic key from the query biometric features. Matching is performed indirectly by verifying the validity of the extracted key. Examples of the

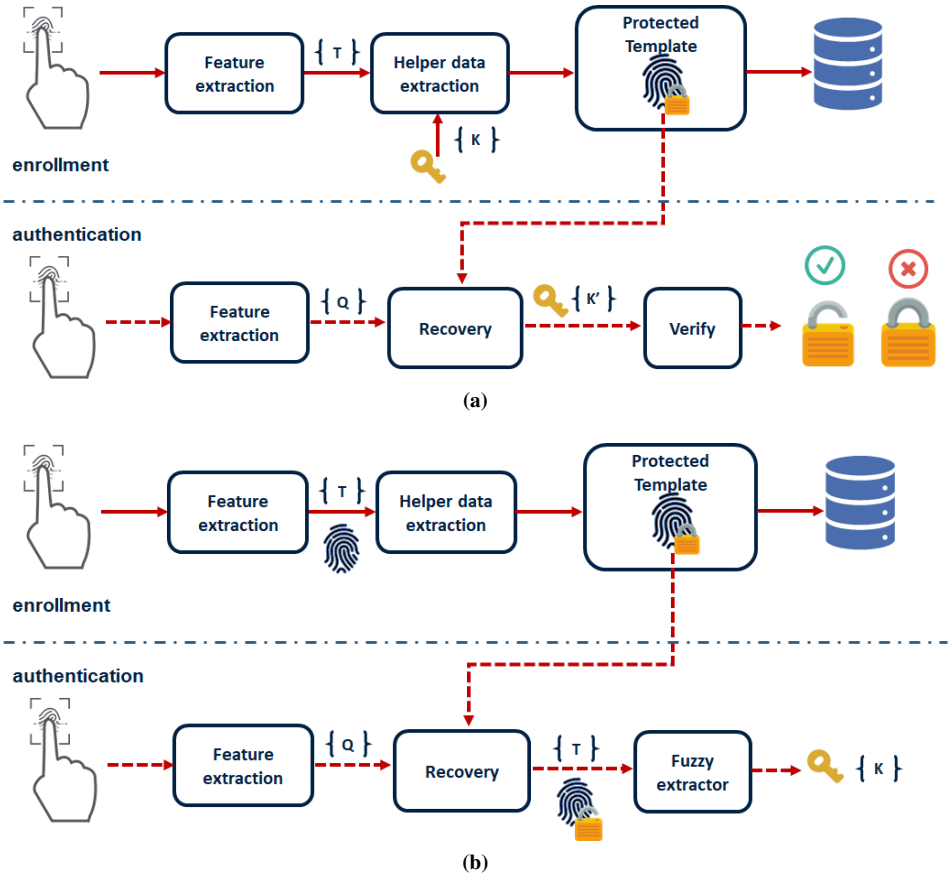


Figure 1.11: *Biometric cryptosystems for template protection: (a) key-binding, (b) key-generating.*

biometric cryptosystem scheme can be seen in Fig. 1.11. Error correction coding techniques are typically used to handle intra-user variations.

Biometric cryptosystems can be classified into two subclasses, depending on how the helper data is obtained: key-binding and key-generating systems [106, 293]. When the helper data is obtained by binding a key, independent of the biometric features, the protection scheme is named as *key-binding* biometric cryptosystem. The aim of a key-binding scheme is to store information obtained by combining biometric data with a randomly generated keys, that is secure some cryptographic key by means of a biometric trait. On the other hand, if the helper data is derived only from the biometric template and the cryptographic key is directly generated from the helper data and the query biometric features, the scheme is referred as *key generation* biometric cryptosystem. A key-generating system, derive a

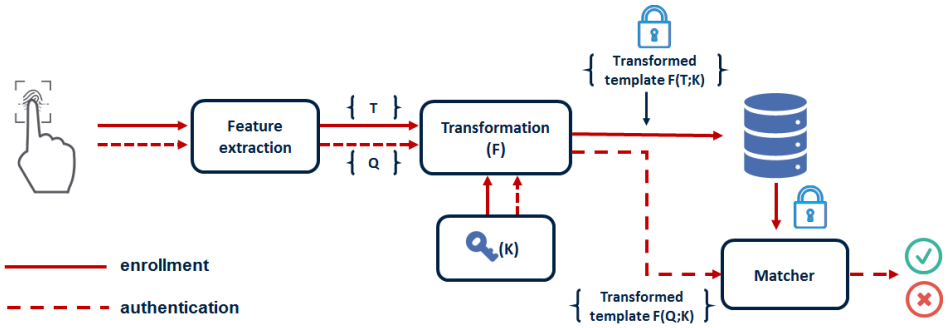


Figure 1.12: *Feature transformation approach for biometric template protection.*

cryptographic binary key directly from the acquired biometrics.

Feature Transformations. In the feature transformation approach, or *cancelable biometrics*, the biometric template or the original biometric data ($\mathcal{T}$) are transformed by applying a transformation function ($\mathcal{F}$). The parameters of the transformation function are typically derived from a random key ($\mathcal{K}$) or password, and only the transformed template or signal ($\mathcal{F}(T;K)$) are stored in the database, as shown in Fig. 1.12. During the matching stage, the same transformation is applied to to query features/images ($\mathcal{Q}$) and the transformed query ($\mathcal{F}(Q;K)$) is directly matched against the template stored in the database ($\mathcal{F}(T;K)$). Depending on the characteristics of the transformation function ($\mathcal{F}$), the feature transform schemes can be further categorised as salting and non-invertible transforms. In *salting* or *biohashing*, the considered function is invertible and the key used to define the user-specific transformation needs to be securely stored or remembered by the user and presented during authentication. The security of salting approaches relies on the protection and secrecy of the password of the function parameters, i.e. the key ($\mathcal{K}$); if the key is compromised the template is no longer secure and an adversary that gain access to the key and the transformed template is able to recover the original biometric template.

In *non-invertible transformation* schemes the transformation function applied on the template is typically a one-way function, that is easy to compute (in polynomial time) but it is computationally hard to invert (given $\mathcal{F}(x)$, the probability of finding x in polynomial time is small); for that reason, even if the key is known, it's computationally hard for an attacker to recover the original biometric template starting from the distorted one and this quality makes this scheme more robust and secure compared to the sating approach. The one-way function must meet the three properties

Table 1.2: *Summary of different template protection schemes. Here, $\mathcal{T}$ represents the biometric template, Q represents the query, and $\mathcal{K}$ is the key used to protect the template. In salting and non-invertible feature transform, $\mathcal{F}$ represents the transformation function, and $\mathcal{M}$ represents the matcher that operates in the transformed domain. In biometric cryptosystems, $\mathcal{F}$ is the helper data extraction scheme and $\mathcal{M}$ is the error correction scheme that allows reconstruction of the helper data $\mathcal{H}$ [123].*

Approach	**What imparts security to the template?**	**What entities are stored?**	**How are intrauser variations handled?**
Key-generating biometric cryptosystem	Level of security depends on the amount of information revealed by the helper data $\mathcal{H}$	Public domain: helper data $\mathcal{H} = F(T)$	Error correction and user specific quantization $\mathcal{K} = M(F(T), Q)$
Key-binding biometric cryptosystem	Level of security depends on the amount of information revealed by the helper data $\mathcal{H}$	Public domain: helper data $\mathcal{H} = F(T; K)$	Error correction and user specific quantization $\mathcal{K} = M(F(T; K), Q)$
Non-invertible transform	Non invertibility of the transformation function $\mathcal{F}$	Public domain: transformed template $\mathcal{F}(T; K)$, key $\mathcal{K}$	Matching in transformed domain $\mathcal{M}(F(T; K), F(Q; K))$
Salting	Secrecy of key $\mathcal{K}$	Public domain: transformed template $\mathcal{F}(T; K)$, Secret: Key $\mathcal{K}$	Quantization and matching in transformed domain $\mathcal{M}(F(T; K), F(Q; K))$

defined by the ISO/IEC 24745 standard. In addition, the application of the one-way function should not significantly degrade the system's recognition performance as well as the recognition performance should not be sensitive to the actual key used during the template protection step [224, 255]. If the non-invertible transformation functions leave the biometric template in the original feature space even after the transformation, the matching criteria used is the same for the original and the transformed feature set; templates that lie in the same space after the application of a non-invertible transform have been referred to as *cancelable templates* [23].

The use of salting approaches typically results in low false accept rates, and, since the key is user-specific, multiple templates for the same user biometric can be generated by using different keys, allowing diversity. Also in case a template is compromised, it is easy to revoke the compromised template and replace it with a new one generated by using a different user-specific key, allowing revocability. On the other hand, if a key gets compromised, the user template is no longer secure due to the invertibility of the transformation, that is, if an adversary gains access to the key and the transformed template, she can recover the original biometric template. Besides, since matching takes place in the transformed domain, the salting mechanism needs to be designed in such a way that the recognition performance does not degrade, especially in the presence of large intra-user variations.

When non-invertible transforms are used, since it is hard to recover the original biometric template, even when the key is compromised, this scheme provides better security than the salting approach and no significant information can be acquired on the template. Moreover, differently from the cryptosystem approaches, the original and transformed templates can remain in the same feature space, being possible to apply standard matching techniques in the transformed domain; this allows to achieve similar performance compared with those achieved with the unprotected approach. In addiction, transformation-based approaches generally result in matching scores that can later be fused through multibiometric methods. Eventually, diversity and revocability can be achieved by using application-specific and user-specific transformation functions, respectively. The main limitation of feature transformations approaches lies in the need to find the trade-off between discriminability and non-invertibility of the transformation function. The transformation function should be designed in order to keep the intra-class and inter-class distances in the transformed domain similar to the corresponding ones in the original domain, in order to preserve the discriminability. On the other hand, the transformation should also be non-invertible, and it is difficult to design transformation functions that satisfy

both the discriminability and non-invertibility conditions simultaneously. Moreover, it is very difficult to perform a rigorous security analysis concerning non-invertibility of the scheme, above all when the transformation algorithm and related keys or parameters are also compromised.

In this thesis, some feature transformation approaches are applied to the vein pattern biometric identifier. More in detail, non-invertible transformation are applied. Details about the proposed work can be found in Section 4.2.

CHAPTER 2

Invisible Biometrics: Hand Vein Pattern

VEIN RECOGNITION refers to a biometric authentication method based on the network of blood vessels of people's hands. In biometric applications where human recognition with high-level security is required, such as border controls, smartphone unlocking, ATM cash withdrawals, and e-commerce to cite a few, vein patterns are nowadays often chosen as solution.

Vein patterns have recently attracted a significant interest from the industrial and the academic communities, thanks to the several advantages this biometric trait can offer with respect to other traditional identifiers. More in detail, the vein pattern, being hidden under the skin and visible only thanks to the help of a near-infrared light and camera, falls under the *invisible biometrics* category. The aforesaid property makes this biometric trait very difficult to be stolen and replicated, and, then, a spoofing attack nearly impossible to be implemented. Besides, the aforementioned biometric modality allows a contactless recording acquisition procedure, thus making it possible to design a vein-based biometric system ensuring users' comfort and ease of use. Additionally, liveness detection is intrinsically provided. On the other hand, vein-pattern images are often characterised

by low contrast and poor definition, due to the sub-cutaneous placement of the veins, thus making vein-related feature extraction a challenging process and then implying a bad impact on the recognition performance of the system.

Many practical systems exploiting different kinds of vein pattern have already been implemented and are commercially available. The main producers of vein-based biometric systems are Hitachi™,[1], Fujitsu™,[2] and IDEMIA™,[3]. The first commercially available application of vein-based biometric systems appeared in Poland, that has become the first country in Europe to introduce a network of ATMs relying on finger-vein based recognition for cash withdrawal. Behind the Polish ATM, there is Hitachi™ Europe[4]. The incorporation of finger-vein biometrics into ATM machines was also developed in China by the fourth largest bank, Bank of China[5]. Some Japanese banks also use the aforementioned technology as a security measure which allows, or prevents, customers from accessing safety deposit boxes in branches[6]. Barclays™ Bank introduced "finger-vein ID" readers that allow customers to junk their pin numbers, passwords and identification codes and instead access their account with just a scan of their finger-vein pattern thought a device plugged into the computer's USB port[7]. Recently vein-based recognition technologies have been integrated in mobile system: LG™ designed its LG G8 using the veins in the palms of the user's hands to unlock the phone, which the company calls "Hand ID". To detect and recognise the veins, the smartphone is provided with a camera that detects a users' veins with infrared light. There's also a 3D sensing "time of flight" (ToF) camera on the front of the phone that detects the shape, thickness, and other individual characteristics of a user's palms. To unlock the LG G8 using Hand ID, a user needs to hover their palm above the LG G8's selfie camera, and raise it up slowly[8].

A standard vein based-biometric system is composed by the following modules: an acquisition system able to capture the vein pattern images, a

[1] http://www.hitachi.co.jp/products/it/veinid/global/products/embedded_devices_u.html

[2] https://www.fujitsu.com/it/solutions/business-technology/security/palmsecure

[3] https://www.idemia.com/morphoaccess-vp

[4] https://www.theguardian.com/money/2014/may/14/fingerprints-vein-pattern-scan-atm

[5] https://findbiometrics.com/finger-vein-authentication-atms-china-502087/

[6] https://www.bbc.com/news/business-29062901

[7] https://www.bbc.com/news/business-29062901https://www.bbc.com/news/business-29062901

[8] https://www.biometricupdate.com/201902/palm-vein-biometrics-and-3d-facial-recognition-built-into-new-lg-g8-thinq-smartphone

pre-processing module, which aim is to enhance the acquired data, a feature extraction module, and, eventually, a matching module that compares the authentication feature with the enrolled ones, returning a score linked to the similarity between the compared features. This chapter aims to give an introduction on vein-pattern-based biometric systems and its modules.

This chapter is organised as follows. Section 2.1 describes the acquisition procedure of the vein pattern, while Section 2.2 gives an overview about the vein-pattern publicly available databases. Section 2.3 gives a survey about the most important works presented in literature in the field of vein based-biometric recognition. Eventually, an overview of the state-of-the-art works regarding deep-neural-network (DNN) approaches applied in the framework of vein-based biometric recognition is provided Section 2.4.

2.1 Vein Pattern Acquisition Systems

The blood vessels are part of the circulatory system and their main function is to carry blood in the body [87]. The blood vessels can be categorized in two types: arteries and veins. Arteries carry oxygenated blood from the heart to various parts of the body, while veins carry deoxygenated blood from all body regions to the heart for purification.

The blood is composed of fluid plasma and cellular parts. The largest percentage of cells in the blood are *erythrocytes*, that is red blood cell, consisting of the protein *hemoglobin* surrounded by a plasma membrane. The property of hemoglobin is to bind oxygen with its iron atoms, so that most of the oxygen needed for the metabolism is transported in the blood to the tissues. The differences in the blood's properties carried by veins and arteries are reflected in the kind of hemoglobin that composes the blood itself: veins contains deoxygenated hemoglobin (Hb) linked to oxygen-depleted blood, while the blood carried by the arteries contains oxygenated hemoglobin (HbO2).

The technique used to capture the vein pattern later used for biometric recognition exploits the differences in the absorption spectra of the blood and the tissues surrounding the blood vessels, mainly composed by water. The aforementioned absorption spectra are depicted in Fig. 2.1. In the range of near infra-red (NIR) the absorption coefficients of oxygenated and deoxygenated hemoglobin are higher than the water's one.

The aforementioned chemical properties are exploited in the acquisition devices used to capture the blood vessels, that are generally composed by the following modules:

- **NIR light source:** the region containing the vein pattern to be cap-

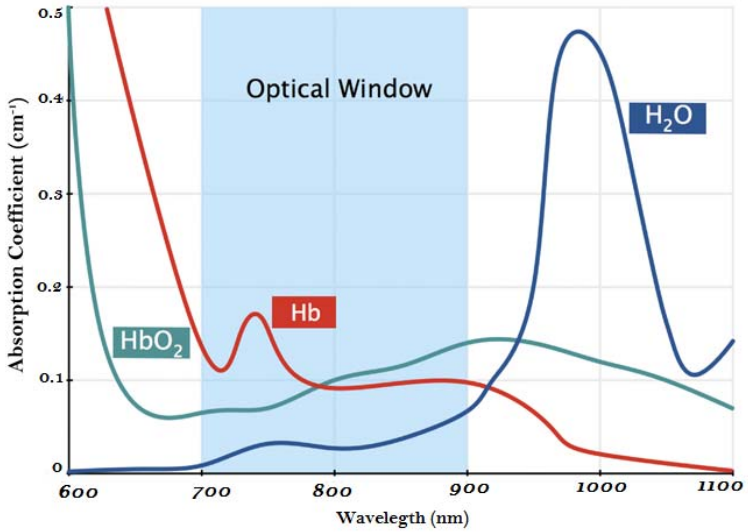


Figure 2.1: *Absorption spectra of oxi- and deoxi-haemoglobin and water.*

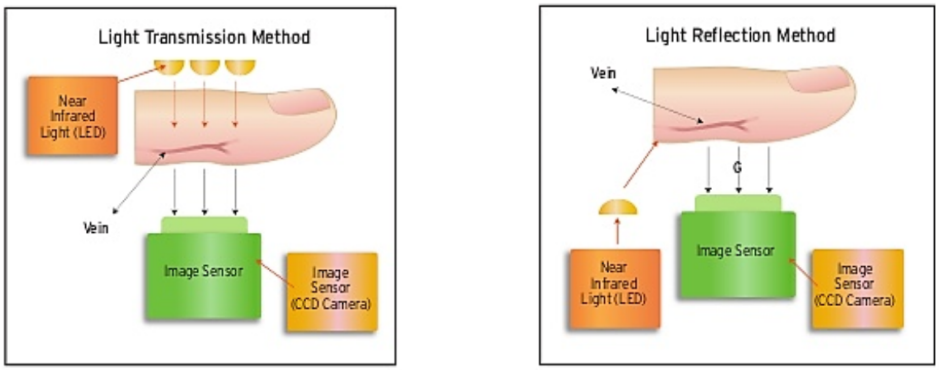
tured is illumined by a light source, that is generally a light emitting diode (LED), or an array of LEDs emitting NIR radiations. The wavelengths of the LEDs used in existing literature vary from 780 nm to 890 nm but the most common wavelength is 850 nm, which allows to reach the largest contrast between tissues and blood vessels.

- **Image sensor:** the vein pattern is captured by devices able to capture NIR image. The images are captured by the use of a charge-coupled device (CCD) or metal-oxide semiconductor (CMOS) image sensor [234]. Both sensors perform closely in terms of their ability to capture an image, but they need to be configured such that they are able to capture an image beyond the visible light spectrum.

- **IR filter:** in order to let only specific wavelength of the IR light to pass through the image sensor a IR filter is used.

- **Light diffuser:** the NIR light source needs to be evenly distributed and this can be achieved by using a light diffuser.

The NIR rays generated by the light source are absorbed by the blood and penetrate the other tissues due to the different absorption properties of the components. This leads to an acquired image where the vein structures appear as dark areas, surrounded by brighter areas related to the other tissues.



(a) (b)

Figure 2.2: *Modalities of acquisition of vein pattern: (a) transmission (b) reflection.*

There are two different methods for capturing vascular images of the hand, as shown in Figure 2.2: *light transmission* and *light reflection* [104]. In the *light transmission* modality (see Fig. 2.2a), the hand is placed between the light source and the image sensor and the light passes through the hand, reaches the sensor and the vein pattern is captured by it. The advantage of this modality is the quality of the image it is possible to achieve, since the light is shone directly through the hand and background light does not have a big influence on the result. In the *light reflection* method (see Fig. 2.2b), the light source and the acquisition sensor are placed on the same side compared to the hand position. The light is reflected by the hand and the reflected light is captured by the camera. Because of the strong reflection from the skin's surface and the shallow penetration of light under the skin, the obtained images will be characterised by a worst contrast compared to the previous modality. On the other hand, the device is usually contactless, making the device more user-friendly.

2.2 Vein Databases for Biometric Purposes

Dataset plays a crucial role in any recognition system. They are composed of a set of images acquired from various individuals using a specific scanning device. The technology used for the production of the vein acquisition devices has not been standardised, given the novelty of the biometric identifier. Every scanning device has therefore different specifications and the resulting captured images are characterised by varying quality. Different kinds of vein pattern have been analysed in literature for recogni-

Table 2.1: *Publicly available vein databases.*

Dataset	Vein Kind	No. of Subjects	No. of samples	No. Images per sample	Image Size	Year
HKPU [152]	Finger	156	2	12	513×256	2011
SDUMLA - HMT [338]	Finger	106	6	6	320×240	2011
MMCBNU_ 6000 [180]	Finger	100	6	10	640×480	2013
UTFVP [289]	Finger	60	6	4	672×380	2013
FV-USM [5]	Finger	123	4	12	640×480	2014
FVRC2016-DS0 [336]	Finger	50	1	5	512×384	2016
DS1, DS2, DS3		1000	1	5		
CASIA [96]	Palm	100	2	6	768×576	2008
PolyU [342]	Palm	250	2	12	704×578	2009
VERA [288]	Palm	110	2	5	480×680	2015
NCUT [349]	Dorsum	102	2	10	640×480	2008
GPDS100Veins CCDcylindrical [78]	Dorsum	102	1	10	-	2009
Bosphorus [339]	Dorsum	100	2	12	300×240	2011
UC3M [223]	Wrist	29	2	6	640×480	2010
PUT [133]	Wrist	50	2	12	1290×960	2011

tion purposes, namely finger vein [273], palm vein [135, 350], hand dorsal vein [175, 307] and wrist vein [223].
Many of the vein pattern datasets are not available globally for research purposes due to the confidentiality concerns. Some details of the most important publicly available databases of the different hand vein structures are reported in the following and summarised in Table 2.1.

Finger-vein

- **Hong Kong Polytechnic University (HKPU) Database [152]**: this database consists of simultaneously acquired finger vein and finger surface texture images from 156 subjects. The volunteers are both male and female, and about 93% of the subjects are younger than 30 years. The finger images, captured using a contactless imaging device, were acquired in two separate sessions with a minimum interval of one month, maximum interval of over six months. In each session, each of the subjects provided 6 image samples from index finger middle finger respectively, and each sample consists of one finger vein image

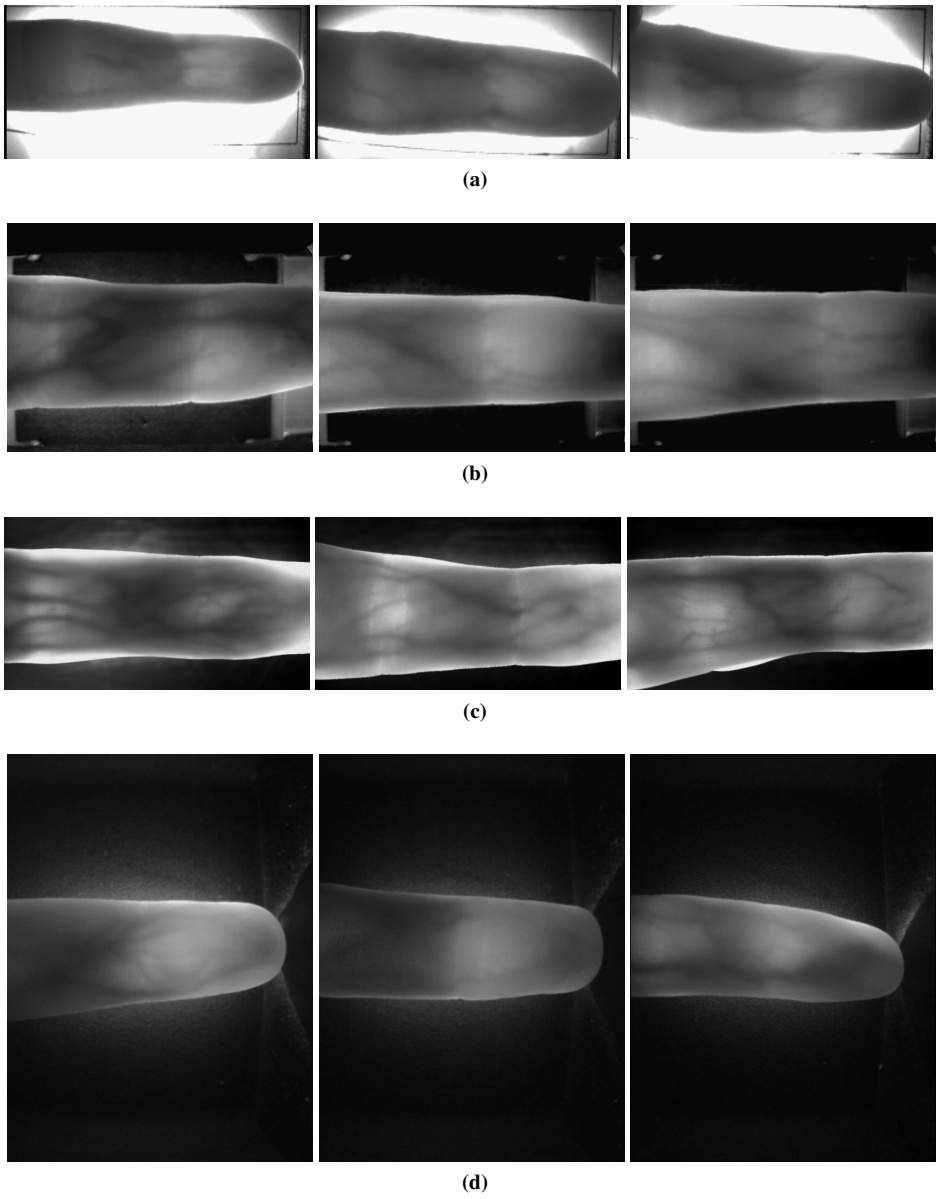


Figure 2.3: *Image samples from the (a) HKPU [152] (b) SDUMLA-HMT [338] (c) UTFVP [289] and (d) FV-USM [5] finger-vein databases.*

and one finger texture image from left hand. Therefore, each subject provided 24 images in one session. The currently available database is then composed by 6264 images stored in bitmap ("BMP") format with the resolution of 513×256. Image samples are shown in Fig. 2.3a.

- **Shandong University Homologous Multi-modal Traits Database (SDUMLA-HMT) Database [338]**: this database was collected at Shandong University, Jinan, China. 106 subjects, including 61 males and 45 females, with age between 17 and 31, participated in the data collection process, in which five biometric traits (face, finger vein, gait, iris and fingerprint) were collected for each subject. The finger vein database is composed by images of index finger, middle finger and ring finger of both hands, and the collection for each of the 6 fingers is repeated for 6 times to obtain 6 finger vein images. Data were acquired in a single session. The finger vein database is composed of 3816 images. Every image is stored in bitmap ("BMP") format with 320×240 pixels in size. Image samples are shown in Fig. 2.3b.

- **Multimedia Lab, Chonbuk Nation University (MMCBNU_6000) Database [180]**: the database consists of finger vein images from 100 volunteers. The age of volunteers in MMCBNU_6000 varies from 16 to 72 years old. During the acquisition process, each subject was asked to provide images of index finger, middle finger, and ring finger of both hands. The collection for each of the 6 fingers is repeated for 10 times to obtain 60 finger vein images for each volunteer. Hence, MMCBNU_6000 is composed of 6000 images. Each image is stored in bitmap ("BMP") format with the resolution of 480×640.

- **University of Twente Finger Vascular Pattern (UTFVP) Database [289]**: the dataset contains 1440 finger vascular pattern images in total which have been collected from 60 volunteers, the 73% of them are male and the age of the volunteers ranges from 19 to 30. Images were captured in two sessions with an average time lapse of 15 days. For each volunteer the vascular pattern of the index, ring and middle finger of both hands was collected twice at each session. This means that each individual finger has been captured four times in total. The captured images have a resolution of 672×380 pixels. The images are stored using the 8 bit grey-scale portable network graphics ("PNG") format. Image samples are shown in Fig. 2.3c.

- **University Sains Malaysia (FV-USM) Database [5]**: the images of the database were collected from 123 volunteers comprising of 83 males and 40 females. The age of the subject ranged from 20 to 52 years old. Every subject provided four fingers: left index, left middle, right index and right middle fingers resulting in a total of 492 finger classes obtained. Each finger was captured six times in one session

and each individual participated in two sessions, separated by more than two weeks' time. In the first session, a total of 2952 images were collected. Therefore, from two sessions, we obtained a total of 5904 images from 492 finger classes. The spatial and depth resolution of the captured finger images were 640×480 and 256 grey levels, respectively. Image samples are shown in Fig. 2.3d.

- **Finger Vein Recognition Competition (FVRC2016) [336] Database**: this database was collected for a competition organized for the purpose of assessing and comparing finger vein recognition algorithms. The FVRC2016 database is composed by four data sets, namely DS0, DS1, DS2 and DS3, and the collection of the data lasted from September 2015 to January 2016. Images for index and middle fingers from both hands are collected. All images are in 8-bit BMP format, 256 gray-scale and 512×384 pixel resolution. The capturing device is D501, manufactured by Beijing Yannan Tech Co., Ltd. The total number of images are 15250, with no overlapping between any two sets. DSO includes 50 fingers and 5 images for each finger. DS1, DS2 and DS3 each contains 1000 fingers and 5 images for each finger. Images in DSO were collected indoor, under full guidance and supervision. Images in DS1 were collected indoor, under slight guidance and strict supervision. Images in DS2 were collected outdoor, without guidance or supervision. DS3 consists of images captured outdoor, without guidance or supervision.

Palm Vein

- **CASIA Multi-Spectral Palmprint Image [96] Database**: this database contains 7200 palm images captured from 100 different people using a self-designed multiple spectral imaging device. All palm images are 8 bit gray-level JPEG files. For each hand, two sessions of palm images are captured. The time interval between the two sessions is more than one month. In each session, there are three samples. Each sample contains six palm images which are captured at the same time with six different electromagnetic spectra. Wavelengths of the illuminator corresponding to the six spectrum are 460nm, 630nm, 700nm, 850nm, 940nm and white light respectively. A certain degree of variations of hand postures is allowed in order to increase diversity of intra-class samples and simulate practical use. No pegs to restrict postures and positions of palms are provided in the device. Subjects are

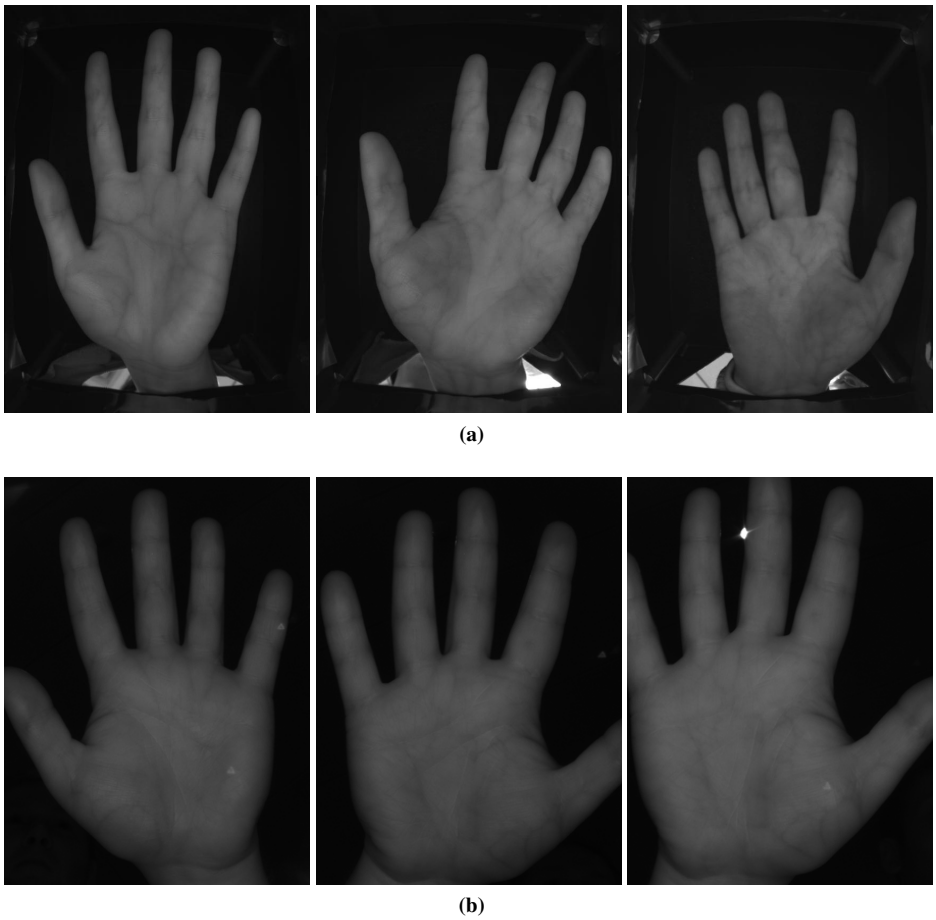


Figure 2.4: *Image samples from the (a) CASIA [96] and (b) VERA [288] palm-vein databases.*

required to put his palm into the device and lay it before a uniform-colored background. The device supplies an evenly distributed illumination and captures palm images using a CCD camera fixed on the bottom of the device. Image samples are shown in Fig. 2.4a.

- **Hong Kong Polytechnic University (PolyU) [342] Multispectral Palmprint Database**: the images of the database have been acquired through a real time multispectral palmprint capture device which can capture palmprint images under blue, green, red and near-infrared (NIR) illuminations, and has used it to construct a large-scale multispectral palmprint database. Multispectral palmprint images were collected from 250 volunteers, including 195 males and 55 females.

The age distribution is from 20 to 60 years old. We collected samples in two separate sessions. In each session, the subject was asked to provide 6 images for each palm. Therefore, 24 images of each illumination from 2 palms were collected from each subject. In total, the database contains 6,000 images from 500 different palms for one illumination. The average time interval between the first and the second sessions was about 9 days.

- **VERA [288] Database**: the VERA Palm vein database consists of 2200 images depicting human palm vein patterns. Palm vein images were recorded from 110 volunteers for both left and right hands. For each subject, images were obtained in two sessions of five pictures each per hand. This database was produced at the Idiap Research Institute in Martigny and at Haute Ecole Spécialisée de Suisse Occidentale in Sion, in Switzerland. All palmvein images have been recorded using palm vein prototype sensor developed by the Idiap Research Institute. The recordings have been performed at 2 different locations, always inside buildings with normal lightening conditions. The database is composed of 40 women and 70 men whose ages are between 18 and 60 with an average at 33. The palm vein images have a resolution of 480x680 pixel and are saved as bitmap image using a PNG format. Image samples are shown in Fig. 2.4b.

Dorsal Hand Vein

- **North China University of Technology (NCUT) [349] Dataset**: it is one of the largest public dorsal hand vein databases, and it is composed by two parts, namely Part A and Part B. There are 2040 Near-Infrared (NIR) images of dorsal hand veins in NCUT Part A, collected from the left and right hands of 102 individuals (10 samples for each hand) including 52 females and 50 males. Similarly, NCUT Part B contains 10 right and 10 left dorsal hand images from 101 individuals, which has the total number of 2020 samples. Both the two parts of the dataset have the same image resolution of 640x480. In contrast to the samples in NCUT Part B, the ones in Part A have better a quality with less noise and higher contrast due to the more expensive acquisition device used.
- **GPDS100VeinsCCDcylindrical [78] Database**: the database consists of 10 different acquisitions of 102 people in two separate weekly sessions. In the first session a short explanation of the procedure is

(a)

Figure 2.5: *Image samples from the (a) PUT [133] wrist-vein database.*

performed, while during the second session the acquisition procedure is repeated without supervision. The images in the second session thus present greater inner class variability of translation, rotation and projection distortion. The users were allowed to wear rings, which increases the variability of the hand position. Watches or bracelets were also permitted, even though they could hide a large portion of the vein pattern. The users range in age from 18 to 40. In the database, 54% of the users are males.

- **Bosphorus [339] Database**: the hand vein pattern is acquired using NIR imaging technology and reflection method. The hand vein database consists of 1200 images of left hands of 100 different people (42 female and 58 male). Each subject underwent four imaging sessions that consisted of the left hand under normal condition, after having carried a bag weighing 3 kg for 1 min, after having squeezed an elastic ball repetitively for 1 min and after having cooled the hand by holding an ice pack on the surface of the back of the hand. In addition to the left-hand images, 300 right-hand images from subjects under normal conditions were collected. Finally, in order to test the time lapse, additional data of left hands under normal conditions from 25 persons after some time ranging from 2 months to 5 months were collected. The images have 300×240 pixel size with a gray-scale resolution of 8-bit.

Wrist Vein

- **UC3M [223] Database**: the focus of this experiment was to evaluate the effect of different illumination intensities on the visibility of veins. For each of the 29 users, 6 images were taken for each hand under three different illumination settings. This results into 348 images in total.

- **PUT Wrist Vein Pattern [133] Database**: data was acquired from both hands of 50 students, with means it has a 100 different patterns for wrist region. Pictures ware taken in 3 series, 4 pictures each, with at least one week interval between each series. A construction allowing to place palm and wrist in comfortable way was used to help positioning the hand during the image acquisition. Images in database have 1280x960 resolution and are saved as 24-bit bitmap. Image samples are shown in Figure 2.5a.

2.3 Feature Extraction and Matching: State-of-the-Art

The feature extraction task is one of the most challenging tasks of every biometric systems. Given an image representing the vein pattern, the state of the art approaches aiming to extract discriminative characteristics from the structure of blood vessels can be categorised into five groups:

- **segmentation-based methods:** since veins can be seen as a network of dark lines on a brighter background, the vein pattern can be segmented through geometric-based algorithms and information as line shapes, minutiae or point information are taken into account as representative features of the users. Generally, the region of interest (ROI) where the vein pattern is contained is extracted, and some alignment steps are performed in order to reduce the errors due to misalignment during the acquisition step. This category of feature extraction methods suffers from negative impact on performance linked to image quality, as well as poor ability in discrimination of scaling, displacement and rotation variations;

- **subspace-based methods:** subspace coefficients computed though subspace-based methods such as principal component analysis (PCA), independents component analysis (ICA), linear discriminant analysis (LDA) or non-negative matrix factorisation (NMF) can be considered as discriminative features;

- **local-invariant-based methods:** local invariant features are extracted directly from the original vein image. This kind of approaches is inspired by techniques employed in computer vision, such as speeded-up robust feature (SURF) and scale-invariant feature transform (SIFT);

- **statistical-based methods:** features are computed employing statistical information, such as moments and local binary histogram, extracted from the vein images. These techniques can be divided into

local and global methods. Global- and local-statistics-based methods differ in their behaviour in terms of invariance to scaling, translation and rotation. Specifically, global statistical-based methods are invariant to the aforementioned variations, while local statics methods are not;

- **deep learning-based methods:** deep neural networks are exploited to learn discriminative representations from vein patterns and perform the classification task. A DNN consists of a sequence of processing layers and can be exploited as feature extractor or classifier module in a vein-based biometric system.

In Sections 2.3.1 - 2.3.4 an overview of the state-of-the-art approaches for the different kind of vein pattern are presented. A detailed overview of deep learning-based methods for vein recognition in presented in Section 2.4.

2.3.1 Finger Vein

As a convenient biometric recognition technology, finger vein recognition has attracted extensive attention over the past few years. Table 2.2 reports a summary of the recent state-of-the-art finger-vein-based biometric systems, briefly described in the following.

Geometry-based approaches

In [304] a Weber local descriptor (WLD) with variable curvature Gabor filters is proposed for finger vein recognition. First, the differential excitation operator in the original WLD is improved by adding directional information, then variable curvature Gabor filters are introduced to extract finger vein features that can simultaneously reflect the directional information and the curvature of the finger veins. The normalised correlation coefficient (NCC) scale matching score is employed to measure the similarity between finger vein features. Gabor filters with morphological processing have been exploited in [152] for feature extraction, with XOR-based similarity scores used for finding similarity between images. Gupta *el al.* [90] used a fusion strategy named variational approach to combine enhanced vein images obtained from both multi-scale matched filtering and line tracking. Similarity scores have been obtained by first registering the two vein images to be compared and then computing the number of overlapping binary pixels between them.

Liu *et al.* [178] used singular value decomposition (SVD) for minutiae extraction, and local extensive binary pattern (LEBP) for removing false pairs. Templates have been compared through Euclidean and Hamming distances. Ong *et al.* [217] proposed a two-stage multi-instance finger-vein identification system, combining minutiae features extracted from multiple instances of finger veins. A genetic algorithm (GA) [191] has been used to select the most reliable minutiae points from the feature point pool-set. A K-modified Hausdorff distance (k-MHD) [70] has been employed to evaluate the closet point set of two minutiae templates for comparison.

In [136] used geometry-based features and tested different feature level fusion techniques in order to increase the performance of the system. Banerjee *et al.* [10] used affine-registration-based template matching (ARTeM) algorithms.

Subspace-based approaches

Yang *et al.* [331] applied $(2D)^2$PCA to extract features of finger veins, later classified through a KNN classifier. Besides, the SMOTE technology is adopted to solve the class-imbalance problem. Van *et al.* [299] used the modified finite Radon transformation (MFRAT) [128] for discriminant orientation feature extraction and GridPCA [298] for further redundant information removal. In the matching stage, enlarging-training-set (ETS)-based comparisons [128] and Euclidean distances have been calculated. Qui *et al.* in [242] used dual-sliding window localisation and pseudo-elliptical transformation, with a two-dimensional principal component analysis (2D-PCA) used to project the transformed image for feature extraction. Euclidean distance has been used for measuring similarity between training and testing images. Bakhtiar *et al.* [5] enhanced finger-vein images using modified Gaussian filter (MGF) [166] and then correcting the image displacements. Band-limited phase only correlation (BLPOC) [286] has been used for measuring the similarity between registered and test images. In [333] authors used anatomy-structure-analysis-based vein extraction (ASAVE) and elastic matching.

Local-invariant-based approaches

In [326], the authors tested feature-component-based extreme learning machines (FC-ELMs). Features have been extracted by a guided filter using the eight block-based average absolute deviation (AAD) directional features. An ensemble component-based ELM network (EC-ELM) has been employed for final decision.

Table 2.2: *State-of-the-art works about finger-vein based biometric system.*

Paper	Database		Biometric System		Performance
	Name	# Classes	Feature Extraction	Matching	
Wang *et al.* [304]	HKPU SDUMLA	156 106	WLD + Gabor Filters	Normalized correlation coefficients	EER = 0.64% EER = 0.79%
Kumar *et al.* [152]	HKPU	105	Gabor filter with morphological processing	X-OR based similarity score	CIR=90.08%
Gupta *et al.* [90]	HKPU	105	Variational approach for vein extraction	Overlapping pixels between registered and binarized templates	EER = 4.47%
Liu *et al.* [178]	HKPU	156	SVD based minutiae extraction LEBP based false removing	Fusion of Euclidean and Hamming Distance	CIR=95.71%
Ong *et al.* [217]	SDUMLA	106	Minutiae	Genetic algorithm & k-modified Hausdorff distance (k-MHD)	CIR = 99.7%
Kauba *et al.* [136]	UTFVP	60	Different feature level fusion	Correlation based comparison	EER = 0.19%
Yang *et al.* [331]	Private	80	$(2D)^2$PCA	KNN	CIR = 99.17%
Banerjee *et al.* [10]	SDUMLA	106	Images, after Fuzzy contrast enhancement + CLAHE + directional dilation (DD)	Affine registration based template matching algorithm (ARTeM)	CIR=90.72%
Van *et al.* [299]	SDUMLA	106	MFRAT [128] & GridPCA	Euclidean distance	CIR=95.67%
Qiu *et al.* [242]	SDUMLA FV-USM	106 123	Dual-sliding window localization + Pseudo-elliptical transformer + 2D-PCA	Euclidean distance	CIR=97.61% CIR=97.02%
Asaari *et al.* [5]	FV-USM	123	Modified Gaussian Filter (MGF) enhanced & displacement corrected images	Band Limited Phase Only Correlation (BLPOC)	EER = 2.34%
Yang *et al.* [333]	HKPU SDUMLA	105 106	Anatomy Structure Analysis based Vein Extraction(ASAVE)	Elastic Matching	EER = 0.38% EER = 1.39%
Xie *et al.* [326]	SDUMLA	106	Block-based average absolute deviation (AAD) features	Ensemble component-based extreme learning machines (EC-ELM) network	CIR=97.76%
Lu *et al.* [179]	SDUMLA	106	Polydirectional Local Line Binary Pattern	Histogram intersection	CIR=99.21%
Rosdi *et al.* [263]	Own	51	LLBP	Hamming Distance	EER = 1.78%
Xi *et al.* [323]	HKPU	105	Discriminative binary codes (DBC)	SVM	EER = 1.44%

Statistical-based approaches

In [179], the authors proposed a polydirectional local line binary pattern (PLLBP) method for extracting vein line patterns, and employed histogram intersection to measure the similarity between two histograms. Xi *et al.* [323] proposed a discriminative binary codes (DBC) learning method, building subject relation graph to capture correlations among subjects and, based on that, generating binary templates according to the graph transform. Eventually, support vector machines (SVMs) have trained as code learners for each bit. Rosdi *et al.* [263] proposed a texture descriptor called local line binary pattern (LLBP) utilized as feature extraction technique. The similarity between the extracted binary codes and the enrolled codes is measured using Hamming Distance (HD).

2.3.2 Palm Vein

Table 2.4 reports a summary of the recent state-of-the-art palm-vein-based biometric systems, briefly described in the following.

Geometry-based approaches

In the context of segmentation-based method, many works have been proposed in literature for palm vein recognition. Han and Lee [95] applied texture-based feature extraction techniques based on Gabor filters to palm vein, with the obtained representation encoded in a bit-string representation, later employed in a matcher based on Hamming distance. An adaptive 2D Gabor filter-based approach has also been proposed by Ma *et al.* [182], which studied a novel template matching algorithm referred to as the minimum normalised Hamming distance. In [346] Zhang *et al.* have employed multi-scale matched filters to extract features and matching scores are obtained through the Hamming distance. Also Chen *et al.* [42] have exploited multi-scale matched filters for palm vein feature extraction, with matching scores generated using the iterative closest point (ICP) algorithm.

Subspace-based approaches

Some works about subspace learning applied to palm-vein-based biometric recognition have been proposed in the last decade. Wang *et al.* [305] used a Laplacian palm representation on fused palm vein and palm print images, exploiting locality-preserving projection (LPP) subspace, then using a k-NN classifier for matching purposes. Zhou and Kumar [350] have designed

Table 2.4: *State-of-the-art works about palm-vein based biometric system.*

Paper	Database		Biometric System		Performance
	Name	# Classes	Feature Extraction	Matching	
Han *et al.* [95]	Private	207	Gabor filter	Hammin distance	CIR = 99.38%
Ma *et al.* [182]	CASIA	100	Gabor filter	Hamming distance	EER = 0.12%
Zhang et al. [346]	Own	24	Matched filters	Hamming distance	EER = 4.00%
Chen *et al.* [42]	?	?	Matched filters	ICP	?
Lee et al. [169]	Own	207	Adaptive Gabor filter	Normalized Hamming distance	EER = 0.44%
Wu et al. [321]	Own	256	Gaussian filter bank	Normalized Hamming distance	EER = 0.518%
Wirayuda [319]	CASIA (left hands)	100	Minutiae features	Weighted Euclidean distance	CIR = 90.87%
Cancian et al. [36]	Own	24	2D Gabor filters	Hellinger distance	EER = 1.84%
Wang *et al.* [305]	Private	120	"Laplacianpalm"	Wavelet-based	CIR = 99.00%
Zhou *et al.* [351]	CASIA	100	Principal direction	Bin matching	CIR = 96.00%
	PolyU	500			CIR = 96.67
	Private	102			CIR = 97.71
Zhou et al. [350]	CASIA	200	NMRT	Hamming distance	EER = 0.51%
			Hessian Phase		EER = 0.44%
	PolyU	500	NMRT	Hamming distance	EER = 0.004%
			Hessian Phase		EER = 0.43%
Lee [171]	Own	207	Modified $(2D)^2$LDA	Euclidian distance	CRR = 99.41%
Ladoux et al. [162]	Own	24	SIFT	Euclidian distance	EER = 0.14%
Kang et al. [134]	CASIA	200	RootSIFT	RootSIFT matching + LBP-based mismatching removal	EER = 0.996%
	Own	210	RootSIFT	RootSIFT matching + LBP-based mismatching removal	EER = 3.112%
Yan *et al.* [329]	CASIA	200	SIFT & Feature-level fusion	Bidirectional matching	EER = 0.16%
	Private	105			EER = 0.73%
Mirmohamadsadeghi *et al.* [203]	CASIA	200	LBP	Histogram Intersection	CIR = 93.20%
			LDP		CIR = 97.00%
Kang et al. [135]	CASIA (left hands)	100	Mutual foreground LBP	Chi-square distance	EER = 2.53%
			Mutual foreground LBP	Chi-square distance + SVM score fusion	EER = 0.27%
Tome et al. [288]	VERA	220	LBP	Histogram intersection	EER = 3.75%
Pratiwi et al. [237]	CASIA	200	LBP Rotation Invariant	Cosine Distance	CIR = 96.00%
					EER = 11.70%

an approach based on the Hessian phase that extracts the structure of the vessels by analysing the second-order derivatives of the normalised palm vein images, and proposed a matching approach based on the Hamming distance. Zhou *et al.* [351] proposed a classification for palm–vein identification based on principal direction features. In the registration process, the Gaussian-Radon transform is adopted to extract the orientation matrix and then compute the principal direction of a palm–vein image based on the orientation matrix. The database can be classified into six bins based on the value of the principal direction. In the identification process, the principal direction of the test sample is first extracted to ascertain the corresponding bin. One-by-one matching with the training samples is then performed in the bin.

Local-invariant-based approaches

Among the local-invariant-based methods proposed for palm vein recognition, Ladoux *et al.* [162] have used SIFT descriptors as users' discriminative features, and Euclidian distance as employed similarity measure. A variant of the SIFT method, namely RootSIFT, has been proposed by Kang *et al.* in [135], with a hierarchical mismatching removal algorithm based on neighbourhood searching and local binary pattern (LBP) histograms adopted to improve the accuracy of feature matching. Yan *et al.* [329] presented palm vein recognition method based on local invariant features, namely SIFT features, based on multi-sampling and feature-level fusion strategy. RootSIFT and bidirectional is adopted to match to establish identification.

Statistical-based approaches

Statistical-based methods have been deeply exploited in the field of palm-vein-based biometric systems. Mirmohamadsadeghi *et al.* [203] have used LBPs and local derivative patterns (LDP) operators to extract features and histogram intersection to obtain the matching scores. Kang *et al.* [134] have proposed an approach based on mutual foreground LBP for palm vein identification, that is, the gradient-based maximal principal curvature (MPC) algorithm, with a k-means method utilised for texture extraction, and an LBP matching strategy adopted for similarity measurements between the mutual foreground of gray-scale images.

2.3.3 Hand Dorsal Vein

In general,dorsal hand veins consist of some significant textures and a lot of minutiae similar to the ridges and branches of palmprint. These features offer stable, unique and reliable biometric for personal identification. Table 2.6 reports a summary of the recent state-of-the-art hand-dorsa-vein-based biometric systems, briefly described in the following.

Geometry-based approaches

Lee *et al.* [168], proposed an adaptive Gabor filter method to extract the dorsal hand vein patterns and encode the vein features in bit string representation, called VeinCodes. The similarity of two VeinCodes is measured by normalised Hamming distance. Lee *et al.* [170] proposed a feature extraction approach that uses the directional filter bank to extract the line-based features from a dorsal hand image called minimum directional filtering response (MDFR). In addition, a robust minimum directional code (MDC) is proposed to encode the dorsal hand vein features into binary code, assigned by the MDFR. A normalised Hamming distance is adopted for the similarity measurement in MDC matching. In the study proposed by Lajevardi *et al.* [164], a modular feature-based technique, called biometric graph matching (BGM), is used to create a spatial graph template from the hand vein pattern which represents the feature locations and the vascular connections between them. BGM uses an efficient graph-based registration process followed by an inexact graph matching algorithm to match the hand vein template. A global graph model which takes into account both shape cues, that is the minutiae of the vein network and their connecting lines, and the holistic texture feature of the patch around each vertex (i.e. its PCA coefficients) was also proposed in [345].

Kumar *et al.* [150] first investigated the extraction and matching of hand vein structure using the key point triangulation. Wang *et al.* [310] proposed a technique utilising the minutiae features extracted from the vein patterns for recognition, which include bifurcation points and ending points. The modified Hausdorff distance (MHD) algorithm is proposed to evaluate the discriminating power of these minutiae for person verification purposes. Chuang [47] proposed a local feature-based vein representation method based on minutiae features from skeleton images of venous networks of the dorsum of the hand. A dynamic pattern tree (DPT) approach is proposed for matching. Huang *et al.* [114] tested a novel local feature-based approach to hand-dorsa vein recognition via matching keypoints localised

through quantities of first to third order gradients closely related to differential geometry.

Subspace-based approaches

Yuksel *et al.* [339] extracted features from image hands using both appearance-based methods and methods reflecting the hand geometry. The appearance-based methods consider the dorsal view of the hand in the near-infrared band and project it to subspaces via independent component analysis (ICA) or non-negative matrix factorisation (NMF) methods. The geometry-based method delineates the veins via skeletonisation and then describes the vein structure as a graph, namely line edge map (LEM). Cosine distance is used to compare ICA and NMF features, while Line segment Hausdorff distance (LHD) is exploited when LEM features are matched. Eventually, score- and decision-level fusion techniques are tested.

Local-invariant-based approaches

Wang *et al.* [313] exploited keypoints extracted from the segmented binary vein images by using SIFT as features and keypoints are matched through the cosine similarity measure. Eventually, in order to reduce incorrect matches, a restriction based on the minimum Euclidean distance is applied. Huang *et al.* [115] proposed a key-point generation pattern, namely centroid-based circular key-point grid (CCKG), localising a certain number of points on the dorsal hand for the following SIFT feature extraction. The matching process makes use of multi-task sparse representation classifier (MtSRC). Wang *et al.* [309] feature extraction algorithm based on polar harmonic transforms(PHT) to realise the hand vein recognition. PCET moments are taken into account as features and the matching method is improved Hausdorff distance (MHD).

Statistical-based approaches

Wang *et al.* [308] proposed a quality-specific vein recognition system based on rotation and affine invariant discriminative LBP (DLBP). An improved LBP coding optimized by Fisher discriminant criterion is realised for exploiting more discriminative and stable binary patterns, followed by different vein feature extraction strategies, with characteristics of being robust to noise, rotation and affine change process, regarding the quality of vein images. Finally, after obtaining uniform histogram vectors of vein images, the matching strategy based on improved Chi-square distance is proposed

Table 2.6: *State-of-the-art works about hand dorsa-vein based biometric system.*

Paper	Database		Biometric System		Performance
	Name	# Classes	Feature Extraction	Matching	
Lee *et al.* [168]	Private	308	Adaptive Gabor filter	Hamming Distance	CIR = 99.27% EER = 0.51%
Lee *et al.* [170]	Private	214	MDC matrix	Normalized Hamming distance	EER = 0.54%
Lajevardi *et al.* [164]	Private	123	Graph	Graph matching	EER = 1.00%
Kumar *et al.* [150]	Private	100	Minutiae	Minutiae Triangulation	EER = 1.14%
Wang *et al.* [310]	Private	47	Minutiea	MHD	EER = 0.00%
Huang *et al.* [114]	NCUT	102	Multi-source keypoints	SIFT matching	CIR = 99.61%
Zhang *et al.* [345]	NCUT	102	Graph	Graph matching	CIR = 98.82%
Yuksel *et al.* [339]	Bosphorus	100	ICA	Cosine distance	CIR = 98.66%
			NMF	Cosine distance	CIR = 96.33%
			LEM	LHD	CIR = 98.00%
Wang *et al.* [313]	Private	204	SIFT	Cosine distance	CIR = 99.29%
Huang *et al.* [116]	NCUT	102	LBP	Chi-square distance	CIR = 97.74%
			BC	Chi-square distance	CIR = 97.05%
			Graph	Graph matching	CIR = 92.72%
			Graph + LBP + BC	Score fusion	CIR = 99.21%
Wang *et al.* [309]	Private	50	PCET	MHD	EER = 0.86%
Wang *et al.* [308]	Private	500	DLBP	Chi-square distance	EER = 0.06%
Huang *et al.* [115]	NCUT Part A	102	CCKG + SIFT	MtSRC	EER = 0.49%
	NCUT Part B	101			EER = 0.61%
Chuang [47]	Private	308	Minutiae	DPT	CIR = 99.76% EER = 0.63%
Trabelsi *et al.* [291]	Bosphorus	100	CDSDP	FMNN	EER = 0.01%
Vairavel *et al.* [297]	NCUT	102	LBP	K-NN	CIR = 96.00%
			HOG		CIR = 98.00%
			WLD		CIR = 99.00%

to calculate the similarity of each feature vector. In the work proposed by Vairavel *et al.* [297] the dorsal veins are extracted by using local binary pattern (LBP), Weber local descriptor (WLD) and histogram of oriented gradients (HOG) with K-NN classifier.

Huang *et al.* [116] integrated both local and holistic analyses. The authors proposed a method for dorsal hand vein recognition, exploring and combining three kinds of features, namely LBP, binary coding (BC), and graph, attributed to the category of local texture, local shape, and global shape, respectively. The LBP and BC features encode the patterns of local texture and shape variations of dorsal hand vein samples, and graph describes their holistic geometric configurations.

Trabelsi *et al.* [291] presented a direction based local descriptor called circular difference and statistical directional patterns (CDSDP) and applied it to dorsal hand vein recognition. In CDSDP, the local gradient orientation information is coded into a weighted number that presents directional information of vein pattern. The classification phase is performed using a feedforward multilayer neural network (FMNN), that is an artificial neural network (ANN) architecture.

2.3.4 Wrist Vein

Though wrist vein is a promising biometric characteristics, there exist very few works on wrist vein recognition system. The most significant details of the state-of-the art works about wrist vein based biometric systems are outlined below and their main features are summarised in table 2.8.

Geometry-based approaches

Akhloufi *et al.* [2] fist demonstrated that the wrist vein pattern can be utilized for biometric user authentication purposes, exploiting a geometry-based algorithm relying on morphological filtering. Most of the works later proposed in literature in the field of wrist vein biometric rely on geometry-based methods. Minutiae extracted for the wrist vein patterns have been largely exploited. Uriarte *et al.* [294], Hartung *et al.* [101, 102] and Pflug *et al.* [225] proposed minutia, feature- and spectral minutia-, chain code fusion-based wrist vein recognition systems, respectively. Additionally, the work in [103] extracted wrist vein features with minutia cylinder codes. A method to extract vein minutiae and transforming them into a fixed-length vector that represents translation, rotation and scale invariant features was proposed in [100]. Uriarte *et al.* [295] explored the possibility of performing wrist vein biometric recognition using crossing number of minutiae.

In [131] and [133], different segmentation techniques relying on the discrete fourier transform (DFT) were, while Gaussian filter was used as feature extraction method in [132].

Mohamed *et al.* [208] proposed a multibiometric system that fuse information from left and right wrist vein patter. The authors used binary images extracted from the pre-processed wrist vein pattern images as features and exploited the Dubois and Parade t-norm based score-level-fusion rules demonstrating the positive impact on recognition performance of the aforementioned fusion approaches.

Subspace-based approaches

In [153], Kurban *et al.* proposed a subspace learning-based approach where FFT-based low pass filtering and PCA are used to extract features from wrist vein images, later used to feed a radial basis function (RBF) network, multi-layer perceptron (MLP) and SVM neural networks.

Local-invariant-based approaches

An approach based on scale invariant feature transform (SIFT) has been proposed by Fernandez *et al.* [77]. The authors presented a low cost, mobile wrist vein authentication system and evaluated the proposed method on the captured wrist vein dataset.

Statistical-based approaches

Between the statistical-based methods, wrist feature extraction based on the dense local binary pattern (D-LBP) has been proposed by Das *et al.* [56] and Support Vector Machines (SVM) are used for classification. Raghavendra *et al.* [246] evaluated the performance of nine different feature extraction methods, belonging to both geometry-based and statistical-based approaches. More in details the considered local features maximum curvature points (MCP) [205] and multi-scale match filter. Seven different global feature extraction schemes are also considered, namely: sparse representation classifier (SRC), local binary patterns (LBP), local phase quantization (LPQ), histogram of gradients (HOG), steerable pyramids, LBP variance (LBPV) and log Gabor (LG) filters. The authors demonstrated that the best results can be achieved when the LG-SRC combination is applied.

Table 2.8: *State-of-the-art works about wrist-vein based biometric system.*

Paper	Database		Biometric System		Performance
	Name	# Classes	Feature Extraction	Matching	
Akhloufi *et al.* [2]	Private	-	Morphological filtering	Distance	CIR = 80.10%
Uriarte *et al.* [294]	UCM3	29	Relative location & orientation of minutiae	Minutiae comparison	EER = 2.27%
Hartung *et al.* [101]	UCM3	29	Spectral Minutiae	SML Correlation SML Fast Rotate	EER = 6.13% EER = 5.90%
Hartung *et al.* [102]	UCM3	29	Chain codes using spacial & orientation properties	Point parallelism	EER = 1.38%
Pflug *et al.* [225]	UCM3	29	Chain codes using spacial & orientation properties	Local Error Distance	EER = 0.67%
Hartung *et al.* [103]	UCM3	29	Minutia Cylinder-Codes	Correlation	EER = 0.31%
Kabacinski *et al.* [131]	PUT	95	DFT	2D Correlation	EER = 3.51%
Uriarte *et al.* [295]	UCM3	29	Crossing number-based minutiae extraction	Minutiae comparison	EER = 15.75%
Hartung *et al.* [100]	UCM3	29	Spectral minutiae	Correlation	EER = 4.48%
Kabacinski *et al.* [133]	PUT	50	Freq. high-pass filtration & local minima analysis	2D Correlation	EER = 2.19%
Kabacinski *et al.* [132]	PUT	50	Gauss filtering	2D Correlation	EER = 3.80%
Mohamed *et al.* [208]	PUT	50	Binary vein images	Debois-Parade t-norm score-level fusion	EER = 0.00%
Kurban *et al.* [153]	Private	34	Enhanced images	RBF MLP SVM	CIR = 94.11% CIR = 94.11% CIR = 96.07%
Fernandez *et al.* [77]	Pivate	30	SIFT	Euclidian Distance	EER = 0.15%
Das *et al.* [56]	PUT	50	D-LBP	SVM	EER = 0.79%
Raghavendra *et al.* [245]	Private	50	9 local and global features	Correlation, SRC	EER = 1.63%

2.4 CNN and Vein-Pattern-Based Biometric Applications: State of the Art

The increasing interest in vein-based recognition arising in both academia and industry communities is entailing a rapid growth of proposed techniques in the field of extraction of features from vein images. Anyway, many of the current state-of-the-art techniques, detailed in Section 2.3, despite the fact that they are able to provide relatively good performance, they are strongly dependent upon the quality of the analysed finger-vein images. Besides, they are very sensitive to rotation and translation of the hand during the acquisition process, requiring the user to place the hand still on the acquisition device for some seconds, reducing the convenience of the device.

In the recent past, there has been an increase in the use of deep learning techniques in the field of biometric recognition, due to the good recognition performance they achieve. In this Section, an overview of the most relevant papers about the application of deep learning methods in the field of vein-based biometrics is presented. The related details are summarised in Table 2.10.

A deep learning approach applied to a finger-vein-based biometric identification system has been first proposed by Radzi *et al.* [243]. The structure of the employed network relies on the one proposed in [276], with the CNN fed with binary images obtained by thresholding the original vein pictures. A more recent work on finger-vein-based identification using CNNs is the one proposed by Das *et al.* [62], where stable and highly-accurate performance is achieved while dealing with finger-vein images of different quality. Hong *et al.* [108] have designed a finger-vein-based verification system exploiting a pre-trained model of VGG-16 [278]. The pre-trained network model is used for fine-tuning, having the difference between two finger-vein images as input. Databases with different image quality are taken into account. A deep CNN (D-CNN) architecture, inspired by the VGG-16 model, has also been exploited for finger-vein based biometric verification by Huang *et al.* [118]. The modified CNN architecture is fed with a two-channel image resulting from the merging of two templates.

An approach for finger-vein-based biometric verification using CNN and supervised discrete hashing (SDH) has been proposed in [325] where different CNN architectures, such a light CNN (LCNN) and a modified version of the VGG-Net-16, have been fed with pairs of vein images. The SDH scheme is also investigated to improve the performance and to reduce the template size. Fang *et al.* [73] have exploited a lightweight deep-learning

Table 2.10: *State-of-the-art works about applications of deep learning algorithms in the field of vein-based biometric recognition systems.*

Paper	Biometric Identifier	Database		CNN Features		Performance
		Name	#Classes	Reference	Aim	
Radzi et al. [243]	Finger-vein	Own	300 (50 users)	[276]	Biometric Identification	CIR = 100%
Das et al. [62]	Finger-vein	HKPU [152]	302 (156 users)	-	Biometric Identification	CIR = 95.32%
		FV-USM [5]	492 (123 users)			CIR = 97.53%
		SDUMLA [338]	636 (106 users)			CIR = 97.48%
		UTFVP [289]	360 (60 users)			CIR = 98.33%
Hong et al. [108]	Finger-vein	Own (Good Quality)	120 (20 users)	VGG-16 [278]	Biometric Verification	EER = 0.396%
		Own (Middle Quality)	198 (33 users)			EER = 1.275%
		SDUMLA [338] (Low Quality)	636 (106 users)			EER = 3.906%
Huang et al. [118]	Finger-vein	Own (Training)	300.000	VGG-16 [278]	Biometric Verification	-
		FVRC2016 - DS1 [336] (Testing)	1000			EER = 0.42%
		FVRC2016 - DS2 [336] (Testing)	1000			EER = 1.41%
		FVRC2016 - DS3 [336] (Testing)	1000			EER = 2.14%
Xie et al. [325]	Finger-vein	HKPU [152]	302 (156 users)	LCNN [322]	Biometric Verification	EER = 0.11%
				VGG-16 [278]		EER = 0.12%
Fang et al. [73]	Finger-vein	MMCBNU_6000 [180]	600 (100 users)	[340]	Biometric Verification	EER = 0.10%
		SDUMLA [338]	636 (106 users)			EER = 0.47%
Jalilian et al. [126]	Finger-vein	UTFVP [289]	360 (60 users)	U-net [262]	Biometric Verification	EER = 4.53%
						EER = 0.41%
				RefineNet [176]		EER = 2.95%
						EER = 1.80%
		SDUMLA [338]	636 (106 users)	SegNet [6]		EER = 2.54%
						EER = 2.17%
Kim et al. [140]	Finger-vein & Finger-shape	SDUMLA [338]	636 (106 users)	ResNet-50 [105], ResNet-101 [105]	Biometric Verification	EER = 2.34%
		HKPU [152]	302 (156 users)			EER = 0.79%
Wang et al. [306]	Hand-dorsal Vein	Own	200 (200 users)	VGG-16 [278]	Biometric Verification	EER = 0.06%

Continuation of Table 2.10

Paper	Biometric Identifier	Database Name	Database #Classes	CNN Features Reference	CNN Features Aim	Performance
Zhang et al. [344]	Palm-Vein	Own	600 (300 users)	Inception ResNet v1 [284]	Biometric Verification	EER = 2.74%
Raghavendra et al. [246]	Finger-vein	Finger video [247]	300 (100 users)	AlexNet [3]	PAD - Inkjet printed artefact	APCER = 3.48%
					PAD - Laserjet printed artefact	APCER = 0.00%
		Finger images [248]	300 (100 users)		PAD - Inkjet printed artefact	APCER = 3.20%
					PAD - Laserjet printed artefact	APCER = 0.40%
Sajjad et al. [268]	Palm Vein	VERA Palm Vein [288]	220 (110 users)	GoogLeNet [285]	Spoofing Attack Detection	AAC = 99.5%
Qin et al. [239]	Finger-vein	FV-USM [5]	492 (123 users)	AlexNet [3]	Feature extraction	EER = 0.7%
		HKPU [152]	302 (156 users)		for quality assessment	EER = 1.5%
Qin et al. [240]	Finger-vein	FV-USM [5]	492 (123 users)	-	Finger-vein segmentation	EER = 1.42%
		HKPU [152]	302 (156 users)		and recovery	EER = 2.70%
Qin et al. [241]	Finger-vein	FV-USM [5]	492 (123 users)	-	Finger-vein image	EER = 0.80%
		HKPU [152]	302 (156 users)		quality assessment	EER = 2.33%

framework for finger-vein verification. Mini-ROIs from the original image are extracted, based on the evaluation of the adopted network, and both the original image and the mini-ROI are integrated through a two-stream network. Jalilian *et al* [126] have used three different fully CNN (FCN) architectures, inspired by the U-net [262], RefineNet [176] and the SegNet [6] networks, in order to extract the finger-vein patterns from NIR finger images. The problem of efficient training and configuration settings for the employed networks has also been taken into account, by training the considered FCN architectures with a varying number of manually- and automatically-generated labeled images. Wang *et al.* [306] have proposed a hand-dorsal vein recognition system constructed by adopting the VGG-16 model, pre-trained on a large-scale database, as a universal feature descriptor. A task-specific selective convolutional features (SCF) model, based on spatial weighting, has been proposed to obtain the discriminative features, and spatial pyramid pooling (SPP) is introduced to obtain the final feature representation.

Kim *et al.* [140] have proposed a multimodal biometric based on finger-veins and finger shapes exploiting CNNs to extract features from the acquired images and compute the matching scores. More in detail, the authors have compared the performance of different CNN configurations, namely ResNet-50, ResNet-101 [105], when fed with the finger-vein image or the spectrogram of the finger shape's ROI. They have shown that it is possible to improve the recognition performance by applying score-level fusion approaches. The ResNet architecture has also been exploited by Zhang *et al.* [344] in the framework of a palm-vein-based verification system. The authors applied a modified version of the Inception ResNet-v1 DNN to extract features, later used for recognition purposes.

CNNs have been exploited also for other applications in the field of biometric systems, although not explicitly for biometric classification purposes. One of the possible usages proposed by Raghavendra *et al.* [246] is presentation attack detection (PAD) in finger-vein based biometric systems. The authors exploited a D-CNN inspired by Alex-Net [3] and they performed a fine-tuning of the model with presentation attack samples in order to classify between bona-fide and artifact samples. Sajjad *et al.* [268] tested CNN-based models trained to detect spoofing in a multimodal biometric system relying on fingerprint, face and palm-vein. The employed CNN model is inspired by GoogLeNet [285] and it used to perform features extraction for anti-spoofing. Qin *et al.* exploited DNNs for quality assessment of finger vein images to be used in a biometric system [239, 241]. In [239] the authors trained a DNN in order to extract feature represen-

tation from finger-vein images and the representation taken from the last hidden layer of DNN and input to P-SVM for finger-vein image quality assessment. In the work proposed in [241], the authors exploited a DNN to automatically label low- and high-quality images and the impact on recognition performance of considering only images classified with high quality by the CNN, while using state-of-the art line-based algorithms for feature extraction [152], has been studied. Qin *et al.* [240] have also proposed the application of a deep learning model in order to segment vein pixels from the background in finger vein images and a fully convolutional network (FCN) for recovering vein patterns in the extracted patterns.

CHAPTER 3

Hand Vein Biometrics: Performance Improvement

THE INCREASING INTEREST in vein-based recognition arising in both academia and industry communities, linked to the convenience of the acquisition process and the robustness of vein-based systems against presentation attacks, is entailing a rapid growth of proposed techniques in the field of extraction of discriminative features from vein images. Anyway, many of the current state-of-the-art techniques, detailed in Section 2.3, despite the fact they are able to provide relatively good performance, they are strongly dependent upon the quality of the analysed finger-vein images. Besides, they are very sensitive to rotation and translation of the hand during the acquisition process, requiring the user to place the hand still on the acquisition device for some seconds, reducing the convenience of the device.

In order to overcome the aforementioned limitations, different approaches are analysed in this Chapter. As first solution, the impact of multibiometric systems in palm-vein recognition is studied. More in detail, a sensor-level fusion approach based on high dinamic range (HDR) techniques is proposed and compared with different biometric fusion techniques [228].

Besides, two studies sharing the exploitation of a recognition architecture based on convolutional neural network (CNN) are reported [62, 157]. More in detail, first a CNN-based-finger-vein identification system is proposed, and the capabilities of the designed network is investigated taking into account four publicly-available databases. The main purpose of this work is to propose a deep-learning method for finger-vein identification, able to achieve stable and highly-accurate performance when dealing with different kinds of finger-vein images, irrespective of their quality. Eventually, the design of a contactless acquisition architecture capable of capturing the finger-vein structure using an array of low-cost cameras, allowing to acquire finger-vein patterns while on the move is presented. Moreover, the exploitation of advantages of HDR techniques and of temporal information are studied.

In this Chapter, some possible solutions aiming to improve the performance of vein-based biometric systems and to deal with the disadvantages related to the poor quality of vein images are studied. More in detail, Section 3.1 propose a solution based on biometric fusion, while Sections 3.2 gives the detail about the study concerning the exploitation of CNNs in the field on finger-vein recognition. Eventually, the "on-the-fly" solution, merging the advantages of biometric fusion and deep learning approaches, is detailed in Section 3.3.

3.1 Palm Vein Recognition using a High Dynamic Range Approach

Several limitations affect biometric systems which operate using any single biometric trait [124]. First, noise could be present in the acquired data, due to sensor's defects or unfavourable ambient conditions, leading to incorrect match between templates. Besides, biometric data could be characterised by a very high intra-class variation, that is, the enrolment and recognition templates might be very different, resulting in problems during the matching stage. Eventually, the system can show intra-class similarities in the features set used to represent the users, restricting the discrimination capability of the biometric characteristic. The aforementioned limitations entail that unimodal biometric systems are unable to provide a high accuracy and security performances.

Several above-mentioned problems can be solved or at least their impact reduced by integrating several biometric information sources, that is implementing a multimodal biometric systems [109]. As detailed in Section 1.5, there are different modalities of fusion of the information that can

be exploited by a multimodal biometric system. Specifically, the information fusion in a biometric system can be classified into several categories, that is sensor-, feature-, score- and decision-level fusion. Performing one or a combination of the aforesaid techniques, the obtained system is able to offer many advantages over unimodal systems, such as:

- significant improvement in the overall accuracy;
- mitigation of the effect of noisy input data;
- population coverage larger than the unimodal system [2];
- greater resistance to spoofing such that they can be more robust than each corresponding uni-modal system, even in the case when all biometric traits are spoofed.

Vein-pattern images are often characterised by low contrast and poor definition, due to the subcutaneous placement of the veins. This shortcoming makes vein-related feature extraction a challenging process. In order to overcome the disadvantages due to the low contrast of the acquired vein patterns, an approach for palm vein recognition relying on high dynamic range (HDR) imaging [11, 40, 196] is proposed in [228], thus following a sensor-level fusion approach. Specifically, it is demonstrated that exploiting multiple-exposure vein images is able to guarantee better recognition performance than a baseline system relying on single-exposure acquisitions. Besides, the effects of feature-, score- and decision-level fusion approaches of a palm-vein-based biometric system are investigated in [227].

In this study, a multiple-exposure dataset is collected. Local binary pattern (LBP) and local derivative pattern (LDP) are employed to extract features from single-exposure images, raw HDR images, and tone-mapped HDR images. The obtained experimental results show that significant performance improvement can be achieved when discriminative features are extracted from HDR contents, with respect to the use of single-exposure images. Besides, better recognition performance can be achieved exploiting such methods, when compared to what can be attained with a unimodal biometric system or with the other biometric fusion techniques.

3.1.1 Vein Patterns and High Dynamic Range

It is well known that blood haemoglobin absorbs NIR light, implying that the vein pattern imaging is carried out through a NIR camera and a NIR illumination system. This latter has to be properly calibrated in order to allow the light to penetrate skin and tissues till reaching the blood vessels.

Unfortunately, vein structures are not evenly positioned under the skin with respect to the imaging and illumination devices, with the result that the obtained images may appear saturated if too much illumination power is employed, or dark if not enough illumination is used. Moreover, the transmittance of the NIR light across the different tissues of the hand is not uniform, due to the different thickness of bones and tissues. This results in veins from the thicker parts of the hand being less distinguishable compared to veins located in the thinner parts. Additionally, being standard camera sensors commonly able to handle only 8-bit images, the full luminance dynamic range cannot be sensed, thus producing low-contrast images with potential loss of details and useful information. Examples of the aforementioned issues are shown in Figs. 3.1 and 3.3. All these undesired effects impact on the quality of the captured vein pattern, leading to a degradation of the recognition accuracy in a biometric system. Thus, improving the quality of the captured vein images is a crucial task in a vein-based recognition system.

Different solutions have been proposed in order to face the aforesaid problems. Contrast enhancement techniques have been deeply investigated in literature as a solution to face the uneven illumination in the acquired vein pattern images [67, 332]. However, image enhancement is not able to recover the information loss due to either overexposure or underexposure of some regions of the picture. Therefore, several works about the adjustment of the illuminance distribution of the lighting system have been proposed [43, 327]. In detail, a uniform illumination in the acquired vein pattern image is obtained by adapting the light source during each image capture. The disadvantages of this kind of approach are that, being the light source modified each time, the system settings vary from acquisition to acquisition and the process of illumination adjustment is typically time consuming. Eventually, multimodal biometric fusion schemes have been extensively proposed in order to increase the system accuracy. Among them, multispectal image-level fusion, that is the combination of palm vein and palmprint images [97, 310], and feature-level fusion approaches [329] can be mentioned.

In order to counteract the aforementioned undesired effects, in this Section a vein pattern recognition system performing information fusion at the sensor level is proposed, relying on HDR imaging techniques able to synthesise images with dynamic range far larger than the one representable in LDR images [40, 256]. The increase of dynamic range can be generated either directly capturing HDR images exploiting specialised devices, which are usually bulky and costly, or merging a set of single-exposure images

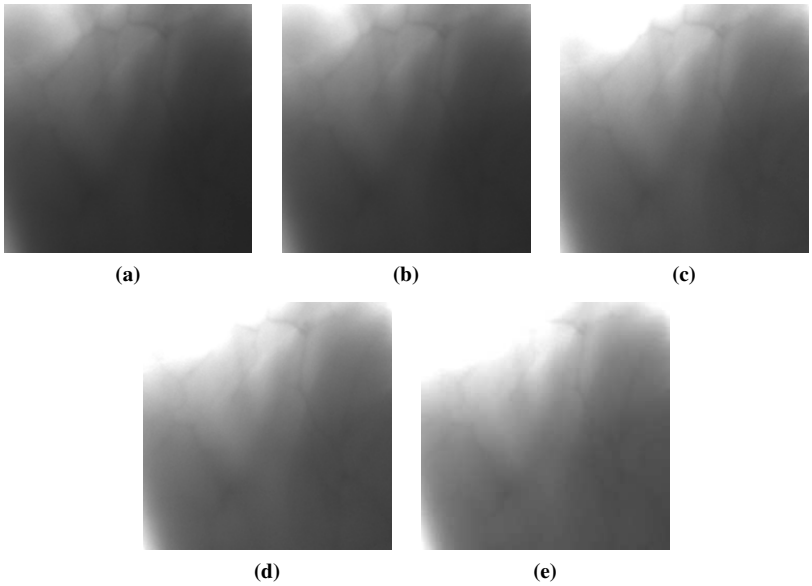


Figure 3.1: *Palm vein LDR images acquired with exposure time to (a) 0.036s (b) 0.042s (c) 0.048s (d) 0.054s (e) 0.060s.*

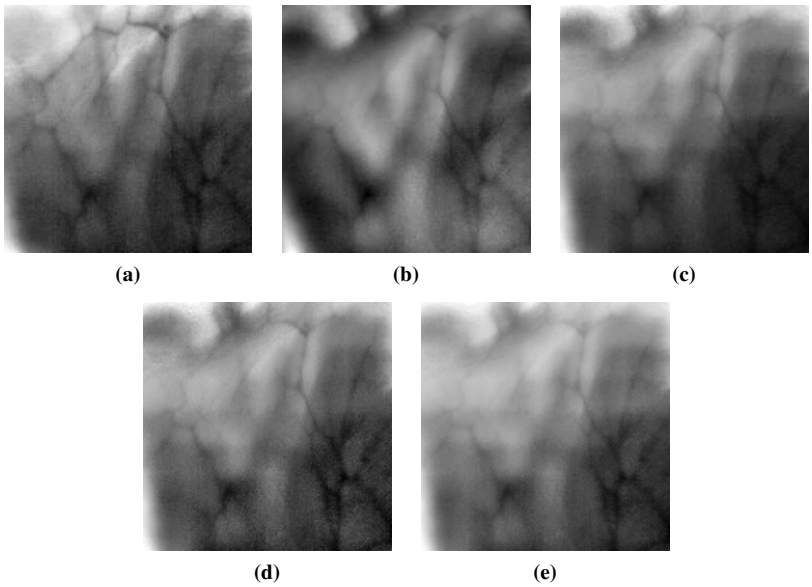


Figure 3.2: *HDR vein images after (a) iCam06 [149] (b) Chiu [44] (c) Drago [69] (d) Ferbman [74] (e) Shan [274] (f) Shibata [275] tone mapping methods applied on the merged LDR images of Fig. 3.1.*

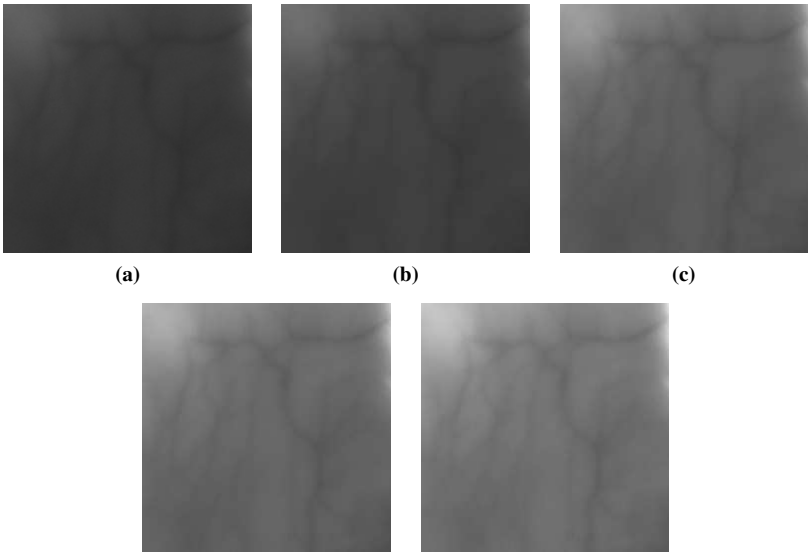


Figure 3.3: *Palm vein LDR images acquired with exposure time to (a) 0.036s (b) 0.042s (c) 0.048s (d) 0.054s (e) 0.060s.*

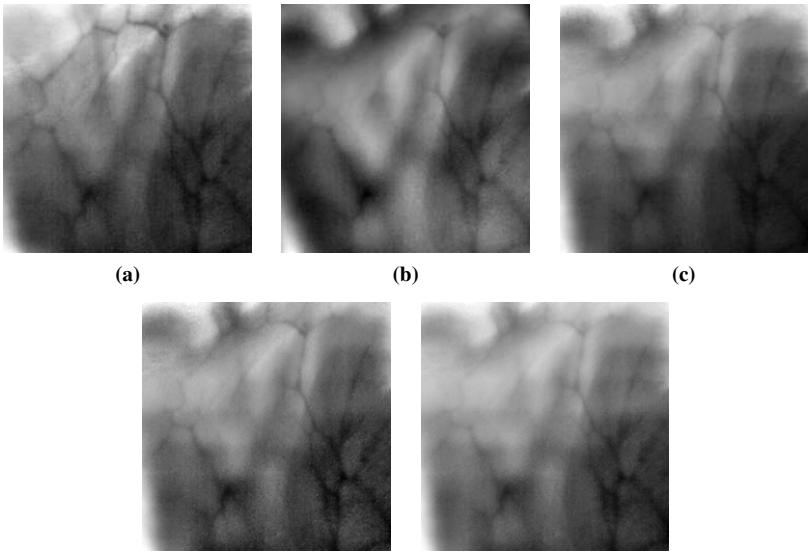


Figure 3.4: *HDR vein images after (a) iCam06 [149] (b) Chiu [44] (c) Drago [69] (d) Ferbman [74] (e) Shan [274] (f) Shibata [275] tone mapping methods applied on the merged LDR images of Fig. 3.3.*

acquired at different shutter speeds, namely using a bracketing-based approach. This latter allows reconstructing the original dynamic range and capturing details from both the image's brightest and darkest areas, taking details pertaining to dark areas from LDR pictures captured with high shutter speeds and contents from very bright regions from low-exposures pictures. The quality of the generated HDR content is therefore typically higher than what is present in its LDR counterpart.

3.1.2 Employed Palm Vein Recognition System

Image Acquisition System. The employed acquisition setup consists of a NIR camera and a NIR illuminator. During each registration a set of N LDR images is acquired at different shutter speeds. Examples of images acquired using the aforesaid experimental setup are given in Figs. 3.1 and 3.3. Implementation details of the employed framework are given in Section 3.1.3.

Preprocessing. A region of interest (ROI) containing the palm vein pattern, with size 240×240, is first extracted from the acquired image. A non-linear image processing is then performed to face the issue of non-uniform background illumination and low contrast in vein pattern images. In detail, the ROI images are divided into blocks of 20×20 pixels, with 4-pixel overlap between two adjacent blocks. For each block the average gray level is computed. The set of obtained mean values is then expanded into 20×20 blocks using a bicubic interpolation, generating the estimated background illumination. This latter is finally subtracted from the considered image, thus obtaining the enhanced vein pattern. The described preprocessing method will be referred to as background-removal (BR) preprocessing in the next sections.

HDR Content Generation. In order to generate the desired HDR vein pattern representation, the N different single-exposure images are combined through a weighted sum of their LDR luminance contents, taking into account the camera response function (CRF) and the exposure time of each picture [11]. The aim of the employed weighting function is to give more importance to middle luminance values while removing possible outliers. The image obtained by combining the LDR sources is here referred as raw HDR image. The obtained HDR content can be later processed in order to properly represent the dynamic range on LDR devices, by means of tone mapping operators (TMO). Specifically, the aim of a TMO is to adapt the

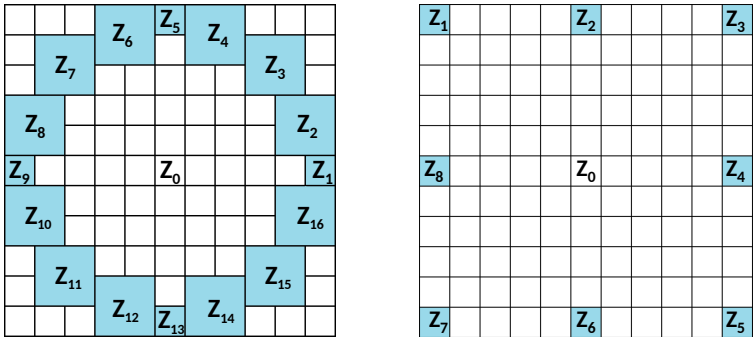


Figure 3.5: *An example of (a) LBP and (b) LDP neighbourood.*

high dynamic range of the merged images to a low dynamic range device, still keeping details and contrast of the raw HDR data.

In this work several TMOs have been applied, specifically iCam06 [149], Chiu [44], Drago [69], Farbman [74], Reinhard [257], Shan [274], and Shibata [275], and evaluate the recognition performance on the so-obtained tone-mapped HDR images. Examples of tone-mapped HDR images, generated from the data shown in Figs. 3.1 and 3.3, are given respectively in Figs. 3.2 and 3.4.

Feature Extraction. Two feature extraction approaches based on local textures, namely local binary pattern and local derivative pattern, are used to obtain palm vein descriptors.

Local Binary Pattern. The LBP operator is a texture descriptor based on the gray level differences and comparisons of a neighbourhood of pixels [202, 215]. Given a central pixel Z_0, an $R \times R$ neighbourhood of P pixels is thresholded by the value of the central pixel and the LBP code for each centre pixel of a grey-scale image I is obtained as:

$$LBP_{P,R}(Z_0) = \sum_{p=1}^{P} f(Z_p, Z_0) 2^{p-1} \tag{3.1}$$

where Z_p is one of the P neighbours of Z_0, as shown in Fig. 3.5a. If the p^{th} neighbour is not a single pixel of the image, a weighted average of the selected pixels is performed, where the weights depend on the distance of the pixels with respect to Z_0. The thresholding function $f(Z_p, Z_0)$ can be represented as:

$$f(Z_p, Z_0) = \begin{cases} 0, & \text{if } I(Z_p) - I(Z_0) < 0 \\ 1, & \text{if } I(Z_p) - I(Z_0) \geq 0. \end{cases} \tag{3.2}$$

Each LBP code represents a micro-pattern of the image and it is saved in a histogram which contains information about the occurrence of the different kind of micro-pattern.

Local Derivative Pattern. The LDP operator is a high-order texture descriptor which extracts the derivative direction variation information [202, 341]. The directions considered to compute derivatives are 0°, 45°, 90° and 135°, where the derivatives along each direction are computed by subtracting pixels of a neighbourhood according to the selected direction. In detail, the first-order derivatives along the four directions, with respect to a given central pixel Z_0, are computed as follows:

$$\begin{aligned} I'_{0^\circ}(Z_0) &= I(Z_0) - I(Z_4) \\ I'_{45^\circ}(Z_0) &= I(Z_0) - I(Z_3) \\ I'_{90^\circ}(Z_0) &= I(Z_0) - I(Z_2) \\ I'_{135^\circ}(Z_0) &= I(Z_0) - I(Z_1) \end{aligned} \tag{3.3}$$

where $Z_1, \ldots, Z_4$ are four of the neighbours around the centre pixel chosen according to the direction of the derivative, as shown in Fig. 3.5b. For a given direction α and central pixel Z_0, the second order LDP code is encoded through the concatenation of the bits corresponding to each neighbour:

$$LDP^2_\alpha(Z_0) = \{f(I'_\alpha(Z_0), I'_\alpha(Z_1)), \ldots, f(I'_\alpha(Z_0), I'_\alpha(Z_8))\} \tag{3.4}$$

where the function $f(I'_\alpha(Z_0), I'_\alpha(Z_i))$ is a binary function providing the type of local pattern, defined as:

$$f(I'_\alpha(Z_0), I'_\alpha(Z_i)) = \begin{cases} 0, & \text{if } I'_\alpha(Z_0) \cdot I'_\alpha(Z_i) > 0 \\ 1, & \text{if } I'_\alpha(Z_0) \cdot I'_\alpha(Z_i) \leq 0\,, \end{cases} \tag{3.5}$$

where $i = 1, 2, \ldots, 8$ is the neighbor's index.

The obtained codes are converted into a decimal value and stored into an histogram which represents the image descriptor. This formulation can be generalized for the n^{th} order LDP, considering the $(n-1)^{th}$ order derivatives in the four directions in the computation of the LDP codes.

3.1.3 Experimental setup

The palm vein database employed for the experimental tests is collected using a Visiosens VFU-V024-M-H-C NIR camera as acquisition device, and an array of NIR leds (wavelength = 850 nm) as illuminator. The resolution of the camera sensor is 752×480 pixels, with 8 bit gray-scale per pixel. The CCD camera sensitive range is between 450 and 900 nm and, in order to eliminate the effect of visible light, the B+W F Pro IR 093 optical infrared filter, with cut-on wavelength at 825 nm, is mounted in front of the camera's lens.

The acquisition process is carried out using a docking device for hand placement and ROI extraction, consisting of a window of the desired dimension and pegs for correct hand positioning, to reduce both misalignment and registration problems. The processed ROI consists of 240×240 pixels, corresponding to a vein width of about 2-8 pixels.

Data from 86 subjects are collected in the employed dataset. The right palm of each subject is acquired five times at exposure time $T \in \{0.036, 0.042, 0.048, 0.054, 0.060\}s$, entailing a total capture time of about $T_{tot} = 0.24s$. This process is iterated twelve times for each palm, thus obtaining a dataset of 86 users × 12 palms × 5 exposures.

Features are extracted from single-exposures images, from raw HDR images, and from tone-mapped HDR images, considering both the LBP and LDP extraction methods. In the experimental tests, LBP features are extracted considering $P = 16$ neighbours and a neighbourhood radius $R = 8$. The LBP operator is applied to 16 non-overlapping 60×60 blocks, with the LBP computed on $(2R+1)$ x $(2R+1)$ sub-blocks centred around each pixel of the block. The histograms resulting from the application of the LBP to each block consist of $P(P-1) = 240$ bins, then concatenated to generate the palm vein template. When LDP is applied, the second-order operator is chosen and a radius of 5 pixels from the central pixel is set when the feature extraction step is performed. The image is divided into 16 non-overlapping 60×60 blocks, and the derivatives in the four directions are computed for all the pixels of each block. The resulting LDP block histograms are concatenated to obtain the palm vein template, resulting of *number of blocks · number of directions* $\cdot 2^8$ = 16384 elements.

Given the so-computed templates, a matching score is obtained through the histogram intersection measure [283], defined as:

$$H(p,q) = \frac{\sum_i \min(p_i, q_i)}{\sum_i q_i} \tag{3.6}$$

where *p* and *q* are the two histograms to be compared, each one consisting

of *i* bins, with $\sum_i q_i$ = 240·240.

HDR images are generated using both all the $N = 5$ images collected at different exposures, as well as only $N = 3$ of them, specifically the image at middle-exposure and the two having lower and the higher exposure time w.r.t. the middle one.

In order to perform a comprehensive analysis, the performance of the proposed approach are compared with what obtained when different image enhancement techniques are applied on the original LDR palm vein images. In detail, some proposed enhancement techniques for vein images are considered:

- histogram equalization (HE),
- contrast limited adaptive histogram equalization (CLAHE),
- circular gabor filter (CGF) [343],
- high-frequency emphasis filtering (HFE) [348],
- local-ridge-enhancement (LRE) [200],
- Retinex method (RM) [195].

The aforementioned techniques are applied to the originally acquired LDR images, which are then further preprocessed through the background-removal (BR) method, described in Section 3.1.2, thus obtaining the desired enhanced images. Finally, in order to have a performance comparison between the proposed approach based on sensor-level fusion and alternative information fusion methods, a score-level fusion approach [264] is also performed in order to combine the outputs obtained when matching features extracted either from the raw single-exposure images or from the enhanced vein images.

3.1.4 Experimental Results - HDR Approach

Tables 3.1 and 3.2 show the equal error rates (EERs) obtained when the LBP and LDP feature extraction methods are applied to the considered images. As first step, features extracted from each single-exposure image are matched in order to evaluate the performance for each considered exposure. The best performance is obtained when considering middle-exposure images, that is the image acquired when the exposure time is set to $T = 0.048$ s, with the achieved values reported in Table 3.1. The scores obtained from

Table 3.1: *EER (%) obtained considering LBP and LDP feature extraction methods applied on the single middle-exposure image, on single-exposure images fused at score level, on raw HDR images and on tone-mapped HDR images.*

	Middle Exposure	Exposures	Score Level Fusion	Raw HDR (no TMO)	iCam06 [149]	Chiu [44]	Drago [69]	Farbman [74]	Reinhard [257]	Shan [274]	Shibata [275]
LBP	3.81	5	3.63	2.97	2.65	**2.51**	2.97	2.99	2.98	2.96	2.59
		3	3.63	3.23	2.86	**2.71**	3.22	3.13	3.21	3.12	**2.51**
LDP	3.17	5	3.24	2.00	1.94	**1.74**	2.00	2.06	2.00	**1.74**	1.75
		3	2.97	2.06	2.19	**1.94**	2.13	2.19	2.07	2.06	2.06

Table 3.2: *EER (%) obtained when LBP and LDP features are extracted only from the middle-exposure image preprocessed with the different image enhancement methods.*

	BR	HE + BR	CLAHE + BR	CGF + BR	HFE + BR	LRE + BR	RM + BR
LBP	3.81	3.70	4.69	3.21	5.49	3.10	3.15
LDP	3.17	3.03	4.46	2.71	4.27	2.90	3.41

Table 3.3: *EER (%) obtained considering the LBP and LDP features extracted from the enhanced LDR vein images and then performing a score-level fusion approach.*

	Exposures	HE + BR	CLAHE + BR	CGF + BR	HFE + BR	LRE + BR	RM + BR
LBP	5	3.42	4.18	3.21	5.02	2.97	2.84
	3	3.38	3.54	3.10	4.90	3.03	2.91
LDP	5	2.71	3.81	2.59	4.16	2.45	2.91
	3	2.67	3.94	2.72	4.07	2.45	3.00

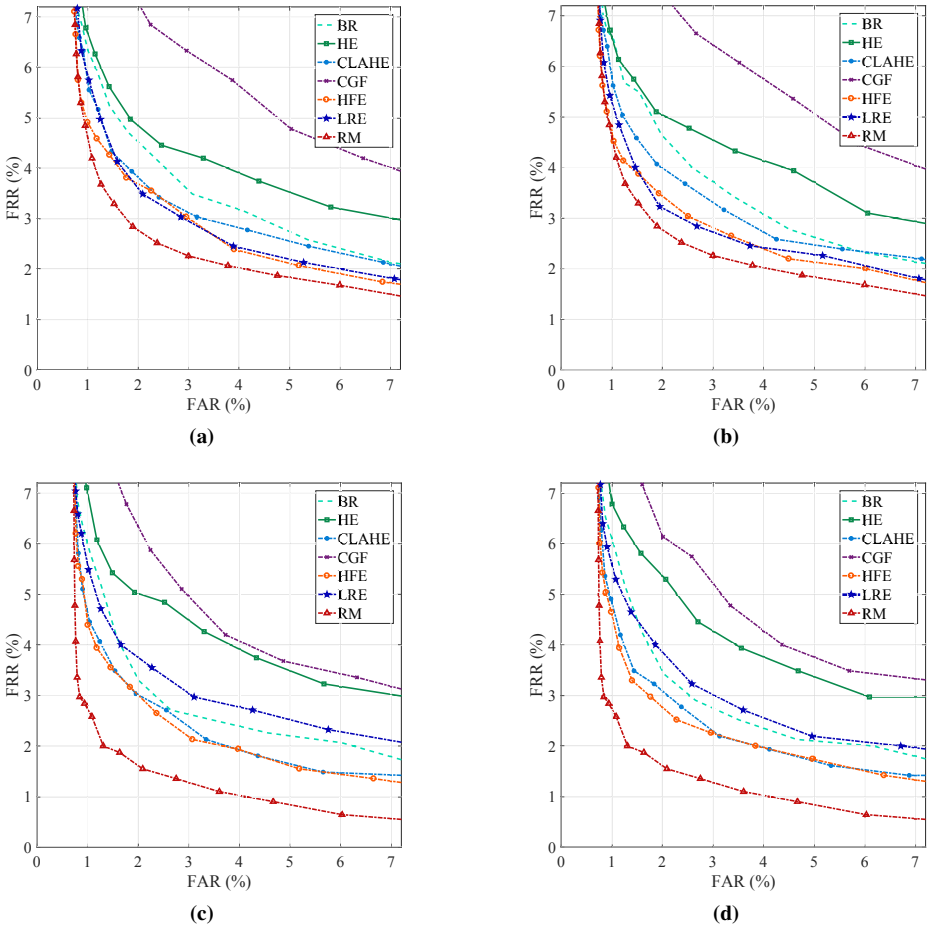


Figure 3.6: *DET curves obtained considering LBP feature extraction method applied on the enhanced vein images and then combining the scores of (a) 3 exposures (b) 5 exposures enhanced pictures and considering LDP feature extraction method applied on the enhanced vein images and then combining the scores of (c) 3 exposures (d) 5 exposures enhanced pictures.*

each single-exposure image are later combined following score-level fusion approaches, namely using the mean, minimum, and maximum rules.

Sets with either 3 or 5 exposures have been exploited. The best results are obtained when the maximum between the scores is considered in the decision step, with corresponding values reported in Table 3.1. LBP and LDP operators are also applied to HDR data, before and after the application of the considered TMOs, with the obtained EERs reported in Table 3.1. Table 3.2 shows the results obtained when different vein image enhance-

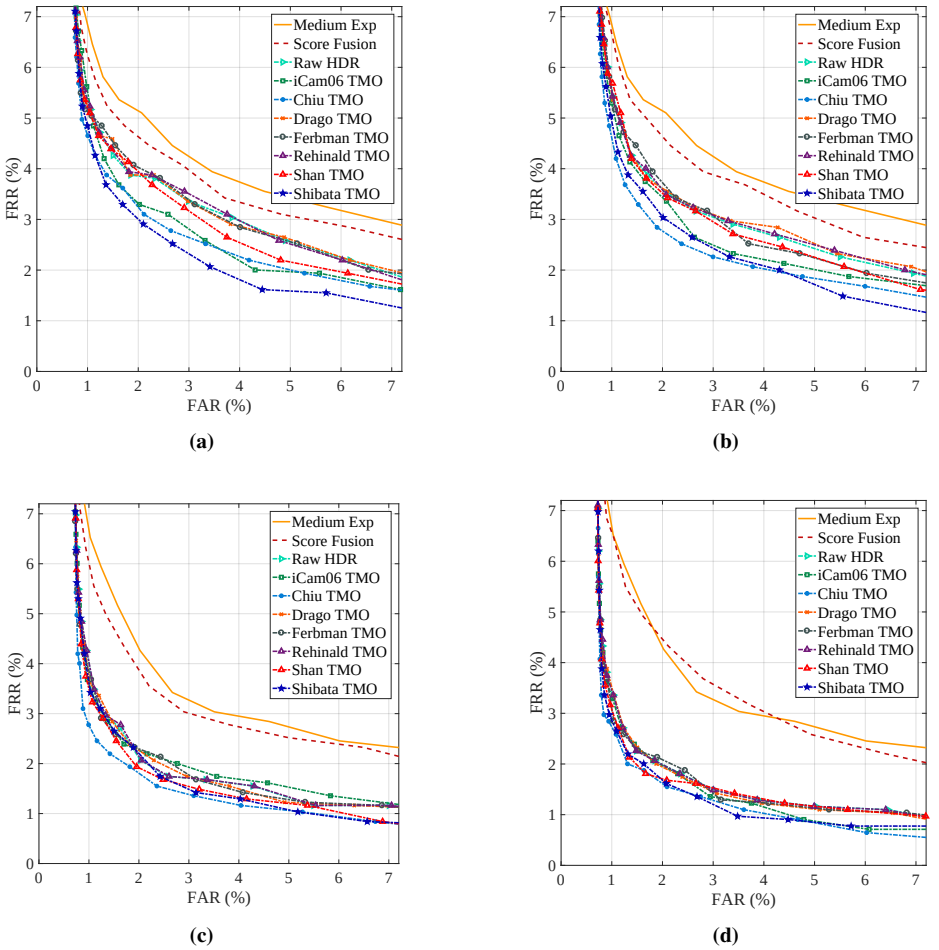


Figure 3.7: *DET curves obtained considering LBP features extracted from sets with (a) 3 and (b) 5 LDR images and considering LDP features extracted from sets with (c) 3 and (d) 5 LDR images.*

ment techniques, listed in Section 3.1.3, are applied to the middle-exposure image and, then, the background-removal (BR) preprocessing is performed. Besides, the performance regarding the score-level fusion approach applied to the results obtained from 3 or 5 exposures images enhanced with the considered pre-processing methods are presented in Table 3.3. In this case, the mean rule is found out to be the best performing score-level fusion strategy. Finally, additional tests are performed in order to analyse the impact on performance of applying the aforementioned enhancement techniques to the LDR images to be fused and then performing the HDR approach apply-

ing the best performing TMO. The results obtained are not reported in this Section because no performance improvement is achieved in comparison to classical HDR approach. This was an expected behaviour since the enhancement techniques modify the contrast of the LDR image and the HDR fusion process may not work correctly.

The detection error trade-off (DET) curves of systems based on middle-exposure images only, on score-level fusion of the available multi-exposure information, on raw HDR content, and on tone-mapped HDR data for LBP and LDP features are also plotted in Fig. 3.7. When no HDR imaging techniques are considered, better recognition accuracy can be achieved when the image enhancement step is performed and the score-level fusion approach is then applied, compared to the performance obtained when no combination of the scores is done. For this reason, only results of Table 3.3 are plotted in the DET curves of Fig. 3.6.

Eventually, the increase in processing time when adopting the sensor-level-fusion strategy in comparison to the baseline system has been evaluated. The experiments are performed on a Core i7-6800K CPU @ 3.40 GHz with 64.0 Gb of RAM and the algorithms are implemented in MATLAB©. In detail, in the proposed HDR-based approach additional time is required in the stages of image acquisition, fusion and possible tone mapping operation. The obtained performance is reported in Table 3.4, where the shown values represent the average times required to process all the images in the considered database.

It can be seen that extracting features from HDR images leads to significant recognition performance improvement, when adopting both LBP and LDP representations, with respect to processing the original single-exposure data. In particular, when the LBP feature extraction method is considered, an EER of 3.81% is obtained when the middle-exposure image is considered, while an EER of 2.51% is reached when the features are extracted from the HDR image built from 5 LDR images fused with the Chiu TMO [44] or when 3 LDR images are combined and then the Shibata TMO [275] is applied. Results concerning LDP features also confirm this behavior, showing an EER of 3.17% obtained when using only middle-exposure images, and an EER of 1.74% with tone-mapped HDR content generated with 5 images taken at different exposures, both considering the Chiu [44] and Shan [274] TMOs. The obtained results also show that generating HDR images considering 5 exposures instead of 3 leads to better results in most of the cases. As shown in Table 3.4, the best performing and less time consuming TMO is the Chiu operator and it is worth highlighting that, when the best performance is achieved, an average total addi-

Table 3.4: *Acquisition and processing time (s) when the HDR imaging approach is considered.*

Step		Required computational time	
		3 Exposures	5 Exposures
Image acquisition		0.1440 s	0.2400 s
Raw HDR image fusion		0.0176 s	0.0278 s
Tone mapping	iCam06 TMO [149]	0.1971 s	
	Chiu TMO [44]	0.3756 s	
	Drago TMO [69]	0.0058 s	
	Farbman TMO [74]	0.1444 s	
	Reinhard TMO [257]	0.0034 s	
	Shan TMO [274]	0.6133 s	
	Shibata TMO [275]	1.1158 s	

tional time of $T_{proc} = 0.192$s (acquisition of four additional LDR-images) + 0.0278s (row HDR image generation) + 0.3756s (Chiu TMO) = 0.5954s is needed.

It is worth remarking that the employed sensor-level fusion approach based on HDR imaging always gives better results compared to the score-level fusion strategy. It is also important to stress out that, with respect to score-level fusion approach, exploiting sensor-level fusion gives additional advantages in terms of required storage space and computational cost. In fact, using HDR content requires extracting features only from a single image, while all the single-exposures images have to be taken into account in the feature extraction and matching stages when score-level fusion is implemented, with the further burden of storing all the derived templates. Finally, it is also demonstrated that, using HDR imaging techniques, it is possible to achieve better results compared to what obtained when the vein images are enhanced by exploiting pre-processing techniques, both considering the score-level fusion approach and not taking it into account.

3.1.5 Experimental Results - Other Fusion Approaches

Different kinds of fusion are then performed, namely feature-, score- and decision-level. When the feature-level fusion approach is applied, features extracted from the same image but exploiting both LBP and LDP feature extraction methods are concatenated and fused into a single feature. In case of score-level fusion approach, the scores derived from the matching of the couples of templates extracted from the same image, but by exploiting the

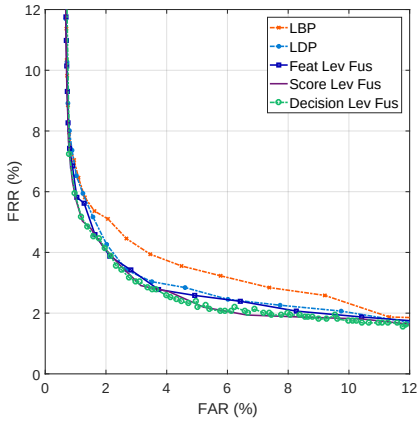


Figure 3.8: *DET curves obtained considering LBP and LDP feature extraction methods and the different adopted fusion techniques*

two considered feature extraction methods, are combined through the *max-rule* method, that is, the fusion rule that has shown the best performance in the performed experimental tests. Finally, a decision level-fusion approach is also carried out. In detail, the effects of the application of different decision rules are investigated, and the *OR* rule has shown to be the best option in the considered case.

The results obtained are shown in the detection error trade-off (DET) curve presented in Fig. 3.8. When LBP codes are considered as features an equal error rate (EER) of 3.81% is reached, while a EER = 3.17% is achieved when the LDP feature extraction method is used to extract discriminative characteristics from palm vein images. When the feature-level fusion approach is performed, the EER of the system decreases to 3.09 %, while an EER of 2.96% is reached when the score-level fusion method is applied. Eventually, an EER of 3.07% can be achieved when the decision-level fusion approach is chosen.

3.1.6 Conclusions

In this section the impact on recognition performance of applying HDR imaging on palm vein recognition systems has been studied, performing a sensor-level fusion approach on images captured at multiple exposures. The obtained results show that significant performance improvement can be reached when HDR content is processed, compared to the use of single-exposure LDR vein images. Besides, the adoption of TMOs allows guaranteeing even further improvements, with performance notably exceed-

ing those achieved when employing vein image enhancement methods and score-level information fusion approaches.

Besides, the performance improvements achievable when exploiting multi-biometrics fusion techniques in a palm-vein-based recognition system leveraging on statistical-based feature extraction methods such as LBP and LDP has been investigated. The obtained results show that a noticeable reduction of the achievable EER can be actually guaranteed with multimodal approached, with respect to unimodal ones. In the considered framework, score-level fusion approach has been highlighted as the best option to be followed when designing palm vein recognition systems based on more than a single template representation. Anyway, the classical fusion approaches are not able to overcome the performance achievable when the HDR-based approach is taken into account.

3.2 Convolutional Neural Network for Finger-Vein-based Biometric Identification

In this Section, finger-vein-based identification by exploiting deep learning techniques is studied. The aim of the proposed work is to achieve good and stable identification performance irrespective of the quality of the considered finger-vein images, their rotation, translation, and scaling. In order to verify the effectiveness of the designed CNN, the considered approach has been tested over four publicly-available finger-vein databases, namely SDUMLA-HTM [338], HKPU [152], FV-USM [5] and UTFVP [289], characterised by different image quality levels. Details about the considered database can be found in Section 2.2.

The achieved performance shows that the proposed method is able to guarantee stable and highly-accurate identification results, irrespective of the quality of the considered finger-vein images. Additionally, the proposed CNN-based identification system requires negligible manual effort for feature selection. In fact, it has been applied without variations to all the four considered databases, without using any application-dependent threshold or any manually-set parameter. This Section is organised as follows: the finger-vein-based biometric system and its modules are described. The experimental settings and tests are then presented, followed by the conclusions.

3.2.1 Employed Finger-Vein Based Biometric System

Preprocessing. The original images, gathered from four publicly-available databases, are pre-processed for ROI extraction and image enhancement. As a first step, the images from all the considered databases, having different sizes, are subsampled to 336×190 pixels in order to guarantee uniformity. Besides, for the databases where the images show a ratio between number of rows and columns different from the target one, marginal background parts are removed, selecting a central area of the image. Eventually, the ROI, i.e. the part of the image which contains the interested finger, is then extracted and a binary mask in which the white pixels correspond to the finger region is obtained. Specifically, the ROI extraction is based on the method proposed by Lee *et al.* [167], where two different masks are used to extract the upper and lower finger's edges respectively. For the HKPU database the aforementioned masks are provided, whereas the aforementioned procedure is applied to the other databases.

Starting from the extracted edges and masks, a normalisation step is performed in order to compensate rotation and vertical translation during the acquisition step. In this work, the approach proposed in [113] has been used, which attempts to fit a straight line between the edges detected in the previous step and estimate the parameters of rotation and vertical translation which are later used to perform an affine transformation. If required, the normalised images may be then enhanced through contrast limited adaptive histogram equalization (CLAHE) [353], which is an adaptive histogram equalization (AHE) method whose aim is to improve the contrast of the image by limiting the contrast amplification in the different considered parts of the image. The preprocessed images are then transposed and resized into 65×153 pixels.

Template Generation. Feature are extracted by a CNN, whose architecture is fully described in [62], summarised in Table 3.5 and depicted in Figure 3.9, fed with the pre-processed finger-vein pattern images. Bigger image size usually leads to a larger CNN with more hidden layers. Hence, in order to have a feasible size network the images are resized into 65×153 in the pre-processing step.

The training and testing templates of the network are either generated by selecting the images from a single session, as proposed by the existing state-of-the-art methods, or otherwise by selecting a combination of images from all available sessions. The reason behind this latter strategy is that different sessions' data may have been acquired with different illuminations for the

Table 3.5: *The proposed CNN configuration.*

Layer Type	Number of Filter	Size of Feature Map	Size of Kernel	Number of Stride	Number of Padding
Image input layer	-	$65 \times 153 \times 1$	-	-	-
CL_1 (Convolutional layer-1)	153	$65 \times 153 \times 1$	5×5	1×1	0×0
M_1 (Max-Pooling Layer-1)	1	$61 \times 149 \times 153$	2×2	2×2	0×0
CL_2 (Convolutional layer-2)	512	$30 \times 74 \times 153$	5×5	1×1	0×0
M_2 (Max-Pooling Layer-2)	1	$26 \times 70 \times 512$	2×2	2×2	0×0
CL_3 (Convolutional layer-3)	768	$13 \times 35 \times 512$	5×5	1×1	0×0
M_3 (Max-Pooling Layer-3)	1	$9 \times 31 \times 768$	2×2	2×2	0×0
CL_4 (Convolutional layer-4)	1024	$4 \times 15 \times 768$	4×15	1×1	0×0
R_1 (ReLu Layer-1)	-	$1 \times 1 \times 1024$	-	-	-
CL_5 (Convolutional layer-5)	U (number of classes)	$1 \times 1 \times 1024$	1×1	1×1	0×0
Softmax Layer	-	$U \times 1$	-	-	-

same finger of a person. Hence, the network may require images captured in different conditions for its proper training, otherwise the identification accuracy may be affected. To find the best combination of templates for training, 1, 2, 3 and 4 images' combinations from all available sessions are considered for training and their identification results allowed to find the best possible combination of templates which can be used for person identification.

CNN Training. The generated templates are passed through the designed CNN and a set of very low-level features are extracted in the first hidden layer. The network gradually builds up over these low-level features in the subsequent convolutional layers, in order to create a set of high-level features for the fully connected layer. For the experiments, each finger of every person is considered as a separate class.

For the HKPU dataset, since 105 subjects have contributed with their index and middle fingers to two sessions, there is a total of 210 classes available for training. The remaining 51 subjects have contributed only to session 1, so they have not been considered for training in the experimental tests and they have been instead only used as impostors while testing. Similarly, for FV-USM database 492 classes (123 subjects with 4 fingers each), 636 classes for the SDUMLA database (106 subjects with 6 fingers each), and 360 classes for the UTFVP database (60 subjects with 6 fingers each) are considered.

For CNN designing and training, the MatConvNet-1.0-beta24 tool [301] is employed. For training purposes 90% finger-vein images are considered, with the remaining 10% used for validation. The learning rate of the CNN

is set at 0.00001 with a batch size of 3 samples for HKPU and FV-USM, 4 for SDUMLA and 2 for UTFVP, so that the loss can be minimised with higher precision through the execution of every epoch or iteration.

As for the number of epochs, higher numbers usually allow the network to be well-trained, so that the weights of different layers are updated with precision. For the experiment, 2500 epochs have been considered for all the experiments. The main purpose of using such a low learning rate and high number of epochs is that it is typically preferable to let a network learn very slowly and converge into the smallest details of every class.

Identification. In the identification stage, for each testing sample, the trained CNN returns a probability value for all the available classes/fingers. The maximum probability value identifies the most similar finger to the testing sample. As each and every finger of an individual represents a different class, it is possible to identify the particular finger with which it is matched and the corresponding subject to whom it belongs to.

It is worth specifying that, similarly to what has been proposed in [152], for the experimental setup a threshold for matching probability of a test image has been introduced, below which the test image is considered as "not-identified". This is for the purpose of genuine impostor testing where no sample images are trained for that particular subject, as they are not associated with any of the enrolled identities. For a given testing sample, if the matching probability value returned by the proposed network is less than 50% for its comparisons with any trained class, then that test image is classified as "not-identified" or "not-present" in the database. For example, this scenario has been tested with the finger-vein images of the 51 subjects captured during a single session in the HKPU dataset. Such images have not been ever employed for training purposes and have been instead used only as testing probes. Each time, when a test sample's result reaches into a maximum matching probability value of lower than 50% for all the trained classes, then it is possible to declare them as "not-identified".

3.2.2 Results and Discussions

In order to evaluate the proposed network, its performance are first compared with several state-of-the-art identification techniques, by using the training and testing strategies adopted in referenced papers for the proposed network as well. Subsequently, an optimal training strategy for the proposed network is designed. Most of the state-of-the-art techniques have used either a single image or images from a single session for their network's training, which may not be ideal for the CNN-based proposed ap-

Table 3.6: *Identification accuracy comparison for the four considered publicly-available databases.*

Database	Training	Testing	State-of-the-art comparison methods									Proposed CNN
			Kumar et al. [152]	Qui et al. [242]	Van et al. [298]	Jia et al. [128]	Van et al. [299]	Xie et al. [326]	Banerjee et al. [10]	MC [206]	RLT [205]	
HKPU	6 images from session 1	6 images from session 2	90.08%	-	-	-	-	-	-	85.24%	78.28%	71.11%
FV-USM	first image from session 1	6 images from session 2	-	97.02%	-	-	-	-	-	90.34%	78.28%	72.97%
SDUMLA	4 images	remaining 2 images	-	-	91.83%	92.50%	95.67%	-	-	86.01%	87.11%	97.48%
	5 images	remaining image	-	-	-	-	-	97.76%	90.72%	97.95%	96.06%	98.90%
UTFVP	1 image from each session	remaining image from each session	-	-	-	-	-	-	-	92.22%	93.05%	95.56%

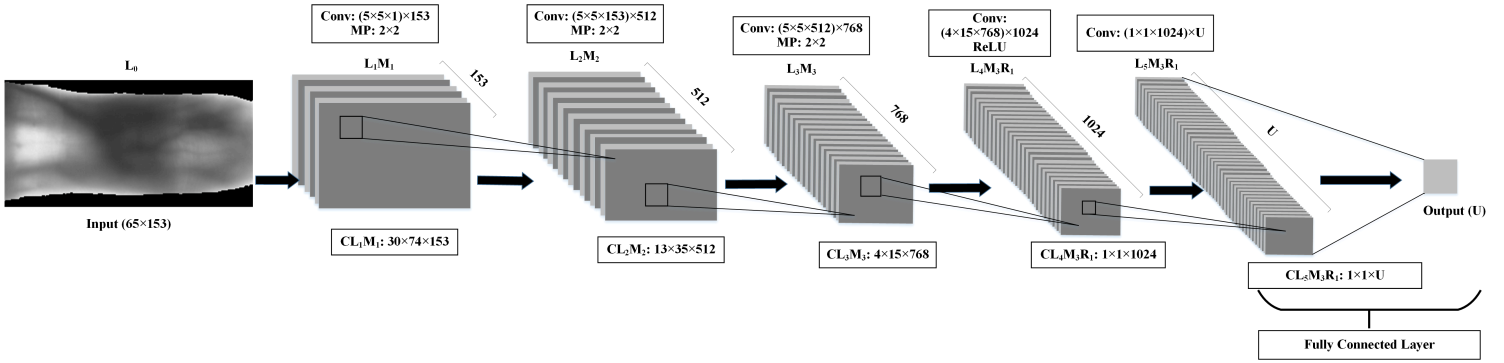


Figure 3.9: *Employed CNN architecture.*

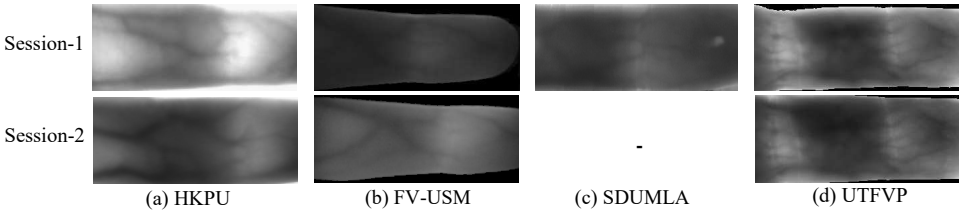


Figure 3.10: *Different luminosity images from different sessions of four publicly-available databases.*

proach. It is in fact well-known that the availability of a single sample of every class, here individual fingers, does not allow a CNN to get trained properly. Eventually, the utility of exploiting image enhancement preprocessing techniques together with the proposed network is evaluated in order to to understand if further performance improvement could be achieved.

All the experiments have been performed in MATLAB® (R2017a) with a system configuration of 64 Gb RAM; Titan X™ (Pascal) graphics card; i7, 3.40GHz processor and Windows® 10 operating system.

Performance Comparisons. The identification accuracies achieved by the most relevant state-of-the-art finger-vein-based biometric systems are reported in Table 3.6, together with the obtained performance with the proposed CNN-based approach, when using the same training and testing strategies. The results obtainable while exploiting two of the most-commonly employed methods for finger-vein recognition, i.e., maximum curvature (MC) [206] and repeated line tracking (RLT) [205], under all the considered settings, are additionally reported for further comparisons.

As it can be seen from the reported accuracies, the CNN-based identification system cannot be properly trained under the experimental setup employed in [152], where only session-1 images from HKPU dataset are used for training, and session-2 images for testing. A similar situation is encountered when comparing the proposed system against the one in [242], where tests over the FV-USM dataset have been performed by considering only the first image of every finger from session-1 for training and the 6 images per finger of session-2 for testing. Again, the reason behind such a low performance depends upon the number of training samples, along with the different quality of finger-vein images that exists in two distinct sessions, as shown by the examples in Fig. 3.10.

A different behaviour of the network has been observed when considering the training/testing settings employed in [299] and the SDUMLA

database, which contains images taken from a single session. In this scenario, the method here proposed is able to achieve identification performance better than those obtained in [326] and [10]. It is therefore reasonable to observe that the proposed CNN-based identification system can work properly when images of similar quality are used for both training and testing purposes, regardless the absolute quality level of the considered images. This assumption is confirmed by the results in Table 3.6, referred to the comparison of the proposed approach against MC and RLT, while taking one finger-vein image from each session of the UTFVP dataset, and using the remaining ones for testing purposes. The proposed CNN-based identification system easily outperforms both MC and RLT.

It is worth remarking that the aforementioned results have been obtained with the proposed CNN-based identification system without performing any kind of enhancement on finger-vein images. Conversely, all the methods used for comparisons use some image enhancement technique and feature selection processes. Therefore, the use of original images without any preprocessing and automatic feature extraction are among the advantages of the proposed network.

Different Training Strategies. The use of a single image or images from a single session for training purposes may not be enough to produce the best accuracy. Therefore, full potentiality of the proposed CNN architecture has been analysed by investigating the variations in identification accuracies that can be achieved while adopting different numbers of images for training. Hence, wherever possible, 1, 2, 3, and 4 original images from all the available sessions of each of the finger-vein patterns from the four databases are considered for the training of the considered network.

Table 3.8 summarises the obtained results and clearly shows that, if the number of training samples from each finger is increased, then the achieved accuracy also increases significantly. Comparing these results with those reported in Table 3.6, a notable improvement in terms of achievable performance can be seen for both HKPU and FV-USM databases. From Table 3.8 it can also be noted that there is not much difference in accuracy, for HKPU and FV-USM database, when 3 or 4 images are used for training. Hence, for further experiments 3 images from each session of these two databases are employed for training. When considering the SDUMLA database, 4 images from each session are needed for training, since low quality images are present in this database, as shown in Figure 3.11. As for the UTFVP database, at most one image from each session can be chosen for training, since only 2 images are available for each finger's acquisition sessions. The

Table 3.7: *CNN-based identification accuracy over the considered publicly-available databases.*

Database	Training	Testing	Methods			
			MC	RLT	CNN with original images	CNN with CLAHE [353] enhanced images
HKPU	3 images from each session	remaining 3 images from each session	83.33%	83.81%	**95.32%**	94.37%
FV-USM	3 images from each sessions	remaining 3 images from each session	92.60%	94.44%	**97.53%**	97.05%
SDUMLA	4 images	remaining 2 images	86.01%	87.11%	**97.48%**	95.13%
UTFVP	1 image from each session	remaining 1 image from each session	92.22%	93.05%	95.56%	**98.33%**

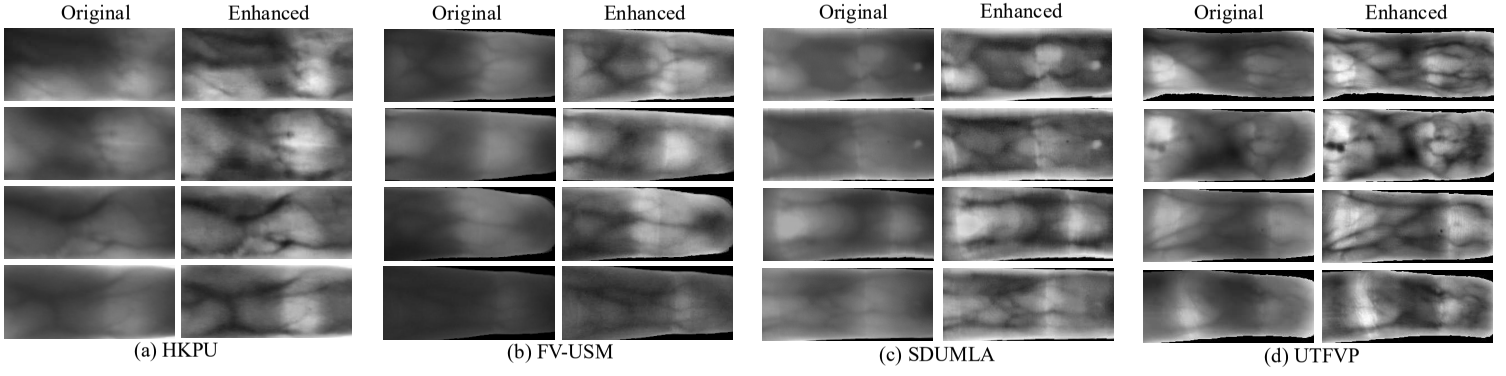


Figure 3.11: *Original and CLAHE [353] enhanced finger-vein image from four publicly-available databases.*

Table 3.8: *Identification accuracy for different training strategies over original images.*

Database	Training (images from each available sessions)			
	1	**2**	**3**	**4**
HKPU	82.19%	92.02%	95.32%	**96.55%**
FV-USM	91.75%	94.82%	97.53%	**98.58%**
SDUMLA	75.25%	77.99%	80.27%	**97.48%**
UTFVP	**95.56%**	-	-	-

obtained results show that identification accuracy greater than 95% can be achieved for all the four considered publicly-available databases.

Table 3.7 shows the improvements that can be obtained when the proposed network is trained according to the best performing training settings. More in detail, Table 3.7 shows rank-1 identification accuracy for the considered databases, when exploiting the proposed CNN-based approach. A performance comparison with standard MC and RLT methods is also provided here. Moreover, it has also been investigated whether the proposed method needs any image enhancement technique to further improve the attainable identification performance, considering the case of the network is fed with CLAHE [353]-enhanced finger-vein images.

According to the obtained results, the proposed CNN-based identification system systematically outperforms MC and RLT approaches. It can also be seen that the contrast-enhanced images achieve performance better than the original ones only when considering the UTFVP dataset. It is evident from Fig. 3.11 that vein patters in images from UTFVP database are significantly more prominent and clearly distinguishable in their enhanced versions rather than in their original ones. Nevertheless, the proposed CNN-based approach is typically able to guarantee a very-high identification rate without using any image enhancement technique.

3.2.3 Conclusions

In this Section, a CNN-based finger-vein identification system which can perform an effective identification irrespective of the environmental conditions is proposed. An exhaustive set of experimental tests performed over the four commonly-used and publicly-available databases is presented. The obtained results show that it is possible to achieve a rank-1 identification accuracy greater than 95% for all the four databases, using the proposed CNN architecture. The present work is one of the first comprehensive

study analysing a finger-vein-based biometric identification system with more than two publicly-available databases, to assess the effectiveness of the proposed network under different image quality conditions, with minimum human intervention. It can also be seen that the identification accuracy of the proposed network significantly increases with the employed number of training images. Moreover, if the finger-vein images are not acquired with the same illumination intensity and ambient lighting conditions for different sessions of acquisition, then the use of multiple session's data for training can be considered as an effective strategy for improving the achievable identification accuracy.

3.3 On-the-fly Finger-Vein-based Biometric Recognition using Deep Neural Networks

The applications where convenience, speed and security are a priority are increasing. Border control at airports, access control at sensitive and high traffic sites, flow management are few examples of a wide range of systems where the aforementioned requirements are crucial. In these situations biometric systems are generally chosen as solution. Besides, recently the convenience and speed of the biometric systems have been improved thanks to the commercial lunch of the "on-the-fly" system, that is a biometric system where it is possible to enrol and authenticate with a fast touchless movement, such a single wave of the hand [1].

The aim of the present study is to merge the intrinsic security and robustness of the vein biometric trait with the speed and convenience of a touchless acquisition device. The approaches proposed in the literature capture the vein-pattern structures either requiring the subject to touch a support, as for Hitachi's "VeinID" finger-vein technology[2], or allowing a contactless acquisition as in Fujitsu's "PalmSecure"[3]. This technology has already beendeployed in real-life applications such as banking or con-sumer products. However, it is worth pointing out that in all cases the user has to hold the hand still during the entire acquisition process. On the contrary, in the proposed approach, as detailed in the following, the user is asked to swipe the hand over the sensor, thus implementing a contactless and "on-the-fly" interaction modality, entailing an increase of the user convenience and of the recognition system's throughput.

[1] https://www.idemia.com/morphowave-desktop

[2] http://www.hitachi.co.jp/products/it/veinid/global/index.html

[3] https://www.fujitsu.com/global/services/security/offerings/biometrics/palmsecure

Deep learning approaches, based on both convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are here exploited to extract discriminative features from the acquired vein pattern videos, and achieve remarkable recognition performance. Eventually, high dynamic range (HDR) techniques [228] are exploited to further improve the recognition performance. The architecture of the proposed identification systems is sketched in Figure 3.12 and its building blocks are detailed in the following paragraphs.

3.3.1 Designed Finger-Vein Identification Pipeline

The architecture of the proposed identification systems is sketched in Figure 3.12. Its building blocks, that is, data acquisition hardware, preprocessing, and classification modules, are detailed in the following.

Data Acquisition Hardware. In order to carry out the finger-vein acquisition by allowing the user to swipe the hand over the sensor while moving and without contact, a novel low-cost acquisition device has been designed, represented in principle in Figure 3.13a. It is composed by four PiNoIR-V2 CMOS cameras, equipped with Sony IMX219 8-megapixel sensors [219] having a NIR sensitivity in the wavelength range $400nm - 1000nm$, each driven by a Raspberry PI-V2 card. CMOS-based cameras have been widely used in the literature for vein-pattern acquisition, essentially because of their lower cost with respect to CCD-based cameras which, on the other side, guarantee higher performance in terms of signal-to-noise ratio (SNR) of the acquired images, especially in the NIR field.

The cameras are arranged in a 2x2 matrix configuration to minimise the parallax effect in the image acquisition process, with respect to the other camera configurations that have been tested and discarded. Each of the four cameras employs a different exposure time, namely 12μs, 16μs, 20μs, and 24μs. The use of different exposures allows generating high dynamic range (HDR) data, which have been proven [228] to allow recognition performance improvement. In addition, a $700nm$ longpass NIR filter is placed over the camera array to cut out the visible light.

The employed illuminator is composed by 20 LEDs, arranged in a rectangular shape of 5×4. Specifically, Osram Opto SFH 4356-UV model IR LEDs has been employed with the following characteristics: *i)* dome shaped lens, *ii)* $80mW$ radiant flux, *iii)* $860nm$ peak wavelength, and *iv)* $850nm$ centred wavelength. It is worth mentioning that most of the systems available in the literature use LEDs operating in the range $[850, 930]$nm. As a matter of fact, LEDs have been tested at $830nm$, $850nm$, and $910nm$. The

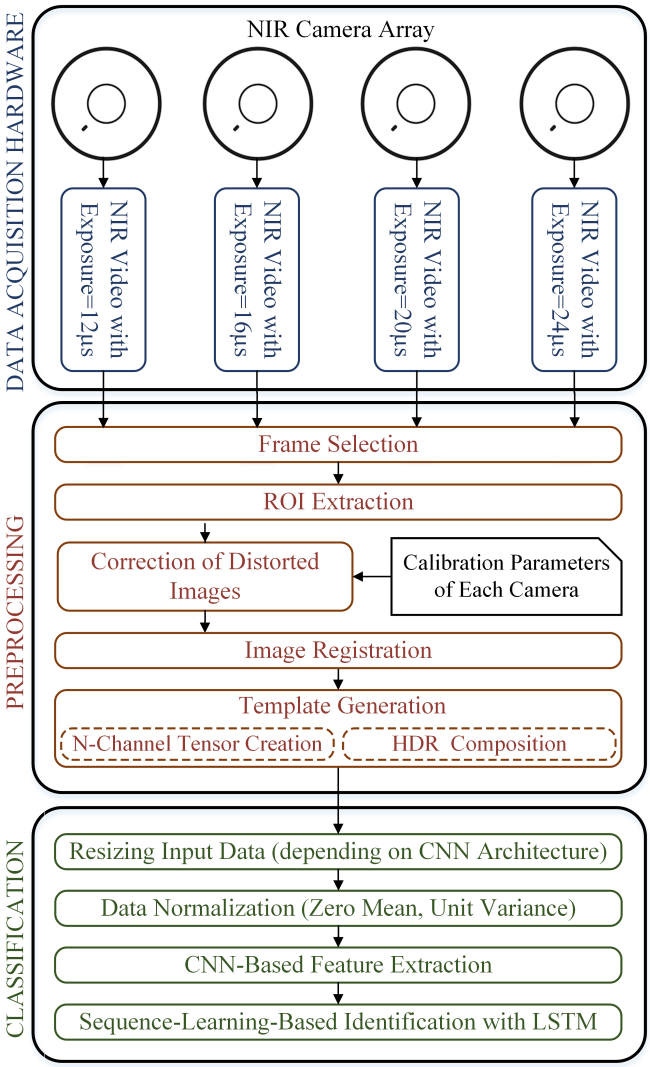


Figure 3.12: *High-level representation of the acquisition and processing pipeline of the proposed system.*

$850nm$ LEDs has proven to be the best performing in the proposed system. The illuminator is fed with a total of $400mA$ of current and $12V$ of voltage. A $3mm$-thick white diffusion glass is placed between the hand and the lighting LEDs to obtain a more uniform light diffusion over the fingers. The acquisition protocol requires the user to swipe the hand between the cameras and the illuminator, thus working in transmission modality. The acquisition system is shown in Figure 3.13b.

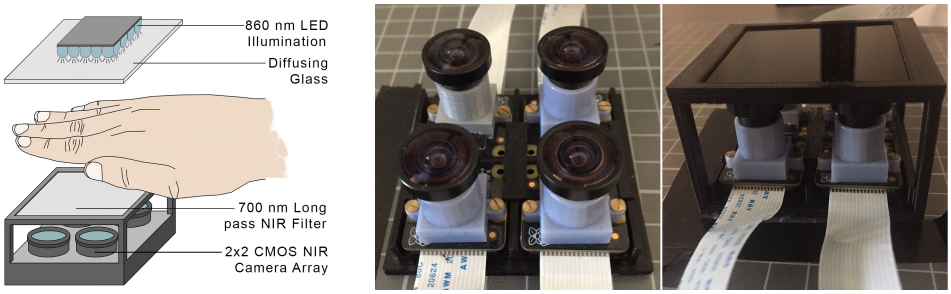


Figure 3.13: *(a) Graphic representation of the propose acquisition system. (b) 2x2 camera array (left), NIR filter on camera array (right).*

It is worth remarking that, in conventional vein-based biometric systems, users have to keep their fingers or hands still during acquisition, and often place them on a support. On the contrary, in the novel approach, this constraint is released, and allow users swiping their hands over the sensor as shown in Figure 3.14. Therefore, during the acquisition, four videos are collected, at the rate of 12 frames per second (fps), each with different exposure times.

Preprocessing. After acquisition, the preprocessing steps sketched in Figure 3.12, and described hereafter, are performed.

- **Frame selection:** passing a hand over the acquisition system may require from one to three seconds, depending on the user's behaviour. During this time, each of the four employed cameras records up to 36 frames, out of which 9 frames, containing all the hand's fingers, are selected for further processing. Specifically, the frame with the overall lower average luminance across the captured videos is chosen as reference and assumed to have the hand in central position with respect to the device. The four frames before and the four frames after it are then selected for each of the four acquired videos.
- **ROI extraction:** a ROI of 720x640 (WxH) pixels is extracted from each image. This choice, made through a trial-and-error process, guarantees that the whole image of the hand is selected, as shown in Figure 3.15. The so-obtained images have been corrected by using a camera-calibration approach to compensate for fish-eye distortion.
- **Image registration:** images acquired from different cameras are misaligned due to the parallax effect, caused by the non-negligible size of

Figure 3.14: *Video acquisition procedure.*

Figure 3.15: *LDR finger-vein templates of a subject using* $12\mu s$*,* $16\mu s$*,* $20\mu s$*,* $24\mu s$ *exposure times (first four images on the left), and resulting tone-mapped HDR vein template (image on the right). Images contrast enhanced for visualization purposes*

the employed cameras. Therefore, image registration is needed. In the proposed approach, the multimodal intensity-based image registration technique proposed in [142] is applied.

- **Template generation:** the images extracted from the acquired videos can be affected by low contrast, due to the difficulty in controlling the employed NIR illumination when capturing moving hand as already pointed out in [228], as well as by blur effect due to hand movement during the acquisition. To mitigate these problems, two different approaches are taken into account:
  - use of HDR imaging techniques [40]: the images captured at different exposure times are fused into a single HDR image which does not suffer from under- or over-exposure issues. The generated HDR content can be then converted into a low dynamic range (LDR) image through the use of a tone mapper. In this study, the iCAM06 tone mapping operator [149] is considered, due to its superior performance within the proposed framework [228];
  - use of a 4-channel tensor: the four images, acquired by the four cameras, at the same time but at different exposures, are represented through a single structure. Specifically, four 1-channel grey-scale images, representing the luminance content at different exposures, are grouped to build a 4-channel image tensor.

Classification. As already mentioned, during each acquisition 9 frames are taken from each of the 4 employed cameras, for a total of 36 frames. A single swipe of a hand therefore generates data characterised by a specific spatial behaviour, given by the properties of the vein patterns of the four fingers captured in each image, as well as a temporal behaviour, represented by the sequences of frames taken at consecutive instants.

In order to take advantage of the collected information, DNNs have been exploited. Specifically, a CNN has been designed to extract reliable features from each processed image. The architecture of the adopted CNN, namely Vein-CNN (V-CNN), has been specifically designed for finger-vein-based recognition tasks, with its configuration (sizes, kernels, etc.) set considering finger geometry and orientation. In addition, an RNN has been used to exploit the availability of multiple frames in the acquired videos. In more detail, a long short-term memory (LSTM) network has been exploited to model the observed temporal course of the hand swipe. The proposed CNN-LSTM architecture is shown in Figure 3.16 and discussed in details

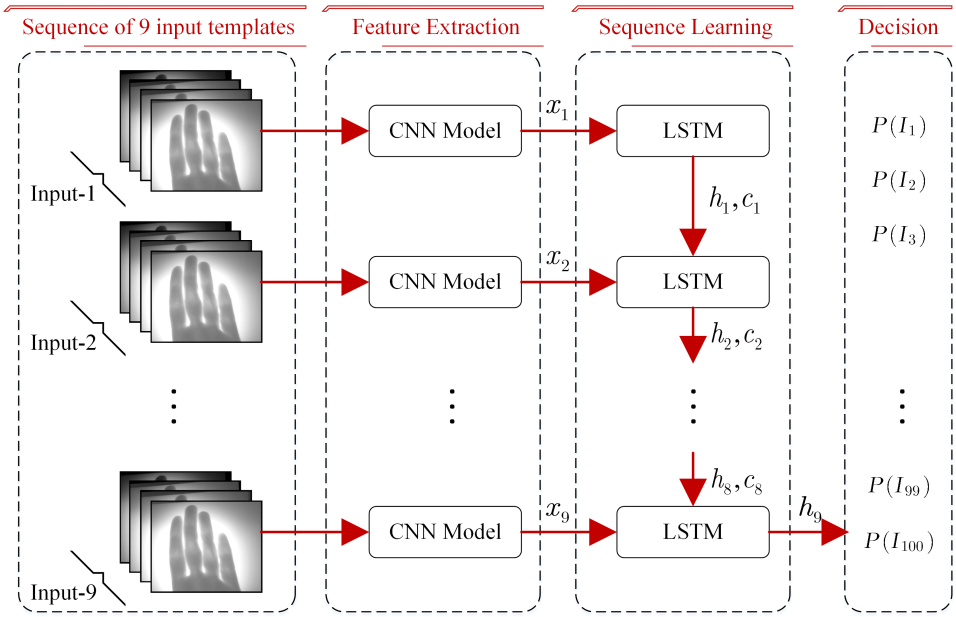


Figure 3.16: *CNN-LSTM architecture of the proposed system.*

in [157]. Besides, the architecture of the CNN adopted in the proposed system and specifically designed for finger-vein-based recognition tasks, is shown in Figure 3.17 and it details of the CNN are reported in Table 3.9.

It is worth pointing out that the considered approach is multimodal in many aspects, as described hereafter:

i) multiple sensors for the same biometric trait: the same finger-vein pattern is acquired using four cameras with different acquisition parameters;

ii) multiple units of the same biometric trait: four fingers, as a whole, are used together for template generation;

iii) multiple biometric traits: both finger veins and finger shape are intrinsically acquired and processed by the proposed system.

3.3.2 Experimental Tests

In order to evaluate the effectiveness of the proposed CNN-LSTM framework for the designed on-the-fly finger-vein-based biometric recognition system, several tests have been performed over a database collected at the Engineering Department of Roma Tre University institution. Specifically,

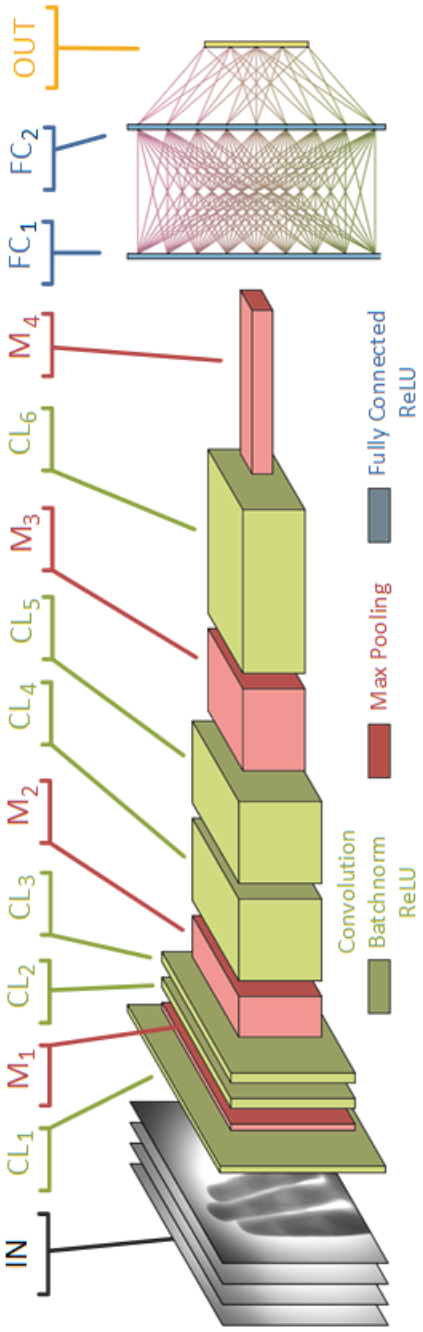


Figure 3.17: *V-CNN Architecture.*

Table 3.9: *Proposed CNN Configuration (V-CNN). In the N_c represents the number of channels for each image (we may have either $N_c = 1$ or $N_c = 4$ in the proposed configurations; N_I the number of the output layer, represent the number of unique identities/subjects in the database.*

Abbreviation	Layer Type	Number of Filter	Size of Feature Map	Size of Kernel	Number of Stride	Number of Padding
IN	Image Input Layer	-	$320 \times 360 \times N_c$	-	-	-
CL_1	Convolutional Layer-1	64	$320 \times 360 \times 64$	3×3	1×1	1×1
M_1	Max-Pooling Layer-1	1	$160 \times 180 \times 64$	2×2	2×2	0×0
CL_2	Convolutional Layer-2	128	$160 \times 180 \times 128$	11×5	1×1	5×2
CL_3	Convolutional Layer-3	128	$160 \times 180 \times 128$	5×5	1×1	2×2
M_2	Max-Pooling Layer-2	1	$80 \times 90 \times 128$	2×2	2×2	0×0
CL_4	Convolutional Layer-4	256	$80 \times 90 \times 256$	5×11	1×1	2×5
CL_5	Convolutional Layer-5	256	$80 \times 90 \times 256$	5×5	1×1	2×2
M_3	Max-Pooling Layer-3	1	$40 \times 45 \times 256$	2×2	2×2	0×0
CL_6	Convolutional Layer-6	512	$40 \times 45 \times 512$	3×3	1×1	1×1
M_4	Max-Pooling Layer-4	1	$8 \times 9 \times 512$	5×5	5×5	0×0
FC_1	Fully-Connected Layer-1	1024×1				
FC_2	Fully-Connected Layer-2	1024×1				
OUT	Output Layer	$N_I \times 1$				

for each subject, 10 acquisitions are made, each consisting of a video sequence of finger-vein images, from the left hands of $N_I = 100$ subjects, 33 female and 67 male. It is worth pointing out that, since the acquisition modality proposed in this study is novel, neither other datasets nor other methodologies are available for performance comparison. Nonetheless, in the tests carried out are included the comparison between the identification performance guaranteed by the proposed V-CNN network, when single-image finger-vein acquisitions are considered. In more detail, the results proving the effectiveness of using multiple-exposure cameras to capture the considered biometric trait, with respect to the standard usage of a single camera, are reported.

During the experiments, preprocessing has been executed using MATLAB ™ (R2017b) and PyTorch 0.4.0 has been chosen to build network architectures with a system configuration of 32Gb RAM, NVIDIA™ Titan V graphics card, i7-3.4GHz processors, and Windows™ 10 operating system.

Usage of Multiple-Exposure Finger-Vein Images. The benefits deriving from using multiple images taken at different exposures for finger-vein-based biometric recognition have been already evaluated in [228]. The system architecture here proposed for on-the-fly finger-vein-based identification stems from the results there reported, which are here further reinforced upon the considered novel scenario.

Specifically, in Table 3.10 the performance achievable when using a single camera are compared against the joint usage of the images taken at the four considered exposures. The reported results are referred to the use of only the V-CNN architecture for identification, selecting, at each iteration:

- images from five acquisitions for training;
- images from one distinct acquisition for validation:
- the finger-vein images from the remaining four acquisitions of each identity for testing,

and performing a 5-fold cross-validation. Out of the five acquisitions selected for training, a different number N_T of samples has been selected at each iteration to evaluate the achievable identification performance at the increasing of the training set size.

From the rank-1 identification accuracy given in Table 3.10 it is possible to confirm the observations in [228], noticing that an HDR image allows achieving performance better than the individual usage of singe-exposure

Table 3.10: *Mean identification accuracy with single exposure inputs (E_i) vs. their joint usages as HDR and 4-layer tensor.*

Training Acquisitions	E_1 ($12\mu s$)	E_2 ($16\mu s$)	E_3 ($20\mu s$)	E_4 ($24\mu s$)	HDR Input	Tensor Input
2	80.95%	81.21%	80.99%	80.75%	82.62%	90.18%
3	88.76%	90.71%	90.33%	89.35%	91.08%	94.89%
4	92.19%	93.86%	93.14%	93.39%	94.65%	96.61%
5	93.94%	95.71%	95.74%	95.81%	96.16%	97.62%

Table 3.11: *Mean identification accuracy comparison of fusion techniques based on DF, SF, and LSTM, over V-CNN features.*

Number of Training Acquisitions	V-CNN		V-CNN+DF		V-CNN+SF		V-CNN+LSTM	
	HDR Input	Tensor Input	HDR Input	Tensor Input	HDR Input	Tensor Input	HDR Input	Tensor Input
2	82.62%	90.18%	86.22%	92.18%	86.22%	92.68%	87.84%	93.85%
3	91.08%	94.89%	93.58%	96.44%	93.78%	96.44%	95.16%	97.10%
4	94.65%	96.61%	96.54%	97.65%	96.94%	97.75%	97.36%	98.29%
5	96.16%	97.62%	97.94%	98.50%	98.04%	98.45%	98.41%	99.13%

images. More interestingly, it can be seen that exploiting the collected data as a 4-layer tensor in the proposed V-CNN guarantees even further improvements. The results obtained when considering images taken at a single exposure also highlight that, in addition to the shape of the captured fingers, which remain the same independently of the exposure, the proposed system exploits the finger-vein patterns for recognition purposes. In fact, the use of finger-vein images acquired at different exposures, which correspond to different imaging of the captured vein patterns, impacts the identification performance. It is worth remarking that all the experiments have been made without applying enhancement techniques on the finger-vein inputs. This is another benefit of the proposed approach.

Exploitation of the Temporal Information. The use of the "on-the-fly" acquisition protocol allows recording also temporal information about the acquisition process related to sliding the hand over the sensors. As a matter of fact, the LSTM network described in [157] has been designed to exploit the temporal evolution of the discriminative features during each acquisition. In order to show the effectiveness of the proposed approach, two alternative methods for fusing the spatial information derived from multiple frames are considered, yet without exploiting any temporal information. Specifically, a score-level fusion (SF) as well as a decision-level fusion (DF) strategy are implemented over the features extracted by the CNN processing individual frames. In more detail, SF is performed by averaging the likelihoods obtained as predictions from the CNN models for each of the nine separate frames of an acquisition. Majority voting is instead performed to implement DF once the predictions for each frame are provided by the CNN.

As reported in Table 3.11, both SF and DF increase the accuracy in the order of 2%-5%, when considering two acquisitions for training. Nevertheless, adding more training samples slightly reduces the relevance of the improvement on the identification accuracy. In addition, both SF and DF result in similar patterns over the identification performance, which means that the two approaches are not significantly different from each other ($p-$ value $= 0.416$ in terms of paired t-tests).

The effects of the exploitation of the temporal information through the proposed CNN-LSTM framework are shown Table 3.11, where the accuracy achieved by V-CNN are reported. The proposed CNN-LSTM network is able to exploit the temporal behaviour of the hand movement better than simple strategies based on SF or DF.

3.3.3 Conclusions

This study proposes, for the first time in literature, an innovative on-the-fly finger-vein-based biometric recognition system that allows a user being identified while swiping the hand over the sensor without requiring any contact. The acquisition module has been built using low-cost sensors and it has been designed to allow free hand movement, with consequent high user convenience during both enrolment and recognition. Multiple cameras with different exposure times, capturing also the dynamic movement of the hand over the sensors, have been used and a database comprising swiping hands of 100 subjects has been collected. The proposed approach has exploited both the still images acquired at different exposure times and the temporal behaviour of the moving hand over the sensors. Deep learning approaches have been used in both scenarios.

In detail, the reported analysis shows that the use of multiple-exposure data increases the recognition accuracy with respect to the use of single exposure images and that the exploitation of multi-channel LDR images taken at different exposure times, as raw input templates, leads to further improvements of the identification accuracy.

In addition, for the first time in the literature, the temporal information related to the user swiping the hand over the sensors has been exploited, and is has been shown that when CNN topologies are used for feature extraction, and LSTM networks are fed by the sequential features based on hand movements, a significant identification accuracy improvement is observed.

CHAPTER 4

Finger-Vein Biometrics: Security and Template Protection

UNPROTECTED STORAGE of biometric reference templates poses severe privacy threats, e.g. identity theft, cross-matching or limited renewability. In fact, biometric data are defined as sensitive data within the European Union (EU) General Data Protection Regulation 2016/679 [49, 83], which means that the use of these data is subjected to the right of privacy preservation. For this reason, most of the commercial biometric systems available today do not store the sensed physical characteristics in their original form but, instead, they store their digital representation, that is the template, in an encrypted format. This serves two purposes. First, the actual physical characteristic cannot be recovered from the digital template thus ensuring privacy. Second, the encryption ensures that only the designated application can use this template. Anyway, also the template storage hides some pose risks to the security of the system, as detailed in Section 1.6.

The invisible biometrics, topic of this thesis, are not immune to security and privacy issues related to the template storage. Biometric template protection schemes have been developed in the last two decades [32, 224],

and several standardisation efforts [120,121,249] have been directed to this topic. As detailed in Section 1.6, there are different template protection techniques which can be exploited in order to secure a biometric system.

In this thesis, the issues of cross-matching between different biometric databases where the template is stored and limited renewability of finger-vein pattern are considered. Specifically, in Section 4.1 the similarity between vein patterns of symmetric fingers of the left and the right hand of a subject is investigated. Demonstrating that the vein patterns from different hands are "sufficiently" different, implies demonstrating that each finger of each user can be considered as different class. In this way, the renewability and the revoke of a compromised template can be considered less challenging because compared to the case of cross-similarity between different fingers.

Besides, an application of biometric template protection schemes is presented. The considered template protection technique belongs to the category of cancelable biometrics. More in detail, non-invertible transformation schemes are considered and applied in the field of finger-vein based biometric recognition. Specifically, Section 4.2 presents the security and privacy issues related to vein recognition and a possible solution to countermeasure the leakage of the aforementioned requirements based on cancelable biometric algorithms.

4.1 Cross-finger Similarity of Vein Patterns

In the field of biometric recognition, it is of crucial importance to demonstrate the individuality of a biometric trait, that is the biometric trait taken into account must be unique to each individual. When talking about vein pattern, it has been empirically proved the uniqueness of the aforementioned identifier for each identity, i.e. person, but it is also important to study what happens when performing the match between the right and left hand vein patterns for the same identity. The study of the similarity between veins of different hand is particularly important when talking about security and revocability: if vein patterns from different hands are "sufficiently" different, that is if the inter-class variation between the considered features is high according to the chosen similarity measure, it means that vein patterns belonging to different hands can be considered as different classes, that is different users.

The aforementioned aspect has not been analysed so far in literature for finger-vein-based biometric applications, while it has been investigated when using palmprint [151, 328]. The research has highlighted the pres-

ence of shared patterns between the palmprints of both hands of a person, allowing a user to be recognised through his/her left palmprint even when only the other one has been recorded during enrolment.

In this Section, the similarity between vein patterns of symmetric fingers of the left and the right hand of a subject is investigated. More in detail, the effects on the recognition performance when using symmetric fingers and geometry- and deep-learning-based feature extraction methods are analysed. The aim of the study is to explicitly asses whether it could be possible, for recognition purposes, to consider pairs of symmetric fingers of a subjects as a single class. In order to perform a comprehensive analysis, the SDUMLA database [338] has been considered in the conducted tests. Moreover, four different methods, belonging to the geometry-and deep-learning-based categories, have been exploited to derive the employed finger-vein feature representations.

4.1.1 Experimental Protocol

In order to verify whether finger-vein patterns of different hands of the same subject have a higher degree of similarity than traits belonging to different persons, several tests have been performed. Each test consists on the estimation of the distributions of scores obtainable by comparing different classes of biometric samples, specifically:

- *genuine* scores are obtained by comparing vein patterns from the same finger of the same hand of the same subject. For instance, vein patterns of the right index of a subject are compared between themselves;

- *impostor* scores are obtained by comparing veins from the same finger of the same hand of different subjects. For instance, patterns of the right index of a subject are compared with those of the right index of a different person;

- *genuine cross-hand (CH)* scores are obtained by comparing veins from the same finger of different hands of the same subject. For instance, the right index of a subject is compared with the left index of the same person;

- *genuine cross-finger (CF)* scores are obtained by comparing veins from different fingers of the same subject. For instance, patterns of the right index are compared with those of the right/left middle finger of the same person;

Table 4.1: *Score distributions evaluated in the performed tests.*

Case	Subject	Hand	Finger	Scores
1	same	same	same	genuine
2	different	same	same	impostor
3	same	different	same	genuine CH
4	same	same/different	different	genuine CF
5	different	same/different	different	impostor CF

- *impostor cross-finger (CF)* scores are obtained by comparing veins from different fingers of different subjects. For instance, the right index of a subject is compared with the left/right middle finger of another person.

Table 4.1 summarises the aforementioned combinations considered for the required scores.

On the basis of the computed distributions, the false rejection rate (FRR) and the false acceptance rate (FAR) related to different scenarios have been evaluated:

- *Test-1*: standard scenario where each finger from each hand is taken as a separate class, FRR and FAR are derived by considering respectively the aforementioned genuine scores and impostor scores;

- *Test-2*: a naïve scenario where an impostor uses a finger different from the one enrolled by the legitimate user is taken into account. FRR and FAR respectively from genuine scores and impostor CF scores are evaluated;

- *Test-3*: in order to verify whether a subject could be recognised by using as authentication probe the same finger of a hand different from the enrolled one, FRR and FAR are derived by considering respectively genuine scores and genuine CH scores;

- *Test-4*: the feasibility of using interchangeably the same finger of different hands to be recognised is further investigated by evaluating the FRR and FAR computed respectively on genuine CH and impostor scores;

- *Test-5*: eventually, the possibility of using as authentication probe fingers different from the enrolled one is also evaluated by deriving FRR and FAR respectively from genuine scores and genuine CF scores.

In order to obtain results from which reliable conclusions could be derived, the aforementioned score distributions have been computed according to several distinct processing methods described in the following paragraph.

4.1.2 Finger-vein Recognition Methods

Score distributions have been computed by considering several different recognition methods, belonging to both geometry- and deep-learning-based approaches.

Geometry-based Finger-vein Recognition

Since the original vein images are typically characterised by low contrast, they are first enhanced in order to improve their quality using a contrast limited adaptive histogram equalization (CLAHE) [353]. Finger boundaries are then obtained by filtering the image with a mask [167]. The finger is then rotated and aligned to the image centre as described in [113]. Eventually, the images of the fingers corresponding to the right hand are flipped in order to achieve a geometry symmetry between fingers of left and right hands.
Finger-vein patterns are extracted from finger areas using the following geometry-based feature extraction methods:

- *Maximum Curvature (MC)* [206]: scores related to veins width and curvature are assigned to positions where vein centres are located, which are then connected using filtering operations. Binary vein images are then obtained by thresholding the computed patterns;
- *Principal Curvature (PC)* [46]: he image gradient field is computed, and noise components filtered out by means of hard thresholding. Values of principal curvature are first computed by considering the eigenvalue corresponding to the eigenvector of the Hessian matrix related to the maximum curvature, and then binarized to generate the desired template;
- *Wide Line Detector (WLD)* [113]: vein positions are extracted by considering circular neighbourhoods of each pixel, and computing differences between the centre and its neighbours. The final binary image

is determined by counting the number of pixels inside this neighbourhood.

The obtained binary vein patterns are trimmed and then compared using the correlation-based method proposed in [205] and [206], with the maximum correlation used as matching score.

Deep-learning-based Finger-vein Recognition

Along with standard geometry-based recognition methods, tests have also been performed exploiting convolutional neural networks (CNNs) to obtain discriminative representations from finger-vein images. An effective CNN, namely Densenet-201 [117], has been employed in the tests. Specifically, the final layers of a Densenet-201 architecture, that is, those performing classification after the extraction of discriminative features, have been substituted with:

- a batch-normalisation layer, followed by a dropout regularisation with 50% of hidden units dropped;
- a fully-connected and a batch-normalisation layers producing C outputs, being C the number of unique identities considered for training.

Densenet's weights have been initialised with those estimated for an image classification task over Imagenet [267], while a unit weight initialisation has been adopted for the batch normalisation layer, and Glorot uniform initialisation preferred for the fully-connected layers. The layers have been then updated using a cross-entropy (CE) loss function for back-propagation, with stochastic gradient descent (SGD) and a batch size of 64. Learning rate has been set to $\epsilon = 0.01$ and divided by 10 after each 30-epoch iteration. Momentum with $\alpha = 0.9$ has been used, as well as an L_2 weight decay regularisation penalty with $\lambda = 0.025$. The maximum number of training epoch is set to 90, with early stopping in case the validation loss is minimised. In the testing phase, the features extracted by the employed network from two input finger-vein samples are compared by evaluating a cosine distance as score, to make genuine/impostor verification.

4.1.3 Results and Discussion

The equal error rates (EERs) achieved with the considered recognition methods for each of the test conditions presented in the previous paragraph

Table 4.2: *EERs (in %) over the SDUMLA database for the performed tests.*

Method	Test-1	Test-2	Test-3	Test-4	Test-5
MC [206]	8.94	8.36	9.93	46.73	9.73
PC [46]	11.70	11.07	12.81	46.64	12.95
WLD [113]	13.66	12.72	14.50	45.56	14.66
CNN	1.02	0.54	1.73	32.62	1.69

(4.1.1) are reported in Table 4.2. It is worth mentioning that, since the geometry-based approaches do not require any specific training, the associated performance has been computed considering all the available $106 \times 3 \times 2 = 636$ classes. Conversely, the results regarding the proposed CNN-based approach have been obtained while reserving the first half of the 106 subjects in the SDUMLA dataset for testing purposes, with the remaining 53 subjects used for CNN training. More in detail, two different training methodologies have been considered:

- to compute the scores associated with the distributions employed for Tests 1-3 and 5, the network has been trained with each finger of each hand of 53 subjects representing a different class. A total of $C = 53 \times 3 \times 2 = 318$ finger-vein classes have been therefore taken into account in this case. For each class, five out of the six available samples have been used for model training, with the remaining one employed for model validation;
- the scores of the distributions used for the considered Test 4, where the feasibility of using interchangeably the same finger of different hands to be recognised is analysed, have been generated considered a network trained with the same fingers of different hands taken as the members of the same class. A total of $C = 53 \times 3 = 159$ classes have been therefore taken into account in this case. As left-right finger samples are put in the same category, each class is now represented with a total of 12 samples, 10 of which are fed into the model for training, while the remaining 2 samples are used for validation. Doing this, the network is trained to look for similarities between same fingers of different hands and associate them to the same class, thus allowing to evaluate the existence of such shared patterns.

The obtained results show that a pair of same fingers from different hands do not posses similarities that allow the user to be recognised when

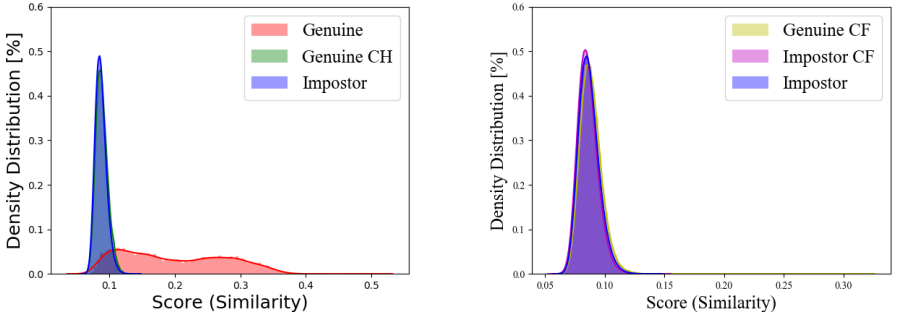


Figure 4.1: *Scores for the **MC**-based method. Left: genuine, impostor, and genuine CH distributions; Right: impostor, impostor CF, and genuine CF distributions.*

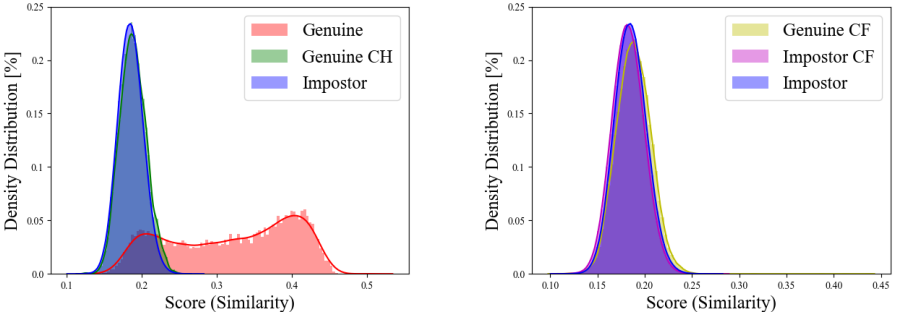


Figure 4.2: *Scores for the **PC**-based method. Left: genuine, impostor, and genuine CH distributions; Right: impostor, impostor CF, and genuine CF distributions.*

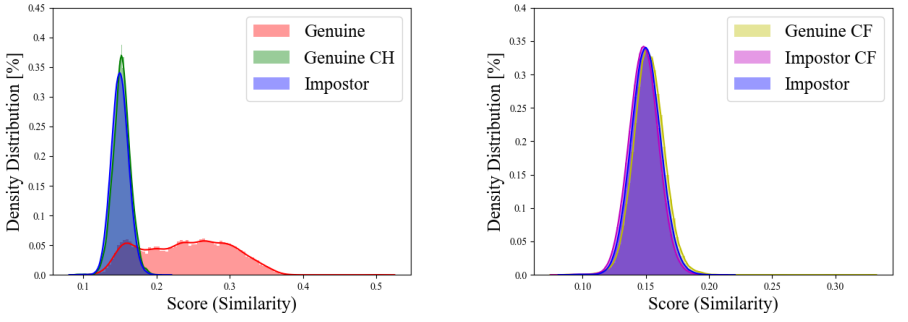


Figure 4.3: *Scores for the **WLD**-based method. Left: genuine, impostor, and genuine CH distributions; Right: impostor, impostor CF, and genuine CF distributions.*

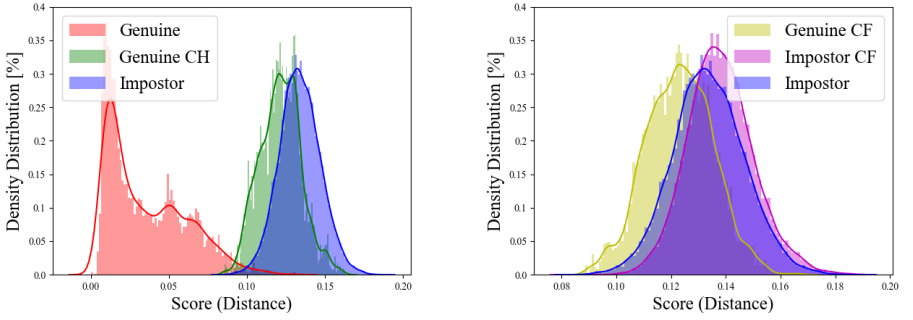


Figure 4.4: *Scores for the **CNN**-based method. Left: genuine, impostor, and genuine CH distributions; Right: impostor, impostor CF, and genuine CF distributions.*

Table 4.3: *Kullback-Leibler divergences with respect to genuine scores.*

Method	impostor	genuine CH	genuine CF	impostor CF
MC [206]	2.232	2.129	2.144	2.297
PC [46]	1.877	1.761	1.747	1.956
WLD [113]	1.718	1.658	1.633	1.790
CNN	0.237	0.106	0.058	0.281

one finger is used for enrolment and the other one for recognition. This is evident by comparing the EERs achieved in Test-1 and Test-3, which are basically the same, meaning that scores generated by comparing same fingers from different hands of the same subject are similar to those obtained when comparing same fingers of different subjects. Actually, the former comparison seems to find some more similarities than the latter, as testified by the slightly worse EERs.

Yet, such similarities cannot be assumed to be significant. Training a CNN while considering CH fingers as belonging to the same class further reinforce this considerations, as shown by the notably-high EER achieved in Test-4, which means the CNN cannot find shared patterns between pairs of fingers associated to the same class.

Interestingly, results in Test-5 also show that different fingers of the same subject share slightly more similarities than the same finger of different persons. Eventually, results in Test-2 suggest that the same fingers of different individuals are more similar than different fingers of different persons. Such resemblance may not necessarily spring from vein patterns, as it may depend on the geometric similarity of same fingers' shapes.

In order to provide further evidence of the observed behaviours, the computed score distributions are reported in Figures 4.1 - 4.4, where a training with 318 classes has been considered for the CNN-based approach. Genuine CH scores show basically the same distribution of impostor scores using geometry-based approaches, while resorting to CNNs highlights the existence of some similarities between pairs of symmetric fingers. CNNs are also able to generate slightly-different distributions for impostor, impostor CF and genuine CF scores, while geometry-based approaches cannot.

The distribution separations are quantitatively evaluated through the Kullback - Leibler divergences reported in Table 4.3, where the values obtained when evaluating the separation of impostor, impostor CF, genuine CH, and genuine CF distributions from that of genuine scores are considered. As can

be seen, for CNNs genuine CH and genuine CF scores are slightly closer to the genuine ones than the impostors, while impostor CF scores are even farther. In conclusion, the obtained results show that, although symmetrical fingers of the same subject show more resemblance than same fingers from different persons, such similarities are not significant enough to be exploited for recognition purposes.

4.2 Towards Practical Cancelable Biometrics for Finger Vein Recognition

The increasing adoption of biometric systems as a solution for reliable user authentication is mainly linked to the several advantages this technology exhibits over traditional authentication methods, such as password- and token-based ones. Together with the rapid growth of computing technologies, there is an increasing need for reliable authentication.

Using biometric data for recognition purposes may also involve severe security and privacy concerns. Due to their uniqueness, biometric characteristics can allow an attacker to track the activities of a subject whose characteristics have been registered in different systems [277]. Moreover, compromised biometric characteristics lose their usefulness and cannot be used anymore, with severe consequent constraints for their owners due to the limited number of usable identifiers. It is worth remarking that biometric characteristics can be often reconstructed from their corresponding templates [81]. Since biometric data cannot be revoked and reissued as it happens for disclosed passwords or stolen keys, proper countermeasures should be taken in order to address the aforementioned issues.

The wide spread use of biometric authentication imposes serious threats to the security and privacy of its users and the system should be able to guarantee the impossibility of leakage of information to unauthorised individuals. Between the possible solutions, as detailed in Section 1.6, biometric template protection (BTP) schemes have been therefore proposed to ensure the secure and private handling of biometric information during the authentication process.

According to the ISO/IEC 24745 standard on biometric information protection [121], a properly defined BTP scheme should satisfy the following properties to protect the privacy of the users:

- *irreversibility*: given a protected template, it should not be possible to reconstruct the original biometric sample;
- *renewability*: from a given biometric characteristic it should be possi-

ble to issue multiple protected templates;

- *unlinkability*: given two protected templates generated form the same sample and enrolled in different applications, it should not be feasible to determine that they belong to the same subject.

BTP schemes are categorised into two main classes: *biometric cryptosystems* and *feature transformation* approaches. The former class can be further separated into *key-binding* methods, whose aim is to secure a cryptographic key by means of biometric data and vice versa [106], and *key-generating* approaches, which derive a cryptographic key from biometric data [293]. Both aforementioned systems share the need of some helper data generated from the employed biometric information. Feature transformation methods instead apply a key-dependent transformation function to the biometric data or templates to be secured. In case invertible transformations are used, with system's security therefore relying only on the secret storage of the employed key, *salting* approaches are defined. Conversely, resorting to non-invertible transformations leads to the definition of *cancelable biometrics* [123]. This latter class of BTP scheme is considered in this study to secure the templates generated from an emerging biometric modality, namely the the vascular patterns of human fingers.

Cancelable biometrics have been successfully applied to many traditional biometrics, including fingerprint [25, 58, 250–252, 312, 330], face [25, 252, 269, 287], iris [94, 231, 232, 254, 354], palmprint [48, 172] and online signature [185, 186, 189] among others. Building upon the work proposed in [229], the effectiveness in generating cancelable biometrics from finger vein templates using three distinct approaches, namely *block remapping*, *image warping*, and *Bloom filters* are here evaluated. The two former methods can be applied to any biometric characteristic captured in the form of an image. They have originally been proposed as BTP scheme in the image domain for face [252] and for iris [94] as well as in the feature domain for fingerprints. Their use for finger veins has been also proposed in [141, 229]. Another advantage of block remapping and warping is that the same comparison algorithm used for the plain templates can be used to compare the protected templates as well. The third considered cancelable-biometrics approach can be applied to binary representations of the treated biometric data, as done for iris in [253] or face in [85].

In summary, the main contributions of the present work are:

- thorough evaluation and comparison, in terms of recognition performance, irreversibility, unlinkability, and renewability, of three distinct

cancelable-biometrics approaches applied to binary features extracted from finger vein patterns;

- application of the block-remapping and image-warping BTP schemes in the feature domain, differently than their usage in [66,229]. Specifically, six different feature representations of finger vein patterns are considered here, in order to evaluate the most appropriate to be used in protected biometric recognition systems;
- proposal of Bloom filters as a mean of template protection for binary finger vein features; in contrast to the previous work in [86], the Bloom filters are applied directly to the binary vein images instead of their application to a vein minutiae based representation.
- proposal of a pre-alignment method for improving the recognition performance attainable by the employed finger vein cancelable biometrics;
- exploitation of a specific attack, based on a square jigsaw puzzle solver algorithm, to quantify the irreversibility of the block remapping approach.

The rest of the Section is structured as follows: Section 4.2.1 briefly outlines the finger vein recognition framework used during our evaluations, including pre-processing, feature extraction, and comparison schemes. In section 4.2.2 the three cancelable biometrics approaches, namely block remapping, image warping, and Bloom filters, are described. The methodology employed for the security and privacy analysis with respect to irreversibility and unlinkablity is discussed in Section 4.2.4. Section 4.2.7 explains the adopted experimental set-up, describing the finger-vein database which has been used in the tests carried out, introducing the recognition and security performance evaluation methodology, and reporting the obtained results. Section 4.2.12 concludes this study and gives an outlook on future work.

4.2.1 Finger Vein Recognition

The standard finger vein recognition processing chain includes: capture of the input vein image, image pre-processing, feature extraction, and finally template comparison.

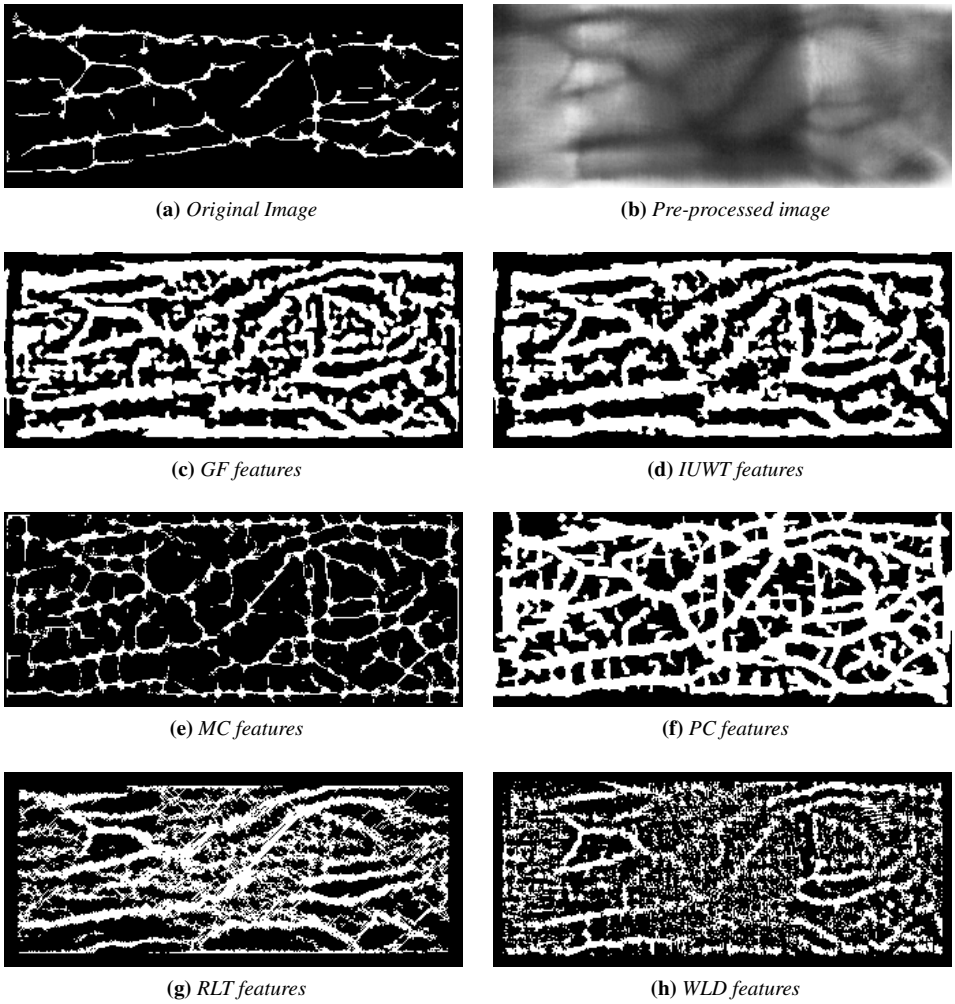
Pre-processing. A region of interest (ROI) detection algorithm is first applied to the acquired image to localise the area containing the relevant finger vein patterns, as proposed in [167]. In order to compensate possible

misalignments due to different finger positioning in distinct acquisitions, a normalisation step is then commonly performed [113]. However, only a coarse alignment can be thus typically achieved, with the need to compensate errors performing shifts in both directions during the comparison step. More importantly, such issue is much more severe when dealing with templates transformed through block-based BTP approaches like the ones considered here, thereby raising the need for a pre-alignment stage, as it is here discussed in Section 4.2.3.

To further improve image quality, the contrast of the ROI image is enhanced, and the non-uniform illumination conditions are compensated by applying contrast limited adaptive histogram equalization (CLAHE) [353], High Frequency Emphasis Filtering (HFEF) [348], and Circular Gabor Filtering (CGF) as proposed in [343].

Feature Extraction. Feature representations are typically derived from pre-processed images, with the aim of extracting discriminative information from them. In order to perform a proper comparison of the performance attainable through the three considered BTP approaches, different feature extraction algorithms have been benchmarked. All of them generate binary templates containing geometric information related to the shape or topological structure of the observed vein patterns. In more detail, the following six methods are considered:

- **Gabor Filtering** (GF) [152], inspired by the human visual system multi-channel processing of visual information. processes vein data through a bank of kernels to obtain distinct filtered images, then fused into a single representation to generate the desired binary template.
- **Isotropic Undecimated Wavelet Transform** (IUWT) [88] is a redundant wavelet transform whose coefficients encode information corresponding to different spatial scales. Levels 2 and 3 of such transform exhibit the best contrast for the blood vessels, and are therefore used to create the sought binary template.
- **Maximum Curvature** (MC) [206] extracts the lines corresponding to the central part of the veins, being therefore insensitive to varying vein widths.
- **Principal Curvature** (PC) [46] computes the eigenvalues of the Hessian matrix given by the gradient at each pixel to extract the principal curvature values of the input image, and use them to get a binary vein template.



(a) *Original Image* **(b)** *Pre-processed image*

(c) *GF features* **(d)** *IUWT features*

(e) *MC features* **(f)** *PC features*

(g) *RLT features* **(h)** *WLD features*

Figure 4.5: *Enhanced image and extracted features for GF, IUWT, MC, PC, RLT and WLD: MC extracts a thinner vein structure than the other methods.*

- **Repeated Line Tracking** (RLT) [205] tracks the veins in an image by computing, for each pixel, a statistical likelihood of belonging to a blood vessel. Thresholding is then applied to obtain a binary vein template.
- **Wide Line Detector** (WLD) [113] is an adaptive thresholding technique, comparing each pixel with its neighbourhood to determine which ones should represent veins in the final binary template.

Figure 4.5 shows a finger vein image with its corresponding pre-processed

version and the extracted features. The publicly available open-source implementation PLUS-OpenVein Toolkit[1] has been used to process finger vein images.

Matching. The binary templates generated from two vein images can be compared using a correlation approach as in [206]. As already mentioned, multiple shifted versions of the available data can be considered in an unprotected system to compensate possible misalignments, with the maximum of the computed correlations used as final comparison score.

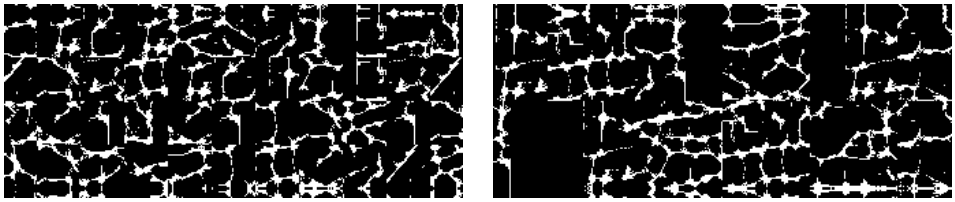
4.2.2 Finger Vein Cancelable Biometrics

The non-invertible transforms considered to generate cancelable biometrics from the finger vein binary templates described in Section 4.2.1 are detailed in the following subsections. All of them are key-dependent. In general, a biometric template protection system can be based on one of two types of keys: system-specific or user-specific keys. For the former type, the same key is used for the whole recognition system, and each user has the same key. In the user-specific key scenario, each user has an individual key, which has to be stored and retrieved whenever the user needs to be authenticated. This requires an additional key handling and secure storage of the user specific keys. For simplicity and due to space constraints only system-specific keys are evaluated in this work.

Block Remapping. A fixed-size region of $N \times W$ pixels, aligned to the centre of the finger area, is first extracted from the binary template. The selected region is then divided into B_T square blocks of $B \times B$ pixels, out of which a subset of B_C blocks are randomly selected. The chosen blocks are remapped according to a system-dependent pre-defined key [252]. The obtained distorted templates can be compared against each other in the transformed domain as described in the previous Section. Examples of cancelable biometrics generated through the block remapping scheme applied to MC features are given in Figure 4.6.

Since only a subset of the available blocks is included in the remapped template, the performed transformation is not a permutation, being therefore non-invertible. The selected blocks can be repeated multiple times in the transformed template, in order to obtain an image whose size is the same as the input one. The decisions about which blocks to consider, and their positions in the remapped template, depend on the employed key. Such key has to be the same for all the templates belonging to the users of a particular

[1] http://wavelab.at/sources/OpenVein-Toolkit/



(a) *Block size 32x32* **(b)** *Block size 64x64*

Figure 4.6: *Block remapping example with different block sizes using the same MC feature image as in Fig. 4.5e*

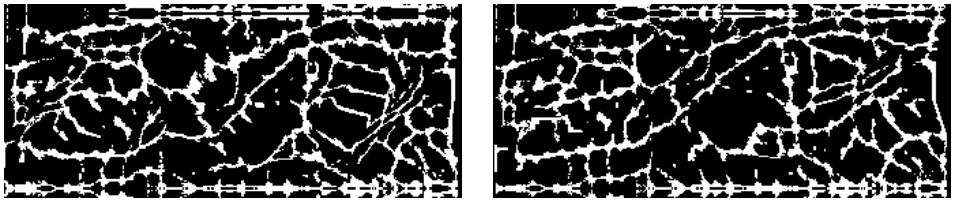
application. The different blocks may not contain the same amount of vein information. Thus, the recognition performance can be affected by such selection. Specifically, due to illumination issues, blocks extracted from the outer part of the finger generally contain less vein information compared to the ones belonging to the central area. Therefore a key-space reduction is performed considering only the selection of blocks belonging to the central area of the finger. The impact on recognition performance, as well as on the unlinkablity of the key-space reduction, is evaluated in Section **??**.

Image Warping. The *mesh warping* [320] algorithm can be applied to the binary vein templates in order to distort them. Specifically, a grid is laid over the representation, and its vertices are offset by amounts linked to the key defining the transformation. Each row and column of the template is then transformed performing a miniaturization or an expansion based on the distorted grid. Miniaturization is performed with the help of a box filter, while linear interpolation is used for expansion. The distorted templates, whose examples are shown in Figure 4.7, can still be compared as described in the previous Section.

Bloom Filters. According to the improved protection scheme proposed in [85], applying Bloom filters to generate cancelable biometrics requires three key steps:

1. *Feature extraction and encoding*: given a two-dimensional binary template originated from the collected biometric sample, it is divided into $nBlocks$ blocks having size $nBits \times nWords$.

2. *Structure-preserving feature re-arrangement*: in order to generate unlinkable templates, it is necessary to dissipate the information of the feature vectors among different blocks, while preserving verification accuracy. To this end, the $nBlocks$ blocks of the original unprotected



(a) *Block size 32, max. offset 12* **(b)** *Block size 64, max. offset 24*

Figure 4.7: *Block warping example with different parameters using the same MC feature image as in Fig. 4.5e*

template are re-grouped into $nGroups$ sets, each consisting of $nBlocks$-$Group$ blocks ($nBlocks = nGroups \times nBlocksGroup$). Then, the rows of the vertical concatenation of the $nBlocksGroup$ blocks are permuted within each set.

3. *Bloom filter computation*: protected templates are extracted by computing one Bloom filter $\mathbf{b}$ from each of the $nBlocks$ blocks, such that the final protected template $\mathbf{C}$ consists of $nBlocks$ Bloom filters of size 2^{nBits}: $\mathbf{C} = \left\{\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(nBlocks)}\right\}$. In order to map one block to a Bloom filter, the entire sequence of columns of each block is successively translated to their decimal value, and the corresponding indices are set to one in the Bloom filter.

The final comparison score s between a probe $\mathbf{C}_q$ and a reference $\mathbf{C}_r$ protected templates is defined as the average Bloom-filter-based dissimilarity score:

$$S(\mathbf{C}_q, \mathbf{C}_r) = \frac{1}{nBlocks} \sum_{i=1}^{nBlocks} \frac{HD(\mathbf{b}_q^{(i)}, \mathbf{b}_r^{(i)})}{|\mathbf{b}_q^{(i)}| + |\mathbf{b}_r^{(i)}|} \tag{4.1}$$

where $|\mathbf{b}|$ denotes the number of bits set to 1 within a Bloom filter $\mathbf{b}$, and $HD(\mathbf{b}_q^{(i)}, \mathbf{b}_r^{(i)})$ is the Hamming distance between two Bloom filters.

4.2.3 Pre-alignment for Template Protection

Misalignment of two templates in terms of shifts and planar rotations does not only cause problems during the comparison, but even more severe problems for all block-based cancelable biometrics schemes. While planar rotations and vertical shifts can be ruled out easily for finger veins (e.g. as described in the pre-processing Section 4.2.1), dealing with horizontal shifts is not as straightforward. Horizontal shifts are usually compensated during

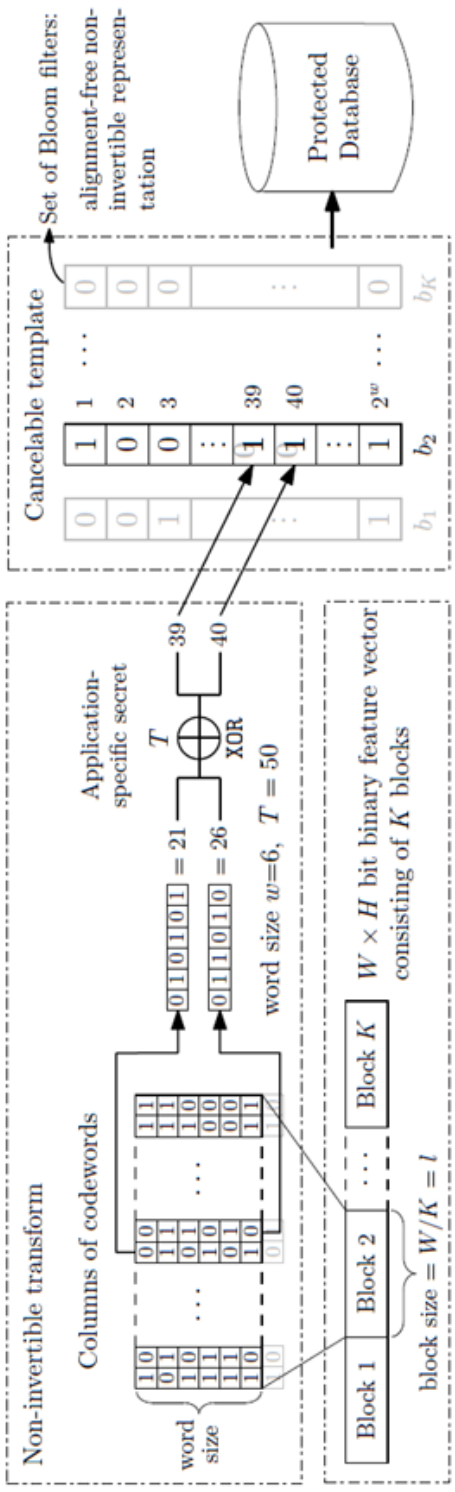


Figure 4.8: *Basic Bloom filter template protection approach, from [85].*

the comparison by shifting one of the templates. However, this strategy cannot be used for protected templates, especially if the shifts become larger than the block size of the cancelable scheme as feature information in different blocks is treated differently, leading to a dissimilar output template, and consequently to a recognition performance drop. Due to the nature of the applied non-invertible transforms, those dissimilarities in the protected templates cannot be compensated by shifts during the comparison. Thus, a suitable pre-alignment prior to the application of the block based cancelable scheme is needed.

Here is therefore proposed a pre-alignment strategy in which all templates originating from the same finger belonging to the same subject are first registered against each other with the help of a correlation-based approach, derived from the comparison scheme in [206] (see Section **??**). Specifically, besides the comparison score, the scheme also outputs the relative position of the two templates to each other. This position encodes the shifts needed during comparison to arrive at the highest score, i.e., the best possible alignment of the two templates. Hence, if the reference template is shifted according to that information, an alignment/registration of the templates can be done. This alignment is only done if the comparison score is above a pre-defined threshold in order to avoid ambiguous alignment results. This kind of pre-alignment is only a proof of concept, which needs to be further analysed and improved. It requires the unprotected templates to be available, hence violating BTP principles. The benefits of the proposed pre-alignment are confirmed during the recognition performance evaluation of the employed cancelable biometrics schemes in Section 4.2.8.

4.2.4 Security Analysis

In this section, the conducted security analysis is described in terms of the two main properties of unlinkability and irreversibility. To quantify the unlinkability, a general approach evaluating comparison scores is employed. Regarding the irreversibility, beyond general considerations about the considered BTP schemes, an exemplary attack against the block remapping approach, based on automated square jigsaw puzzle solvers is presented and evaluated and the irreversibility for block warping and the Bloom filter approach is briefly discussed.

4.2.5 Security Analysis: Unlinkability

Template unlinkability is evaluated according to the protocol proposed in [84], for which a public implementation is available[2]. Two protected templates $\mathbf{T}_1$ and $\mathbf{T}_2$ generated from the same biometric sample are defined as linkable if an attacker can determine that they were extracted from mated instances, and hence conceal a unique identity.

To accomplish such goal, the attacker compares the two protected templates by computing a linkage score $s = LS(\mathbf{T}_1, \mathbf{T}_2)$, upon which he should decide whether the considered templates actually stem from mated instances. Following Kerckhoffs's principle [139], it is assumed that the attacker knows how the system works and, in particular, the *mated-instance* and *non-mated-instance* score distributions generated by comparing protected templates. These distributions can be quantitatively compared by means of two different measures:

- a local measure $\mathrm{D}_{\leftrightarrow}(s)$, evaluating the linkability of templates on a score-wise basis. A measure $\mathrm{D}_{\leftrightarrow}(s_1) = 1$ for a specific linkage score s_1 means that an attacker will be able to link the considered templates to the same instance with full certainty. On the other hand, $\mathrm{D}_{\leftrightarrow}(s_0) = 0$ should be interpreted as full unlinkability for templates giving linkage score s_0. Intermediate values of $\mathrm{D}_{\leftrightarrow}(s)$ report an increasing degree of linkability;
- a global measure $\mathrm{D}_{\leftrightarrow}^{sys}$, giving an overall evaluation of the whole BTP scheme unlinkability. A system with $\mathrm{D}_{\leftrightarrow}^{sys} = 1$ should be fully linkable, meaning that mated-instance and non-mated-instance score distributions having no overlap, with local measures $\mathrm{D}_{\leftrightarrow}(s) = 1$ for linkage scores computed from any pair of mated samples. Similarly, $\mathrm{D}_{\leftrightarrow}^{sys} = 0$ means that the system is fully unlinkable, with mated and non-mated score distributions completely overlapping. All intermediate values of $\mathrm{D}_{\leftrightarrow}^{sys}$ report a decreasing degree of unlinkability.

As detailed in [84], given a linkage score s, the local measure $\mathrm{D}_{\leftrightarrow}(s)$ should indicate whether it is more likely that the two considered templates stem from mated instances, whose probability is $p(H_M|s)$ for hypothesis H_M, than from non-mated instances, characterised by probability $p(H_{NM}|s)$ for hypothesis H_{NM}. Therefore, $\mathrm{D}_{\leftrightarrow}(s)$ can be expressed as the difference of conditional probabilities for each hypothesis:

$$D_{\leftrightarrow}(s) = p(H_M|s) - p(H_{NM}|s). \tag{4.2}$$

[2] https://github.com/dasec/unlinkability-metric

The unknown conditional probabilities can be expressed through the known probabilities of observing s given templates belonging to mated or non-mated samples, that is, $p\left(s|H_M\right)$ and $p\left(s|H_{NM}\right)$, rewriting eq. 4.2 in terms of the likelihood ratio between them, $LR\left(s\right) = p\left(s|H_M\right)/p\left(s|H_{NM}\right)$, as:

$$\mathrm{D}_{\leftrightarrow}\left(s\right) = \begin{cases} 0 & \text{if } LR\left(s\right)\cdot\omega \leq 1 \\ 2\frac{LR(s)\cdot\omega}{1+LR(s)\cdot\omega} - 1 & \text{if } LR\left(s\right)\cdot\omega > 1 \end{cases} \tag{4.3}$$

where $\omega = p\left(H_M\right)/p\left(H_{NM}\right)$ denotes the ratio between the prior probabilities of the mated and non-mated samples distributions. As discussed in [84], this latter constant ratio can be assumed to be known for operating systems with registered mated and non-mated access attempts. Alternatively, it can be set to $\omega = 1$ as in the present analysis.

The global linkability measure is instead computed measuring how likely it is to get a linkage score stemming from the mated samples distribution, being then defined in [84] as

$$\mathrm{D}_{\leftrightarrow}^{sys} = \int_{s_{min}}^{s_{max}} p\left(s|H_m\right)\cdot\mathrm{D}_{\leftrightarrow}\left(s\right)\mathrm{d}s \tag{4.4}$$

4.2.6 Security Analysis: Irreversibility

Measuring the irreversibility of a BTP scheme means evaluating the amount of information regarding the original biometric template or sample which the protected one leaks. There are two different scenarios, one is if the template protection key is known to the attacker, i.e. it has been compromised, and the other one is if the key is not known to the attacker.

As for the block remapping scheme, irreversibility has an upper bound given by the number of blocks selected from the original template to produce the protected one. A coalition attack, where an attacker tries to overtake this limit exploiting the knowledge of multiple protected templates, each containing different parts of the original biometric data, has been proposed in [127] and applied to iris representations protected with this BTP scheme.

Yet, the original template could be partially reconstructed even from only one of its protected version through a simple brute-force attack, where all possible block permutations are tested until the correct block order is found. However, in order to conduct such attack without the knowledge of the original content, an indicator is needed to establish whether a block is set at the correct position or not. This could be done by comparing the border pixels of the available blocks, and searching for the best match

among possible combinations. Such task essentially looks like solving a square jigsaw puzzle, an issue for which several automated procedures can be found in literature [45, 220] As the protected templates do not contain all blocks of the original template, an appropriate reconstruction approach should be able to deal with missing parts. This can be done using the greedy placement strategy and the prediction-based dissimilarity metrics proposed in [220], for which a public implementation is available[3]. This approach does not require any prior knowledge about the original data, and is able to handle puzzles with missing pieces, pieces of unknown orientation, and unknown overall size, as well as pieces from multiple input puzzles. The performance of a square jigsaw puzzle solver can be measured according to either global and local metrics [45]. The former compare original and reconstructed contents quantifying the number of blocks at the correct position. The latter focus on clusters of blocks, rating either the biggest correct block cluster or the number of correct block pairs, that is, blocks with at least one correct neighbour. Given that, in the scenario here considered, the amount of information leakage does not depend on the absolute block positions but on the continuity of the vein lines (local clusters), the two mentioned local metrics are employed in the evaluation of the block remapping robustness against such attack reported in Section 4.2.10.

Regarding the block warping scheme, in case the key is not known to the attacker he can try to derive it, or at least some hints about it, by analysing the interpolation artefacts using image forensic methods [19]. Depending on the strength of the applied warping and the number of available protected templates, the key can be restored with a certain probability. An exact quantification as it is performed for the block remapping is out of the scope of this work though. However, also in case the key is known to the attacker, the applied mesh warping transformation can be considered irreversible, since interpolation strategies are applied. Thus, it is not possible to completely recover the original data even if the warping parameters are known. The level of irreversibility is higher if miniaturization is applied, due to overlay effects [94].

Regarding Bloom filters, as shown in [85], in a full disclosure model where the attacker knows both the stored templates and the employed key, the number $nSeq$ of possible original binary representations which could result in the same protected block features is given by

$$nSeq = \sum_{i=1}^{|\mathbf{b}|} (-1)^{|\mathbf{b}|-i} \binom{|\mathbf{b}|}{i} i^{nWords}, \tag{4.5}$$

[3] https://github.com/ZaydH/sjsu_thesis

This number reaches rapidly for even for small values of the number $|\mathbf{b}|$ of bits set to 1 within the considered block. This in turn leads to the desired irreversibility.

4.2.7 Experimental Tests

In this section, the finger vein database exploited for the conducted experimental tests is detailed, the parameters of the employed BTP schemes are described, and the employed recognition performance evaluation protocol is outlined.

Finger Vein Dataset. The University of Twente Finger Vascular Pattern Database (UTFVP) dataset [289] has been used for our experimental evaluations. UTFVP contains 1440 images in total, captured from 60 subjects, 6 fingers per subject (index, middle and ring finger) and 4 images per finger. The images have a resolution of 672×380 pixels, 8-bit greyscale, and are stored in the PNG format. The width of the visible vein lines inside the images is between $4 - 20$ pixels. The binary templates extracted from the pre-processed images have a size of $N \times W = 336 \times 142$ pixels.

Block Remapping. The input feature representation is divided into square blocks of size $B \times B$ pixels, with $B = 32$ and $B = 64$ considered in the performed tests. Cropping is performed in case the template dimensions are not multiples of the employed block size. As for the percentage of blocks to be included in the transformed template, it has been empirically evaluated that keeping 75% of the original blocks leads to the best trade-off between recognition performance and privacy. The selected blocks are rearranged according to a system-dependent key in the remapped template, whose size is the same as the original input.

Block Warping. In order to perform the mesh warping transformation, a regular grid of $G \times G$ pixels is laid on the image, and the maximum offset O of the vertices that the transformation is allowed to perform is selected. The maximum offset is limited by the block size, and should be smaller than half the block size to achieve a usable output. As for block remapping, cropping is performed in case the size of the original template does not allow to lay on it a grid with all equal elements. In particular, $G = 32$ with $O = 12$ and $G = 64$ with $O = 24$ have been tested.

Bloom Filters. The settings of the Bloom-filter-based protection scheme are determined according to the procedure detailed in [86]. Specifically,

$nBits$ is set to 10 to maximize irreversibility, while $nWords$ is selected within the allowed range as 48 to maintain proper recognition capabilities. The number of blocks in horizontal direction $nBlocksX$ is then set to 7, and the number of blocks in vertical direction $nBlocksY$ to 14, with $nBlocks = nBlocksX \cdot nBlocksY$. In addition, 4 different key permutations are used to add unlinkability to the Bloom filter approach.

4.2.8 Recognition Performance Evaluation

This Section presents the recognition performance obtained for baseline unprotected systems, and the performance achievable with the proposed cancelable biometric approaches, without and with the proposed pre-alignment procedure are outlined.

The system recognition performance is quantified in terms of the false non-match rate (FNMR) and corresponding false match rate (FMR). The variance of the obtained equal error rate (EER, the point where the FMR equals the FNMR) has been used as an indicator of the dependence of the recognition performance on the employed transformations key parameter. In addition, the FMR1000 (the lowest FNMR for FMR = 0.1%) and the ZeroFMR (the lowest FNMR for FMR = 0%) are used to quantify the recognition performance. For their calculation the test protocol of the FVC2004 [183] is followed.

Baseline Recognition Performance Results. The baseline results in terms of EER, FMR1000 and ZeroFMR for the six employed feature representations (IUWT, GF, MC, PC, RLT, and WLD) are listed in Table 4.4. The reported values show that MC performs best in terms of all three performance indicators, followed by PC, IUWT, GF, and WLD, while RLT performs worst.

Cancelable Schemes Recognition Performance Results. Table 4.5 reports the performance results regarding the considered cancelable biometrics schemes for the considered transformation parameters, expressed in terms of mean EER with the corresponding 95% confidence intervals. The DET curves showing the aforementioned results are reported in Figure 4.9. Fig. 4.10 also shows the impact of parameter selection for block remapping and block warping, when applied to MC feature representations. For block remapping, bigger block sizes are preferable for recognition purposes, with MC performing overall best. The same holds for block warping, where bigger block sizes lead to a better performance despite the higher maximum

Table 4.4: *Baseline results on UTFVP dataset.*

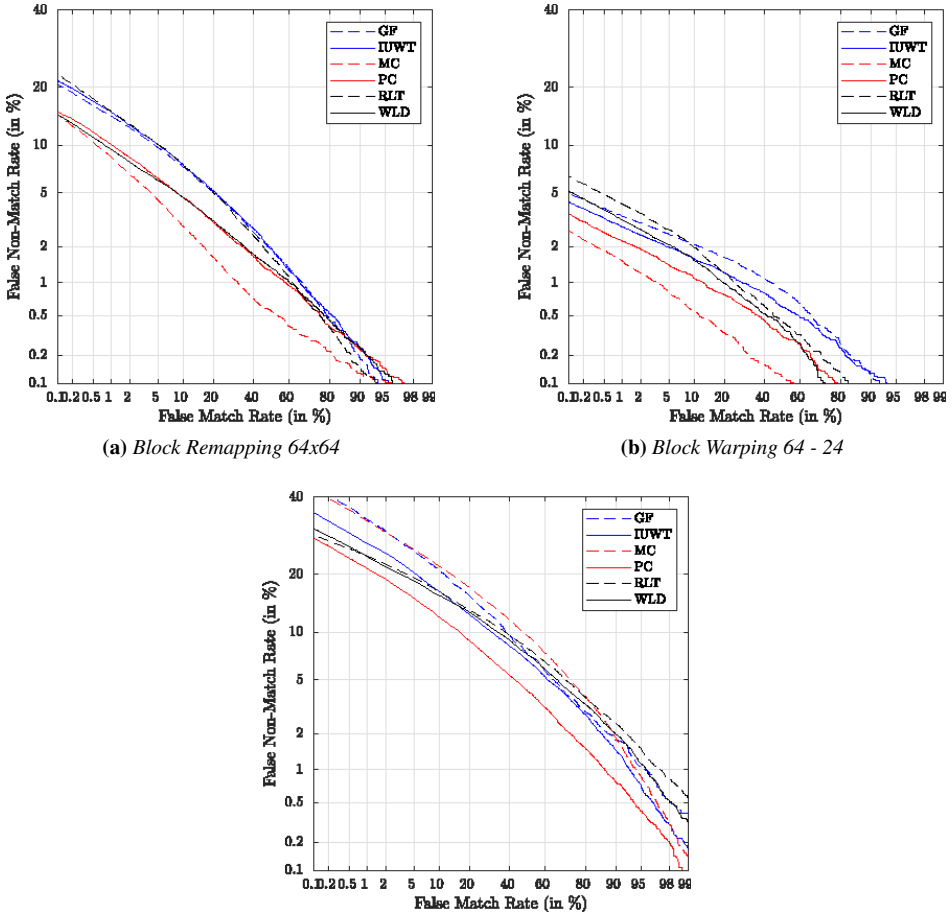
	GF	IUWT	MC	PC	RLT	WLD
EER	0.77%	0.77%	**0.36%**	0.57%	2.1%	0.72%
FMR1000	1.33%	1.28%	**0.51%**	0.92%	4.05%	1.28%
ZeroFMR	4.26%	4.41%	**2.15%**	4.21%	11.9%	6.1%

Table 4.5: *Recognition performance results in terms of EER and 95% confidence intervals for cancelable biometrics schemes applied on the UTFVP database, using 10 different transformation keys for each template.*

Scheme	Parameters	GF	IUWT	MC	PC	RLT	WLD
Block Remapping	32x32	6.58%±0.66%	6.99%±0.66%	5.26%±0.58%	**5.05%±0.57%**	8.33%±0.68%	5.47%±0.58%
	64x64	7.36%±0.55%	7.69%±0.56%	**3.96%±0.41%**	4.91%±0.47%	7.64%±0.55%	5.2%±0.69%
Block Warping	32 - 12	3.38%±0.43%	2.59%±0.39%	**1.25%±0.31%**	2.18%±0.36%	3.66%±0.42%	2.6%±0.39%
	64 - 24	1.57%±0.34%	1.25%±0.31%	**0.73%±0.23%**	1.29%±0.27%	2.04%±0.35%	1.11%±0.31%
Bloom Filters	48/10/7/14	16.6%±0.79%	13.5%±0.73%	18.3%±0.8%	**11.4%±0.67%**	14.7%±0.74%	14.2%±0.73%

Table 4.6: *Recognition performance results in terms of EER and 95% confidence intervals for cancelable biometrics schemes applied on the UTFVP database with feature pre-alignment, using 10 different transformation keys for each template.*

Scheme	Parameters	GF	IUWT	MC	PC	RLT	WLD
Block Remapping	32x32	1.76%±0.39%	2.22%±0.41%	**0.69%±0.25%**	1.34%±0.29%	0.83%±0.25%	1.3%±0.3%
	64x64	1.99%±0.33%	2.17%±0.34%	**0.7%±0.21%**	1.02%±0.24%	1.47%±0.26%	1.2%±0.26%
Block Warping	32 - 12	0.69%±0.18%	0.74%±0.18%	**0.27%±0.11%**	0.56%±0.16%	0.83%±0.18%	0.37%±0.14%
	64 - 24	0.74%±0.18%	0.6%±0.18%	**0.23%±0.12%**	0.41%±0.16%	0.6%±0.18%	0.32%±0.14%
Bloom Filters	48/10/7/14	7.04%±0.753%	2.39%±0.31%	2.23%±0.33%	**1.25%±0.26%**	2.22%±0.32%	3.62%±0.4%



(a) *Block Remapping 64x64*

(b) *Block Warping 64 - 24*

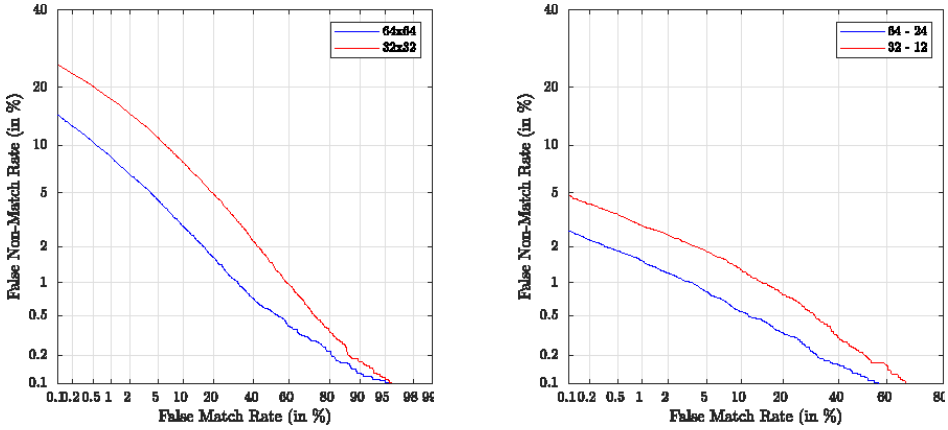
(c) *Bloom Filter*

Figure 4.9: *DET curves for all six feature types and the different template protection schemes: (a) block remapping, (b) block warping and (c) Bloom Filters on the UTFVP database.*

offset. The best performance using Bloom filters is achieved for the PC-based features, yet with overall results far worse than those achieved with the block remapping and the block warping BTP schemes.

It can be observed that template protection significantly degrades the achievable recognition performance, while the employed transformation keys introduce only limited variability in the obtained results.

Cancelable Schemes Performance Results with Pre-Alignment. The effectiveness of the alignment approach proposed in Section 4.2.3 for the employed cancelable schemes is confirned by the results given in Table 4.6. It



(a) *Block Remapping* **(b)** *Block Warping*

Figure 4.10: *DET curves showing the impact of different block remapping (left) and block warping (right) parameters for MC on the UTFVP dataset.*

is evident that pre-aligning the considered finger vein patterns significantly improves the results attainable for block-remapping, block-warping, and Bloom-filter BTP approaches, allowing to obtain recognition performance close to baseline unprotected systems. When adopting pre-alignment, applying Bloom filters to PC features even outperforms almost all block-remapping combinations, with the only exception related to the use of MC features. For block warping, a consistent improvement in terms of recognition accuracy can be achieved for all six feature types and all employed transformation parameters.

4.2.9 Unlinkability Analysis

The local unlinkability measure $D_{\leftrightarrow}(s)$ is computed for selected combinations, with the obtained results depicted as blue curves in Figure 4.11 for the considered cancelable biometrics approaches. The global measures $\mathrm{D}_{\leftrightarrow}^{sys}$ obtained for all the considered parameter combinations are instead listed in Table 4.7. As it may be observed, especially in the case of block warping and MC features, there is no big overlap between the mated (green) and non-mated (red) score distributions, being $p(H_m|s) > p(H_{nm}|s)$ for $s > 0.15$. The same happens in case of block remapping for $s > 0.22$. Accordingly, for those intervals $\mathrm{D}_{\leftrightarrow}(s) = 1$, as the templates are fully linkable. For block warping, since most of the weight of the mated instances score distributions lies in the aforementioned score interval, the global linkability of the systems $\mathrm{D}_{\leftrightarrow}^{sys}$ is 0.57, thereby showing that the con-

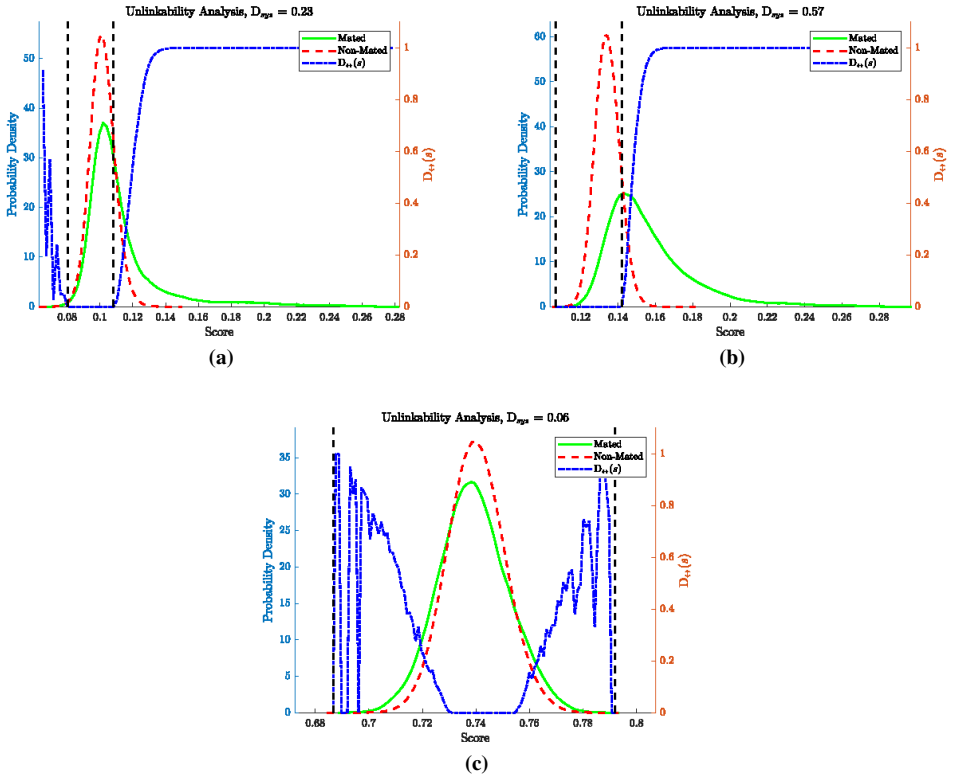


Figure 4.11: *Mated-sample (solid green) and non-mated-sample (dashed red) score distributions for protected templates generated from the UTFVP dataset. The blue curve represents the score-wise linkability measure* $\mathrm{D}_{\leftrightarrow}(s)$*, and* $\mathrm{D}_{\leftrightarrow}^{sys}$ *gives an estimation of the overall linkability level of the whole system. (a): Block remapping using MC and* $B = 64$*; (b): Block warping using MC,* $G = 64$ *and* $O = 24$*; (c): Bloom filters.*

sidered scheme do fulfil the unlinkability requirement only partially. The block remapping scheme instead shows proper unlinkability for smaller block sizes, with a notable worsening of the obtained performance for larger block values. The lowest linkability measures are obtained when employing Bloom filters as BTP scheme. The unlinkability requirement is therefore satisfied only by the block-remapping and the Bloom-filter BTP schemes, whereas the block-warping approach is not suitable for template protection in terms of unlinkability.

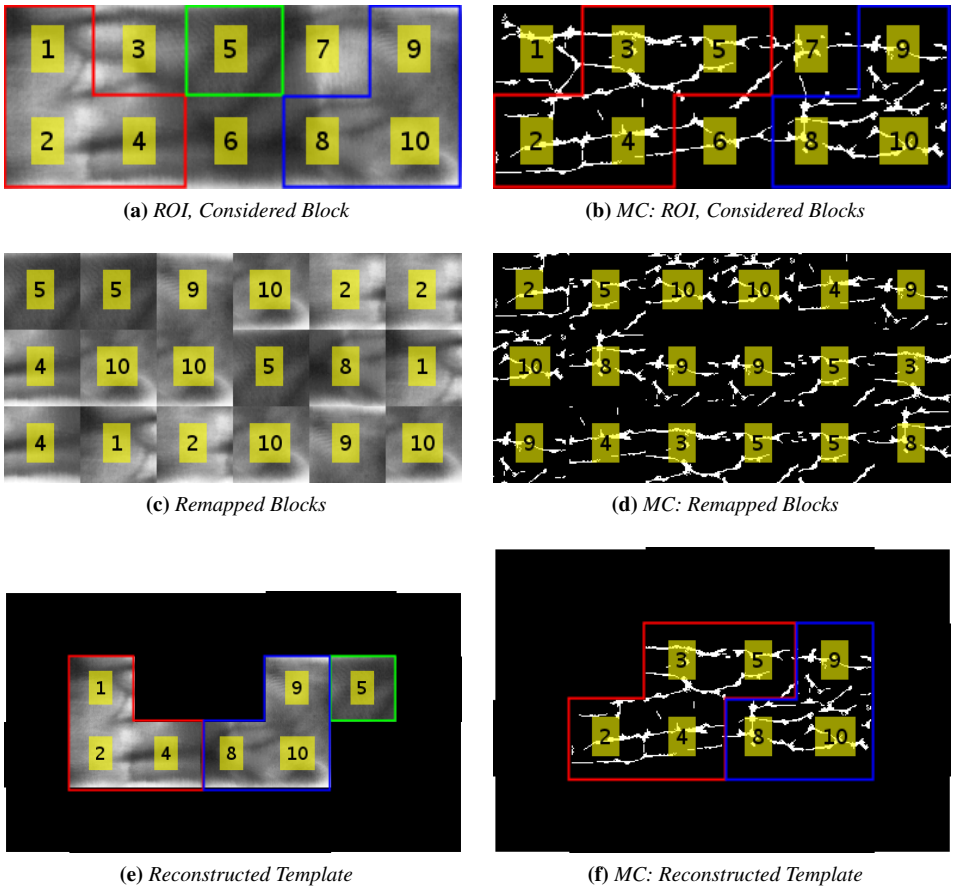
Table 4.7: *Unlinkability results in terms* $\mathrm{D}_{\leftrightarrow}^{sys}$ *for all cancelable schemes and all six feature types on the UTFVP dataset.*

Scheme	GF	IUWT	MC	PC	RLT	WLD
BM 32	0.016	0.02	0.079	0.09	0.033	0.076
BM 64	0.112	0.116	0.229	0.078	0.11	0.176
BW 32/12	0.435	0.498	0.658	0.572	0.52	0.501
BW 64/24	0.412	0.441	0.566	0.502	0.48	0.468
Bloom Filter	0.344	0.029	0.063	0.024	0.027	0.031

4.2.10 Irreversibilty Analysis

As already discussed in the previous section, the block-remapping approach shows proper unlinkability performance to be considered as BTP scheme. Its irreversibility is therefore here also analyzed in detail, exploiting the automated square jigsaw puzzle solver algorithm introduced in Section 4.2.6, of which Figure 4.12 depicts the process. Specifically, the top row of Figure 4.12 reports the original template. The blocks considered during remapping are grouped into regions of connected blocks, and their outline is highlighted. The middle row shows the block remapped images. This image consists only of considered blocks. The bottom row shows the square jigsaw puzzle solver reconstruction results. Again, this image consists only of the considered blocks. In a successful reconstruction, all block regions are restored. Due to the omitted blocks, an exact arrangement of the regions is not always possible. The blocks are marked with the same numbers across all three images. The amount of information from the original template (considered blocks) which can be restored hereby is directly linked to the irreversibility property of the template protection scheme, with the highest possible amount of reconstructed data r is $r = \frac{B_{pt}}{B_{ot}}$, where B_{pt} is the number of blocks considered in the protected template, and B_{ot} is the number of blocks contained in the original template.

Table 4.8 lists the results of the puzzle solver approach, averaged over each single run (key) and then again over all the 10 different keys. The values in the table are relative to the maximum possible amount of data that can be reconstructed (r), e.g., if only 7 out of 10 blocks are considered and the value in the table is 100%, this means that $100\% \cdot \frac{7}{10} = 70\%$ of the total unprotected template has been successfully reconstructed. For both kinds of considered local metrics, the best reconstruction results were obtained for PC features, where more than 90% of the information can be restored



(a) *ROI, Considered Block*

(b) *MC: ROI, Considered Blocks*

(c) *Remapped Blocks*

(d) *MC: Remapped Blocks*

(e) *Reconstructed Template*

(f) *MC: Reconstructed Template*

Figure 4.12: *Template reconstruction using PuzzleMultisolver [220], extracted features (MC) in the right column. Top row: ROI, labeled blocks and used blocks, middle row: remapped images, bottom row: puzzle solver reconstruction.*

Table 4.8: *Irreversibilty Analysis for Block Remapping on the UTFVP dataset*

Reconstructed Pairs						
Block size	GF	IUWT	MC	PC	RLT	WLD
32x32	72.4%	71.5%	47.3%	69.3%	47.5%	42.0%
64x64	93.1%	94.0%	87.9%	94.3%	90.4%	84.7%
Max. Reconstructed Region Size						
Block size	GF	IUWT	MC	PC	RLT	WLD
32x32	54.9%	53.7%	30.6%	51.4%	32.2%	28.1%
64x64	93.4%	95.1%	88.1%	95.7%	93.2%	84.7%
Perfect Reconstruction						
Block size	GF	IUWT	MC	PC	RLT	WLD
32x32	2.58%	2.33%	0.07%	1.99%	0.03%	0.03%
64x64	76.7%	82.4%	61.5%	85.2%	71.5%	54.4%

for all block size/considered blocks combinations. The lowest results were achieved using MC, for which however nearly 80% of the maximum possible template information could be always restored. This means that the level of irreversibility depends solely on the number of blocks considered, as almost all the considered blocks can be reconstructed to match the unprotected template. The difference in the reconstruction performance can be explained by the used methods: MC extracts only the centre lines of the recognized veins, whereas PC and WLD extracts wider vein patterns.

Considering the block size of 64×64, for $55 - 85\%$ of the images a perfect reconstruction was possible. Thus, if an attacker gets the original template and the protected one, he is able to reconstruct the key (i.e. the mapping information), which poses another threat for this kind of template protection scheme if a system-dependent key is used.

For BTP schemes based on Bloom filters, the success probability for an attack trying to recover the original unprotected features from their protected representation can be estimated as $\overline{nSeq}^{nBlocks}$, being $\overline{nSeq}$ the average number of possible sequences resulting in a single Bloom filter, defined in eq. (4.5). In the considered tests, the success probabilities for guessing the original unprotected templates range from 10^{-192} to 10^{-23}, therefore confirming the irreversibility of the employed Bloom filters for template protection.

4.2.11 Results Discussion and Summary

Regarding recognition capabilities, the block warping scheme performed best, as it achieved the highest recognition performance in terms of EER, followed by the block remapping scheme, with the Bloom filters achieving the worst performance. For both block remapping and warping there is a general trend of recognition performance improving with increasing block sizes. The proposed pre-alignment approach turned out to be beneficial for the all the block-based cancelable biometrics schemes, as it was able to improve the performance considerably.

Compared to the Bloom filter based approach for the spectral minutiae representation as presented in [86] our results for the protected templates without performing the pre-alignment are inferior. However, the detection and extraction of reliable minutiae points for finger veins is a difficult and error prone task. Hence, usually the binary representation are used as they achieve more reliable and stable results (cgf. our baseline EER of 0.36% on the UTFVP dataset compared to the baseline EER of 1.5% as reported in [86]. On the other hand, those binary representation have the aforementioned, inherent alignment problems in combination with all block based template protection schemes. This is confirmed by the fact that our results, in case the pre-alignment strategy is employed, are superior to the ones reported in [86] (cgf. 0.23% EER for Block Warping 64 - 24, and 1.25% EER for Bloom filters, instead of 2.1% EER reported in [86]).

In terms of unlinkability, block warping achieves the lowest security, thereby unveiling the method inadequacy as BTP scheme.

The employed puzzle-solver attack also showed that the block remapping scheme is not secure enough, since its irreversibility solely relies on the amount of blocks which are not considered, and even the key can be reconstructed under certain circumstances. Hence, in terms of security, the Bloom filter approach remains as the only effective solution.

The different feature types had an impact on the recognition performance and the security as well, but the general trend remained the same among all six tested feature types. There is a trade-off between recognition performance and security (in terms of unlinkability and irreversibility) observed for both the block remapping and the block warping approaches: changing the employed transformation parameters, the higher the recognition performance, the lower the level of security, and vice versa. Consequently, the decision regarding which kind of template protection scheme should be used depends on the current requirements: if recognition performance is more important, block remapping or warping should be applied,

whereas if security is the main concern, the Bloom filter approach has to be chosen. The results for the pre-alignment method confirmed that an alignment of the templates to be protected should be done in any case as it helps to retain a higher recognition performance. The proposed proof of concept alignment method does not comply with BTP principles. Therefore, it is necessary to further investigate a universal alignment method, based on some kind of finger landmarks.

4.2.12 Conclusion

In this study three different cancelable biometrics schemes for finger vein recognition are evaluated, namely block remapping, block warping and Bloom filters. Six different feature extractors of well-established vein recognition schemes, outputting binary templates were utilised to generate the plain templates. These templates were then protected using the three tested cancelable biometrics schemes. In addition, a pre-alignment approach prior to the application of the cancelable schemes is proposed and tested as well. The evaluation was conducted on two well known finger vein datasets, the UTFVP and the SDUMLA-HMT. Recognition performance, unlinkability and irreversibility were evaluated.
Block remapping and block warping in combination with the pre-alignment achieved the best results in terms of recognition performance. However, block remapping is not secure enough as it turned out that its unlinkability as well as irreversibility is rather low. Block warping has a low unlinkability as well.
Hence, only the Bloom filter approach is suitable in terms of security. In combination with the pre-alignment it achieves an acceptable recognition performance, although this recognition performance is still much worse than the baseline one. Without the pre-alignment, the resulting recognition is not usable at all.

Thus, an accurate, universal pre-alignment, which does not require the unprotected templates to be present in the system, is necessary in order to employ a well performing (in terms of recognition accuracy) template protection scheme.

In our future work we will aim for further performance improvements of the Bloom filter approach as this method turned out to be the most beneficial one. Furthermore, we will evaluate other variants of block remapping and block warping, like remapping including shifts in the blocks and recursive remapping, as well as other strategies to derive the warped grid in the warping approach.

CHAPTER 5

Invisible Biometrics: Brain Waves

"COGNITIVE BIOMETRIC" refers to biometric traits which are detected during cognitive and/or emotional brain states. More in detail, cognitive biometrics are based on the measurement of signals directly or indirectly generated by the "way the individual thinks" as a distinctive characteristic for automatic user recognition [33]. In the last years, the interest in this field is growing, and many studies exploiting biological signals, like electroencephalogram (EEG), electrodermal response (EDR), electrocardiogram (ECG), blood pulsevolume (BPV), electromyogram (EMG), to cite a few, for the purpose of automatic user recognition have been proposed. In particular, the evidence that the brain activity carries distinctive information about the individual identity has led to an increasing interest in the research on cognitive biometrics.

The brain activity during specific mental states can be captured by means of different methodologies. Specifically, brain activity can be recorded either by measuring the blood flow in the brain or by measuring the neuron electrical activity. To the first category belong approaches like functional magnetic resonance imaging (fMRI), which measures the concentration of oxygenated and de-oxygenated haemoglobin in response to magnetic fields;

the near-infrared spectroscopy (NIRS), which measures the concentration of oxygenated and deoxygenated hemoglobin by means of the reflection of infrared light by the brain cortex through the skull; the positron emission tomography (PET), which measures neuron metabolism through the injection of a radioactive substance in the subject. To the second category belong approaches like magneto-encephalography (MEG), which is sensitive to the small magnetic fields induced by the electric currents in the brain, and electroencephalography, which is sensitive to the electrical field generated by the electric currents in the brain.

Since EEG recordings are acquired with portable and relatively inexpensive devices when compared to the other brain imaging techniques, it is more feasible to develop an EEG-based biometric system instead of designing automatic recognition based in the other acquisition methodologies. In this Section the use of EEG signals for the purpose of user recognition will be presented. EEG signals belong to the category of *invisible biometrics*: since EEG signals are produced by ionic current flows within the brain's neurons, they're inherently "secret". In addition, EEG-based biometric systems are robust against sensor spoofing. Unlike conventional biometrics, an attacker can't covertly acquire EEG signals in physical form or synthetically generate them at a later time and feed them to sensors. Also, there's no need for specially designed sensors to provide liveness detection [34].

This Chapter is structured as follows. In Section 5.1 a characterization of the brain signals and rhythms is given. Section 5.2 presents an overview about a generic EEG signal acquisition system. In Section 5.3 EEG signal acquisition protocols used in biometric oriented applications are described. Section 5.4 gives an introduction on the use of EEG signals in the field of biometric recognition. Eventually, Section 6.1 presents a study concerning an EEG-based biometric system exploiting visual stimuli.

5.1 EEG Signals: Brain Activity and Brain Rhythms

Since the early research on EEG analysis, it has been observed that the regions of a healthy human cortex have their own intrinsic rhythms in the range of 0.5-40Hz. In general, five main rhythms can be detected from an EEG recording: Delta (δ $[0.5-4]$Hz), Theta (θ $[4-8]$Hz), Alpha (α $[8-14]$Hz), Beta (β $[14-30]$Hz) and Gamma (γ over 30Hz). In Figure 5.1 examples of δ,θ, α, β and γ rhythms acquired through a single channel are depicted.

The amount of activity in different EEG frequency bands can be quantified employing spectral analysis techniques. The contribution of the dif-

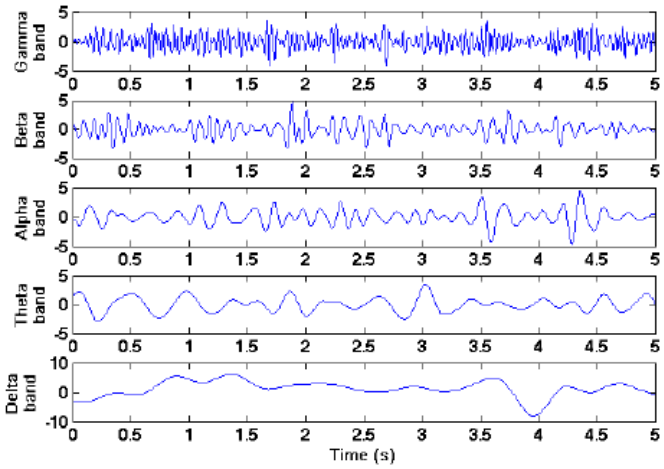


Figure 5.1: *Examples of Delta, Theta, Alpha, Beta, and Gamma waves acquired from a single channel.*

ferent rhythms to the EEG depends mainly on the level of alertness, on the age and behavioral state of the subject. Moreover an EEG pattern is influenced by neuro-pathological conditions, metabolic disorders, and drug action. The different brain rhythms or some combination of them become dominant oscillations during specific mental states, which can be induced by the performance of a proper acquisition protocol. Specifically, Delta and Theta frequency bands are considered to represent slow oscillating neural synchronization, or slow wave (SW) activity, while Beta and Gamma bands represent fast wave (FW) activity [13]. Brain oscillations in these frequency bands has been linked to various physio-psychological states and cognitive functions, as reported for instance in [147], [72] and [14]. A more detailed characterization of the subbands is given in the following.

- **Delta** $[0.5-4]$Hz: Delta rhythm is a predominant oscillatory activity in EEGs recorded during the so called deep or slow wave sleep (SWS). In this stage, Delta waves usually have large amplitudes $(75-200\mu V)$ and show strong coherence all over the scalp. In newborns, slow Delta rhythms predominate. An increase in Delta EEG activity during the performance of a mental tasks has shown to be related to an increase in subjects' attention to internal processing [98].

- **Theta** $[4-8]$Hz: In human scalp EEG, changes in Theta rhythms are very difficult or almost impossible to detect without the help of computational methods. If EEG power in a resting condition is compared with a test condition, Theta power synchronizes, that is an increased

Theta activity is observed. In particular Theta band power increases in response to memory demands, selectively reflecting the successful encoding of new information [144].

- **Alpha** $[8 - 14]$Hz: The oscillatory Alpha band activity is the most dominant rhythm which emerges in normal subjects, most pronounced in the parieto-occipital region. It is manifested by a peak in spectral analysis. The Alpha brain oscillations may present amplitudes large enough to be clearly seen in raw EEG traces acquired in specific mental states. It is characteristic of a relaxed but wakeful state primarily with closed eyes, and attenuates with eyes opening or mental exertion due to event-related Alpha power desynchronization. These changes in the Alpha band reflect an increase in tonic energetic levels related to increased arousal caused by basic processing of visual information [12]. Moreover there is evidence that attentional and semantic memory demands lead to a selective suppression of Alpha in different subbands and that the well described effects of visual stimulation represent just a special class of sensory-semantic task demand [143]. According to these evidences, the Alpha oscillations play an important role in suppression of processing for inputs in the brain. This confirms the evidence that Theta and Alpha band power are related to each other, although in an opposite way.

- **Beta** $[14 - 30]$Hz: Phase synchrony in Beta frequency band is enhanced for consciously perceived stimuli [198], and detectable mainly from the involved cortical areas, including somatosensory, frontal, parietal and motor regions, depending on the performed task. Specifically, Beta activity is characteristic for the states of increased alertness and focused attention.

- **Gamma** (over 30Hz): Neuronal synchronization in the Gamma band is considered important for the transient functional integration of neural activity across brain areas, in order to achieve various functions involving active information processing, e.g., recognition of sensory stimuli, and the onset of voluntary movements [300]. Gamma components are difficult to record by scalp electrodes and their frequency usually does not exceed 45Hz. Components up to 100Hz, or even higher, may be registered in electrocorticogram (ECoG).

In general, it can be assumed that the slowest brain rhythms are dominant during an inactive state and the fastest are typical of information processing performance.

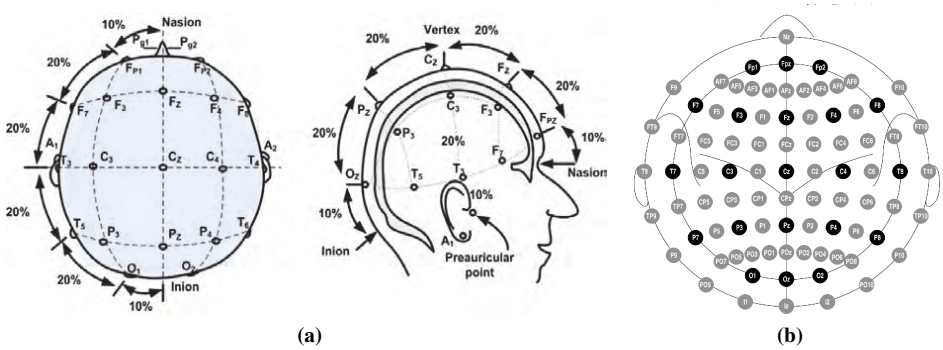


Figure 5.2: *(a) The 10-20 international system seen from left (right) and above the head (left). The letters F, T, C, P and O stand for frontal, temporal, central, parietal, and occipital lobes. Even numbers identify electrodes on the right hemisphere, odd numbers those on the left hemisphere, and "z" (zero) refers to electrodes placed on the midline. (b) Location and nomenclature of the intermediate 10% electrodes, as standardized by the American Electroencephalographic Society [190]*

5.2 Acquisition of Brain Signals

EEG signals are usually acquired using superficial scalp electrodes, placed according to the 10-20 international system depicted in Figure 5.2 and recommended by the International Federation of Societies for Electroencephalography and Clinical Neurophysiology [190]. The "10' and "20" refer to the percentage of the distance between the landmark points, namely the *inion*, the *nasion*, and the preaurical points used to draw the lines at which intersections the electrodes are positioned. In other words, given the landmark points, the electrodes positioning is made by considering the intersections between lines which are sagittally and coronally drawn, spaced at 10 or 20% of the distance between the landmark points.

EEG recordings are acquired with portable and relatively inexpensive devices when compared to the other brain imaging techniques. Specifically, signal amplifiers with high sensitivity and high noise rejection are used to measure the voltage fluctuation on the scalp surface resulting from the electric field generated by the firing of collections of pyramidal neurons of the cortex [33].

The EEG amplitude of a normal subject in the awake state, recorded with scalp electrodes, is in the range $[10, 200]\mu V$ and the most relevant cerebral activity falls in the range of $[0.5, 40]Hz$. The EEG based brain imaging technique presents a limited spatial resolution due to the physical dimension, in the range of several millimeters, of the surface electrodes,

usually employed in the acquisition setup, which limits the possible number of the electrodes covering the whole scalp. A limited spatial resolution is also due to the dispersion of the signals, generated by the sources on the cortex, within the head structures before they reach the scalp. On the contrary, EEG techniques have a high temporal resolution, in the range of milliseconds, which allows dynamic studies to understand the underlying mechanisms by means of complex computational methods. In fact, information concerning for instance psycho-physiological state, neurological and neuromuscular health, emotions, memory, the course of concentration, attention, levels of arousal, mental fatigue or workload during special tasks, sensitiveness to external stimulation can be extracted from EEG inspection and manipulation. Such a kind of evidence has led in last decades to use brain signals to convey conscious volition in EEG-based systems, like brain computer interface (BCI), and brain machine interface (BMI), aiming at controlling remote devices by means of the interpretation of the brain electrical activity.

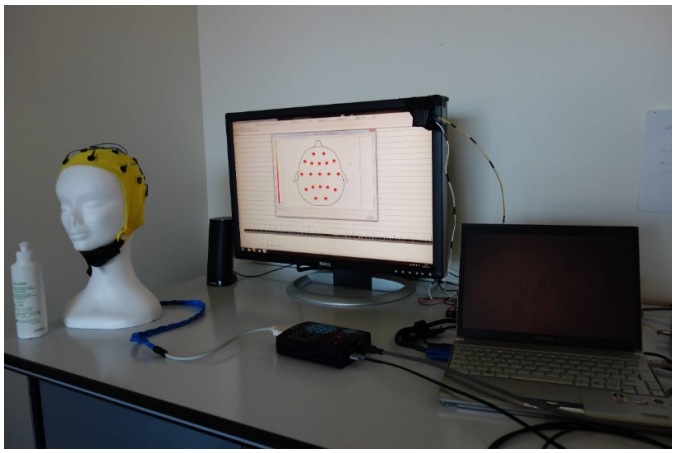
Some examples of EEG acquisition devices are detailed in the following:

- **EB Neuro Galileo BE Light**[1]: it is a system allowing to acquire and amplify bio-electrical signals when connected to a system of analysis and data storage. The aforementioned device is a non-intrusive medical device meant for acquisition, amplification, digital conversion and data tranfer to a host pc of bio-electrical signals produced by the human body. The BW Light acquisition system is composed by three main modules, as shown in Figure 5.3a: the headset, the BE Light Amplifier module and the BE NET Interface, that connects, thought an optical fiber, the amplifier to a computer provided with the "Galileo NT" software. The amplifier allows to acquire an EEG signal from an headset composed by a maximum of 28 channels. All input EEG channels (referred to the REF channel) acquired from the electrodes are amplified, sampled at 256 Hz, converted with a 16 bit analog to digital converter by the amplifier and sent to the pc.

- **g.Tec g.Nautilus**[2]: g.Nautilus is a g.tec's wireless biosignal acquisition system. It consists of the following components, as shown in Figure 5.3b: g.Nautilus Research Headset, that is a wireless biopotential amplifier with prefixed electrode strands and a cap (g.GAMMAcap[2]); besides, the device is provided with a g.Nautilus Research Base Sta-

[1] http://www.ebneuro.com/en/

[2] http://www.gtec.at/Products/Hardware-and-Accessories/g.Nautilus-Specs-Features



(a)

(b)

(c)

Figure 5.3: *Different acquisition devices for EEG signals: (a)* ***EB Neuro Galileo BE Light*** *(b)* ***g.tec g.Nautilus*** *(c)* ***OpenBCI Ultracortex "Mark IV"***

tion, that is a stationary receiver unit with USB cable to connect to a PC and it allows the communication with the g.Nautilus Research Headset. The g.Nautilus Headset can be equipped with gel-based active electrodes (g.LADYbird technology) or with dry active electrodes (g.SAHARA technology). All input channels (referred to the REF channel) acquired from the electrodes are amplified, sampled at 250 Hz or 500 Hz, and converted with a 24 bit analog to digital converter by the amplifier. The digitized data are sent via the wireless data link to the Base Station and the PC.

- **OpenBCI Ultracortex "Mark IV" EEG Headset**[3]: The Ultracortex, as it can be seen from Figure 5.3c, is an open-source, 3D-printable headset intended to work with any OpenBCI Board, whose characteristics are to be low-cost, high-quality biosensing hardware for brain computer interfacing. It is capable of recording research-grade brain activity (EEG), muscle activity (EMG), and heart activity (ECG). It is not designed for transcranial stimulation, or for sending any kind of signals. This headset is designed to receive EEG signals only. The Ultracortex Mark IV is capable of sampling up to 16 channels of EEG placed in 10-20 locations. The OpenBCI board generally adopted for EEG recording is the Cyton Biosensing Board[4], an Arduino- compatible, 8-channel neural interface with a 32-bit processor. At its core, the OpenBCI Cyton Board implements the PIC32MX250F128B microcontroller, giving it lots of local memory and fast processing speeds. The board comes pre-flashed with the chipKIT^{TM} bootloader, and the latest OpenBCI firmware. Data is sampled at 250Hz on each of the eight channels. The board communicates wirelessly to a computer via the the OpenBCI USB dongle using RFDuino radio modules. It can also communicate wirelessly to any mobile device or tablet compatible with Bluetooth Low Energy (BLE).

5.3 Acquisition Protocols

EEG signals can be acquired through portable devices that sense the electric field generated by the brain while resting or during a variety of cognitive tasks, such as response to audio or visual stimuli, real or imagined body movements, imagined speech, etc. More specifically, when a small change in the electrical activity of the brain, time-locked to a meaningful externally (exogenous) or internally (endogenous) generated event is produced, the brain response takes the name of “event related potentials” (ERP) [27]. ERP signals convey information on changes which occur when similarly oriented pyramidal neurons of both individual and different local networks fire in synchrony. For endogenous ERPs, time-locked to a mental event such as the recognition of a target stimulus, the activity of the cortex reflects functional coordination during neurocognitive information processing [272]. ERP components can be described in terms of latency time, polarity, and topography. Large individual differences exist for the ERP

[3] https://shop.openbci.com/collections/frontpage/products/ultracortex-mark-iv

[4] https://shop.openbci.com/collections/frontpage/products/cyton-biosensing-board-8-channel?variant=38958638542

components, while a certain stability is observed within a subject [21].

Other largely studied brain signals are the "slow cortical potentials" (SCPs), also used as control signals in BCI context. They represent slow voltage shifts in EEG, which are involved in the modulation of the excitability level of underlying cortical regions, and in the preparatory allocation of resources for cortical processing [26]. SCPs last from 300ms to several seconds and can be self-regulated with different purposes using immediate feedback.

The spatial distribution of brain activations as reflected in scalp EEG signals strongly depends either on the mental state of the subject or on the performed task during the acquisition session. For each designed protocol it can be identified a suitable electrode configuration in terms of number of sensors, their placement on the scalp as well as their density, according to the aim of the analysis, selecting a subset of channels in the 10-20 extended system shown in Figure 6.1. A specific minimal set of electrodes selected considering neurophysiological evidences and optimization criteria is generally employed in each experimental setup. Usually, the reduction of data dimension due to a selection of electrodes helps in improving the effectiveness of the data analysis.

Several data acquisition protocols have been proposed in the literature specifying the data acquisition conditions, the task definition, and the sensing electrodes configuration related to the neurophysiological function under analysis. Some acquisition protocols employed in EEG studies are described hereafter. Topographic information on source activation are reported depending on the performed task and guidelines for efficient scalp electrodes configurations are provided.

Resting State. Since the earliest applications of EEG signals, particular interest has been shown in the study of cerebral activity during a state of rest, due mainly to the simplicity of the acquisition process. Therefore, the resting state protocol, with eyes closed or open, has been widely studied for different purposes. Within this paradigm the enrolled subjects are typically seated in a comfortable chair with both arms resting, in a dimly lit or completely dark room. Generally, external sounds and noise are minimized to favor the relaxed state of the subjects. Participants are asked to perform few minutes of resting state with eyes closed or eyes open, avoiding any focusing or concentration, but staying awake and alert. Brain activity during resting state without performing any task carries interesting information as contained in EEG specific patterns. Eyes closed and eyes open resting conditions are usually employed in EEG research studies for baseline esti-

mates, although they represent different processes related to global arousal and focal activation.

In the closed eyes resting condition the predominant Alpha oscillations can be detected especially in the parieto-occipital region of the scalp. They reflect the default mechanisms of synchronization of cortical neurons activity [72]. Therefore a description of the ability of the central nervous system to transmit signals to and from the cerebral cortex can be carried out focusing on signals from parieto-occipital electrodes. On the other hand, a widespread reduction in activity is commonly observed turning to open eyes resting conditions, which reflects neuronal Alpha desynchronization. Further topographic changes occurring across frequency bands can be detected considering a full-scalp montage in the analysis of open eyes resting state.

Mental Calculations. EEG patterns have shown significant differences, specially related to the spectral analysis, between rest and several cognitive tasks, and even between different cognitive tasks themselves, involving distinct neural systems. In order to infer about the properties of neural activation in the involved brain regions, math, logical, and spatial cognitive operations have been considered in the development of suitable acquisition protocols. Changes in neuronal activation patterns due to specific components of mental calculation tasks can be observed from the analysis of each frequency band, as they seem to be related to oscillatory activity of different neural networks. In this regard, different EEG patterns have been examined by testing healthy subjects in different conditions of mental calculation through properly designed protocols. In these protocols the mental task period is usually preceded by a rest period in order to provide a baseline. During the mental task interval, the subject is asked to solve a problem providing an answer. The features of such kind of brain patterns reflect inter-individual variability due to different abilities, aptitudes, innate mechanisms of habit, brain plasticity, etc.

Various configurations can be employed for the effective detection of different EEG activation patterns during the performance of different mental calculation tasks. Some significant differences between mental calculation tasks, related to change in power between task and rest conditions, have been observed in the δ and β bands in the frontal lobe, reflecting different selective processes during focusing on relevant information, depending on the complexity of the task and the specific cognitive function involved. In general an increase of δ, θ and β activity in frontal leads during subject's internal concentration has been observed. This is in accordance with the

evidence that among the various functions of the human brain directing and allocating brain resources are governed by the frontal lobe. In particular decision making, reasoning and complex calculation require the integration of multiple processes, specific of each task. It results in differences of frontal lobe activity between tasks, reflecting activation of different neural networks. Therefore frontal leads can be effectively employed for the analysis of such specific functions.

ERP: P300. The most explored protocols involve the elicitation of the above mentioned ERPs. Task-related ERPs, as well as background EEG, are associated to different behavioral and cognitive traits. ERP signals can be elicited using different stimulation paradigms involving for instance sensory, cognitive or motor events. Usually, the exogenous eliciting events are repetitively modulated sensory stimuli such as a visual flicker. The so elicited evoked potentials strongly depend on the physical parameters of the stimuli. On the other hand endogenous ERPs depend on internal cognitive events reflecting the way the subject evaluates a stimulus. A largely studied and employed brain potential is the P300 ERP, especially used in BCI context. The P300 ERP is a positive deflection of the scalp potential which occurs around 300 ms after the onset of a task-relevant stimulus, with a centro-parietal focus [233]. The most effective paradigm for inducing a P300 response is the oddball task. In this paradigm an infrequent but task-relevant stimulus is presented among frequent irrelevant stimuli. Different kind of stimuli can be employed to carry out such paradigm, and the characteristics of the P300 response will change with the type of stimulation, its timing, and with the task difficulty. An example of the oddball paradigm consists in presenting different geometric shapes, and the subject is asked to detect just one specific shape among the others.For this particular case, a good brain response can be detected in central and parietal electrodes, as a much larger P300 amplitude related to target stimuli stands out from a baseline measure obtained by averaging non-target responses. The P300 individual differences relate to amplitude, latency, waveform and scalp potential distribution and reflect psychophysiological aspects of individual central nervous system reactivity. Several studies in literature addressed the effectiveness of different electrode configurations used to detect the P300 brain response. A trade off between user friendly solutions employing few electrodes and accuracy in terms of correct classification of brain responses is needed for the suitability of such P300-based systems.

ERP: Visual Evoked Potentials. Another typically employed ERP stimulation protocol during EEG acquisitions is the elicitation of Visual Evoked

Potentials (VEP), performed in order to analyze the way the brain perceives and processes visual inputs, to control BCI applications and to support neurological diagnosis. VEPs are evoked potentials that occur in the visual cortex, time-locked to a repeated sensory stimulation related to a subject's visual field. Within VEP protocols no response or cognitive processing by the subject is required. The visual stimulation can consist for instance of checkerboard pattern reversal, flashing black/white images, pattern on set stimuli or photic stimulation (light). In a typical setup to elicit VEPs a flashing stimulus is displayed either at the center of a screen or through light-emitting diodes (LEDs) in the central visual field, since it causes a greater response amplitude.

A subset of the electrodes montage is considered when studying VEP signals related to specific kind of visual attention stimulation. In these cases typically EEG signals are recorded from electrodes located in the posterior region of the scalp, mostly over the left and right hemispheres of the primary visual cortex. Indeed either periodic or transient brain responses to stimulation involving the visual system can be detected just considering electrodes O1 and O2 within the 10-20 international system [138].

Motor Imagery. Some interesting evidences have been obtained from the analysis of μ and β EEG rhythms recorded over sensorimotor cortex within the so-called motor imagery paradigm. Typically, during each acquisition session, subjects are asked to imagine moving for instance either a hand or a foot for few seconds when the cue representing the movement instruction appears on a screen. As reported in [197] it has been observed that the patterns of μ and β rhythms desynchronization over sensorimotor cortical areas during motor imagery are similar to those during real performed movement. Moreover, in the same work principal components analysis on sample average signals has shown marked individual differences in motor-related EEG patterns, topographically and spectrally focused.

In the analysis of rhythm topographies during motor imagery protocols, a subset of the extended 10-20 international system is often employed, considering sensors placed over the sensory-motor cortical area, namely FC3, FC1, FCz, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4. In fact it has been repeatedly shown that both movement and motor imagery are accompanied by desynchronization in μ and β bands over the centro-lateral side of the scalp, showing hemispheric asymmetries for specific conditions and frequency ranges [197]. Results showed marked individual-specific traits regarding topographic and spectral effects of movement and motor imagery, also indicating that movement, imagery

and bands are dissociated in terms of individual differences.

Speech Imagery. More recently, EEG acquisitions have been performed during the so called "speech imagery", aiming at recognizing the neural activities associated with speech production. In some protocols, enrolled subjects are instructed to imagine continuous vowel vocalization for few seconds from the onset of a specific cue which can be an acoustic signal or a task-representative image appearing on a screen. In BCI context this kind of tasks are designed in order to discriminate differences in brain activity during vowel or syllables speech imagery providing a control scheme for communication based on the interpretation of individual speech correlates in EEG.

Signals from motor cortex are employed for the performance of speech imagery protocols. In has been shown that neural activation detected over medial and posterior regions occur during imaginary lip movement and vocalization of vowels [63] or their mental repetition. The electrodes that are distant from the active regions may not provide relevant information. Therefore, a more effective analysis of EEG features during speech imagery could be obtained discarding them.

SCP. When using SCPs, the training for their voluntary control can be carried out within an acquisition protocol where a thought-translation device provides a feedback cursor on a screen, whose position constantly reflects the voltage shifts. Typically the subjects are asked to move a cursor which appears at the center or at the periphery of a screen toward a target, by modulating their brain electrical activity. A preparatory phase in which the cursor remains stationary on the screen is followed by an active phase in which it moves in a direction, either horizontal or vertical, with constant speed, and in the perpendicular one according to the user's SCP amplitude.

Effective negative and positive SCP shifts can be controlled selecting the best per-forming channel. In this regard, in [209] it is shown that self-regulation skills differ between subjects, but that the Cz derivation could be generally used for an effective SCP feedback learning. Interestingly, in that study it was shown that many subjects generated a maximal SCP differentiation at other, often neighboring, electrodes than Cz.

5.4 EEG signal for Biometric Recognition

Brain signals have been deeply investigated and exploited for medical and brain-computer interface (BCI) purposes since the beginning of the twentieth century [13]. In recent years, the interest in using such physiological

characteristic also for biometric recognition is rapidly increasing. Many studies in such research field have in fact been focused on the use of electroencephalography (EEG) signals, showing that the brain response to specific tasks can be exploited to extract discriminative features able to guarantee high levels of recognition accuracy [33]. The reason for the interest in using EEG data for biometric purposes is linked to some advantages the aforementioned signals possess, compared to other traditional biometric identifiers: universality is in fact guaranteed, and robustness to spoofing attacks and privacy compliance can be easily achieved.

In [279], back in 1980, the basis for automatic people recognition using EEG signals were posed. However, only in the last decade the study of EEG based recognition systems has received a significant development. EEG signals present some peculiarities, which are not shared by the most commonly used biometrics, like face, iris, and fingerprints, and that make the investigation on the use of EEG signal as biometric identifier not a mere academic exercise but an analysis with potential dramatic effects on the design of the next generation biometric systems, namely the cognitive biometrics based systems. Specifically, brain signals can be considered an *invisible biometric* trait: they are more privacy compliant than commonly used biometrics like face, iris, and fingerprints, since they are not exposed and therefore cannot be captured at a distance. Moreover, they cannot be left on a crime scene, not even a digital one, and being brain signals the result of a cerebral activity, they are less likely to be synthetically generated and fed to a sensor to spoof it, like it can happen when using gummy fingers to spoof a fingerprint sensor. This also helps in addressing the liveness detection issue. Furthermore, when using EEG based recognition systems, it is impossible for an intruder to force a user to authenticate. In fact stress signals would be present in the measured brainwaves, thus resulting in a denial of access [146].

On the other hand, the use of brain signals poses new challenges. In fact, being the brain continuously and spontaneously active, there is a background electrical activity upon which the signals of interest, which come from the firing of specific collections of neurons responding accordingly to a variety of tasks, are superimposed. Part of this difficulty is the understanding of the brain areas where the response originates. These findings would drive an optimal or sub-optimal choice about the number of electrodes to use and their location. Furthermore, due to the weakness of the signal detected on the scalp while generated on the cortex, the EEG acquisition process results very sensitive to endogenous and exogenous noise, that is artifacts generated by physiological processes and by external sources

respectively. Therefore, the basic mechanisms which are behind the physiological process of brain signal generation, the signal stability in time, the acquisition protocols, the optimal sensors location depending on the employed acquisition protocol, the amount of the discriminative information, as well its frequency localization, need a much deeper understanding.

In this Section, the different characteristics of a biometric identifier, namely universality, uniqueness, permanence, collactability, performance, acceptability, and circumvention, are detailed with respect to EEG biometrics. It is worth pointing out that the analysis that follows has different depth levels for the different desired characteristics, since EEG biometrics is still in its infancy and an exhaustive analysis of the aforementioned issues is still missing in literature. Nevertheless, in the following we draw some considerations, which, in some cases, have been borrowed from physiological studies on EEG signals made for clinical applications and that can be applied to the field of EEG biometrics.

- *Universality.* The level of universality of brain signals is very high. In fact people with no pathological conditions affecting the brain, can make use of EEG biometrics.

- *Uniqueness.* The uniqueness of EEG signals is a complex issue which has several facets to be considered and that has not captured the necessary attention within the biometric community so far. Nevertheless, some early studies in neurophysiology, see for example [18], [303], [163] have demonstrated that EEG is a highly individual characteristic. Of course, the level of individuality is also related to the specific acquisition protocol, subband analyzed, and to the extracted features. Moreover, it is worth pointing out that the uniqueness and the permanence issues can be considered as two facets of the same medal, being related to the intra-individual and inter-individual distances, and that these distances get some meaning when related to each other.

- *Permanence.* The issue of the reproducibility of EEG biometrics in different acquisition sessions, in other words the intra-individual EEG stability, has been object of scientific investigation within the neurophysiology field in the past. In fact, also in clinical applications, it would be desirable not to have significant variations of an individual EEG pattern when no alterations, due for example to new pathological conditions, occur. In the clinical field these studies are known as “test-retest reliability” or as “longitudinal” studies. Of course the aforementioned issue is strictly dependent on the features which are

extracted to summarize the EEG and on the irreliability. It is worth pointing out that a significant effort has been done for the test-retest reliability analysis of EEG in resting conditions as well as, in the recent years, when performing cognitive and sensory tasks. Despite the effort that has been done in the neurophysiology field, the repeatability issue of EEG biometrics has not received the necessary attention from researchers in the bio-metric scientific community. Nevertheless, its understanding is propedeutic towards the deployment of EEG biometrics in real life.

- *Collectability.* Collectability of EEG signals is dependent on many factors like the number of electrodes to be used, the need to use conductive paste or saline liquid to lower the skin impedance to acceptable levels, and the acquisition time needed to be able to collect relevant information for the recognition process. All these issues can limit the collectability of EEG biometrics. However, recent advances have shown that interesting performance can be achieved also limiting the number of used electrodes thus making the signal collection more user convenient. Moreover, the latest technological developments have shown that the aforementioned issues can be mitigated.

- *Performance.* Promising recognition rates have been achieved in the last years. A detailed analysis of the recognition performance of state-of-the-art EEG-based biometric systems is given in Section 5.5.

- *Acceptability.* Acquisition of EEG signals may present some drawbacks in terms of user acceptability being related to brain activity thus potentially evoking ancestral worries related to "mindreading" and emotion analysis from the data controller. This may generate a sense of discomfort in the users. Also privacy issues can be seen as an obstacle towards the acceptability of EEG based biometric applications in real life due to the potentiality to detect existing pathologies or disposition towards pathologies, as possible also for other biometrics. This could potentially lead to discrimination and undermine the human dignity. However, no specific studies on the acceptability issue of EEG biometrics have been performed yet.

- *Circumvention.* Brain signals, as a result of cerebral activity, are not exposed biometrics, that is they are *invisible biometrics*. Therefore, as internal traits, they are less prone to spoofing attacks than other external biometrics [259], since they are "secret" by their nature, being

impossible to capture them furtively at a distance, while this is possible for face and iris, which can be then synthetically generated. Besides, EEG biometrics cannot be acquired in absence of the user, since they are not left on objects like it might happen with fingerprints that instead can be used at a later time in order to spoof the employed acquisition sensor. This is virtually impossible with brain signals since they are the result of ionic current flows within the neurons of the brain in response to a specific task or during a specific mental state. Therefore, an attacker should be able to synthetically generate resulting EEG waveforms and feed them to a sensor to spoof it. Hence, the problem of liveness detection, which needs to be addressed when using conventional biometrics, is naturally overcome without the need to resort to specifically designed sensors.

5.5 EEG Biometrics: State-of-the-Art

The use of EEG data for person recognition has been explored since 1998 and more than 100 papers in this field are published. Most of these papers were focused on reporting the performance using conventional accuracy metrics. However, to objectively assess an EEG biometric recognition system and establish its potential suitability for real-life applications, the performance of EEG-based biometrics system needs to be evaluated based on more factors than the recognition accuracy only [334]. Besides the conventional measures of recognition accuracy, there are other factors which should be taken into consideration for assessing the practical usability of any reported EEG-based recognition systems:

- number of the subjects for which the system was designed or tested on (N_S);
- number of the electrodes employed (N_C);
- number of considered sections (S);
- the stimulation protocol used to elicitate the brain response;

The number of electrode(s) used for data collection has a considerable impact on the usability of the EEG-based biometric system: a large number of electrodes may increase the difficulty in deploying the system in real-life scenarios. Number of the subjects employed indicates the effectiveness of the biometric system in large scale deployments. The number of session is a crucial experimental setting that needs to be considered to deploy

EEG-based biometric systems in real-life applications: if signals recorded in a single session are used to test the designed biometric system an important concern arises on whether the obtained recognition accuracy is representative of the distinctiveness of subjects based on EEG features, or it is representative of the uniqueness of each acquisition session [188]. Eventually, the stimulation protocol is linked to the convenience of the acquisition procedure.

In this Section, an overview of the most relevant state-of-the-art works related to biometric recognition based on EEG signals is given. The proposed work are divided according to the considered acquisition protocol and summarised in Table 5.1.

Resting State (RS).

Some mental tasks are more appropriate to be performed for person recognition than others being intrinsically able to highlight distinctive traits of individuals. Several studies investigate EEG traits during brain ongoing activity for user recognition, which does not require any mental task at all. Specifically, in [236] a closed eyes in resting condition protocol was employed to acquire data using the O2 channel from the occipital region of the head. The α rhythm, predominant in the parieto-occipital region during rest, was extracted and overlapping subbands were individually considered for feature extraction. The performed tests were aimed at verifying four authorised users against a single class of non-authorised users and at their identification. The obtained classification scores in terms of genuine acceptance rate (GAR) ranged between 80% and 100% depending on the individual, the frequency band, and the test performed, while the correct identification rate (CIR) related to the identification tests ranged between 80% and 96%. In general, different frequency bands showed to be more performant for different individuals. The same protocol was tested in [236]. A different analysis of the same rest EEG signals yield to a GAR ranging from 72% to 84%. In [222] the EEG activity was recorded from 40 subjects while resting both with eyes open (EO) and with eyes closed (EC). Although eight sensors were employed for the acquisition, only the signals acquired using the channel P4, from the parietal region of the head, were considered in the study. An analysis was performed for user identification in the EO condition and GAR ranging from 49% to 82%, depending on the modeling parameters, was obtained. In [260] a closed eyes resting condition was used to acquire EEG signals from 51 subjects using two forehead electrodes (FP1 and FP2). Through discriminant analysis the best achieved result was an EER=3.4% . In [281] the influence of the diet and circadian

effects on the identification performance was investigated. In the considered protocol, segments of 5 minute EEG signals, acquired by an FP1 electrode, were recorded during rest with closed eyes. Signals were acquired on two separate days (sessions) in which subjects had water in one session and coffee in the other one. In each session, 6 EEG runs were recorded. A database of 40 subjects was collected. The classification accuracy achieved for subject identification was of 95%. In the same study an implementation of the Covert Warning (CW) concept to enhance the security of the EEG-based biometric system was presented. Muscle signals from clenching the teeth, shown to produce robust signals, were used to send the covert message. 24 volunteers were enrolled and performed 3 minutes of resting with closed eyes, while clenching the teeth 3 times. Authors showed that CW messages were detected perfectly, while a small decrease in the identification performance with respect to the scenario without CW was observed. In [1] signals from 10 subjects in 5 different sessions over two weeks, using 8 electrodes to obtain bipolar signals at C3, P3, C4, P4, were collected. In each session subjects performed resting state with closed eyes and open eyes, repeating each task in 5 runs of 30 seconds. Different spatial arrangements were evaluated in order to identify users using a suitable electrodes configuration. Best performance of CRR=97% was obtained employing all 4 channels in the eyes closed condition, while configurations in the right hemisphere (C4, C4-P4) produced the highest CRR compared to the other arrangements relying on an equal number of electrodes. Such result was in accordance with the significant role of the right hemisphere, involved in processes like imagination, creativity and feeling, which are dominant during resting. This supports the idea that brain activity detected in the right hemisphere shows distinctive information during rest. Brain ongoing activity in EC condition was investigated in [35] for user identification. EEG signals were recorded from 48 subjects employing 56 scalp electrodes. An analysis on suitable scalp configurations was carried out considering different sets of symmetrically placed electrodes. Signals filtered in the range 0 – 33Hz were analysed and a best CRR=96.98% was obtained considering channels T7, Cz, T8. In [158] signals from 45 subjects in EC resting conditions, acquired through 56 electrodes, were analysed. Signals were filtered in order to extract the different brain rhythms ($\delta, \theta, \alpha, \beta$) , so that the different frequency bands were individually analysed, as well as combined together. Different channel configurations were considered to perform user identification and a best CRR=98.73% was obtained from a set of 3 parieto-occipital channels. A comparison between EC and EO condition for user identification was carried out in [50] on a smaller dataset. Longitudinal

recordings allowed addressing the repeatability of EEG features, which is a very important issue for the application of biometric systems in real life scenarios. A perfect identification of users enrolled in a previous acquisition session was obtained for the EC condition considering the subband 0.5 – 30Hz and a set of 3 electrodes placed in the posterior part of the head. An extensive analysis was also performed in [159] in order to find the most appropriate set of parameters involved in the analysis.

Signals from 20 people, recorded during 2 sessions at a median distance of 15 months, and represented through PSD characteristics,have been used in [210] to estimate a rank-1 correct identification rate (CIR) at about 88%. A database collected from 9 subjects during 2 one-year-apart sessions has been considered in [148], applying a frequency-zooming auto-regressive (FZ-AR) modelling to 53 channels to achieve CIR=87.1%. Signals from 9 subjects have been recorded during 2 sessions spanning up to 3 weeks in [159], and exploited to achieve perfect IR for EEG data acquired in eyes-closed (EC) conditions, and IR=90.53% for the eyes-open (EO) scenario. Signals recorded from 4 subjects in EC conditions during 2 one-week-apart sessions have been processed through continuous wavelet transform (CWT) in [314] guaranteeing CIR=92.58%. Parsimonious representations in the frequency domain have been proposed in [187], where CIR=87.9% and IR=75.4% have been respectively achieved in EC and EO conditions, using EEG signals taken from 30 subjects during 2 recording sessions spanning one month.

Perfect accuracy has been achieved in [266] applying the system proposed in [265] to 20 subjects whose EEG signals have been recorded during 2 sessions at an average distance of 9 months. The most detailed analysis on permanence so far performed for EEG-based biometric recognition systems has been presented in [188], where the performance behaviour achievable when comparing data captured from 50 subjects during 3 different sessions spanning a 1-month period, and represented through auto-regressive (AR), PSD and spectral coherence (COH) features, has been discussed. IR at 90.8% comparing signals captured in EC conditions, and CIR=85.6% for the EO scenario, have been reported almost regardless of the sessions being compared out of the 3 available ones. Maiorana *et al.* [184] performed experimental tests on a database comprising 45 subjects, whose EEG signals have been collected during five to six distinct sessions spanning a total period of three years, using four different elicitation protocols. Between the proposed protocols, EC and EO are included. Statistical and performance-related analysis are conducted, using different EEG features and hidden Markov models as classifiers.

Motor Imagery (MI).

A brief summary of the state-of-the-art works using MI-elicited EEG signals for biometric recognition is here provided. In [112], authors have exploited MI EEG data captured through six channels from three subjects, elicited to perform hands, feet and tongue movements during a single acquisition session. Auto-regressive (AR) and moving average (ARMA) coefficients have been used for feature extraction, while classification has been performed through multi-layer back propagation neural network (BP-NN). Identification accuracies ranging from 81.9% to 83.9% have been achieved. In [324] authors have achieved 80% correct recognition rate (CRR) using Fisher distance as characteristic feature and BP-NN as classifier, for the same number of subjects, session and imaginary tasks as the previous case, while 60 EEG channels have been instead exploited. In [192], EEG data elicited through imaginary left and right hand movements have been acquired from nine subjects in three different sessions, using eight central and parietal channels to extract power spectral density (PSD) features. A half total error rate (HTER) of 7.1% has been achieved using Gaussian mixture models (GMMs) as classifier. EEG data from nine subjects have been collected in two different days in [292], using four channels and according to four standard MI tasks. Cepstral values have been used as features to achieve 94.72% accuracy with Mahalanobis distance as classifier. In [335], EEG signals have been acquired through nine channels in a single acquisition session from 108 subjects performing two protocols, based on either left/right fist or both fists and feet's imaginary movements. A 94.72% accuracy has been achieved for the former protocol, while 93.1% has been obtained for the latter, using wavelet packet decomposition for feature extraction and linear discriminant analysis (LDA) as classifier. Das et al. [61]proposed am EEG-based biometric identification system, using a motor imagery task, specifically performing imaginary arms and legs movements. Deep learning methods such as convolutional neural network (CNN) is used for automatic discriminative feature extraction and person identification. An extensive set of experimental tests, performed on a large database comprising EEG data collected from 40 subjects over two different sessions taken at a week distance, shows the existence of repeatable discriminative characteristics in individuals' brain signals.

It can be observed that the studies carried out so far report recognition performance evaluated over either EEG data collected from a very small number of subjects, or data recorded during a single acquisition session, which cannot provide any convincing evidence for considering EEG sig-

nals as a stable biometric identifier. In fact, under such conditions it is hard to state whether the reported recognition performance depend only on the characteristics of each subject's neural activity, or also on session-specific exogenous conditions, such as the capacitative coupling of electrodes and cables with lights or computer, induction loops created between the employed equipment and the body, power supply artifacts, and so on.

Speech Imagery (SI).

Some authors report the use of imagined speech, for example [28] used EEG signals from a small population of 6 subjects while imagining the syllables /ba/ and /ku/. The collected database consisted of 6 sessions and for each one 20 trials per subject from 128 channels with a sampling frequency of 1024 Hz. using Electrical Geodesics device. For feature extraction they used the PSD for each EEG signal, then auto-regressive (AR) model coefficients were computed for each electrode using the Burg method. The classification stage was performed using the linear kernel of Support Vector Machine (SVM) classifier and using1-Nearest-Neighbor (k-NN). For these two syllables they obtained 99.76% and 99.41% of accuracy respectively. Maiorana *et al.* [184] explored the speech imagery protocol on a database composed of 45 subjects. AR reflection coefficients, MFCC coefficients and Bump model are used as features and GMM and HMM as classifier, achieving an EER of 9%. In the work presented in [213], resting-states were used for subject identification using Linear SVM. The dataset used consisted of 40 subjects, and 192 instances per subject. The sampling frequency was 256Hz with 64 channels. First, for pre-processing a band-pass filter (0.5-40 Hz) and then the Common Average Reference were applied. For feature extraction the Morlet Wavelet was used to extract power spectrum of 7 frequency bands, to finally apply a downsampling to 32 Hz. The accuracies obtained in the best cases were 100%, 96% and 72% respectively for 3 lengths of the signal (300, 60 and 30 seconds). However, in a real application, the registry of 300, 60 or even 30 seconds of a signal can be impractical and with high computational cost for real-time. In addition the use of 128 or 64 channels does not support the portability of the device. In [207] Empirical Mode Decomposition (EMD) is used to decompose EEG signals during imagined speech in order to use it as a biometric marker for creating a biometric recognition system. For each EEG channel, the most relevant Intrinsic Mode Functions (IMFs) are decided based on the Minkowski distance,and for each IMF 4 features are computed: Instantaneous and Teager energy distribution and Higuchi and Petrosian Fractal Dimension. To test the proposed method, a dataset with 20 subjects who

imagined 30 repetitions of 5 words in Spanish, is used. Four classifiers are used for this task - random forest, SVM, naive Bayesandk-NN- and their performances are compared. The accuracy obtained (up to 0.92 using LinearSVM) after 10-folds cross-validation suggest that the proposed method based on EMD can be valuable for creating EEG-based biometrics of imagined speech for Subjects identification.

Visually Evoked Potentials (VEP).

A brief synopsis of the state-of-the-art works on the use of visual-stimuli-elicited EEG signals for biometric identification is presented in this section. This approach, for individual identification, has been first proposed in [221], where VEP signals have been recorded from 20 subjects by presenting black and white images of common objects, using 61 channels and exploiting the gamma ([30 : 40]Hz) band, with spectral power ratio as features. A back-propagation neural network (BPNN) has been used to identify individuals with 99.6% accuracy while performing ANOVA tests on each individual channel. In [290] EEG responses have been collected from 5 different subjects during 5 sessions on the same day. In a particular session, a sequence of 9 images has been randomly shown for 20 times to each subject, while asking him to focus on one or more pre-selected target images and ignore the rest. Principal component analysis (PCA) has been applied on the obtained time sequences for feature extraction, and linear discriminant analysis (LDA) used for classification. A performance accuracy of 97.6% has been achieved by considering only one channel for both target and non-target stimuli. The significance of irrelevant stimuli has been studied in [89] using rapid serial visual paradigm (RSVP) stimuli on 8 different subjects. EEG signals elicited from 8 channels have been acquired in a solo session, and P300 waves used as features. A threefold cross-validation using Bayesian LDA has been performed to obtain a maximum correct recognition rate (CRR) of 97%. In [57] VEP data from 20 subjects have been collected by exhibiting face and car images for 40ms each. SVM and LDA have been applied to discriminate individuals. A 94% classification accuracy has been achieved by selecting the best performing post-stimulus set, and using a k-nearest-neighbors (KNN)- based classification technique. It is worth specifying that all the above mentioned works have considered EEG data acquired on a single day to achieve high performance accuracy.

Conversely, in [337] EEG signals have been collected from 10 subjects during 2 separate sessions on different days, using a random sequence of

self-face and other's-face images as visual stimuli. Each performed session has included 2 distinct runs, each comprising 50 trials, where in each trial a total of 20 images (10 self-face and 10 other's-face images) has been presented. A total of 180 trials has been selected for training with the remaining 20 used for testing, therefore mixing data from the two available sessions for enrolment purposes. An adaptive discriminant feature method has been used for extracting features, and non-linear-SVM for classification purposes, achieving a CRR of 86.1%. Two different schemes have been instead considered in [4]: first, EEG signals have been collected from 15 subjects at an inter-session temporal distance of one week. Then, only 8 subjects' signals have been recorded at a time span of 6 months. CRRs at 89.0% and 93.0% have been achieved by considering event related potentials (ERPs) as features, and signal correlation as classifier. Visual-evoked potentials (VEPs) to both target and non-target stimuli have been evaluated in [59] to provide equal error rates (EERs) respectively at about 18% and 13%, over a database comprising signals acquired from 50 users during 3 sessions taken during a period of 1 month. Das *at al.* [60] studied the use of (EEG) signals, elicited by means of visual stimuli, for biometric identification. A deep learning method such as convolutional neural network (CNN), is used for automatic discriminative feature extraction and individual identification. Experiments are performed on a longitudinal database comprising of EEG data acquired from 40 subjects over two distinct sessions separated by a week time. The experimental results testify the existence of repeatable discriminative characteristics in individuals' EEG signals

As can be noticed, it is worth remarking that works performing tests on EEG data collected during acquisition sessions spanning different days typically report recognition performance much lower than those obtained exploiting EEG signals recorded during a single acquisition session.

Table 5.1: *State-of-the-art works about EEG-based biometric recognition.*

Paper	Database				Biometric System		Performance
	N_S	N_C	S	Protocol	Feature Extraction	Matching	
Poulos *et al.* [236]	4	1	-	EC	α spectrum	NN	CRR = 80 - 95%
Poulos *et al.* [235]	4	1	-	EC	AR($8^{th} - 12^{th}$ Order)	NN	GAR = 72 - 84%
Riera *et al.* [260]	51	2	4	EC	AR, DFT, COH, Cross-corr.	Discriminant analysis	EER = 3.4%
Su *et al.* [281]	40	1	2	EC	PSD	k-NN	CIR = 95%
Campisi *et al.* [35]	48	3	1	EC	Burg's refl. coeff.	Polynomial Regression	CIR = 96.98%
La Rocca *et al.* [158]	45	2,3,5	1	EC	Burg's refl. coeff.	Polynomial Regression	CIR = 98.73%
La Rocca *et al.* [159]	9	3,5	2	EC/EO	Burg's refl. coeff.	Linear Classifier	CIR = 100%
Abdullah *et al.* [1]	10	4	5	EC/EO	AR	NN	CIR = 97%
Paranjape *et al.* [222]	40	1	1	EO	AR ($3^{rd} - 21^{st}$ Order)	DA	GAR = 49% - 82%
Näpflin *et al.* [210]	20	60	2	EC	PSD	Linear Regression	CIR = 88%
Kostílek *et al.* [148]	9	53	2	EC	FZ-AR ($7th$ Ord.)	Mahalanobis Dist.	CIR = 87.1%
Wang *et al.* [314]	4	128	2	EC	CWT	L2 Dist.	CIR = 92.58%
Maiorana *et al.* [187]	30	19	2	EC	EigenBrains	L1, L2, cosine Dist.	CIR = 87.9%
				EO			CIR = 75.4%
Maiorana *et al.* [188]	50	19	3	EC	AR, PSD, COH	L1, L2, cosine Dist.	CIR = 90.8%
				EO			CIR = 85.6%
Maiorana *et al.* [184]	45	19	45	EC	AR, MFCC, Bump Model	GMM, HMM	EER = 6.5%
				EO			EER = 10.8%
Marcel *et al.* [192]	9	8	3	MI	PSD	GMM	HTER = 19.3% - 42.6%
Hu [112]	3	6	1	MI	AR/ARMA cofficients	Multi-layer BP-NN	CIR = 81.9 - 83.9%
Xiao *et al.* [324]	3	60	1	MI	Fisher Distance	BP-NN	CIR = 80%
Tsuru *et al.* [292]	9	4	2	MI	Cepstral Values	Mahalanobis	EER = 0.17%
Yang *et al.* [335]	108	9	1	MI	Wavelet coefficients	LDA	CIR = 93.1 - 94.72%
Das *et al.* [61]	40	17	2	MI	ERP	CNN	CIR = 99.3 %

Continuation of Table 5.1

Paper	Database				Biometric System		Performance
	N_S	N_C	S	Protocol	Feature Extraction	Matching	
Bringham *et al.* [28]	6	128	4	SI	AR (2^{nd} Order)	SVM	CIR = 78.6 - 99.8%
Moctezuma *et al.* [207]	20	14	1	SI	EMD	Random forest, naive Bayes, SVM k-NN	CIR = 92%
Maiorana *et al.* [184]	45	19	5	SI	AR, MFCC, Bump	HMM	EER = 9.0%
Nishimoto *et al.* [213]	40	64	1	SI	Morlet Wavelet	Linear SVM	CIR = 100%
Das *et al.* [59]	50	19	3	VEP	Evoked Potential	Cosine Dist.	EER = 13 %
Armstrong *et al.* [4]	15	1	2	ERP	ERP Signal	Correlation	CIR = 89.0%
	8	1					CIR = 93%
Ruiz-Blondet *et al.* [266]	20	26	2	ERP	ERP Signal	Correlation	CIR = 100%
Palaniappan [221]	20	61	1	VEP	Spectral Power Ratio	BPNN	CIR = 99.6%
Touyama [290]	5	1	5	VEP/ERP	PCA	LDA	CIR = 97.6%
Gupta *et al.* [89]	8	8	1	VEP/ERP	P300	LDA	CIR = 97%
Das *et al.* [57]	20	20	1	VEP	LDA related features	KNN	CIR = 94%
Yeom *et al.* [337]	10	8	2	VEP/ERP	Adaptive disciriminative features	Non-linear SVM	CIR = 86.1%
Das *et al.* [60]	40	17	2	VEP	ERP	CNN	CIR = 98.8%

CHAPTER 6

EEG Biometrics: Performance Improvement

BRAIN SIGNALS, being invisible biometrics, are able to provide secure and reliable user recognition. The "secretness" of the signal entails difficulty in stealing and replicating the data, making a presentation attack almost impossible to be implemented. Besides, brain waves are sensitive to mental stress, so the EEG-based recognition modality can detect a legitimate user who is forced by a fraudulent person to be involved in the recognition process. Eventually, brain signals are able to provide continuous authentication and they can be employed in applications where this requirement is needed. The aforementioned advantages make EEG biometrics particularly suitable for application where high level of security is required, such as governmental, forensics or military use cases.

On the other hand, there are still some open problems that make the use of EEG-based biometric systems difficult to be accepted in real life scenarios. Specifically:

- EEG recordings are highly susceptible to various forms and sources of *noise*, such as environmental sources of noise (AC power lines,

lighting and a large array of electronic equipment) and physiological noise (cardiac signal, movement artefacts caused by muscle contraction and ocular signal caused by eyeball movement). The aforesaid noise's sources lead to a very small signal-to-noise ratio of EEG signals [258];

- noise negatively impact of the recognition *performance* of the system, entailing a decrease of correct recognition percentage;
- the *cost* of the acquisition device is high;
- the *convenience* of the acquisition procedure is low;
- there are not enough studies in literature demonstrating the *permanence* of EEG signals, that is performing the enrolment and authentication phases utilising signals acquired in temporally separated sessions [188].

In this Chapter a solution aiming to deal with the limitations of EEG-based biometric systems is proposed. Specifically, in Section 6.1, a system where steady-state visual evoked potentials (SSVEPs), that is brain waves recorded as response to a flickering visual stimuli, is chosen in order to exploit the high signal-to-noise ratio the considered signals show. Biometric fusion techniques are chosen as a solution to improve the performance of the system and to reduce the undesired impact of the noise on performance. Besides, of multiple elicitation frequencies are jointly used in order to further improve the recognition rates, thus allowing to reduce the number of electrodes needed during EEG collection and entailing to face the issues of cost and convenience of the acquisition procedure of brain signals. Eventually, EEG signals have been recorded during two temporally separated sessions, with the second session carried out after an average temporal distance of 15 days from the first session; in this way the issue of permanence across time of SSVEP-based biometric systems and the stability of the signals across time is addressed.

6.1 Steady-State Visual Evoked Potentials for EEG-Based Biometric Identification

EEG signals used for biometric purposes can be recorded as a response to different kind of stimuli. In this Section, an EEG-based biometric recognition system where discriminative features are extracted from steady-state visual evoked potentials (SSVEPs) is proposed. SSVEPs are a particular

kind of VEPs that consist of stationary periodic oscillations observed in brain activity as response to a repetitive visual stimulus in the range of 4 Hz to 60 Hz. When an individual focuses his attention on a flickering stimulus within this frequency range, typically presented on an LED setup or LCD display, an increased oscillatory activity, with spectral peaks at the stimulus frequency and its harmonics, can be observed in brain signals [244].

SSVEPs exhibit a high signal-to-noise ratio and a stable spectrum, properties which have led to their widespread use for the investigation of cognitive processes such as visual attention and working memory, and clinical conditions such as schizophrenia, autism and epilepsy [302]. These characteristics have also led SSVEPs to being widely adopted in BCI systems, that is, systems allowing an individual to communicate or control equipments solely through their brain activity [352]. The consistent, rapid and prominent response of SSVEPs also makes these signals particularly appealing for EEG-based biometric applications. In contrast with their use in BCI systems, where the primary aim is distinguishing between different visual targets for a given individual, in a biometric system the main challenge lies in identifying features that are sufficiently distinct across individuals, whilst ensuring their stability across multiple recording sessions of the same subject [188]. The use of SSVEP in biometric applications has been so far investigated only in [226], where an analysis based on the peak magnitude and frequency of the short-term Fourier transform has been exploited to identify five users, whose signals have been yet recorded during a single acquisition session.

In this Section, a novel approach for EEG recognition based on SSVEPs is proposed. Being the issue of permanence across time of paramount importance for real-life applications of EEG-based biometric systems, the stability of SSVEPs is also specifically addressed. The Section is structured as follows. Section 6.1.1 gives an overview of the employed acquisition protocol and the tools used to acquire EEG data. Section 6.1.2 describes the proposed biometric recognition system, while the achieved performance and permanence results are reported in Section 6.1.4. Some conclusions are eventually drawn in Section 6.1.5.

6.1.1 Employed Acquisition Protocol

EEG signals from $U = 25$ healthy volunteers are recorded and used for experimental tests. The device employed to elicit SSVEPs consists of a square array of 9 green leds, whose flickering frequency can be manually tuned. Four different elicitation frequencies are exploited, namely $f_S \in$

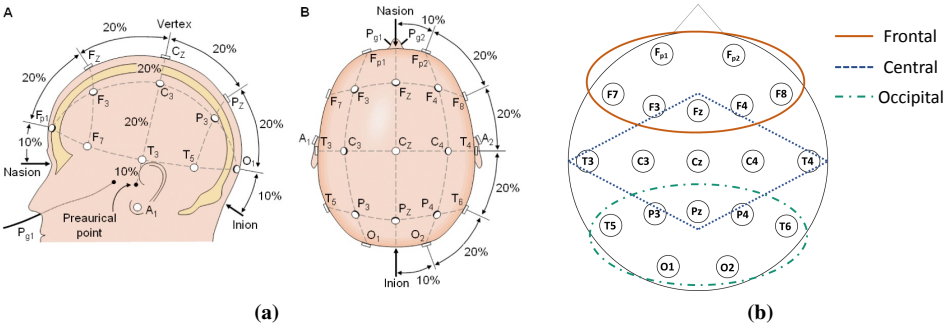


Figure 6.1: *(a) Montage of electrodes used during the acquisition stage. (b) Brain regions.*

$\mathcal{F}_S = \{6, 12, 18, 24\}\ Hz$. During each EEG data acquisition, subjects were comfortably seated on a chair in a dimly lit room, and asked to concentrate on the flickering target for one minute for each considered frequency. The involved subjects were asked to perform the proposed experiment during two temporally separated sessions, referred in the following as S1 and S2. The second session S2 is carried out after an average temporal distance of 15 days from the first session. EEG signals are acquired using a GALILEO BE Light amplifier operating at a sampling rate of $256Hz$. Brain activity is recorded from 19 electrodes placed on the scalp according to the 10-20 international system, as shown in Fig. 6.1a, with potentials referred to an electrode placed at the middle of the central region. At the beginning of each acquisition, the electrical impedance between each electrode and the scalp is kept under $30k\Omega$ using conductive gel. The recorded EEG signals are later preprocessed in order to remove noise and improve signal-to-noise (SNR) ratio, before distinctive features are first extracted and then matched for recognition purposes, as described in Section 6.1.2.

6.1.2 Employed SSVEP-based Recognition System

A scheme of the employed SSVEP-based recognition system is shown in Fig. 6.2 and the proposed approach is detailed hereafter.

Preprocessing. In order to improve the quality of the acquired EEG signals, a spatial filter, namely a common average referencing (CAR) filter, is first applied to the recorded data. The aim of such filter is to reduce artefacts related to inappropriate reference choices in monopolar recordings [271] or unexpected reference variations. Having indicated as $\mathbf{v}_m^{(u)}$, with $u = 1, ..., U$ and $m = 1, ..., M$, the u-th user's potential between the

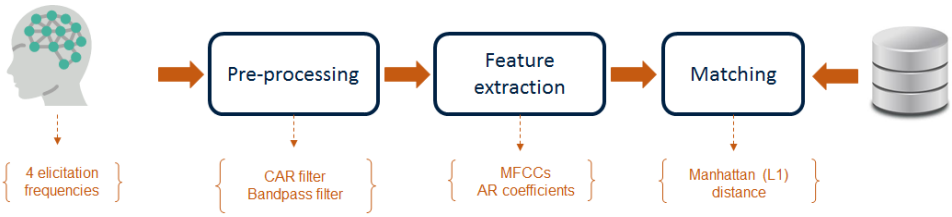


Figure 6.2: *Scheme of the proposed SSVEP-based biometric system.*

m-th electrode and the reference electrode, filtered data are obtained by computing the difference between the considered EEG signal and the mean of the entire electrode montage:

$$\mathbf{c}_m^{(u)} = \mathbf{v}_m^{(u)} - \frac{1}{M}\sum_{m=1}^{M}\mathbf{v}_m^{(u)} \tag{6.1}$$

A band-pass filtering is then performed on the CAR-filtered signals. Specifically, since EEG data are characterised by a frequency spectrum with significant elements mainly below 40 Hz, the signals are filtered in the $[0.5, 40]$ Hz band. In order to analyse the brain response behaviour, different combinations of the subbands related to the main brain rhythms, that is Delta (δ, $[0.5-4]$ Hz), Theta (θ, $[4-8]$ Hz), Alpha (α, $[8-14]$ Hz), Beta (β, $[13-30]$ Hz) and Gamma (γ, over 30 Hz) are also considered in the performed experimental tests when defining the applied band-pass filter. The obtained data are then downsampled at 128 Hz when the frequency interval of interest comprises the γ subband, otherwise the signals are downsampled at 64 Hz. The so-obtained data are then segmented into R consecutive overlapping frames $\mathbf{y}_m^{(u,r)}$, $r = 1, \ldots, R$, lasting $D = 5$ s with a normalised overlapping factor of $O = 75\%$ between each frame and the previous one.

Feature Extraction. After EEG data have been preprocessed, discriminative features are evaluated to generate a template from each user u's recording. In this work two different representations are exploited, namely mel-frequency cepstral coefficients (MFCCs) and auto-regressive (AR) coefficients, respectively detailed the following.

Mel Frequency Cepstral Coefficients (MFCCs). MFCCs are a parametric representation of the signal based on the Fourier spectrum, widely used in speech-based biometric systems [82] and recently applied to EEG data [212] too. The following steps detail the processing carried out for MFCCs extraction:

1. **power spectral estimate**: the power spectral density (PSD) $\mathbf{Y}_m^{(u,r)}$ of each signal $\mathbf{y}_m^{(u,r)}$, $m = 1, ..., M$ and $r = 1, \ldots, R$, is computed through the Welch's averaged modified periodogram approach, using 1-s sliding Hanning windows with 0.5-s overlap;

2. **mel-filter bank processing**: a bank of B mel-filters is used to warp the computed spectrum bins into the mel-scale, defined as:

$$Mel(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right). \tag{6.2}$$

   The generated mel-spectrum is indicated in the following as: ${}^{MEL}\mathbf{Y}_m^{(u,r)}[b], \quad b = 1, \ldots, B;$

3. **log compression**: the range of the values of the mel-spectrum is reduced through a logarithmic transformation, that is:

$${}^{LOG}\mathbf{Y}_m^{(u,r)} = ln({}^{MEL}\mathbf{Y}_m^{(u,r)}); \tag{6.3}$$

4. **discrete cosine transform**: MFCCs are computed from the log-compressed mel-spectrum using the discrete cosine transform (DCT):

$$\mathbf{d}_m^{(u,r)}[l] = \sum_{b=1}^{B} {}^{LOG}\mathbf{Y}_m^{(u,r)}[b] \cos\left[l\left(b - \frac{1}{2}\right)\frac{\pi}{B}\right], l = 1, \ldots, L, L < B. \tag{6.4}$$

In the adopted implementation, $B = 18$ mel-filters are employed, and $L = 12$ DCT coefficients are used to generate the representation of each considered signal. The template associated to the r-th frame of user u's recording, having length $P = M \cdot L$, is eventually obtained by combining the M representations of each channel:

$$\mathbf{f}^{(u,r)} = [\mathbf{d}_1^{(u,r)}, \ldots, \mathbf{d}_M^{(u,r)}]. \tag{6.5}$$

AR Reflection Coefficients. Each EEG frame $\mathbf{y}_m^{(u,r)}$ extracted from the pre-processed signals can be modeled as a realisation of an AR stochastic process of order Q, with $Q = 10$ in the adopted implementation. According to such assumption, the available signals can be expressed as:

$$\mathbf{y}_m^{(u,r)}[n] = -\sum_{q=1}^{Q} a_{m,Q,q}^{(u,r)} \mathbf{y}_m^{(u,r)}[n-q] + \mathbf{w}_m^{(u,r)}[n] \tag{6.6}$$

where $\mathbf{w}_m^{(u,r)}[n]$ is a realisation of a white noise process having standard deviation $\sigma_{m,Q}^{(u,r)}$, and $a_{m,Q,q}^{(u,r)}$ are the autoregressive coefficients representing the model. The Yule-Walker equation [137] is used to estimate the Q autoregressive coefficients, employing the recursive Levinson algorithm and introducing the concept of reflection coefficients. In detail, being $a_{m,Q,q}^{(u,r)}$ a generic AR coefficient, we have:

$$\begin{cases} a_{m,Q,q}^{(u,r)} = a_{m,Q-1,q}^{(u,r)} + K_{m,Q}^{(u,r)} \cdot a_{m,Q-1,Q-q}^{(u,r)}, \quad q = 1, ..., Q-1 \\ \sigma_{m,Q}^{(u,r)} = \sigma_{m,Q-1}^{(u,r)} \sqrt{1 - (K_{m,Q}^{(u,r)})^2} \end{cases} \tag{6.7}$$

where the term $K_{m,Q}^{(u,r)}$ is referred to as reflection coefficient of order Q. In this study, the reflection coefficients are estimated through the Burg method [137], and employed as representative features of each user u's EEG data. For the generic r-th frame $\mathbf{y}_m^{(u,r)}$ extracted from the m-th channel of the EEG signal belonging to the user u, we therefore generate a feature vector $\mathbf{K}_m^{(u,r)}$ composed of the Q estimated AR reflection coefficients. The overall template associated to a given frame is obtained by combining the M representations generated for each channel into a single vector having size $P = M \cdot Q$, as:

$$\mathbf{f}^{(u,r)} = [\mathbf{K}_1^{(u,r)}, \ldots, \mathbf{K}_M^{(u,r)}]. \tag{6.8}$$

6.1.3 Identification

During the identification stage, the Manhattan (L1) distance is used to evaluate the similarity between features extracted during enrolment, and those obtained from an identification probe. In more detail, having indicated as $\mathbf{f}^{(u,e)}$ the template associated with the e-th frame extracted from user u's enrolment, $e = 1, \ldots, E$, and with $\mathbf{f}^{(x,i)}$ the representation generated from the i-th frame taken from the probe of an unknown subject x, $i = 1, \ldots, I$, the distance between such identification frame and the whole set of enrolment frames is evaluated as:

$$d_i^{(u)} = \min_e \left\{ \sum_{p=1}^{P} \left| \mathbf{f}^{(x,i)}[p] - \mathbf{f}^{(u,e)}[p] \right| \right\}, \tag{6.9}$$

that is, selecting the minimum among the distances computed between the i-th identification frame and all the recorded enrolment data. A decision $\hat{x}_i = \arg\min_u \{d_i^{(u)}\}$ is then taken for each available identification frame, with the final decision $\hat{x}$ regarding the identity of the presented subject taken according to a majority voting rule, selecting the identity with the highest number of occurrences among the votes $\hat{x}_i$, $i = 1, \ldots, I$.

6.1.4 Experimental Results

The aim of the present work is to analyze the recognition performance of an EEG-based recognition system exploiting an SSVEP protocol as stimulus for the involved users, taking into account issues regarding repeatability and stability across time of EEG signals. For this purpose, as remarked in Section 6.1.1, the collected database comprises EEG recordings taken, for each user, during two disjoint sessions, separated by an average time distance of 15 days. Data from the first session (S1) are considered as enrolment samples, while testing data are selected from the second session (S2). Comparing EEG samples taken during two distinct sessions allows estimating performance depending only on the peculiar characteristics of subject-specific neural activity. This way, session-specific exogenous conditions, such as the capacitative coupling of electrodes and cables with lights or computer, induction loops between the employed equipment and the body, and so on, cannot affect either inter- and intra-class variability of EEG recordings, as instead it may happen when performing tests by comparing EEG data collected during a single acquisition session [188].

In order to estimate statistically-significant results, a cross-validation procedure is carried out. Specifically, 30 different runs are performed for each of the scenarios described in the following, with 75% of the frames extracted from S1 employed as enrolment dataset for each considered user, and 75% of the frames generated from S2 randomly selected and employed as testing probes at each run.

The performance obtained when exploiting the considered elicitation frequencies $f_S \in \mathcal{F}_S = \{6, 12, 18, 24\}\ Hz$, and taking into account all the available channels for template generation ($M = 19$), is reported in terms of average correct recognition rate (CRR) in Tables 6.1 and 6.3, respectively for MFCC- and AR-based templates. Besides using the considered stimuli separately, they are also jointly employed by fusing their contributions at:

- **feature level**, by concatenating the templates $\mathbf{f}^{(u,r)}$ generated from the r-th frame of user u's EEG collected at different elicitation frequencies, during both enrolment and identification phases;
- **score level**, summing the distances $d_i^{(u)}$ obtained for each i-th identification frame matched with user u's EEG, for signals collected at different elicitation frequencies;
- **decision level**, adopting a majority voting rule over the final decisions $\hat{x}$ individually taken considering EEG data collected at different elicitation frequencies.

Table 6.1: *Average correct recognition rate (CRR %) obtained over 30 cross-validation runs, using MFCCs as features. The considered subbands are reported in terms of range of associated frequencies, thus in terms of brain rhythms:* $[0.5, 40]Hz = \{\delta \cup \theta \cup \alpha \cup \beta \cup \gamma\}$, $[0.5, 30]Hz = \{\delta \cup \theta \cup \alpha \cup \beta \cup \gamma\}$, $[4, 40]Hz = \{\theta \cup \alpha \cup \beta \cup \gamma\}$, $[4, 30]Hz = \{\theta \cup \alpha \cup \beta\}$, $[8, 40]Hz = \{\alpha \cup \beta \cup \gamma\}$, $[8, 30]Hz = \{\alpha \cup \beta\}$.

Channels	**SSVEP**	**EEG subband**					
		$[0.5, 40]Hz$	$[0.5, 30]Hz$	$[4, 40]Hz$	$[4, 30]Hz$	$[8, 40]Hz$	$[8, 30]Hz$
All ($M = 19$)	$f_S = 6\ Hz$	70.93	76.67	85.07	86.80	73.60	71.87
	$f_S = 12\ Hz$	92.67	93.73	94.40	92.67	84.80	87.20
	$f_S = 18\ Hz$	**94.53**	90.80	88.27	87.07	84.13	80.00
	$f_S = 24\ Hz$	89.73	88.93	87.87	89.33	85.33	85.73
	$f_S \in \mathcal{F}_S$, feat. fus.	97.47	94.67	**99.73**	98.67	99.33	95.73
	$f_S \in \mathcal{F}_S$, score fus.	95.73	**99.33**	96.27	97.33	97.33	93.33
	$f_S \in \mathcal{F}_S$, dec. fus.	99.47	97.33	99.87	**100.00**	98.87	98.00

Table 6.3: *Average correct recognition rate (CRR %) obtained over 30 cross-validation runs, using AR reflection coefficients as features. The considered subbands are reported in terms of range of associated frequencies, thus in terms of brain rhythm's:* $[0.5, 40]Hz = \{\delta \cup \theta \cup \alpha \cup \beta \cup \gamma\}$, $[0.5, 30]Hz = \{\delta \cup \theta \cup \alpha \cup \beta\}$, $[4, 40]Hz = \{\theta \cup \alpha \cup \beta \cup \gamma\}$, $[4, 30]Hz = \{\theta \cup \alpha \cup \beta\}$, $[8, 40]Hz = \{\alpha \cup \beta \cup \gamma\}$, $[8, 30]Hz = \{\alpha \cup \beta\}$.

Channels	**SSVEP**	**EEG subband**					
		$[0.5, 40]Hz$	$[0.5, 30]Hz$	$[4, 40]Hz$	$[4, 30]Hz$	$[8, 40]Hz$	$[8, 30]Hz$
All ($M = 19$)	$f_S = 6\ Hz$	72.93	74.67	78.53	79.07	66.93	64.40
	$f_S = 12\ Hz$	88.40	88.93	85.47	82.93	79.47	78.67
	$f_S = 18\ Hz$	93.60	93.47	**94.80**	93.47	86.93	83.87
	$f_S = 24\ Hz$	79.73	79.73	88.80	87.20	82.27	88.80
	$f_S \in \mathcal{F}_S$, feat. fus.	96.27	91.60	**98.93**	96.53	93.87	94.13
	$f_S \in \mathcal{F}_S$, score fus.	94.27	92.27	98.93	**99.33**	96.67	88.27
	$f_S \in \mathcal{F}_S$, dec. fus.	**99.73**	98.80	99.60	97.87	98.27	98.53

Table 6.5: *Average correct recognition rate (CRR %) obtained when different spatial configurations are selected and 30 cross-validation runs are performed. The considered subbands are reported as the range of frequencies they span, thus in terms of brain rhythms:* $[0.5, 40]Hz = \{\delta \cup \theta \cup \alpha \cup \beta \cup \gamma\}$, $[0.5, 30]Hz = \{\delta \cup \theta \cup \alpha \cup \beta\}$, $[4, 40]Hz = \{\theta \cup \alpha \cup \beta \cup \gamma\}$, $[4, 30]Hz = \{\theta \cup \alpha \cup \beta\}$.

Channels	**SSVEP fusion**	**MFCCs**			**AR reflection coefficients**		
		$[0.5, 30]Hz$	$[4, 40]Hz$	$[4, 30]Hz$	$[0.5, 40]Hz$	$[4, 40]Hz$	$[4, 30]Hz$
Frontal ($M = 7$)	Feature	81.73	85.46	79.60	81.87	75.20	74.00
	Score	**94.13**	90.67	84.67	89.46	**91.20**	82.27
	Decision	87.33	91.06	88.80	89.07	88.13	85.6
Central ($M = 7$)	Feature	78.13	88.27	90.00	80.4	83.60	84.40
	Score	86.00	90.13	89.60	84.00	84.53	88.13
	Decision	89.33	95.33	**95.87**	86.93	93.6	**94.53**
Occipital ($M = 7$)	Feature	84.67	85.87	84.40	90.8	86.40	86.27
	Score	74.80	78.13	80.93	77.60	79.47	79.47
	Decision	88.16	86.87	**89.33**	88.13	**90.93**	83.47
$\mathcal{M}$ ($M = 5$)	Feature	92.13	89.6	86.26	91.47	88.13	81.06
	Score	90.40	88.80	86.60	83.60	82.13	72.67
	Decision	**96.00**	94.80	93.73	**91.47**	88.13	84.67

As can be seen, for systems employing a single SSVEP elicitation frequency as stimulus, $f_S = 18\ Hz$ guarantees the best achievable identification rates, with CRR = 94.53% obtained using MFCCs to represent EEG data in the [0.5, 40] Hz subband, and CRR = 94.80% employing AR features estimated from EEG signals in the [4, 40] Hz subband. The considered fusion strategies allow to significantly improve such performance, being able to offer a perfect recognition rate (CRR = 100.00%) when a decision-level fusion is performed on information generated through MFCCs, while CRR = 99.73% when exploiting decision-level fusion with AR features.

Given the high accuracy obtained when exploiting all the available 19 channels, further tests are carried out to check whether similar results can be obtained while lowering the number of employed channels. It is worth remarking that minimizing the number of employed electrodes is an issue of paramount importance to reduce user inconvenience. In this regard, Table 6.5 reports the performance obtained when considering only $M = 7$ electrodes placed in either frontal, central and occipital areas, according to the montages shown in Fig. 6.1.(b), together with the rates obtained with an even smaller set $\mathcal{M} = \{F_z, C_z, P_z, O_1, O_2\}$ with $M = 5$ electrodes, comprising only midline and occipital channels. Only the recognition rates achieved exploiting all the considered elicitation frequencies through fusion approaches, and taking into account the best-performing subbands according to the results shown in Tables 6.1 and 6.3, are reported in Table 6.5. From the obtained accuracies it can be seen that the central area of the scalp seems guaranteeing the best performance achievable with a reduced number of electrodes, achieving CRR = 95.87% for MFCC and CRR = 94.53% for AR representations, when considering EEG recordings filtered in the $\theta \cup \alpha \cup \beta$ subband. An even better result is obtained when considering only the set $\mathcal{M}$ with $M = 5$ in the $\delta \cup \theta \cup \alpha \cup \beta$ subband, for which a CRR = 96.00% is achieved using MFCCs, while AR features provides CRR = 91.47%.

6.1.5 Conclusions

This study evaluates the feasibility of designing an automatic biometric recognition system exploiting EEG signals elicited through protocols generating steady-state visual evoked potentials (SSVEPs). The use of flickering stimuli at specific frequencies and the representation of the acquired EEG data through either MFCC or AR templates, allows achieving high identification rates, thanks to the proved existence of permanent character-

istics in SSVEP brain responses across different acquisition sessions. According to the reported experimental tests, the joint use of multiple elicitation frequencies guarantees a notable improvement in recognition rates, thus allowing to reduce the number of electrodes needed during EEG collection, a relevant property to foster the adoption of EEG-based biometric identifiers in practical recognition systems.

CHAPTER 7

Visible Beyond Invisible Biometrics: Hidden Information, Emotions and Working Memory

THE ADVANTAGES OF BIOMETRIC SYSTEMS linked to ease and comfort in interaction with the authentication device is leading to an always increasing success of biometric-based recognition systems. Additionally, biometric technologies requiring very little cooperation or participation from the users may be perceived as being more convenient by the subjects. On the other hand, biometric characteristics that do not need user participation can be captured without the knowledge of the user, and this is perceived as a threat to privacy by many individuals.

Because biometric technologies are based on measurements of physiological or behavioural characteristics of the human body and the collection and storage of personal data, they raise a host of ethical concerns related to the protection of individual values such as privacy, autonomy, bodily integrity, dignity, equity, and personal liberty [282]. The main ethical concerns about the application of biometrics are related to privacy issues. Since huge amounts of data are collected and kept in databases, there is a conse-

quent threat to the security of persons' identity and data protection: the process of biometric recognition leaves behind trails of private information. For example, if a person is identified each time she makes a purchase, information about where this person shops and what she buys can be simply collected and used by telemarketers to invade her privacy. The issue of privacy becomes more serious with biometric-based recognition systems because biometric characteristics may provide additional information about the background of an individual. In fact, some biometric identifiers may provide medical information about the users of the system. A health insurance company may use this information in an unethical way for economic gains by denying benefits to a person determined to be of high risk. Eventually, people fear that biometric identifiers could be used for linking personal information across different systems or databases.

The invisible biometric identifiers considered in this thesis are not immune from the aforesaid shortcomings. Diseases related to the cardiovascular system and medical state of a human can be revealed from vein pattern [99]. In general the diameter and the position of the veins are of medical interest. An example is thrombosis, where a blood clot (thrombus) blocks the blood flow in the cardio-vascular system. Diseases changing the position and the structure of the vein network affect all feature extraction methods resulting in a vein pattern. The appearance of the hand vein pattern are changed by two kind of diseases: arteriovenous malformation (AVM), that is a congenital disorder where veins and arteries are connected in an abnormal way; the second abnormality is the hypothenar hammer syndrome (HHS) which is also identifiable throughout the vein pattern of the hand. HHS is a thrombosis of the superficial palmar arch of the ulnar artery and is caused by repeated mechanical force, as seen in fighting sports or the work with vibrating tools (e.g. a hammer) [161].

The visible health information that can be revealed by the EEG invisible biometric is huge. In fact, EEG is a valuable instrument and it has been first used in the diagnosis and treatment of spinal cord injuries, strokes, and brain disorders including epilepsy, Alzheimer's disease, schizophrenia, and Parkinson's disease [34]. Consequently, EEG-based recognition systems, if they are not provided with a privacy-protection mechanisms, can reveal personal health information. In addiction, the human body reveals emotional states through measurable physiological responses, such as heart rate, blood pressure, skin conductivity, muscle tension, facial expressions, pupil diameter, voice, body movements and posture. In addition to periphery biosignals, signals captured from the brain in central nervous system (CNS) have been proved to provide informative characteristics in

responses to the emotional states. EEG has been used in cognitive neuroscience to investigate the regulation and processing of emotions for the past decades [177]. Besides, mental stress is also detectable through the analysis of EEG signals [129].

Emotional and stress states might be induced by a specific acquisition protocol employed by an EEG-based biometric system, thus implying that the recorded biometric data reveal further personal information of the user. Eventually, one of the most important brain tasks is the processing of sentences for both understanding and knowledge updating. More in detail, humans continuously make predictions about the contents a speaker is about to convey next, on the basis of information already available in the foregoing discourse. Moreover, participants build expectations that forthcoming contents are presented in ways coherent with their having been already introduced or not, and with their relevance to the communicative task at hand. The process of the brain that is responsible of the understanding and prediction of sentences is the working memory. When a person is presented a sentence, one possible information detectable from EEG signals is the brain processing cost associated to misalignment with respect to the way information is expected to be organised within utterances [160].

In this Chapter, an example of visible beyond the EEG invisible biometric trait is investigated. More in detail, in Section 7.1 the analysed visible information is the brain cost associated to the processing of sentences containing linguistic misalignment. The sentence are presented as audio stimuli, so the adopted scenario could be seen as an EEG-based biometric system where the brain waves are collected as response to audio stimuli.

7.1 Information Structure Effects on the Processing of Nouns and Verbs: Evidence from Event-Related Potentials and Brain Oscillatory Dynamics

Thanks to the great temporal resolution that characterises them, electroencephalographic (EEG) signals have been often analysed to gain insights into the brain processes which are carried out during language processing tasks. In more detail, investigations on language processing have been performed considering event-related potentials (ERPs) since the early 1980s [154,156, 296]. ERPs are voltage changes of the electrical activity of the brain, and can be induced by sensory or cognitive events [181]. Two ERP signatures, N400 and P600, have been found to strongly correlate with the brain response to linguistic inputs. Specifically, N400 is a negative component peaking between 300 and 500 ms after stimulus onset, and its elicitation

has been associated with difficulties in lexical-semantic retrieval [154,165], semantic integration/unification mechanisms [92], the processing of more or less expected information structural patterns [50, 194, 311], and the decoding of non-literal meanings [155, 316]. P600, a component peaking between 500 and 800 ms, has been originally observed in parsing difficulties caused by syntactic violations or garden path sentences [91, 130, 218], yet its functional role has been also associated to mechanisms of context update [29, 107] and new information decoding [30, 68].

Other recent studies have shown revived interest in the exploration of brain oscillatory rhythms (power spectrum density levels, henceforth PSD levels) in language comprehension "as a result of the view that they might provide a window on the dynamics of the coupling and uncoupling of functional networks involved in cognitive processing" [16]. Analyzing the PSD behavior allows measuring brain activity beyond limited phase-locked responses, which provide only partial windows on the inner workings of the brain. Language-related activity in oscillatory rhythms generally manifests as power increases (Event-Related Synchronization) or decreases (Event-Related Desynchronization) in different frequency bands. The frequency ranges that have been observed to be mostly involved in language processing are theta (4-8 Hz), delta (1-4 Hz), alpha (8-12 Hz) and beta (13-30 Hz). Semantic and syntactic violations have been reported to correlate with increases in theta power [16], also elicited by the processing of object-relative vs. subject-relative clauses [318]. Differences between semantically congruous and incongruous sentence endings have been instead reflected in larger coherence in the gamma band [317]. Conversely, power decreases in the beta band have been elicited by syntactically complex constructions. In an insightful view put forward by [16], neuronal dynamics in the theta and alpha frequency ranges would be revealing of more demanding retrieval operations, which impose larger demands on working memory activity, while oscillations in the gamma and beta bands would be generally indicative of semantic and syntactic unification mechanisms. Also in the domain of information structure processing, stronger synchronization and desynchronization effects in the theta and alpha bands, respectively, have been found to correlate with topics conveying new information and focused constituents conveying given information [160].

The present work aims at assessing the contribution of both ERPs and PSD levels in exploring how the brain deals with a special type of language interface, namely the one between the micropragmatic and the word class level of a sentence. Notably, variations in the brain response, in terms of both evoked potentials and PSD levels in different frequency bands, will be

inquired when more or less expected combinations between word classes (mainly noun and verb) and distinct patterns of information structure (i.e. Topic-Focus articulations) are processed.

This Section is organised as follows. Section 7.1.1 reports the state of the art regarding the aspects of language processing whose correlation is here investigated, namely noun and verb processing and informational dichotomies decoding. In Section 7.1.2 a working definition of information structure units is provided, and their relation to noun and verb classes in language use is canvassed. Building on this, the prediction that there should be some sort of "processing preference" for topical nouns over focal nouns, and for focal verbs over topical verbs, is formulated in Section 7.1.3. Section 7.1.4 describes the experimental design adopted to test our predictions on the neurophysiological response to distinct patterns of associations between noun and verb categories and information structure units. Results from both ERP measurements and brain oscillations are then presented in Section 7.1.5 and discussed in Section 7.1.6, while conclusions eventually drawn in Section 7.1.7.

7.1.1 Literature Overview

The literature regarding noun and verb processing and experimental findings on Information Structure processing are outlined in the following paragraphs.

Noun and Verb Processing

Brain response to nouns and verbs has been the object of several neurophysiological investigations over the last two decades [37, 52, 238]. Both fMRI and ERP studies report fairly consistent topographic specialisations of these two word classes in the human brain, with nouns mainly activating visual (cortical regions) and verbs chiefly involving pre-frontal and frontal motor regions [37]. Different processing patterns, though, have appeared less consistent and less robust in other works in which grammatical class detection produced a more remarkable response only when extended sentence contexts were taken into account in experimental stimuli [173]. ERP measurements have also proved useful to unravel how word class processing taps into the construal of other levels of analysis, and earlier and more recent studies in this respect have revealed that semantic and grammatical distinctions between nouns and verbs is bound to emerge even earlier than the canonical N400 time interval [211, 347]. In other experiments, more taxing processing has been observed for verbs due to their greater

morphological and semantic complexity since they designate events, which necessarily involve other participants [7]. These findings however appear less systematic when it comes to ambiguous verbs and nouns such as the English *cut*, *kiss*, *head*, etc., which can function either as verbs or nouns depending on their context of occurrence. Indeed, using English words of this kind, Federmeier *et al.* [76] conducted an ERP study to assess the extent to which manipulation of prior contextual information made the processing of nouns and verbs more or less costly. Notably, presenting short texts with ambiguous nouns and verbs alternatively embedded in verb-predicting and noun-predicting contexts, the authors noticed that more prominent N400 deflections were elicited by both nouns and verbs in less expected contexts (i.e. nouns embedded in verb-predicting contexts and verbs embedded in noun-predicting contexts). They thus concluded that rather than correlating with neatly delimited patterns of neural activation, word class distinctions "rather emerge in real-time from an interaction of semantic and syntactic properties at both the single-word and the discourse level" [76]. It should be noted that, differently from the grammatical and semantic representation of word classes, their interplay with discourse structure has been less extensively investigated within the neurophysiological purview, and even less is the correlation between different parts of speech and their information structure profile. The present work intends to contribute to this line of research by further developing Federmeier *et al.*'s premises on the role played by discourse in facilitating word class differentiation.

Information Structure processing

Most of what we know about information structure processing comes from behavioural and EEG studies [22, 110, 270, 280], among others. In the behavioural domain, the psychological processes underlying the mental encoding of topical vs. focal information have mainly been investigated throu- gh reading times and eye movement measures, which yielded overall greater processing demands elicited by focused information, as opposed to topical information [22]. Possibly due to the adoption of more extensively contextualised stimuli, subsequent neurolinguistic experiments revealed quite deflecting processing trends of information units, in that increasing costs were not only observed in association to information statuses per se, but also - and even more conspicuously - as conditional upon more or less expected syntactic realisations [31], phonological profiles [17,50], and activation degrees in discourse [311]. In these accounts, Topics conveying new information [311] or realised by object dislocation strategies [31] are reported to cost more than topics carrying given information and realised

by syntactic subjects. By the same token, focused phrases lacking intonational prominence appear to be costlier than Focuses conveying given content and displaying typically prominent intonation contours. These and other findings on the whole converge on the involvement of both N400 and P600 responses which, as discussed in the mainstream literature, respectively reflect mismatches detection at both the semantic and the discourse level [68, 154, 194] as well as difficulties in context updating [30]. Phonological, syntactic or context-dependency features inconsistently matching with information statuses generally elicited greater N400 responses, sometimes accompanied by subsequent positive deflections. Brain response to more or less expected information structural patterns has also been inquired in the frequency domain in which increases of the Power Spectrum Density have been found in the theta band with corresponding decreases in the alpha band while processing information structures inconsistently matching with the activation state of contents [160].

So, much of what is at play in information structure processing is profoundly contingent on the level of expectations interlocutors entertain on the distribution information receives in an utterance and the types of interactions it displays with other levels of sentence representation. In the present study, the level of expectations we propose to look into concerns the relation between Topic and Focus units and the lexical categories filling them in a sentence.

7.1.2 Theoretical Views: Information Structure and Word Classes

Since its very discovery as an independent level of utterance organization (related to- but not subsumed by semantics or syntax), Information Structure was defined in terms of predicativity. The founding remarks by the Second Prague School, beginning at the half of the XX century, cf. [53–55, 79, 80], led to calling as *Theme* and *Rheme* the fundamental units of what was then called an utterance's *Functional Sentence Perspective*, with the first seen as "what the utterance is about", and the second as "what the utterance tells about the Theme". Even etymologically, and absolutely not by chance, the Theme is conceived as typically encoding *reference to some object or entity*, while the Rheme is the *predication*, the part of the utterance encoding what is actually said.

Halliday [93] introduces *Thematic Structure* as a feature of the clause. In accordance with the Prague School terminology, he defines the Theme as "the element which serves as the point of departure of the message" expressed by the clause, "that with which the clause is concerned". The

Rheme, conversely, is defined as "the remainder of the message, the part in which the Theme is developed". A similar definition suggests a strong affinity between the Theme and nominal constituents on the one side, between the Rheme and verbal or in general predicative constituents on the other. Halliday himself remarks that "a Predicator is rarely thematic". In sum, Theme and Rheme seem to present themselves (by definition, and in actual utterances) as two complementary parts of any message encoded by a clause, which is made of an entity ("what the message is about", the Theme) and a predication ("what is told", the Rheme), exactly as the clause is made - syntactically - of a nominal and a verbal part. What is thematic has the nature of an entity, what is rhematic that of a process.

Currently, the terms Theme and Rheme have been replaced in most of the literature by Topic and Focus respectively, but the concepts remain essentially the same. Emanuela Cresti's pathbreaking work [51] has shown that Topics have their typical (mainly ascending-descending) "Topic-contours", while Focuses are produced under the various contours which describe the utterances' illocutions. In an assertion, the Topic will be prosodically produced as a Topic, and the Focus will carry an assertive contour. In a question, the Topic will again have its Topic contour, but the Focus will carry an interrogative (ascending) contour. The same for an illocutionary act of command, protest and so on. In other words, prosody crucially shows that while the Topic of the utterance only encodes the entity to which the illocutionary act will apply, the Focus is responsible for the illocution, that is, for the particular kind of predication encoded by the utterance.

Among others, [51] have shown pretty well, on huge amounts of data belonging to corpora of spontaneous speech, that an information unit carrying the function of a Topic can actually be made of any kind of syntactic constituent, and the same holds for a Focus. Therefore, information structure is largely independent from syntax. Considering for example the following sentences:

(1)
A. Is John in town?
B. John went to China.

(2)
A. Who is representing us in China now?
B. JOHN went to China.

the clauses contained in the "B" utterances in (1) and (2) are actually different, despite the apparent syntactic identity. In fact, in (1) the utterance

is about John, and it predicates that he went to China. Hence, *John* is the Topic of the message, and *went to China* is the Focus. In (2), conversely, the utterance is about going to China, and it predicates that it is John who did it. In other words, *went to China* is the Topic of the message, and *John* is the Focus. Now, the case represented by (2) is possible and even frequent, but (1) is the default case. It is more natural for nominal constituents to realize the nomination of entities, and for verbal constituents to realize the predication of the message. In the mentioned examples, this can be seen from the fact that language is organized to express the first case by means of the unmarked, default construction, while the second case requires marked, contrastive intonation.

In more detail, it has been shown from vaste corpora of spontaneous speech that nominals are more frequently associated to Topics, while verbal constituents more frequently realize Focuses. For example, [204] has shown that in the C-ORAL-BRASIL Brazilian Portuguese corpus, nominal Topics are more than twice as frequent as verbal Topics, while the ratio found by [38] in a vast American English corpus was 7:1[1]. [51] report that, in a representative corpus of Italian spontaneous speech, Topic units are filled nearly 60% by noun phrases and nearly 40% by other constituents, including adverbial phrases, adjectival phrases, prepositional phrases, and subordinate as well as main clauses. By the same token, Focus units are filled nearly 62% by verb phrases and nearly 38% by adverbial phrases, adjectival phrases, prepositional phrases and noun phrases[2].

These observations lead us to formulate the prediction that the associations between the syntactic categories of Noun and Verb and the information categories of Topic and Focus may not be inter-independent, but oriented. More precisely, the processing of nominal Topics and verbal Focuses should be more natural and less costly in terms of required brain processing, being the most frequent and more "homogeneous" option: Nouns are already made for denoting entities and Verbs are already made for predicating about entities. On the contrary, the processing of verbal Topics and nominal Focuses should be less natural and more costly, being the less frequent and less "homogeneous" option: Verbs must be presented in a syntactically/prosodically marked way in order for them to denote an entity, and nouns must be treated in a marked way if they are to express a predication.

[1]These figures are extracted by Mittman's and Cavalcante's data by considering utterances whose information structure does not involve more than one clause. They do not consider those cases where, in a complex sentence, the Topic of the utterance can be an entire clause, possibly made of both nominal and verbal constituents.

[2]In their terminology, the Focus is called "Comment". The quoted data are actually drawn on the PhD thesis of Sabrina Signorini, *Topic e soggetto in corpora di italiano parlato spontaneo*, discussed under their guidance at the University of Florence

The aim of the present study consists in verifying the plausibility of such predictions, analysing the costs required by the brain when processing sentences with different kinds of associations between the syntactic categories of Noun and Verb and the information categories of Topic and Focus. Specifically, electroencephalographic (EEG) signals, giving information on the electrical activity of the brain, are exploited to perform such analysis. Both EEG event-related potentials (ERPs), that is, time- and phase-locked brain responses measured as the direct result of specific cognitive events, as well as EEG power spectral density (PSD), representing non-phase-locked activity, are used as descriptors of the brain workload in the considered scenarios.

7.1.3 Predictions

Capitalising on the findings above discussed, we expect differences between Topic-Noun/Focus-Verb and Topic-Verb/Focus-Noun combinations to strikingly emerge in modulations in the N400 signature. Notably, a stronger negative response is expected to be elicited by less homogeneous information structure/word class matchings, represented by topical verbs and focused nouns. An N400 response would be consonant with previous accounts on the expectation-related nature of this component [8, 155] and, particularly, with unmet predictions on information packaging strategies [50]. In the experimental paradigm used, no given-new opposition has been measured for the critical information, so we should not expect potential P600 effects to be driven by the activation status parameter [39]. Nonetheless, P600 involvement can be predicted to possibly hint at costs of discourse model enrichment mechanisms, once the mismatch - assumingly correlating with an N400 response - has been solved. Zooming in on the categorial combinations considered, it is reasonable to expect more remarkable positivity effects for the Topic-Verb condition, being it on the whole less frequent than the Focus-Noun condition in conversation.

In the frequency domain, greater amplitudes in the N400 component are expected to be accompanied by increasing synchronization effects in the theta band, as suggested in previous studies highlighting the correlation between this band and negativity patterns in conditions of working memory overload [317].

A less expected effect of information structure/word class combinations may also regard a power decrease in the alpha and beta bands, hinting at increasing attentional demands related to the recognition of unexpected stimuli [71, 317].

7.1.4 Methods

The performed experimental tests are described in the following. Specifically, the adopted experimental design, the employed stimuli, the collected data and the performed data processing are outlined in the following paragraphs.

Experimental Design

In order to collect a proper number of brain responses to all the interesting combinations between information structure and word class (Noun and Verb), a set of 60 texts composed of three-sentence passages has been created, with each passage made of a two- sentence context, followed by a target sentence. The critical region in the target sentence contains a Noun or a Verb either realised as Topic or as Focus. To avoid potential overlapping with other discourse phenomena, mainly indefinite phrases have been considered for the Noun set (Table 7.1), since definite noun phrases would have been interpreted as triggering a presupposition, thus blurring topicalization and focalization effects. As for the Verb set, mainly infinitives have been used, since they can be flexibly moved from Topic to Focus position without remarkable infelicity effects (at least in Italian, the same would hardly obtain with fully-inflected verbs).

As can be seen from the examples in Table 7.1, texts have been arranged in pairs, so that the same two-sentence context can be followed by two different target sentences, with a Noun (or a Verb) in either Focus or Topic condition. To test the predictions outlined in Section 4, the design has been constructed so as to assess the interaction between the two main independent variables of the study, i.e. *Type* (Noun, Verb) and *Condition* (Topic, Focus), and how such interaction is reflected in the two dependent variables considered, ERP measurements and PSD levels.

Stimuli

In more detail, to isolate the effects of information packaging and word class variation from those related to the discourse availability (givenness vs. newness, [39]) of contents, which strongly modulates sentence processing [15, 29], we have chosen to keep all regions of interest equally new. Therefore, the critical Nouns or Verbs, in Topic or in Focus condition, are always context-independent. Differently from other studies such as [17], [160], and [110], among others, where expectations on information structure processing have been measured relative to the degree of activation

Table 7.1: *Examples of the experimental stimuli (target nouns and verbs are bold-typed).*

	Topic	**Focus**
	Context	*Context*
	Stamattina, Carlo si è svegliato molto presto.	Stamattina, Carlo si è svegliato molto presto.
	Entrato in macchina, è uscito dal passo carrabile.	Entrato in macchina, è uscito dal passo carrabile.
	[Eng. This morning, Carlo has woken up very early.	[Eng. This morning, Carlo has woken up very early.
	He got into the car and left through the garage.]	He got into the car and left through the garage.]
NOUN SET		
	Target Sentence	*Target Sentence*
	In quel momento, un **gatto** è passato davanti al	In quel momento, davanti al cancello è passato un
	cancello e sembrava molto impaurito.	**gatto**, e sembrava molto impaurito.
	[Eng. At that moment, a **cat** crossed the	[Eng. At that moment, in front of the entrance
	entrance gate and seemed really scared.]	gate, there came a **cat**, who seemed really scared.]
	Context	*Context*
	Di recente, Paolo è stato lasciato dalla sua	Di recente, Paolo è stato lasciato dalla sua
	fidanzata. Era molto innamorato di lei.	fidanzata. Era molto innamorato di lei.
	[Eng. Paolo has been recently dropped by his	[Eng. Paolo has been recently dropped by his
	girlfriend. He was deeply in love with her.]	girlfriend. He was deeply in love with her.]
VERB SET		
	Target Sentence	*Target Sentence*
	Voleva dimenticarla per sempre, così **viaggiare**	Voleva dimenticarla per sempre così l'unica soluzione
	è stata l'unica soluzione: voleva conoscere gente nuova.	è stata **viaggiare**: voleva dimenticarla per sempre.
	[Eng. He wanted to forget her for good, so **travelling**	[Eng. He wanted to forget her for good, so the only solution
	was the only solution: he longed to meet new people.]	solution was **travelling**: he longed to meet new people.]

of the contents carried by topical or focal units, in this study we are mainly interested in brain responses to topicalisations and focalisations as realised by different word classes which, to us, makes the unvaried information status parameter even more compelling.

The position of the target word has been carefully determined for both the Condition and the Type factors. Particularly, for the Topic condition, the average position of critical nouns in the target sentence is 5 (SD= 1), whereas for verbs it is 4.5 (SD = 1.7). In the Focus condition, the position of nouns is approximately fixed at 10 (SD = 1.8), while for verbs it is 9 (SD = 2.3). Overall, the distribution of critical nouns and verbs is therefore fairly homogeneous within and between the Topic and Focus conditions, meaning that the effects of Topic vs. Focus packaging should not be distorted by unsystematic positional oscillations of the target words. As a result, the syntactic encoding of critical words as Topic or Focus, at least in terms of sentential position, is expected to be more comparable between the Noun and the Verb set. In the target sentences, the mean length of critical words did not significantly differ for the chosen Noun and Verb sets, nor did their overall frequency in common language uses, as the resulting mean values show (NOUN = 25,85; VERB = 20,83). Furthermore, in compliance with standard normalising measures in experiments utilising context-target pairs as stimuli, the naturalness of all texts has been judged on a 5-point Likert scale by another group of subjects in an offline questionnaire. A two-away ANOVA on the collected responses showed no significant interaction between the Condition (Topic, Focus) and Type (Noun, Verb) parameters ($F < 1$). This suggests that any effect to be foreseen at the electrophysiological level should not be put down to unnatural or implausible features of the stimuli.

Data Collection

Thirty-five students (7 men, mean age = 22.8, SD = 3.5) from the University of Roma Tre have taken part in the experiment. Data from three participants were excluded because of excessive number of artefacts. All subjects were right-handed (mean laterality = 0.81, SD = 0.16, cf. [216], native Italian speakers, with normal or corrected-to-normal vision. None of them reported history of neurological or psychiatric disorders. Informed consent was obtained from all subjects prior to each experimental session.

During the experiment, participants sat in a dimly-lit, sound-attenuated room. The stimuli described in the previous paragraph have been submitted as audio tracks. During each track, subjects were asked to look at a fixation cross in the centre of a computer screen.

The 60 pairs of texts have been arranged into two randomised lists according to a Latin Square design, so that each participant was presented with only one occurrence of the two-sentence context whose target sentence contain either a Noun or a Verb in Topic or in Focus condition. Both lists also contained further 30 fillers, randomly interspersed between the experimental trials, with no marked topicalizing or focalizing constructions.

In order to make sure that the investigated Type $\times$ Condition interactions had no significant effect on the comprehensibility of the designed texts, all experimental stimuli were accompanied by two verification questions presented visually on the computer screen. After reading each question, subjects had to press a TRUE/FALSE button on the keyboard.

During the presentation of the stimuli, EEG signals of the participants have been acquired using a 19-channels system GALILEO Be Light Amplifier, with an original sampling rate of $S_r = 256$ Hz. The electrodes were placed on the scalp according to the 10-20 standard montage, and the electrical impedance was kept under 10 kΩ using conductive gel at the beginning of each acquisition. The EEG measures are referenced to the AFz position, and represented as potentials $v^{(c)}[t]$ between the c-th electrode and the reference electrode, with $c = 1, \ldots, C = 19$. EEG recordings have been time-locked to the presentation of the target words, represented by the head noun of the indefinite phrase for the Noun set (see Table 1), and by the infinitive verb for the Verb set. The obtained synchronisation signal has been used to lock the raw EEG traces to the occurrence of the words of interest.

Data Processing

The acquired EEG signals are filtered through the application of a band-pass filter in order to retain spectral components in the range $[0.5 - 40]Hz$, containing the main EEG rhythms of interest for the present study. Subsequently, EEG signals are segmented into epochs time-locked to the words under analysis. More in detail, being t_0 the instant relative to the stimulus' end, an epoch of length $T = T_{preS} + T_{postS}$ is selected for each EEG signal, considering $T_{preS} = 500$ ms before t_0, that is the pre-stimulus, and $T_{postS} = 1500$ ms after t_0. The result of the aforementioned process is a set of $N_T = 60$ trials of ERPs $v_n^{(c)}[t]$, $n = 1, \ldots, N_T$, for each of the considered participant. Starting from the segmented signals, artefacts related to eye-related activity are removed through an approach based on independent component analysis (ICA, [199]), thus generating the filtered samples $\hat{v}_n^{(c)}[t]$, $n = 1, \ldots, N_T$ and $c = 1, \ldots, C$.

Subsequently, the epochs' windows of length T_{postS} ms starting at t_0 and ending T_{postS} ms later are referenced and normalised to the pre-stimulus baseline. More in detail, the normalisation procedure can be formalised as follows:

$$\bar{v}_n^{(c)}[t] = \hat{v}_n^{(c)}[t] - \frac{1}{T_{preS} * S_r} \sum_{\xi=-T_{preS}*S_r}^{0} \hat{v}_n^{(c)}[\xi], \tag{7.1}$$

with $1 \leq t \leq S = T_{postS} * S_r$.

The normalised trials of ERP are then processed through the multivariate outlier detection procedure defined within the t-test continuous wavelet transform (t-CWT) method [24], originally introduced as a solution for the problem of single-trial ERP classification. The aim of this step is to remove possible outlier ERP trials from the subsequent analysis, which should be therefore performed only on N_{Tout} survival samples instead of the original N_T ones. The adopted approach iteratively performs a principal component analysis (PCA) on a frequency representation of the available dataset of ERP trials. At each step, the samples whose distance from the total mean is greater than a threshold depending on the computed PCA eigenvalues are marked as to be deleted, and not considered for further PCA evaluations, till the same set of trials is marked in two consecutive iterations. The outliers thus determined are discarded from the following analysis, aimed at characterising the brain activity during the considered language processing tasks through two distinct descriptors, namely the averaged ERPs and the PSD as illustrated in the following.

In more detail, representing a single n-th ERP trial as a $L = C \cdot S$-dimensional row vector $\mathbf{v}_n = \left[\bar{v}_n^{(1)}[1], \ldots, \bar{v}_n^{(1)}[S], \ldots, \bar{v}_n^{(C)}[1], \ldots, \bar{v}_n^{(C)}[S]\right]$, the whole dataset of N_T trials could be described by the N_T-by-L matrix:

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_{N_T} \end{bmatrix} \tag{7.2}$$

As detailed by [24], performing the outliers detection procedure in the frequency domain is computationally convenient. The ERP trials could be therefore represented through their discrete Fourier transform (DFT) as $\mathbf{V}_f = \mathbf{VW}$, being $\mathbf{W}$ the L-by-L orthogonal DFT matrix. Yet, since only the frequencies contributions in the range of interest are needed, that is, with $f \leq 2f_c$ being f_c the cut-off frequency, the columns of $\mathbf{W}$ corresponding to

frequencies $f > 2f_c$ are set to zero, thus defining the "reduced" DFT matrix $\hat{\mathbf{W}}$. Furthermore, in order to smooth the cut-off, the vector components corresponding to the frequencies of the last octave, that is, $f_c < f \leq 2f_c$, can be gradually attenuated through the use of a L-by-L diagonal matrix $\mathbf{R}_f$, whose diagonal elements are given by the function:

$$r(f) = \begin{cases} 1 & \text{for } f \leq f_c \\ 2 - \frac{f}{f_c} & \text{for } f_c < f \leq 2f_c. \end{cases} \tag{7.3}$$

This way, a filtered representation of the treated trails can be obtained as $\tilde{\mathbf{V}}_f = \mathbf{V}\hat{\mathbf{W}}\mathbf{R}_f$, using $f_c = 30$ Hz as cut-off frequency for the considered ERPs.

Indicating now with $\mathbf{\Omega}$ the covariance matrix of $\tilde{\mathbf{V}}_f$, and with $\mathbf{T}_p$ the matrix corresponding to the principal component transform, after performing the PCA we obtain:

$$\mathbf{V}_p = \tilde{\mathbf{V}}_f \mathbf{T}_p, \; \mathbf{\Omega}_p = \mathbf{T}_p^T \mathbf{\Omega} \mathbf{T}_p, \tag{7.4}$$

where $\mathbf{\Omega}_p$ is the diagonalised covariance matrix. The eigenvalues of $\mathbf{\Omega}_p$ are the squared standard deviation $\sigma_p^2[i]$ of the components $v_p[i]$ obtained after the transformation, with $i = 1, 2, \ldots, L$. Only Q_p components linked to the greatest eigenvalues explaining a certain percentage P_v of the variance are retained after the transformation. In our work, we choose $P_v = 99\%$. The other components are temporarily removed and the remaining variables are then normalised by the transformation:

$$x[i] = v_p[i]/\sigma_p[i], \quad i = 1, 2, \ldots, Q_p. \tag{7.5}$$

From the normalised variables, for each of the N_T trials, the Mahalanobis distance D_n from the n-th single-trial ERP $\mathbf{x}_n$ to the total mean $\bar{\mathbf{x}}$ is computed as:

$$D_n^2 = D^2\left(\mathbf{x}_n, \bar{\mathbf{x}}\right) = \sum_{i=1}^{Q_p} \left(x_n^{[}i] - \bar{x}[i]\right)^2 \tag{7.6}$$

The n-th ERP is temporally marked as an outlier if

$$D_n > \bar{D} + K_{\sigma_D}, \tag{7.7}$$

where $\bar{D}$ is the mean and σ_D is the standard deviation of D, and K is a heuristically chosen coefficient.

The aforementioned steps are repeated iteratively and the trials, that is, the rows of the matrix $\mathbf{V}_p$, marked as outliers are excluded from $\mathbf{V}_p$ and from the computation of $\bar{D}$ and σ_D in the next iteration, but they are tested again by (7.7) together with all other trials. In order to facilitate convergence, the exclusion of principal components for one iteration is performed only until their number remains unchanged through two consecutive iterations. Besides, in order to prevent oscillatory behaviour, if the number of marked outliers does not increase after the current iteration, the outliers detected in the previous iteration are marked again together with those detected in the current iteration. The procedure ends when the set of detected outliers does not change any more, i.e., the same trials are marked in two consecutive iterations.

The result of the outlier detection procedure is a reduction of dimensionality both for $\tilde{\mathbf{V}}_f$ and for $\mathbf{T}_p$ because rows of $\tilde{\mathbf{V}}_f$ and columns of $\mathbf{T}_p$ representing the outlier components are deleted. Indicating with $\hat{\mathbf{T}}_p$ the reduced principal component transform matrix, the "reduced" ERP matrix is:

$$\hat{\mathbf{V}}_p = \tilde{\mathbf{V}}_f \hat{\mathbf{T}}_p, \tag{7.8}$$

In the frequency domain the reduction of dimensionality is obtained through the inverse transform:

$$\tilde{\mathbf{V}}_f^{(p)} = \hat{\mathbf{V}}_p \hat{\mathbf{T}}_p^T = \tilde{\mathbf{V}}_f \hat{\mathbf{T}}_p \hat{\mathbf{T}}_p^T, \tag{7.9}$$

where the outlier have been removed from $\mathbf{V}_f$. Eventually, the filtered and free from outliers ERP sample $\tilde{\mathbf{V}}_f^{(p)}$ can be represented back in the time domain:

$$\tilde{\mathbf{V}}^{(p)} = \tilde{\mathbf{V}}_f^{(p)} \hat{\mathbf{T}}_f^T. \tag{7.10}$$

The outliers thus determined are discarded from the following analysis, aimed at characterising the brain activity during the considered language processing tasks through two distinct descriptors, namely the averaged ERPs and the PSD as outlined in the following.

- **Event-Related Potentials:** The exploited ERPs are obtained from the set of trials available after removing the outliers. Specifically, for each user and for each possible combination of Type $\times$ Condition interaction, the remaining samples are averaged in order to generate a single ERP signal. In more detail the proposed analysis is focused on the behaviour of the N400 and P600 components, isolated considering time windows starting $t_{N400Start} = 250$ ms and $t_{P600Start} = 500$ ms after the stimulus' end respectively, and lasting 250 ms. Within

this time lapse, different features are extracted and taken into account as indicator if the cost of processing the different sentences, namely the peak, the latency, and the mean value of the obtained ERP components. Such characteristics, separately evaluated for each of the C considered channels, are employed in the statistical analysis outlined in Section 7.1.5.

- **Power Spectral Density:** In addition to the characteristics of ERPs, also the PSD is employed to evaluate the brain's processing cost of linguistic mismatches for different EEG rhythms. Indicating with $\tilde{v}_n^{(c)}[t]$ the n-th EEG signal belonging to the set of epochs that do not contain the outliers acquired from the c-th channel, with $n = 1, \ldots, N_{Tout}$ and $c = 1, 2, \ldots, C$, its PSD representation $\tilde{V}_n^{(c)}[f]$, with f indicating the frequencies of interest selected with a frequency resolution of 1 Hz, is estimated through the Welch's periodogram method using a sliding Hanning window of 0.5s with an overlap of 0.25s.

  The results reported in Section 7.1.5 are obtained considering the range of frequencies $f \in [1; 30]$Hz. The highest EEG frequency subband γ is therefore not taken into account, being typically less relevant for the processing of unexpected events in language comprehension. Each of the remaining subbands, that is, δ, θ, α and β, is considered separately, by evaluating the average value of the PSD $\tilde{V}_n^{(c)}[f]$ over the interested frequency band, thus generating multiple indicator of language processing cost for each occurrence of a given Type $\times$ Condition combination.

7.1.5 Experimental Results

The results obtained with our tests are here reported, together with a discussion on the observed ERPs and brain rhythmic changes in response to different patterns of associations between information units and the two word classes considered in the present study.

In order to verify the hypotheses stated in Section 7.1.3, several statistical hypothesis testing procedures have been carried out. In more detail, first the results obtained from an analysis performed in order to address whether the usage of different Type $\times$ Condition combinations may affect the understandability of the experimental texts are reported. The outcomes obtained from the tests performed on the obtained ERP and PSD samples to evaluate the effects of different Type $\times$ Condition interactions on the brain cost of the corresponding language processing are then illustrated. Eventually, the

obtained results are discussed in Section 7.1.6.

Understandability analysis

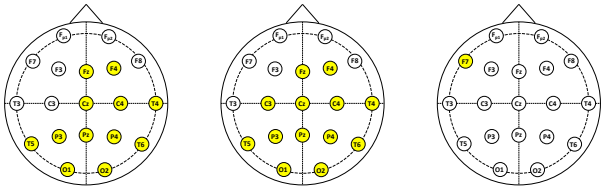
A preliminary analysis of subjects' responses to verification questions yielded an overall accuracy of 95% (SD= 0.07) which suggests that all texts have been carefully read by the subjects. A two-way ANOVA crossing Condition (Topic, Focus) and Type (Noun, Verb) and verification accuracy displayed no significant interaction ($F < 1$), indicating that texts belonging to both the Noun-Verb and Topic-Focus sets have been understood equally well. Another two-way ANOVA has been performed on the interaction between the two factors for the subjects' response times to verification questions, showing again no statistically significant result ($F < 1$), which implies that subjects took more or less the same amount of time to answer verification questions, irrespective of the Condition or Type manipulations carried out in the target sentences.

Text complexity has also been evaluated by measuring the length of the submitted texts, designed with a range between 33.3 and 36 and an $SD = 5$) words. A two-way ANOVA run on the Condition $\times$ Type interaction has shown no significant result ($F < 1$), suggesting that (a) all texts displayed on the whole the same length and that (b) the length parameter did not affect the brain response to the experimental passages listened to by the subjects. In other works, text complexity has also been gauged by calculating readability indexes (see Gulpease index for Italian written texts, [230]) which, given the auditory presentation modality of our stimuli, we have preferred not to consider for the present study.

ERP Results

The ERP features mentioned in Section 7.1.3 are considered as dependent variables in different *within-subjects t-test* statistical hypothesis scenarios, each evaluating the effect of using different word class types in a specific condition of information structure. The taken *null hypotheses* assume that the brain cost of processing a Topic or a Focus expressed through distinct word classes is the same, and is rejected in our tests when *p-values* < 0.5.

The brain regions where the performed tests have highlighted significant differences between the processing of nouns and verbs are reported in Fig. 7.1 for the Focus condition. The results show that the most significant differences are linked to the mean and peak values of the N400, mostly involving central, parietal and occipital brain regions.



(a) (b) (c)

Figure 7.1: *Regions with significant differences (p-values ≤ 0.05) for Focus Noun/Focus Verb comparisons when (a) mean value, (b) peak value and (c) latency of the ERP are evaluated in the N400 response.*

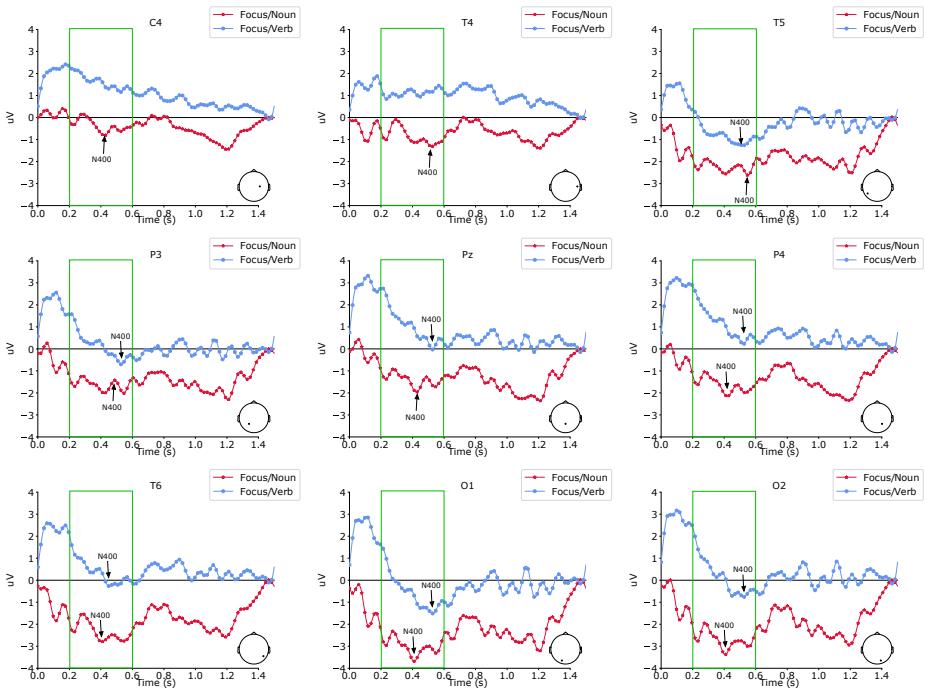


Figure 7.2: *Grand average ERP for the Focus Noun/Focus Verb comparison.*

The grand average ERPs related to the electrodes displaying more prominent N400 deflections between nominal and verbal Topics are shown in Fig. 7.2.

Analogously, the results related to the ERP response to nominal and verbal Topics are shown in Fig. 7.3. Also in this case, the scenarios highlighting the most significant results involve the mean amplitude and the peak of the N400 responses, with a mainly central, parietal and occipital distribu-

tion. In this case, some significant differences are also linked to the P600 component, and mainly involve the frontal brain region as shown in Figure 7.4. The grand average ERPs related to the electrodes showing significant differences in the N400 features are shown in Fig. 7.5.

PSD Results

A further analysis on PSD levels has been performed, separately for each channel, in order to evaluate the differences in brain processing of matched and non-matched word class-information units pairs. Fig. 7.6 shows the results for the tests between Focus-Noun vs. Focus-Verb conditions, while Fig. 7.7 shows the results of the tests comparing PSD features for the Topic-Verb vs. Topic-Noun conditions.

The PSD results for the Focus condition show that the most significant differences between the Noun and the Verb types are more robust in the *delta* subband, with a parieto-occipital distribution (electrodes T5, P3, O1 and O2), and in the *theta* band with a mainly frontal distribution (electrodes F7 and F8). No significant differences are instead observed in the *alpha* and *beta* subbands.

The PSD results for the Topic condition show that there are differences in the processing of nouns and verbs for all the considered frequency subbands, that is, *delta*, *theta*, *alpha* and *beta*. The most remarkable behavior is shown by the *theta* subband, where the parieto-occipital brain area shows significant differences in the PSD features. Differences in the frontal area appear in the *delta* (electrodes F7 and F3) and *alpha* (electrodes Fp1, F4 and Cz) bands, while the parieto-occipital region (electrodes C4, P4, T6 and O2) shows significant differences in the *beta* subband.

7.1.6 Discussion

The results obtained from the ERP analysis confirm the expectations about the N400 component, associated with difficulties in lexical-semantic retrieval and the processing of less expected information structural patterns [9, 154, 311]. In our study, the less expected structural patterns are represented by the Focus-Noun and Topic-Verb combinations, whose statistical distributions of the peak and mean of the N400 response in the centro-parietal and occipital regions are significantly different from the ones related to the expected structures, that is Focus-Verb and Topic-Noun.

Our results are in line with the expectation-based processing of word classes suggested by [75], who found modulations in the N400 signature during the online processing of English nouns and verbs in more or less

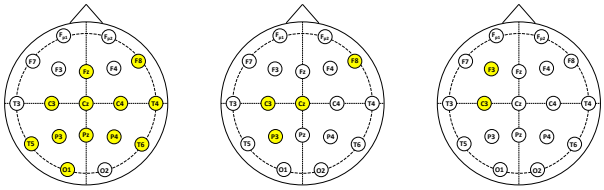


Figure 7.3: *Regions with significant differences (p-values ≤ 0.05) for Topic Noun/Topic Verb comparisons when (a) mean value, (b) peak value and (c) latency of the ERP are evaluated in the N400 response.*

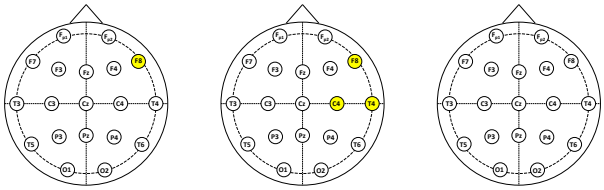


Figure 7.4: *Regions with significant differences (p-values ≤ 0.05) for the Topic Noun/-Topic Verb comparisons when (a) mean value, (b) peak value and (c) latency of the ERP are evaluated in the P600 response.*

predictable syntactic positions or discourse functions. Our findings also confirm those discussed by [75] on the processing cost imposed by information structures more or less consistently aligned with different activation states of sentence contents. In this study we sought to demonstrate that, besides interactions with the prosodic level [50, 111] and with degrees of activation of information in the receiver's short-term memory [194], the processing of information structures is also sensitive to expectations associated with the word class being selected by the speaker to package some information as Topic or Focus of the sentence.

Quite in line with our predictions, the word class/information structure comparison also yielded positivity effects for the Topic Verb condition. This trend appears more remarkable in frontal and central regions, as displayed in the plots in Fig. 7.4. As shown in the grand averages in Fig. 7.5, positivity develops in a slightly more delayed time window than a canonical P600 interval. The presence of a distinct N400-P600 pattern for topicalized verbs can be explained as reflecting integration and reanalysis costs imposed by mentally construing a more predicative type of informa-

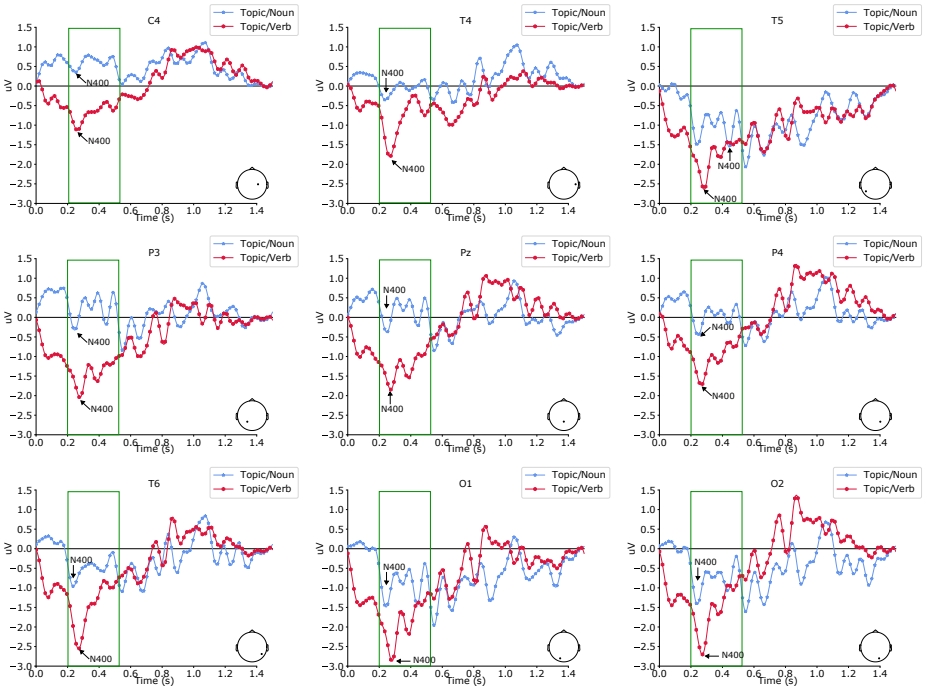


Figure 7.5: *Grand average ERP for the Topic Noun/Topic Verb comparison.*

tion (which is the essence of verbs) in topical function (typically associated with nouns). In this sense, if N400 hints at an effort imposed on the attentional system dealing with a misaligned matching between word class and information unit, this mismatch is eventually solved through the integration of this pattern in the receiver's current representation of the discourse model, which correlates with P600 modulations. This behaviour is partly in line with previous findings on the involvement of P600 during the processing of less expected information structural configurations [31]. Other electrophysiogical studies on unexpected packaging criteria of new information [194] did not report P600 modulations. For example, [194] noticed that a new presupposition (e.g. *the migration*) in a context-embedded sentence elicits only a N400 response compared to when the same item of new information is asserted e.g. *a migration*). The authors interpreted this effect as hinting at linking mechanisms, i.e. anaphorical operations by which some content is "linked" to the foregoing discourse. The absence of P600 effects in their study may be indicative of less effortful updating and parsing mechanisms on the part of the receiver once the mismatch has been detected. Put more simply, new presuppositions are probably easier to de-

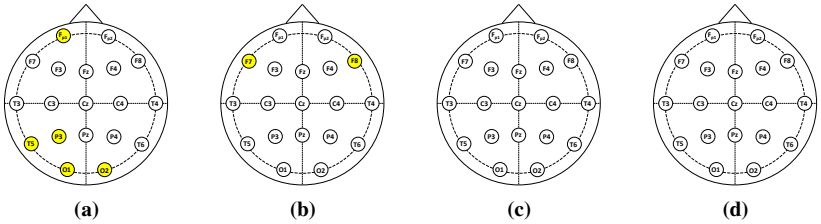


Figure 7.6: *Regions with significant difference (p-values ≤ 0.05) in the PSD levels for the Focus Noun/Focus Verb comparisons when different subbands, namely (a) delta, (b) theta, (c) alpha and (d) beta, are considered.*

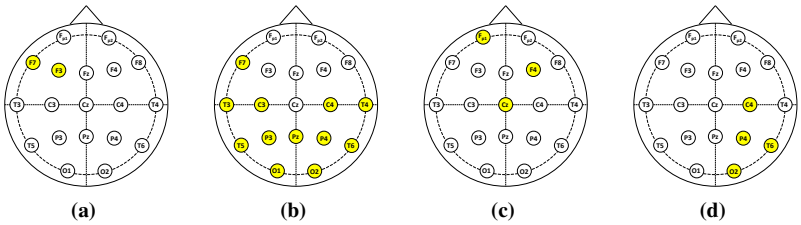


Figure 7.7: *Regions with significant difference (p-values ≤ 0.05) in the PSD levels for the Topic Noun/Topic Verb comparisons when different subbands, namely (a) delta, (b) theta, (c) alpha and (d) beta, are considered.*

code than new Topics expressed by verbal categories. In our scenario, P600 can be regarded as the signal of additional updating costs [119] due to a less common syntactic pattern (verbs fronted to realise topics) which in turn imposes a revision of previous expectations on the informational architecture of the prior linguistic context.

Evidence from the PSD results gives us a further explanation of increasing costs elicited by processing less expected word class-information structure matchings [261]. Previous studies on the activity of frequency bands in sentence processing demonstrated that oscillation amplitudes of *delta* and *theta* bands are directly related to processing demands, in that amplitude increases (synchronization) correlate with working memory load [145]. On the contrary, *alpha* and *beta* bands show an activity that is inversely related to processing demands, with amplitude decreases (desynchronization) indicating increased involvement of neural resources for sentence processing [65, 145]. Given the previous assumptions, the obtained results for the Focus condition show that the *theta* band is significantly involved in the processing of words displaying a less expected packaging, indicating an effort linked to working memory load when the non-matched condition is

presented. Similar considerations can be put forth for the Topic condition, where the main frequency bands involved are *theta* and *beta*, indicating a working memory load and an involvement of neural resources for sentence processing when the mismatched structures are processed by the brain.

7.1.7 Conclusions

In this study, the interplay of utterances' information structure and the word-class level has been inquired looking into their brain correlates through ERP and PSD level measurements. Data showed that the processing of informational hierarchies is indeed sensitive to the word class selected to realise the Topic or the Focus unit of the sentence. More particularly, both modulations in the N400 time-window and increases in the power spectrum density levels in the *theta* and *delta* bands has been observed in response to Verbs encoded as Topic, and Nouns encoded as Focus, with more robust effects elicited by topicalized verbs. These findings are in line with two main predictions set forth for the present research: (a) the cost associated with information structure processing follows discourse-driven expectations also with respect to the word-class level, and (b), as put forth by [75], the cognitive cost of mentally representing verbal and nominal classes is not only conditional on the evaluation of category-related features, but follows an expectation-driven path, that is, it responds to the receiver's anticipation of the information packaging properties that a word is expected to exhibit based on the discursive function it is called upon to perform.

In this Section an invisible biometric identifier, namely brain signals, is considered. The aforementioned signals can be seen as biometric traits acquired as a response to audio stimuli. The proposed work is an example of visible information beyond and invisible biometric identifier. More in detail, if a EEG-based biometric system relies on a stimulation protocol consisting of audio stimuli, information relative to the processing cost and working memory usage can be easily obtained, mining privacy-related information of the user.

CHAPTER 8

Conclusions and Future Work

In this thesis, a new classification of biometric identifiers has been introduced: visible and invisible biometrics. With the term *invisible biometrics* we refer to all those biometric systems based on physiological characteristics hidden in the human body and not easy to be captured at a distance with traditional acquisition devices and methodologies. A biometric trait hidden inside the human body is a requirement of paramount importance in biometric systems where a high level of security is needed. In fact, some kind of biometric data are inherently public and can be stolen and replicated, and once an attacker obtains a person's biometric identifier, it is easy for him to gain access to the user's account requiring his biometric recognition. The choice of designing a biometric system based on an invisible identifier, being inherently more difficult to steal and replicate, allows to reduce the vulnerabilities and security risks of the systems. On the other hand, because of the "hiddenness" of the invisible traits, the acquisition procedure is generally affected by noise that makes the feature extraction process challenging, resulting in negative impact on the recognition performance of the system. Besides, some security vulnerabilities of biometrics systems are not overcome by using invisible biometrics. Eventually, being

invisible biometrics related to physiological characteristics, they may reveal some personal information, threatening the privacy of the users of the system.

8.1 Conclusions

In this thesis, two invisible biometric identifiers are investigated: vein pattern and electroencephalogram. The *vein pattern*, that is the network of blood vessels of people's hands, falls into the invisible biometrics category because it is hidden under the skin and visible only thanks to the help of a near-infrared light and camera. The aforesaid property makes this biometric trait very difficult to steal and replicate, and, then, a spoofing attack nearly impossible to be implemented. On the other hand, vein-pattern images are often characterised by low contrast and poor definition, due to the sub-cutaneous placement of the veins, thus making vein-related feature extraction a challenging process and then impacting on the recognition performance of the system. In this thesis, some possible solutions aiming to improve the performance of vein-based biometric systems and to deal with the disadvantages related to the poor quality of vein images are studied.

The fist proposed solution, aiming at improving the overall accuracy, mitigating the effect of noisy input data and providing greater resistance to spoofing, is based on the integration of several biometric information sources, that is the implementation of a multimodal biometric systems. More in detail, an approach for palm vein recognition relying on high dynamic range (HDR) imaging is proposed, thus following a sensor-level fusion approach. Besides, the effects of feature-, score- and decision-level fusion approaches of a palm-vein-based biometric system are investigated. In this study, a multiple-exposure dataset is collected. Local binary pattern (LBP) and local derivative pattern (LDP) are employed to extract features from single-exposure images, raw HDR images, and tone-mapped HDR images. The obtained experimental results show that significant performance improvement can be achieved when discriminative features are extracted from HDR contents, with respect to the use of single-exposure images. Besides, better recognition performance can be achieved exploiting such methods, when compared to what can be obtained with the other fusion techniques.

As alternative solution to the issues related to the feature extraction step in vein-based biometric system, a recognition architecture based on convolutional neural network (CNN) is proposed in this thesis. More in detail, a CNN-based-finger-vein identification system has been proposed,

and the capabilities of the designed network has been investigated taking into account four publicly-available databases. The main purpose of this work is to propose a deep-learning method for finger-vein identification, able to achieve stable and highly-accurate performance when dealing with different kinds of finger-vein images, irrespective of their quality. The achieved performance shows that the proposed method is able to guarantee stable and highly-accurate identification results, irrespective of the quality of the considered finger-vein images. Additionally, the proposed CNN-based identification system requires negligible manual effort for feature selection. In fact, it has been applied without variations to all the four considered databases, without using any application-dependent threshold or any manually-set parameter.

Given the results obtained by the application of sensor-level fusion and CNN techniques in the field of vein-based recognition, in this thesis the advantages of HDR and deep learning techniques have been merged with the aim of further improving the recognition performance of a finger-vein-based biometric system. Besides, the intrinsic security and robustness of the vein biometric trait, the speed and convenience of a touchless acquisition device and, eventually, the affordability of a low cost capturing device have been exploited. In the proposed approach, an "on-the-fly" system, where the user is asked to swipe the hand over the sensor composed by low-cost cameras is studied. In order to compensate the quality loss due to the low-cost sensors, and to the free hand movements during acquisition, multiple cameras with different exposure times, capturing also the dynamic movement of the hand over the sensors, have been used and both the still images acquired at different exposure times and the temporal behaviour of the moving hand over the sensors have been exploited. Deep learning approaches have been used in both scenarios. The reported analysis shows that the use of multiple-exposure data increases the recognition accuracy with respect to the use of single exposure images and that the exploitation of multi-channel LDR images taken at different exposure times, as raw input templates, leads to further improvements of the identification accuracy. In addition, the temporal information related to the user swiping the hand over the sensors has been used; the obtained results show that when CNN topologies are used for feature extraction and LSTM networks are fed by the sequential features based on hand movements, a significant identification accuracy improvement is observed.

As all the biometric systems, also vein-based biometric systems are vulnerable to attacks at different levels, entailing weakness in data protection and privacy if proper countermeasures are not adopted. More in detail,

security issues of paramount importance are the revocability of biometric template and cross-matching of data stored in different database. In the light of the aforementioned problems, two different studies are proposed. The first work analyses the cross-finger similarity of finger vein patterns, studying if vein patterns acquired from fingers of different hands are "sufficiently" different, that is if the inter-class variation between the considered features is high, in terms of the considered feature extraction and matching methods. It has been demonstrated that a pair of same fingers from different hands do not posses similarities that allow the user to be recognised when one finger is used for enrolment and the other one for recognition. This feature improves the security of a vein based biometric system, reducing the possibility of cross-matching between different databases if different fingers are used. Besides, it increases the possibility of revoking the biometric trait.

The other proposed solution in the field of security in vein-based biometric systems relies on non-invertible transforms. More in detail, block remapping, image warping and Bloom filters are applied to finger-vein images. The recognition performance of the proposed approaches are evaluated and an evaluation of security in terms unlinkability for all three cancelable biometrics approaches is performed. Besides, the irreversibility analysis in case of block remapping transformation is applied by exploiting an automated square jigsaw puzzle solver algorithm as a possible attack. The achieved results show that best recognition performance can be achieved by the block remapping and block warping approaches in combination with the pre-alignment. However,block remapping is not secure enough as it turned out that its unlinkability as well as irreversibility is rather low. Block warping has a low unlinkability as well.Hence, only the Bloom filter approach is suitable in terms of security. In combination with the pre-alignment it achieves an acceptable recognition performance, although this recognition performance is still much worse than the baseline one. Without the pre-alignment, the resulting recognition is not usable at all.Thus, an accurate, universal pre-alignment, which does not require the unprotected templates to be present in the system, is necessary in order to employ a well performing (in terms of recognition accuracy) template protection scheme.

The second considered invisible biometric identifiers are *brain waves*. Electroencephalographic (EEG) signals are more privacy compliant than commonly used biometrics because they are not exposed and therefore cannot be captured at a distance. Besides, they are less likely to be synthetically generated and fed to a sensor to spoof and intrinsically provide liveness detection. On the other hand, the use of brain signals poses new challenges.

First, EEG signals are very sensitive to noise, entailing a bad impact on recognition performance. Besides, the user-convenience of the acquisition procedure is generally low. In this thesis, EEG-based biometric recognition system where discriminative features are extracted from steady-state visual evoked potentials (SSVEPs) is proposed. SSVEPs exhibit a high signal-to-noise ratio and a stable spectrum, properties which make these signals particularly appealing for EEG-based biometric applications. The issues of performance, permanence and acceptability are addressed. The use of flickering stimuli at specific frequencies and the representation of the acquired EEG data through either MFCC or AR templates, allows achieving high identification rates, thanks to the proved existence of permanent characteristics in SSVEP brain responses across different acquisition sessions. According to the reported experimental tests, the joint use of multiple elicitation frequencies guarantees a notable improvement in recognition rates, thus allowing to reduce the number of electrodes needed during EEG collection, a relevant property to foster the adoption of EEG-based biometric identifiers in practical recognition systems.

Eventually, the problem of privacy information that can be obtained by processing invisible biometrics is investigated for the case of brain responses. EEG signals may provide additional information about the background of an individual, such as diseases and heath information, emotional status and mental states. In this thesis an example of visible information beyond invisible biometrics is presented; the analysed information is the brain cost associated to the processing of sentences containing linguistic misalignment. The sentence are presented as audio stimuli, so the adopted scenario could be seen as an EEG-based biometric system where the brain waves are collected as response to audio stimuli. Data showed that the processing of informational hierarchies is indeed sensitive to linguistic misalignment, visible through an analysis pf ERP and PSD extracted from the EEG signals.

8.2 Future Work

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

- Both for the vein-based and EEG-based biometric recognition proposed approaches, due to the novelty of the experimental settings, the experiments rely on the apparatus used and data capture protocols, thus implying difficulties in making comparisons with other techniques reported in literature. As part of future work, we aim to collect

different datasets following the same acquisition protocols but varying the acquisition device, thus making possible a more statistical significant analysis.

- The datasets used for most of the experiments are not publicly avalaible. We aim to make them publicly available.
- An application of convolutional neural network on the study proposed in Section 3.1 will be carried on.
- In order to further improve the performance obtained in Section 3.2, the application of DNN preprocessing techniques such as data augmentation will be taken into account.
- We aim to apply deep learning techniques in the studies concerning EEG-based biometric recognition.

Author's Contributions

Journals

J. 1 **Piciucco, E.**, Maiorana E., and Campisi, P. (2018),
"Palm vein recognition using a high dynamic range approach",
in IET Biometrics, Vol. 7(5), pp. 439–446.

J. 2 Das, R., **Piciucco, E.**, Maiorana, E., and Campisi, P. (2018),
"Convolutional neural network for finger-vein-based biometric identification",
in IEEE Transactions on Information Forensics and Security, Vol. 14(2), pp. 360–373.

J. 3 Kuzu, R.S, **Piciucco, E.**, Maiorana, E., and Campisi, P. (2020),
"On-the-fly Finger-Vein-based Biometric Recognition using Deep Neural Networks",
in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 2641-2654.

J. 4 **Piciucco, E.**, Masia, V., Maiorana, E., Vallauri, E. L. and Campisi, P. (2020),
"Information Structure Effects on the Processing of Nouns and Verbs: Evidence from Event-Related Potentials and Brain Oscillatory Dynamics",
submitted to Language, Cognition and Neuroscience.

J. 5 Kauba, C., **Piciucco, E.**, Maiorana, E., Gomez-Barrero, M., Prommegger, B., Campisi, P., Uhl, A. (2020),

"Towards Practical Cancelable Biometrics for Finger Vein Recognition",
submitted to IEEE Transactions on Dependable and Secure Computing.

Conferences

C. 1 **Piciucco, E.**, Maiorana, E., Kauba, C., Uhl, A., and Campisi, P. (2016),
"Cancelable biometrics for finger vein recognition",
in 2016 First International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE), 2016, pp. 1–5.

C. 2 Das, R., **Piciucco, E.**, Maiorana, E., and Campisi, P. (2016),
"Visually evoked potentials for EEG biometric recognition",
in 2016 First International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE), 2016, pp. 1–5.

C. 3 Kauba, C., **Piciucco, E.**, Maiorana, E., Campisi, P., and Uhl, A., (2016),
"Visually evoked potentials for EEG biometric recognition",
in 2016 International Conference of the Biometrics Special Interest Group (BIOSIG), 2016, pp. 1–7.

C. 4 **Piciucco, E.**, Maiorana E., and Campisi, P. (2017),
"Biometric fusion for palm-vein-based recognition systems",
in International Tyrrhenian Workshop on Digital Communication, 2017, pp. 18–28.

C. 5 **Piciucco, E.**, Maiorana, E., Falzon, O., Camilleri, K.P., and Campisi, P., (2017),
"Steady-state visual evoked potentials for EEG-based biometric identification",
in 2017 International Conference of the Biometrics Special Interest Group (BIOSIG), 2017, pp. 1–5.

C. 6 **Piciucco, E.**, Kuzu, R. S., Maiorana E., and Campisi, P. (2019),
"On the Cross-Finger Similarity of Vein Patterns",
in International Conference on Image Analysis and Processing (ICIAP), 2019, pp. 12–20.

Bibliography

[1] Muhammad Kamil Abdullah, Khazaimatol S Subari, Justin Leo Cheang Loong, and Nurul Nadia Ahmad. Analysis of effective channel placement for an eeg-based biometric system. In *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pages 303–306. IEEE, 2010.

[2] Moulay Akhloufi and Abdelhakim Bendada. Hand and wrist physiological features extraction for near infrared biometrics. In *2008 Canadian Conference on Computer and Robot Vision*, pages 341–344. IEEE, 2008.

[3] Ilya Alex Krizhevsky, Sutskever and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[4] Blair C Armstrong, Maria V Ruiz-Blondet, Negin Khalifian, Kenneth J Kurtz, Zhanpeng Jin, and Sarah Laszlo. Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for erp biometrics. *Neurocomputing*, 166:59–67, 2015.

[5] Mohd Shahrimie Mohd Asaari, Shahrel A Suandi, and Bakhtiar Affendi Rosdi. Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics. *Expert Systems with Applications*, 41(7):3367–3382, 2014.

[6] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (12):2481–2495, 2017.

[7] M.C. Baker. *Lexical Categories: Verbs, Nouns and Adjectives*. Cambridge University Press, Cambridge, 2003.

[8] Valentina Bambini, Chiara Bertini, Walter Schaeken, Alessandra Stella, and Francesco Di Russo. Disentangling metaphor from context: an erp study. *Frontiers in psychology*, 7:559, 2016.

[9] Valentina Bambini, Donatella Resta, and Mirko Grimaldi. A dataset of metaphors from the italian literature: exploring psycholinguistic variables and the role of context. *PloS one*, 9(9):e105634, 2014.

[10] Anupam Banerjee, Sumana Basu, Subhadip Basu, and Mita Nasipuri. Artem: a new system for human authentication using finger vein images. *Multimedia Tools and Applications*, 77(5):5857–5884, 2018.

[11] Francesco Banterle, Alessandro Artusi, Kurt Debattista, and Alan Chalmers. *Advanced High Dynamic Range Imaging: Theory and Practice.* CRC Press, Natrick, MA, USA, 2011.

[12] Robert J Barry, Adam R Clarke, Stuart J Johnstone, Christopher A Magee, and Jacqueline A Rushby. Eeg differences between eyes-closed and eyes-open resting conditions. *Clinical Neurophysiology*, 118(12):2765–2773, 2007.

[13] E. Başar. *Brain Function and Oscillations: Integrative brain function. Neurophysiology and cognitive processes*. Springer, 1999.

[14] Erol Başar, Canan Başar-Eroglu, Sirel Karakaş, and Martin Schürmann. Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology*, 39(2-3):241–248, 2001.

[15] Canan Başar-Eroglu, Erol Başar, Tamer Demiralp, and Martin Schürmann. P300-response: possible psychophysiological correlates in delta and theta frequency channels. a review. *International Journal of Psychophysiology*, 13(2):161–179, 1992.

[16] Marcel Bastiaansen, Ali Mazaheri, and Ole Jensen. Beyond erps: oscillatory neuronal dynamics. In *The Oxford handbook of event-related potential components*, pages 31–50. Oxford University Press, 2012.

[17] Stefan Baumann and Petra B Schumacher. (de-) accentuation and the processing of information status: evidence from event-related brain potentials. *Language and Speech*, 55(3):361–381, 2012.

[18] Jan Berkhout and Donald O Walter. Temporal stability and individual differences in the human eeg: An analysis of variance of spectral values. *IEEE Transactions on Biomedical Engineering*, (3):165–168, 1968.

[19] Tiziano Bianchi and Alessandro Piva. Image forgery localization via block-grained analysis of jpeg artifacts. *IEEE Transactions on Information Forensics and Security*, 7(3):1003–1017, 2012.

[20] Simone Bianco and Paolo Napoletano. Biometric recognition using multimodal physiological signals. *IEEE Access*, 7:83581–83588, 2019.

[21] Niels Birbaumer. Slow cortical potentials: plasticity, operant control, and behavioral effects. *The Neuroscientist*, 5(2):74–78, 1999.

[22] Stacy Birch and Keith Rayner. Linguistic focus affects eye movements during reading. *Memory & Cognition*, 25(5):653–660, 1997.

[23] Ruud M Bolle, Jonathan H Connell, and Nalini K Ratha. Biometric perils and patches. *Pattern Recognition*, 35(12):2727–2738, 2002.

[24] Vladimir Bostanov. Multivariate assessment of event-related potentials with the t-cwt method. *BMC neuroscience*, 16(1):73, 2015.

[25] Terrance E. Boult, Walter J. Scheirer, and Robert T. E. Woodworth. Revocable fingerprint biotokens: accuracy and security analysis. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2007.

[26] Yu A Boytsova and SG Danko. Eeg differences between resting states with eyes open and closed in darkness. *Human physiology*, 36(3):367–369, 2010.

[27] Steven L Bressler. Event-related potentials of the cerebral cortex. In *Electrophysiological recording techniques*, pages 169–190. Springer, 2011.

[28] Katharine Brigham and BVK Vijaya Kumar. Subject identification from electroencephalogram (EEG) signals during imagined speech. In *4th International Conference on Biometrics: Theory Applications and Systems (BTAS)*, pages 1–8. IEEE, 2010.

[29] Petra Burkhardt. Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, 98(2):159–168, 2006.

[30] Petra Burkhardt. The p600 reflects cost of new information in discourse memory. *Neuroreport*, 18(17):1851–1854, 2007.

[31] Juliane Burmester, Katharina Spalek, and Isabell Wartenburger. Context updating during sentence comprehension: The effect of aboutness topic. *Brain and language*, 137:62–76, 2014.

[32] Patrizio Campisi. *Security and privacy in biometrics*, volume 24. Springer, 2013.

[33] Patrizio Campisi and Daria La Rocca. Brain waves for automatic biometric-based user recognition. *IEEE Transaction on Information Forensics and Security*, 9(5):782–800, 2014.

[34] Patrizio Campisi, Daria La Rocca, and Gaetano Scarano. Eeg for automatic person recognition. *Computer*, 45(7):87–89, 2012.

[35] Patrizio Campisi, Gaetano Scarano, Fabio Babiloni, F DeVico Fallani, Stefania Colonnese, Emanuele Maiorana, and Laura Forastiere. Brain waves based user recognition using the "eyes closed resting conditions" protocol. In *2011 IEEE International Workshop on Information Forensics and Security*, pages 1–6. IEEE, 2011.

[36] Pierandrea Cancian, GW Di Donato, Vincenzo Rana, and Marco D Santambrogio. An embedded gabor-based palm vein recognition system. In *International Conference on Biomedical & Health Informatics (BHI)*, pages 405–408. IEEE, 2017.

[37] Stefano F Cappa and Daniela Perani. The neural correlates of noun and verb processing. *Journal of Neurolinguistics*, 16(2-3):183–189, 2003.

[38] Frederico Amorim Cavalcante. *The topic unit in spontaneous American English: a corpus-based study*. Universidade Federal de Minas Gerais, 2015.

[39] Wallace Chafe. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. *Subject and topic*, 1976.

[40] Alan Chalmers, Patrizio Campisi, Peter Shirley, and Igor García Olaizola. *High Dynamic Range Video: Concepts, Technologies and Applications*. Academic Press, 2016.

[41] R. Chellappa, C.L. Wilson, and S. Sirohey. Human and machine recognition of faces: a survey. *Proceedings of the IEEE*, 83(5):705–741, 1995.

[42] Haifen Chen, Guangming Lu, and Rui Wang. A new palm vein matching method based on icp algorithm. In *2nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, pages 1207–1211. ACM, 2009.

[43] Liukui Chen, Jing Wang, Shiyu Yang, and Haibo He. A finger vein image-based personal identification system with self-adaptive illuminance control. *IEEE Transactions on Instrumentation and Measurement*, 66(2):294–304, 2016.

[44] Ken Chiu, Michael Herf, Peter Shirley, S. Swamy, Changyaw Wang, Kurt Zimmerman, et al. Spatially nonuniform scaling functions for high contrast images. In *Graphics Interface*, pages 245–245. Canadian Information Processing Society, 1993.

[45] Taeg Sang Cho, Shai Avidan, and William T. Freeman. A probabilistic image jigsaw puzzle solver. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 183–190. IEEE, 2010.

[46] Joon Hwan Choi, Wonseok Song, Taejeong Kim, Seung-Rae Lee, and Hee Chan Kim. Finger vein extraction using gradient normalization and principal curvature. In *Image Processing: Machine Vision Applications II*, volume 7251, pages 725111–725111. International Society for Optics and Photonics, 2009.

[47] Shang-Jen Chuang. Vein recognition based on minutiae features in the dorsal venous network of the hand. *Signal, Image and Video Processing*, 12(3):573–581, 2018.

[48] Tee Connie, Andrew Teoh, Michael Goh, and David Ngo. Palmhashing: a novel approach for cancelable biometrics. *Information processing letters*, 93(1):1–5, 2005.

[49] European Council. Regulation of the european parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (general data protection regulation). *document Regulation (EU) 2016/679*, 2016.

[50] H Wind Cowles, Robert Kluender, Marta Kutas, and Maria Polinsky. Violations of information structure: An electrophysiological study of answers to wh-questions. *Brain and Language*, 102(3):228–242, 2007.

[51] Emanuela Cresti and Massimo Moneglia. Informational patterning theory and the corpus-based description of spoken language: The compositionality issue in the topic-comment pattern. *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*, pages 13–45, 2010.

[52] Antonio R Damasio and Daniel Tranel. Nouns and verbs are retrieved with differently distributed neural systems. *Proceedings of the National Academy of Sciences*, 90(11):4957–4960, 1993.

[53] František Daneš. *A three-level approach to syntax*. Travaux linguistiques de Prague, 1964.

[54] František Daneš. Order of elements and sentence intonation. *To Honor Roman Jakobson*, 1:499–512, 1967.

[55] František Daneš. Functional sentence perspective and the organization of the text. *Papers on functional sentence perspective*, pages 106–128, 1974.

[56] Abhijit Das, Umapada Pal, Miguel Angel Ferrer Ballester, and Michael Blumenstein. A new wrist vein biometric system. In *2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, pages 68–75. IEEE, 2014.

[57] Koel Das, Sheng Zhang, Barry Giesbrecht, and Miguel P Eckstein. Using rapid visually evoked eeg activity for person identification. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2490–2493. IEEE, 2009.

[58] Priyanka Das, Kannan Karthik, and Boul Chandra Garai. A robust alignment-free fingerprint hashing algorithm based on minimum distance graphs. *Pattern Recognition*, 45(9):3373–3388, 2012.

[59] Rig Das, Emanuele Maiorana, and Patrizio Campisi. EEG biometrics using visual stimuli: A longitudinal study. *IEEE Signal Processing Letters*, 23(3):341–345, 2016.

[60] Rig Das, Emanuele Maiorana, and Patrizio Campisi. Visually evoked potential for eeg biometrics using convolutional neural network. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 951–955. IEEE, 2017.

[61] Rig Das, Emanuele Maiorana, and Patrizio Campisi. Motor imagery for eeg biometrics using convolutional neural network. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2062–2066. IEEE, 2018.

[62] Rig Das, Emanuela Piciucco, Emanuele Maiorana, and Patrizio Campisi. Convolutional neural network for finger-vein-based biometric identification. *IEEE Transactions on Information Forensics and Security*, 14(2):360–373, 2018.

[63] Charles S DaSalla, Hiroyuki Kambara, Makoto Sato, and Yasuharu Koike. Single-trial classification of vowel speech imagery using common spatial patterns. *Neural networks*, 22(9):1334–1339, 2009.

[64] J.G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1148–1161, 1993.

[65] Douglas J Davidson and Peter Indefrey. An inverse relation between event-related and time–frequency violation responses in sentence processing. *Brain Research*, 1158:81–92, 2007.

[66] L. Debiasi, S. Kirchgasser, B. Prommegger, A. Uhl, G. Artur, and K. Marcin. Proceedings of the biometric template protection in the image domain using non-invertible grey-scale transforms, 2019.

[67] A Djerouni, H Hamada, A Loukil, and N Berrached. Dorsal hand vein image contrast enhancement techniques. *International Journal of Computer Science Issues (IJCSI)*, 11(1):137–142, 2014.

[68] Filippo Domaneschi, Paolo Canal, Viviana Masia, Edoardo Lombardi Vallauri, and Valentina Bambini. N400 and p600 modulation in presupposition accommodation: The effect of different trigger types. *Journal of Neurolinguistics*, 45:13–35, 2018.

[69] Frédéric Drago, Karol Myszkowski, Thomas Annen, and Norishige Chiba. Adaptive logarithmic mapping for displaying high contrast scenes. In *Computer Graphics Forum*, volume 22, pages 419–426. Wiley Online Library, 2003.

[70] M-P Dubuisson and Anil K Jain. A modified hausdorff distance for object matching. In *12th International Conference on Pattern Recognition (ICPR)*, volume 1, pages 566–568. IEEE, 1994.

[71] Andreas K Engel and Pascal Fries. Beta-band oscillations signalling the status quo? *Current opinion in neurobiology*, 20(2):156–165, 2010.

[72] Niedermeyer Ernst and FHL Da Silva. Electroencephalography: Basic principles, clinical applications and related fields. *Lippincott Williams &Wilkins*, 2005.

[73] Wu Fang, Yuxun, Qiuxia, and Wenxiong Kang. A novel finger vein verification system based on two-stream convolutional network learning. *Neurocomputing*, 290:100–107, 2018.

[74] Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM Transactions on Graphics (TOG)*, volume 27, page 67. ACM, 2008.

[75] Kara D Federmeier, Devon B McLennan, Esmeralda De Ochoa, and Marta Kutas. The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An erp study. *Psychophysiology*, 39(2):133–146, 2002.

[76] Kara D Federmeier, Jessica B Segal, Tania Lombrozo, and Marta Kutas. Brain responses to nouns, verbs and class-ambiguous words in context. *Brain*, 123(12):2552–2566, 2000.

[77] Pol Fernández and Rainhard Dieter Findling. Mobile wrist vein authentication using sift features. In *EUROCAST*, page 140, 2017.

[78] Miguel A. Ferrer, Aythami Morales, and Lourdes Ortega. Infrared hand dorsum images for identification. *Electronics letters*, 45(6):306–308, 2009.

[79] Jan Firbas. On defining the theme in functional sentence analysis. *Travaux Linguistiques de Prague*, 1:267–280, 1966.

[80] Jan Firbas. On the delimitation of the theme in functional sentence perspective. *Functionalism in linguistics*, pages 137–156, 1987.

[81] Javier Galbally, Arun Ross, Marta Gomez-Barrero, Julian Fierrez, and Javier Ortega-Garcia. Iris image reconstruction from binary templates: An efficient probabilistic approach based on genetic algorithms. *Computer Vision and Image Understanding*, 117(10):1512–1525, 2013.

[82] Todor Ganchev, Nikos Fakotakis, and George Kokkinakis. Comparative evaluation of various mfcc implementations on the speaker verification task. In *Proceedings of the SPECOM*, volume 1, pages 191–194, 2005.

[83] Marta Gomez-Barrero, Javier Galbally, Christian Rathgeb, and Christoph Busch. General framework to evaluate unlinkability in biometric template protection systems. *IEEE Transactions on Information Forensics and Security*, 13(6):1406–1420, 2017.

[84] Marta Gomez-Barrero, Javier Galbally, Christian Rathgeb, and Christoph Busch. General framework to evaluate unlinkability in biometric template protection systems. *IEEE Transactions on Information Forensics and Security*, 13(6):1406–1420, 2018.

[85] Marta Gomez-Barrero, Christian Rathgeb, Javier Galbally, Christoph Busch, and Julian Fierrez. Unlinkable and irreversible biometric template protection based on bloom filters. *Information Sciences*, 370-371:18–32, 2016.

[86] Marta Gomez-Barrero, Christian Rathgeb, Guoqiang Li, Raghavendra Ramachandra, Javier Galbally, and Christoph Busch. Multi-biometric template protection based on bloom filters. *Information Fusion*, 42:37–50, 2018.

[87] Henry Gray and Susan Standring. *Gray's anatomy: the anatomical basis of clinical practice*. Churchill Livingstone, 2008.

[88] Haitao Guo and C Sidney Burrus. Convolution using the undecimated discrete wavelet transform. In *International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings (ICASSP)*, volume 3, pages 1291–1294. IEEE, 1996.

[89] Cota Navin Gupta, Ramaswamy Palaniappan, and Raveendran Paramesran. Exploiting the p300 paradigm for cognitive biometrics. *International Journal of Cognitive Biometrics*, 1(1):26–38, 2012.

[90] Puneet Gupta and Phalguni Gupta. An accurate finger vein based verification system. *Digital Signal Processing*, 38:43–52, 2015.

[91] P. Hagoort, C. Brown, and J. Groothusen. The syntactic positive shift as an erp measure of syntactic processing. *Journal of Cognitive Neuroscience*, 8:337–364151–165, 1993.

[92] Peter Hagoort and Jos van Berkum. Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):801–811, 2007.

[93] Michael AK Halliday. *An introduction to functional grammar*, volume 1. London: Edward Arnold, 1985.

[94] Jutta Hämmerle-Uhl, Elias Pschernig, and Andreas Uhl. Cancelable iris biometrics using block re-mapping and image warping. In *Information Security*, pages 135–142. Springer, 2009.

[95] Wei-Yu Han and Jen-Chun Lee. Palm vein recognition using adaptive gabor filter. *Expert Systems with Applications*, 39(18):13225–13234, 2012.

[96] Ying Hao, Zhenan Sun, Tieniu Tan, and Chao Ren. Multispectral palm image fusion for accurate contact-free palmprint recognition. In *2008 15th International Conference on Image Processing (ICIP)*, pages 281–284. IEEE, 2008.

[97] Ying Hao, Zhenan Sun, Tieniu Tan, and Chao Ren. Multispectral palm image fusion for accurate contact-free palmprint recognition. In *2008 15th IEEE International Conference on Image Processing-(ICIP)*, pages 281–284. IEEE, 2008.

[98] Thalía Harmony, Thalía Fernández, Juan Silva, Jorge Bernal, Lourdes Díaz-Comas, Alfonso Reyes, Erzsébet Marosi, Mario Rodríguez, and Miguel Rodríguez. Eeg delta activity: an indicator of attention to internal processing during performance of mental tasks. *International journal of psychophysiology*, 24(1-2):161–171, 1996.

[99] Daniel Hartung and Christoph Busch. Why vein recognition needs privacy protection. In *5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 1090–1095. IEEE, 2009.

[100] Daniel Hartung, M Aastrup Olsen, Haiyun Xu, H Thanh Nguyen, and Christoph Busch. Comprehensive analysis of spectral minutiae for vein pattern recognition. *IET biometrics*, 1(1):25–36, 2012.

[101] Daniel Hartung, Martin Aastrup Olsen, Haiyun Xu, and Christoph Busch. Spectral minutiae for vein pattern recognition. In *2011 International Joint Conference on Biometrics (IJCB)*, pages 1–7. IEEE, 2011.

[102] Daniel Hartung, Anika Pflug, and Christoph Busch. Vein pattern recognition using chain codes spatial information and skeleton fusing. *SICHERHEIT 2012–Sicherheit, Schutz und Zuverlässigkeit*, 2012.

[103] Daniel Hartung, Massimo Tistarelli, and Christoph Busch. Vein minutia cylinder-codes (v-mcc). In *2013 International Conference on Biometrics (ICB)*, pages 1–7. IEEE, 2013.

[104] Junichi Hashimoto. Finger vein authentication technology and its future. In *Symposium on VLSI Circuits. Digest of Technical Papers.*, pages 5–8. IEEE, 2006.

[105] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on computer vision and pattern recognition (CVPR)*, pages 770–778. IEEE, 2016.

[106] Gabriel Emile Hine, Emanuele Maiorana, and Patrizio Campisi. A zero-leakage fuzzy embedder from the theoretical formulation to real data. *IEEE Transactions on Information Forensics and Security*, 12(7):1724–1734, 2017.

[107] JC Hoeks, Harm Brouwer, and T Holtgraves. Electrophysiological research on conversation and discourse. *The Oxford Handbook of Language and Social Psychology*, pages 365–386, 2014.

[108] Hyung Hong, Min Lee, and Kang Park. Convolutional neural network-based finger-vein recognition using nir image sensors. *Sensors*, 17(6):1–21, 2017.

[109] Lin Hong, Anil K. Jain, and Sharath Pankanti. Can multibiometrics improve performance? In *Proceedings AutoID*, volume 99, pages 59–64. Citeseer, 1999.

[110] Claudia Hruska and Kai Alter. Prosody in dialogues and single sentences: How prosody can influence speech perception. *Information Structure: Theoretical and Empirical Aspects*, pages 221–226, 2004.

[111] Claudia Hruska and Kai Alter. Prosody in dialogues and single sentences: How prosody can influence speech perception. *Information structure: Theoretical and empirical aspects*, pages 221–226, 2004.

[112] Jian-feng Hu. New biometric approach based on motor imagery eeg signals. In *2009 International Conference on Future BioMedical Information Engineering (FBIE)*, pages 94–97. IEEE, 2009.

[113] Beining Huang, Yanggang Dai, Rongfeng Li, Darun Tang, and Wenxin Li. Finger-vein authentication based on wide line detector and pattern normalization. In *20th International Conference on Pattern Recognition (ICPR)*, pages 1269–1272. IEEE, 2010.

[114] Di Huang, Yinhang Tang, Yiding Wang, Liming Chen, and Yunhong Wang. Hand-dorsa vein recognition by matching local features of multisource keypoints. *IEEE Transactions on Cybernetics*, 45(9):1823–1837, 2014.

[115] Di Huang, Renke Zhang, Yuan Yin, Yiding Wang, and Yunhong Wang. Local feature approach to dorsal hand vein recognition by centroid-based circular key-point grid and fine-grained matching. *Image and Vision Computing*, 58:266–277, 2017.

[116] Di Huang, Xiangrong Zhu, Yunhong Wang, and David Zhang. Dorsal hand vein recognition via hierarchical combination of texture and shape clues. *Neurocomputing*, 214:815–828, 2016.

[117] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *IEEE conference on computer vision and pattern recognition (CVPR)*, pages 4700–4708. IEEE, 2017.

[118] Houjun Huang, Shilei Liu, He Zheng, Liao Ni, Yi Zhang, and Wenxin Li. Deepvein: Novel finger vein verification methods based on deep convolutional neural networks. In *International Conference on Identity, Security and Behavior Analysis (ISBA)*, pages 1–8. IEEE, 2017.

[119] Yu-Chen Hung and Petra B Schumacher. Topicality matters: position-specific demands on chinese discourse processing. *Neuroscience letters*, 511(2):59–64, 2012.

[120] ISO/IEC FDIS 30136, ISO/IEC JTC1 SC37 Biometrics. *Information Technology — Performance Testing of Biometric Template Protection Schemes*. International Organization for Standardization, 2017.

[121] ISO/IEC JTC1 SC27 Security Techniques. *ISO/IEC 24745:2011. Information Technology - Security Techniques - Biometric Information Protection*. International Organization for Standardization, 2011.

[122] Anil Jain, Ruud Bolle, and Sharath Pankanti. *Biometrics: personal identification in networked society*, volume 479. Springer Science & Business Media, 2006.

[123] Anil K. Jain, Karthik Nandakumar, and Abhishek Nagar. Biometric template security. *EURASIP Journal on Advances in Signal Processing*, 2008:113, 2008.

[124] Anil K. Jain, Arun Ross, Salil Prabhakar, et al. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1), 2004.

[125] Anil K Jain, Arun A Ross, and Karthik Nandakumar. *Introduction to biometrics*. Springer Science & Business Media, 2011.

[126] Ehsaneddin Jalilian and Andreas Uhl. Finger-vein recognition using deep fully convolutional neural semantic segmentation networks: The impact of training data. In *IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–8. IEEE, 2018.

[127] Stefan Jenisch and Andreas Uhl. Security analysis of a cancelable iris recognition system based on block remapping. In *2011 18th IEEE International Conference on Image Processing*, pages 3213–3216. IEEE, 2011.

[128] Wei Jia, De-Shuang Huang, and David Zhang. Palmprint verification based on robust line orientation code. *Pattern Recognition*, 41(5):1504–1513, 2008.

[129] Guo Jun and Kavallur Gopi Smitha. Eeg based stress level identification. In *International Conference on Systems, Man, and Cybernetics (SMC)*, pages 003270–003274. IEEE, 2016.

[130] Edith Kaan and Tamara Y Swaab. Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, 15(1):98–110, 2003.

[131] Rafał Kabaciński and Mateusz Kowalski. Human vein pattern segmentation from low quality images–a comparison of methods. In *Image Processing and Communications Challenges 2*, pages 105–112. Springer, 2010.

[132] Rafał Kabaciński and Mateusz Kowalski. Human vein pattern correlation-a comparison of segmentation methods. In *Computer Recognition Systems 4*, pages 51–59. Springer, 2011.

[133] Rafal Kabacinski and Mateusz Kowalski. Vein pattern database and benchmark results. *Electronics Letters*, 47(20):1127–1128, 2011.

[134] Wenxiong Kang, Yang Liu, Qiuxia Wu, and Xishun Yue. Contact-free palm-vein recognition based on local invariant features. *PloS one*, 9(5):97548, 2014.

[135] Wenxiong Kang and Qiuxia Wu. Contactless palm vein recognition using a mutual foreground-based local binary pattern. *IEEE Transactions on Information Forensics and Security*, 9(11):1974–1985, 2014.

[136] Christof Kauba, Emanuela Piciucco, Emanuele Maiorana, Patrizio Campisi, and Andreas Uhl. Advanced variants of feature level fusion for finger vein recognition. In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–7. IEEE, 2016.

[137] Steven M Kay. *Modern spectral estimation*. Pearson Education India, 1988.

[138] Simon P Kelly, Edmund C Lalor, Ciarán Finucane, Gary McDarby, and Richard B Reilly. Visual spatial attention control in an independent brain-computer interface. *IEEE Transactions on Biomedical Engineering*, 52(9):1588–1596, 2005.

[139] Auguste Kerckhoffs. La cryptographic militaire. *Journal des sciences militaires*, pages 5–38, 1883.

[140] Wan Kim, Jong Song, and Kang Park. Multimodal biometric recognition based on convolutional neural network by the fusion of finger-vein and finger shape using near-infrared (nir) camera sensor. *Sensors*, 18(7):2296, 2018.

[141] Simon Kirchgasser, Christof Kauba, and Andreas Uhl. Cancellable biometrics for finger vein recognition - application in the feature domain. In Andreas Uhl, Christoph Busch, Sebastien Marcel, and Raymond Veldhuis, editors, *Handbook of Vascular Biometrics*, chapter 16, pages 507–525. Springer Nature Switzerland AG, Cham, Switzerland, 2019.

[142] Stefan Klein, Marius Staring, Keelin Murphy, Max A Viergever, and Josien PW Pluim. Elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, 2010.

[143] W Klimesch, M Doppelmayr, Th Pachinger, and B Ripper. Brain oscillations and human memory: Eeg correlates in the upper alpha and theta band. *Neuroscience letters*, 238(1-2):9–12, 1997.

[144] W Klimesch, M Doppelmayr, Th Pachinger, and H Russegger. Event-related desynchronization in the alpha band and the processing of semantic information. *Cognitive Brain Research*, 6(2):83–94, 1997.

[145] Wolfgang Klimesch. Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29(2-3):169–195, 1999.

[146] Juris Klonovs, Christoffer Kjeldgaard Petersen, Henning Olesen, and Allan Hammershoj. Id proof on the go: Development of a mobile eeg-based biometric authentication system. *IEEE Vehicular Technology Magazine*, 8(1):81–89, 2013.

[147] Gennady G Knyazev. Motivation, emotion, and their inhibitory control mirrored in brain oscillations. *Neuroscience & Biobehavioral Reviews*, 31(3):377–395, 2007.

[148] Milan Kostílek and Jakub Šťastnỳ. Eeg biometric identification: Repeatability and influence of movement-related eeg. In *2012 International Conference on Applied Electronics*, pages 147–150. IEEE, 2012.

[149] Jiangtao Kuang, Garrett M Johnson, and Mark D. Fairchild. icam06: A refined image appearance model for hdr image rendering. *Journal of Visual Communication and Image Representation*, 18(5):406–414, 2007.

[150] Ajay Kumar and K Venkata Prathyusha. Personal authentication using hand vein triangulation and knuckle shape. *IEEE Transactions on Image processing*, 18(9):2127–2136, 2009.

[151] Ajay Kumar and Kuo Wang. Identifying humans by matching their left palmprint with right palmprint images using convolutional neural network. In *First Int. Workshop on Deep Learning and Pattern Recognition*, 2016.

[152] Ajay Kumar and Yingbo Zhou. Human identification using finger images. *IEEE Transactions on Image Processing*, 21(4):2228–2244, 2012.

[153] Onur Can Kurban, Özden Nıyaz, and Tülay Yildirim. Neural network based wrist vein identification using ordinary camera. In *2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 1–4. IEEE, 2016.

[154] Marta Kutas and Kara D Federmeier. Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12):463–470, 2000.

[155] Marta Kutas and Kara D Federmeier. Thirty years and counting: Finding meaning in the n400 component of the event-related brain potential (erp). *Annual Review of Psychology*, 62:621–647, 2011.

[156] Marta Kutas and Steven A Hillyard. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427):203–205, 1980.

[157] Ridvan S. Kuzu, Emanuela Piciucco, Emanuele Maiorana, and Patrizio Campisi. On-the-fly finger-vein-based biometric recognition using deep neural networks. *IEEE Transactions on Information Forensics and Security*, 15:2641–2654, 2020.

[158] Daria La Rocca, Patrizio Campisi, and Gaetano Scarano. Eeg biometrics for individual recognition in resting state with closed eyes. In *2012 BIOSIG-Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG)*, pages 1–12. IEEE, 2012.

[159] Daria La Rocca, Patrizio Campisi, and Gaetano Scarano. On the repeatability of eeg features in a biometric recognition framework using a resting state protocol. In *BIOSIGNALS*, pages 419–428, 2013.

[160] Daria La Rocca, Viviana Masia, Emanuele Maiorana, Edoardo Lombardi Vallauri, and Patrizio Campisi. Brain response to information structure misalignments in linguistic contexts. *Neurocomputing*, 199:1–15, 2016.

[161] Jeanne M. LaBerge, Roy L Gordon, Robert K. Kerlan, and Mark W. Wilson. Interventional radiology essentials. *Lippincott Williams & Wilkins*, 2000.

[162] Pierre-Olivier Ladoux, Christophe Rosenberger, and Bernadette Dorizzi. Palm vein verification system based on sift matching. In *International Conference on Biometrics (ICB)*, pages 1290–1298. Springer, 2009.

[163] GC Lairy. *Handbook of Electroencephalography and Clinical Neurophysiology: Vol 6*. Elsevier Scientific Publishing Company, 1975.

[164] Seyed Mehdi Lajevardi, Arathi Arakala, Stephen Davis, and Kathy J Horadam. Hand vein authentication using biometric graph matching. *IET Biometrics*, 3(4):302–313, 2014.

[165] Ellen F Lau, Colin Phillips, and David Poeppel. A cortical network for semantics:(de) constructing the n400. *Nature Reviews Neuroscience*, 9(12):920–933, 2008.

[166] Eui Chul Lee, Hyunwoo Jung, and Daeyeoul Kim. New finger biometric method using near infrared imaging. *Sensors*, 11(3):2319–2333, 2011.

[167] Eui Chul Lee, Hyeon Chang Lee, and Kang Ryoung Park. Finger vein recognition using minutia-based alignment and local binary pattern-based feature extraction. *International Journal of Imaging Systems and Technology*, 19(3):179–186, 2009.

[168] J-C Lee, C-H Lee, C-B Hsu, P-Y Kuei, and K-C Chang. Dorsal hand vein recognition based on 2d gabor filters. *The Imaging Science Journal*, 62(3):127–138, 2014.

[169] Jen-Chun Lee. A novel biometric system based on palm vein image. *Pattern Recognition Letters*, 33(12):1520–1528, 2012.

[170] Jen-Chun Lee, Tsung-Ming Lo, and Chien-Ping Chang. Dorsal hand vein recognition based on directional filter bank. *Signal, Image and Video Processing*, 10(1):145–152, 2016.

[171] Y. P. Lee. Palm vein recognition based on a modified $(2D)^2$LDA. *Signal, Image and Video Processing*, 9(1):229–242, 2015.

[172] Lu Leng and Jiashu Zhang. Palmhash code vs. palmphasor code. *Neurocomputing*, 108:1–12, 2013.

[173] W.J.M. Levelt, A. Roelofs, and A.S. Meyer. A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22:1–38, 1999.

[174] Ming Li and Shrikanth Narayanan. Robust ecg biometrics by fusing temporal and cepstral information. In *20th International Conference on Pattern Recognition (ICPR)*, pages 1326–1329. IEEE, 2010.

[175] Chih-Lung Lin and Kuo-Chin Fan. Biometric verification using thermal images of palm-dorsa vein patterns. *IEEE Transactions on Circuits and systems for Video Technology*, 14(2):199–213, 2004.

[176] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian D. Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, page 5. IEEE, 2017.

[177] Yuan-Pin Lin, Chi-Hong Wang, Tzyy-Ping Jung, Tien-Lin Wu, Shyh-Kang Jeng, Jeng-Ren Duann, and Jyh-Horng Chen. Eeg-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, 57(7):1798–1806, 2010.

[178] Fei Liu, Gongping Yang, Yilong Yin, and Shuaiqiang Wang. Singular value decomposition based minutiae matching method for finger vein recognition. *Neurocomputing*, 145:75–89, 2014.

[179] Yu Lu, Shan Juan Xie, Sook Yoon, and Dong Sun Park. Finger vein identification using polydirectional local line binary pattern. In *2013 International Conference on ICT Convergence (ICTC)*, pages 61–65. IEEE, 2013.

[180] Yu Lu, Shan Juan Xie, Sook Yoon, Zhihui Wang, and Dong Sun Park. An available database for the research of finger vein recognition. In *2013 6th International Congress on Image and Signal Processing (CISP)*, volume 1, pages 410–415. IEEE, 2013.

[181] Steven J Luck and Emily S Kappenman. *The Oxford handbook of event-related potential components*. Oxford university press, 2011.

[182] Xin Ma, Xiaojun Jing, Hai Huang, Yuanhao Cui, and Junsheng Mu. Palm vein recognition scheme based on an adaptive gabor filter. *Iet Biometrics*, 6(5):325–333, 2016.

[183] Dario Maio, Davide Maltoni, Raffaele Cappelli, Jim L Wayman, and Anil K Jain. Fvc2004: Third fingerprint verification competition. In *International Conference on Biometric Authentication*, pages 1–7. Springer, 2004.

[184] Emanuele Maiorana and Patrizio Campisi. Longitudinal evaluation of eeg-based biometric recognition. *IEEE Transactions on Information Forensics and Security*, 13(5):1123–1138, 2017.

[185] Emanuele Maiorana, Patrizio Campisi, Julian Fierrez, Javier Ortega-Garcia, and Alessandro Neri. Cancelable templates for sequence-based biometrics with application to on-line signature recognition. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 40(3):525–538, 2010.

[186] Emanuele Maiorana, Patrizio Campisi, Javier Ortega-Garcia, and Alessandro Neri. Cancelable biometrics for hmm-based signature recognition. In *International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6. IEEE, 2008.

[187] Emanuele Maiorana, Daria La Rocca, and Patrizio Campisi. Eigenbrains and eigentensorbrains: Parsimonious bases for eeg biometrics. *Neurocomputing*, 171:638–648, 2016.

[188] Emanuele Maiorana, Daria La Rocca, and Patrizio Campisi. On the permanence of EEG signals for biometric recognition. *IEEE Transaction on Information Forensics and Security*, 11(1):163–175, 2016.

[189] Emanuele Maiorana, Marcos Martinez-Diaz, Patrizio Campisi, Javier Ortega-García, and Alessandro Neri. Template protection for hmm-based on-line signature authentication. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–6. IEEE, 2008.

[190] Plonsey Malmivuo, Jaakko Malmivuo, and Robert Plonsey. *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, USA, 1995.

[191] Kim-Fung Man, Kit-Sang Tang, and Sam Kwong. Genetic algorithms: concepts and applications [in engineering design]. *IEEE transactions on Industrial Electronics*, 43(5):519–534, 1996.

[192] Sebastien Marcel and José del R Millán. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4), 2007.

[193] Ivan Martinovic, Kasper B Rasmussen, Marc Roeschlin, and Gene Tsudik. Pulse-response: Exploring human body impedance for biometric recognition. *ACM Transactions on Privacy and Security (TOPS)*, 20(2):6, 2017.

[194] Viviana Masia, Paolo Canal, Irene Ricci, Edoardo Lombardi Vallauri, and Valentina Bambini. Presupposition of new information as a pragmatic garden path: Evidence from event-related brain potentials. *Journal of Neurolinguistics*, 42:31–48, 2017.

[195] John McCann. Lessons learned from mondrians applied to real images and color gamuts. In *7th Color and Imaging Conference*, pages 1–8. Society for Imaging Science and Technology, 1999.

[196] John J. McCann and Alessandro Rizzi. *The art and science of HDR imaging*, volume 26. John Wiley & Sons, 2011.

[197] Dennis J McFarland, Laurie A Miner, Theresa M Vaughan, and Jonathan R Wolpaw. Mu and beta rhythm topographies during motor imagery and actual movements. *Brain topography*, 12(3):177–186, 2000.

[198] Kimford J Meador, Patty G Ray, Javier R Echauz, David W Loring, and George J Vachtsevanos. Gamma coherence and conscious perception. *Neurology*, 59(6):847–854, 2002.

[199] Maarten Mennes, Heidi Wouters, Bart Vanrumste, Lieven Lagae, and Peter Stiers. Validation of ica as a tool to remove eye movement artifacts from eeg/erp. *Psychophysiology*, 47(6):1142–1150, 2010.

[200] Goh Kah Ong Michael, Tee Connie, and Andrew Beng Jin Teoh. A contactless biometric system using palm print and palm vein features. *Advanced Biometric Technologies*, 2011.

[201] Benjamin Miller. Vital signs of identity [biometrics]. *Spectrum, IEEE*, 31(2):22–30, 1994.

[202] Leila Mirmohamadsadeghi and Andrzej Drygajlo. Palm vein recognition with local binary patterns and local derivative patterns. In *2011 International Joint Conference on Biometrics (IJCB)*, pages 1–6. IEEE, 2011.

[203] Leila Mirmohamadsadeghi and Andrzej Drygajlo. Palm vein recognition with local texture patterns. *IET Biometrics*, 3(4):198–206, 2014.

[204] Maryualê Malvessi Mittmann. O c-oral-brasil e o estudo da fala informal: um novo olhar sobre o tópico no português brasileiro. 2012.

[205] Naoto Miura, Akio Nagasaka, and Takafumi Miyatake. Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. *Machine Vision and Applications*, 15(4):194–203, 2004.

[206] Naoto Miura, Akio Nagasaka, and Takafumi Miyatake. Extraction of finger-vein patterns using maximum curvature points in image profiles. *IEICE Transactions on Information and Systems.*, 90(8):1185–1194, aug 2007.

[207] Luis Alfredo Moctezuma and Marta Molinas. Eeg-based subjects identification based on biometrics of imagined speech using emd. In *International Conference on Brain Informatics*, pages 458–467. Springer, 2018.

[208] Cheniti Mohamed, Zahid Akhtar, Boukezzoula Naceur Eddine, and Tiago H Falk. Combining left and right wrist vein images for personal verification. In *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE, 2017.

[209] Klaus-Robert Müller, Michael Tangermann, Guido Dornhege, Matthias Krauledat, Gabriel Curio, and Benjamin Blankertz. Machine learning for real-time single-trial eeg-analysis: from brain–computer interfacing to mental state monitoring. *Journal of Neuroscience Methods*, 167(1):82–90, 2008.

[210] Markus Näpflin, Marc Wildi, and Johannes Sarnthein. Test–retest reliability of resting eeg spectra validates a statistical signature of persons. *Clinical Neurophysiology*, 118(11):2519–2524, 2007.

[211] H. Neville, J. Nicol, A. Barss, K.I. Forster, and M.F. Garrett. Syntactically-based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3:151–165, 1991.

[212] Phuoc Nguyen, Dat Tran, Xu Huang, and Dharmendra Sharma. A proposed feature extraction method for EEG-based person identification. In *International Conference on Artificial Intelligence (ICAI)*, page 1, 2012.

[213] Takashi Nishimoto, Yoshiki Azuma, Hiroshi Morioka, and Shin Ishii. Individual identification by resting-state eeg using common dictionary learning. In *International Conference on Artificial Neural Networks*, pages 199–207. Springer, 2017.

[214] Ikenna Odinaka, Po-Hsiang Lai, Alan D Kaplan, Joseph A O'Sullivan, Erik J Sirevaag, and John W Rohrbaugh. Ecg biometric recognition: A comparative analysis. *IEEE Transactions on Information Forensics and Security*, 7(6):1812–1824, 2012.

[215] Timo Ojala, Matti Pietikainen, and David Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *12th International Conference on Pattern Recognition*, volume 1, pages 582–585. IEEE, 1994.

[216] Richard C Oldfield. The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia*, 9(1):97–113, 1971.

[217] Thian Song Ong, Jackson Horlick Teng, Kalaiarasi Sonai Muthu, and Andrew Beng Jin Teoh. Multi-instance finger vein recognition using minutiae matching. In *2013 6th International Congress on Image and Signal Processing (CISP)*, volume 3, pages 1730–1735. IEEE, 2013.

[218] Lee Osterhout and Phillip J Holcomb. Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, 31(6):785–806, 1992.

[219] Mary A Pagnutti, Robert E Ryan, George J Cazenavette, Maxwell J Gold, Ryan Harlan, Edward Leggett, and James F Pagnutti. Laying the foundation to use raspberry pi 3 v2 camera module imagery for scientific and engineering purposes. *Journal of Electronic Imaging*, 26(1):013014, 2017.

[220] Genady Paikin and Ayellet Tal. Solving multiple square jigsaw puzzles with missing pieces. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4832–4839. IEEE, 2015.

[221] Ramaswamy Palaniappan. Method of identifying individuals using vep signals and neural network. *IEE Proceedings-Science, Measurement and Technology*, 151(1):16–20, 2004.

[222] RB Paranjape, J Mahovsky, L Benedicenti, and Z Koles. The electroencephalogram as a biometric. In *Canadian Conference on Electrical and Computer Engineering 2001. Conference Proceedings (Cat. No. 01TH8555)*, volume 2, pages 1363–1366. IEEE, 2001.

[223] J. Enrique Suarez Pascual, Jaime Uriarte-Antonio, Raul Sanchez-Reillo, and Michael G. Lorenz. Capturing hand or wrist vein images for biometric authentication using low-cost devices. In *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, pages 318–322. IEEE, 2010.

[224] Vishal M Patel, Nalini K Ratha, and Rama Chellappa. Cancelable biometrics: A review. *IEEE Signal Processing Magazine*, 32(5):54–65, 2015.

[225] Anika Pflug, Daniel Hartung, and Christoph Busch. Feature extraction from vein images using spatial information and chain codes. *Information security technical report*, 17(1-2):26–35, 2012.

[226] Montri Phothisonothai. An investigation of using ssvep for EEG-based user authentication system. *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, (APSIPA ASC)*, (December):923–926, 2016.

[227] Emanuela Piciucco, Emanuele Maiorana, and Patrizio Campisi. Biometric fusion for palm-vein-based recognition systems. In *International Tyrrhenian Workshop on Digital Communication*, pages 18–28. Springer, 2017.

[228] Emanuela Piciucco, Emanuele Maiorana, and Patrizio Campisi. Palm vein recognition using a high dynamic range approach. *IET Biometrics*, 7(5):439–446, 2018.

[229] Emanuela Piciucco, Emanuele Maiorana, Christof Kauba, Andreas Uhl, and Patrizio Campisi. Cancelable biometrics for finger vein recognition. In *2016 First International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE)*, pages 1–5. IEEE, 2016.

[230] Maria Emanuela Piemontese. *Capire e farsi capire: teorie e tecniche della scrittura controllata*. Tecnodid, 1996.

[231] Jaishanker K. Pillai, Vishal M. Patel, Rama Chellappa, and Nalini K. Ratha. Sectored random projections for cancelable iris biometrics. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1838–1841. IEEE, 2010.

[232] Jaishanker K. Pillai, Vishal M. Patel, Rama Chellappa, and Nalini K. Ratha. Secure and robust iris recognition using random projections and sparse representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(9):1877–1893, 2011.

[233] John Polich. Neuropsychology of p300. *Oxford handbook of event-related potential components*, 159:88, 2012.

[234] Christoph Posch. Detectors, pixels, and signal processing. In *Smart Cameras*, pages 53–79. Springer, 2009.

[235] M Poulos, M Rangoussi, V Chrissikopoulos, and A Evangelou. Person identification based on parametric processing of the eeg. In *ICECS'99. Proceedings of ICECS'99. 6th IEEE International Conference on Electronics, Circuits and Systems (Cat. No. 99EX357)*, volume 1, pages 283–286. IEEE, 1999.

[236] Marios Poulos, Maria Rangoussi, and Nikolaos Alexandris. Neural network based person identification using eeg features. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)*, volume 2, pages 1117–1120. IEEE, 1999.

[237] Annisa Yuditya Pratiwi, W Tjokorda Agung Budi, and Kurniawan Nur Ramadhani. Identity recognition with palm vein feature using local binary pattern rotation invariant. In *4th International Conference on Information and Communication Technology (ICoICT)*, pages 1–6. IEEE, 2016.

[238] Friedemann Pulvermüller, Werner Lutzenberger, and Hubert Preissl. Nouns and verbs in the intact brain: evidence from event-related potentials and high-frequency cortical responses. *Cerebral cortex*, 9(5):497–506, 1999.

[239] Huafeng Qin and Mounîm A. El-Yacoubi. Finger-vein quality assessment by representation learning from binary images. In *International Conference on Neural Information Processing*, pages 421–431. Springer, 2015.

[240] Huafeng Qin and Mounîm A. El Yacoubi. Deep representation-based feature extraction and recovering for finger-vein verification. *IEEE Transaction on Information Forensics and Security*, 12(8):1816–1829, 2017.

[241] Huafeng Qin and Mounîm A. El Yacoubi. Deep representation for finger-vein image quality assessment. *IEEE Trans. on Circuits and Systems for Video Technology*, 28(8):1677–1693, 2018.

[242] Shirong Qiu, Yaqin Liu, Yujia Zhou, Jing Huang, and Yixiao Nie. Finger-vein recognition based on dual-sliding window localization and pseudo-elliptical transformer. *Expert Systems with Applications*, 64:618–632, 2016.

[243] Syafeeza Ahmad Radzi, Mohamed Khalil-Hani, and Rabia Bakhteri. Finger-vein biometric identification using convolutional neural network. *Turkish Journal of Electrical Eng. & Comp. Sciences*, 24(3):1863–1878, 2016.

[244] Günter Rager and Wolf Singer. The response of cat visual cortex to flicker stimuli of variable frequency. *European Journal of Neuroscience*, 10(5):1856–1877, 1998.

[245] R Raghavendra and Christoph Busch. A low cost wrist vein sensor for biometric authentication. In *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*, pages 201–205. IEEE, 2016.

[246] R Raghavendra, Sushma Venkatesh, Kiran B Raja, and Christoph Busch. Transferable deep convolutional neural network features for fingervein presentation attack detection. In *5th International Workshop on Biometrics and Forensics (IWBF)*, pages 1–5. IEEE, 2017.

[247] Ramachandra Raghavendra, Manasa Avinash, Sébastien Marcel, and Christoph Busch. Finger vein liveness detection using motion magnification. In *2015 IEEE 7th Internarnational Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7. IEEE, 2015.

[248] Ramachandra Raghavendra and Christoph Busch. Presentation attack detection algorithms for finger vein biometrics: A comprehensive study. In *11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 628–632. IEEE, 2015.

[249] Shantanu Rane. Standardization of biometric template protection. *IEEE MultiMedia*, 21(4):94–99, 2014.

[250] Nalini Ratha, Jonathan Connell, Ruud M. Bolle, and Sharat Chikkerur. Cancelable biometrics: A case study in fingerprints. In *18th International Conference on Pattern Recognition (ICPR)*, volume 4, pages 370–373. IEEE, 2006.

[251] Nalini K Ratha, Sharat Chikkerur, Jonathan H Connell, and Ruud M Bolle. Generating cancelable fingerprint templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):561–572, 2007.

[252] Nalini K. Ratha, Jonathan H. Connell, and Ruud M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM systems Journal*, 40(3):614–634, 2001.

[253] C. Rathgeb, F. Breitinger, C. Busch, and H. Baier. On application of bloom filters to iris biometrics. 3(4):207–218, 2014.

[254] Christian Rathgeb and Andreas Uhl. Secure iris recognition based on local intensity variations. In *Image Analysis and Recognition*, pages 266–275. Springer, 2010.

[255] Christian Rathgeb and Andreas Uhl. A survey on biometric cryptosystems and cancelable biometrics. *EURASIP Journal on Information Security*, 2011(1):1–25, 2011.

[256] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[257] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. *ACM Transactions on Graphics (TOG)*, 21(3):267–276, 2002.

[258] Grega Repovs. Dealing with noise in eeg recording and data analysis. In *Informatica Medica Slovenica*, volume 15, pages 18–25, 2010.

[259] Kenneth Revett. Cognitive biometrics: a novel approach to continuous person authentication. In *Continuous Authentication Using Biometrics: Data, Models, and Metrics*, pages 105–136. IGI Global, 2012.

[260] Alejandro Riera, Aureli Soria-Frisch, Marco Caparrini, Carles Grau, and Giulio Ruffini. Unobtrusive biometric system based on electroencephalogram analysis. *EURASIP Journal on Advances in Signal Processing*, 2008(1):143728, 2007.

[261] Dietmar Roehm, Ina Bornkessel-Schlesewsky, Matthias Schlesewsky, et al. *The internal structure of the N400: Frequency characteristics of a language related ERP component*. PhD thesis, Nova Science Publishers Incorporated, 2007.

[262] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[263] Bakhtiar Affendi Rosdi, Chai Wuh Shing, and Shahrel Azmin Suandi. Finger vein recognition using local line binary pattern. *Sensors*, 11(12):11357–11371, 2011.

[264] Arun A. Ross, Karthik Nandakumar, and Anil K. Jain. *Handbook of multibiometrics: human recognition systems*. Springer, 2006.

[265] Maria V Ruiz-Blondet, Zhanpeng Jin, and Sarah Laszlo. Cerebre: A novel method for very high accuracy event-related potential biometric identification. *IEEE Transactions on Information Forensics and Security*, 11(7):1618–1629, 2016.

[266] Maria V Ruiz-Blondet, Zhanpeng Jin, and Sarah Laszlo. Permanence of the cerebre brain biometric protocol. *Pattern Recognition Letters*, 95:37–43, 2017.

[267] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[268] Muhammad Sajjad, Salman Khan, Tanveer Hussain, Khan Muhammad, Arun Kumar Sangaiah, Aniello Castiglione, Christian Esposito, and Sung Wook Baik. Cnn-based anti-spoofing two-tier multi-factor authentication system. *Pattern Recognition Letters*, 2018.

[269] Marios Savvides, BVK Vijaya Kumar, and Pradeep K Khosla. Cancelable biometric filters for face recognition. In *17th International Conference on Pattern Recognition (ICPR)*, volume 3, pages 922–925. IEEE, 2004.

[270] Petra B Schumacher and Yu-Chen Hung. Positional influences on information packaging: Insights from topological fields in german. *Journal of Memory and Language*, 67(2):295–310, 2012.

[271] Mark S. Schwartz and Frank Andrasik. *Biofeedback: A practitioner's guide*. Guilford Publications, 2015.

[272] Sidney J Segalowitz and Kerry L Barnes. The reliability of erp components in the auditory oddball paradigm. *Psychophysiology*, 30(5):451–459, 1993.

[273] Kashif Shaheed, Hangang Liu, Gongping Yang, Imran Qureshi, Jie Gou, and Yilong Yin. A systematic review of finger vein recognition techniques. *Information*, 9(9):213, 2018.

[274] Qi Shan, Jiaya Jia, and Michael S Brown. Globally optimized linear windowed tone mapping. *IEEE Transactions on Visualization and Computer Graphics*, 16(4):663–675, 2010.

[275] Takashi Shibata, Masayuki Tanaka, and Masatoshi Okutomi. Gradient-domain image reconstruction framework with intensity-range and base-structure constraints. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2745–2753, 2016.

[276] Patrice Y Simard, David Steinkraus, John C Platt, et al. Best practices for convolutional neural networks applied to visual document analysis. In *ICDAR*, volume 3, pages 958–962, 2003.

[277] Koen Simoens, Pim Tuyls, and Bart Preneel. Privacy weaknesses in biometric sketches. In *Proceedings of the 30th IEEE Symposium on Security and Privacy*, pages 188–203. IEEE, 2009.

[278] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[279] HH Stassen. Computerized recognition of persons by eeg spectral patterns. *Electroencephalography and Clinical Neurophysiology*, 49(1-2):190–194, 1980.

[280] Patrick Sturt, Anthony J Sanford, Andrew Stewart, and Eugene Dawydiak. Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic bulletin & review*, 11(5):882–888, 2004.

[281] Fei Su, Huangling Zhou, Zhiyin Feng, and Junshui Ma. A biometric-based covert warning system using eeg. In *2012 5th IAPR International Conference on Biometrics (ICB)*, pages 342–347. IEEE, 2012.

[282] Margit Sutrop and Katrin Laas-Mikko. From identity verification to behavior prediction: Ethical implications of second generation biometrics. *Review of policy research*, 29(1):21–36, 2012.

[283] Michael J. Swain and Dana H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[284] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *31st AAAI Conference on Artificial Intelligence*, volume 4, page 12, 2017.

[285] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *IEEE Conference on computer vision and pattern recognition (CVPR)*, pages 1–9. IEEE, 2015.

[286] Kenji Takita, Takafumi Aoki, Yoshifumi Sasaki, Tatsuo Higuchi, and Koji Kobayashi. High-accuracy subpixel image registration based on phase-only correlation. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 86(8):1925–1934, 2003.

[287] Andrew B. J. Teoh, Alwyn Goh, and David C. L. Ngo. Random multispace quantization as an analytic mechanism for biohashing of biometric and random identity inputs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):1892–1901, 2006.

[288] Pedro Tome and Sébastien Marcel. On the vulnerability of palm vein recognition to spoofing attacks. In *2015 8th International Conference on Biometrics(ICB)*, pages 319–325. IEEE, 2015.

[289] Bram T. Ton and Raymond N.J. Veldhuis. A high quality finger vascular pattern dataset collected using a custom designed capturing device. In *2013 International Conference on Biometrics (ICB)*, pages 1–5. IEEE, 2013.

[290] H Touyama. Eeg-based personal identificatio. *Biomedical Engineering*, 2009.

[291] Randa Boukhris Trabelsi, Alima Damak Masmoudi, and Dorra Sellami Masmoudi. Hand vein recognition system with circular difference and statistical directional patterns based on an artificial neural network. *Multimedia Tools and Applications*, 75(2):687–707, 2016.

[292] Koji Tsuru and Gert Pfurtscheller. Brainwave biometrics: a new feature extraction approach with the cepstral analysis method. *Transactions of Japanese Society for Medical and Biological Engineering*, 50(1):162–167, 2012.

[293] Umut Uludag, Sharath Pankanti, Salil Prabhakar, and Anil K. Jain. Biometric cryptosystems: issues and challenges. *Proceedings of the IEEE*, 92(6):948–960, 2004.

[294] Jaime Uriarte-Antonio, Daniel Hartung, J Enrique Suarez Pascual, and Raul Sanchez-Reillo. Vascular biometrics based on a minutiae extraction approach. In *2011 Carnahan Conference on Security Technology*, pages 1–7. IEEE, 2011.

[295] Jaime Uriarte-Antonio, J Enrique Suarez-Pascual, Michael Garcia-Lorenz, and Raul Sanchez-Reillo. Parametrical study of a vascular biometric system. In *2011 International Conference on Hand-Based Biometrics*, pages 1–6. IEEE, 2011.

[296] Bambini V. Neurolinguistics. In *Handbook of pragmatics*. Amsterdam/Philadelphia, John Benjamins, 2012.

[297] KS Vairavel, N Ikram, and S Mekala. Performance analysis on feature extraction using dorsal hand vein image. *Soft Computing*, pages 1–10, 2019.

[298] Hoang Thien Van, Phat Quang Tat, and Thai Hoang Le. Palmprint verification using gridpca for gabor features. In *2nd Symposium on Information and Communication Technology*, pages 217–225. ACM, 2011.

[299] Hoang Thien Van, Thanh Tuan Thai, and Thai Hoang Le. Robust finger vein identification base on discriminant orientation feature. In *7th International Conference on Knowledge and Systems Engineering (KSE)*, pages 348–353. IEEE, 2015.

[300] Francisco Varela, Jean-Philippe Lachaux, Eugenio Rodriguez, and Jacques Martinerie. The brainweb: phase synchronization and large-scale integration. *Nature reviews neuroscience*, 2(4):229, 2001.

[301] Andrea Vedaldi and Lenc Karel. Matconvnet: Convolutional neural networks for matlab. In *Proceeding of the 23rd ACM international conference on Multimedia*. ACM, 2015.

[302] François-Benoît Vialatte, Monique Maurice, Justin Dauwels, and Andrzej Cichocki. Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. *Progress in Neurobiology*, 90(4):418–438, 2010.

[303] Friedrich Vogel. The genetic basis of the normal human electroencephalogram (eeg). *Humangenetik*, 10(2):91–114, 1970.

[304] Huabin Wang, Mengli Du, Jian Zhou, and Liang Tao. Weber local descriptors with variable curvature gabor filter for finger vein recognition. *IEEE Access*, 7:108261–108277, 2019.

[305] Jian-Gang Wang, Wei-Yun Yau, Andy Suwandy, and Eric Sung. Person recognition by fusing palmprint and palm vein images based on "laplacianpalm" representation. *Pattern Recognition*, 41(5):1514–1527, 2008.

[306] Jun Wang, Zaiyu Pan, Guoqing Wang, Ming Li, and Yulian Li. Spatial pyramid pooling of selective convolutional features for vein recognition. *IEEE Access*, 6:28563–28572, 2018.

[307] Jun Wang and Guoqing Wang. Hand-dorsa vein recognition with structure growing guided cnn. *Optik-International Journal for Light and Electron Optics*, 149:469–477, 2017.

[308] Jun Wang and Guoqing Wang. Quality-specific hand vein recognition system. *IEEE Transactions on Information Forensics and Security*, 12(11):2599–2610, 2017.

[309] Jun Wang, Guoqing Wang, Ming Li, and Wenkai Du. Hand vein recognition based on pcet. *Optik*, 127(19):7663–7669, 2016.

[310] Lingyu Wang, Graham Leedham, and David Siu-Yeung Cho. Minutiae feature analysis for infrared hand vein pattern biometrics. *Pattern recognition*, 41(3):920–929, 2008.

[311] Luming Wang and Petra B Schumacher. New is not always costly: evidence from online processing of topic and contrast in japanese. *Frontiers in Psychology*, 4:363, 2013.

[312] Song Wang and Jiankun Hu. Design of alignment-free cancelable fingerprint templates via curtailed circular convolution. *Pattern Recognition*, 47(3):1321–1329, 2014.

[313] Yiding Wang, Ke Zhang, and Lik-Kwan Shark. Personal identification based on multiple keypoint sets of dorsal hand vein images. *IET Biometrics*, 3(4):234–245, 2014.

[314] Yunqi Wang and Laleh Najafizadeh. On the invariance of eeg-based signatures of individuality with application in biometric identification. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4559–4562. IEEE, 2016.

[315] James L Wayman. Fundamentals of biometric authentication technologies. *International Journal of Image and Graphics*, 1(01):93–113, 2001.

[316] Hanna Weiland, Valentina Bambini, and Petra B Schumacher. The role of literal meaning in figurative language comprehension: Evidence from masked priming erp. *Frontiers in Human Neuroscience*, 8:583, 2014.

[317] Sabine Weiss and Horst M Mueller. The contribution of eeg coherence to the investigation of language. *Brain and Language*, 85(2):325–343, 2003.

[318] Sabine Weiss, Horst M Mueller, Baerbel Schack, Jonathan W King, Martha Kutas, and Peter Rappelsberger. Increased neuronal communication accompanying sentence comprehension. *International Journal of Psychophysiology*, 57(2):129–141, 2005.

[319] Tjokorda Agung Budi Wirayuda. Palm vein recognition based-on minutiae feature and feature matching. In *International Conference on Electrical Engineering and Informatics (ICEEI)*, pages 350–355. IEEE, 2015.

[320] George Wolberg. Image morphing: a survey. *The visual computer*, 14(8):360–372, 1998.

[321] Kuang-Shyr Wu, Jen-Chun Lee, Tsung-Ming Lo, Ko-Chin Chang, and Chien-Ping Chang. A secure palm vein recognition system. *Journal of Systems and Software*, 86(11):2870–2876, 2013.

[322] Xiang Wu, Ran He, Zhenan Sun, and Tieniu Tan. A light cnn for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security*, 13(11):2884–2896, 2018.

[323] Xiaoming Xi, Lu Yang, and Yilong Yin. Learning discriminative binary codes for finger vein recognition. *Pattern Recognition*, 66:26–33, 2017.

[324] Dan Xiao and Jianfeng Hu. Identification of motor imagery eeg signal. In *2010 International Conference on Biomedical Engineering and Computer Science*, pages 1–4. IEEE, 2010.

[325] Cihui Xie and Ajay Kumar. Finger vein identification using convolutional neural network and supervised discrete hashing. *Deep Learning for Biometrics*, pages 109–132, 2017.

[326] Shan Juan Xie, Sook Yoon, Jucheng Yang, Yu Lu, Dong Sun Park, and Bin Zhou. Feature component-based extreme learning machines for finger vein recognition. *Cognitive Computation*, 6(3):446–461, 2014.

[327] Jia Xu, Cui Jianjiang, Xue Dingyu, and Pan Feng. Near infrared vein image acquisition system based on image quality assessment. In *2011 International Conference on Electronics, Communications and Control (ICECC)*, pages 922–925. IEEE, 2011.

[328] Yong Xu, Lunke Fei, and David Zhang. Combining left and right palmprint images for more accurate personal identification. *IEEE transactions on Image Processing*, 24(2):549–559, 2014.

[329] Xuekui Yan, Wenxiong Kang, Feiqi Deng, and Qiuxia Wu. Palm vein recognition based on multi-sampling and feature-level fusion. *Neurocomputing*, 151:798–807, 2015.

[330] Bian Yang, Daniel Hartung, Koen Simoens, and Christoph Busch. Dynamic random projection for biometric template protection. In *4th International Conference on Biometrics: Theory Applications and Systems (BTAS)*, pages 1–7. IEEE, 2010.

[331] Gongping Yang, Xiaoming Xi, and Yilong Yin. Finger vein recognition based on $(2D)^2$PCA and metric learning. *BioMed Research International*, 2012, 2012.

[332] Jinfeng Yang and Jinli Yang. Multi-channel gabor filter design for finger-vein image enhancement. In *2009 Fifth International Conference on Image and Graphics*, pages 87–91. IEEE, 2009.

[333] Lu Yang, Gongping Yang, Yilong Yin, and Xiaoming Xi. Finger vein recognition with anatomy structure analysis. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(8):1892–1905, 2017.

[334] Su Yang and Farzin Deravi. On the usability of electroencephalographic signals for biometric recognition: A survey. *IEEE Transactions on Human-Machine Systems*, 47(6):958–969, 2017.

[335] Su Yang, Farzin Deravi, and Sanaul Hoque. Task sensitivity in eeg biometric recognition. *Pattern Analysis and Applications*, 21(1):105–117, 2018.

[336] Yapeng Ye, Liao Ni, He Zheng, Shilei Liu, Yi Zhu, Deng Zhang, Weilai Xiang, and Wenxin Li. Fvrc2016: The 2nd finger vein recognition competition. In *2016 International Conference on Biometrics (ICB)*, pages 1–6. IEEE, 2016.

[337] Seul-Ki Yeom, Heung-Il Suk, and Seong-Whan Lee. Person authentication from neural activity of face-specific visual self-representation. *Pattern Recognition*, 46(4):1159–1169, 2013.

[338] Yilong Yin, Lili Liu, and Xiwei Sun. Sdumla-hmt: A multimodal biometric database. In *2011 6th Chinese Conference on Biometric Recognition (CCBR)*, pages 260–268. Springer Berlin Heidelberg, 2011.

[339] Aycan Yuksel, Lale Akarun, and Bulent Sankur. Hand vein biometry based on geometry and appearance methods. *IET computer vision*, 5(6):398–406, 2011.

[340] Sergey Zagoruyko and Nikos Komodakis. Learning to compare image patches via convolutional neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4353–4361. IEEE, 2015.

[341] Baochang Zhang, Yongsheng Gao, Sanqiang Zhao, and Jianzhuang Liu. Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor. *IEEE Transactions on Image Processing*, 19(2):533–544, 2010.

[342] David Zhang, Zhenhua Guo, Guangming Lu, Lei Zhang, and Wangmeng Zuo. An online system of multispectral palmprint verification. *IEEE Transactions on Instrumentation and Measurement*, 59(2):480–490, 2009.

[343] Jing Zhang and Jinfeng Yang. Finger-vein image enhancement based on combination of gray-level grouping and circular gabor filter. In *2009 International Conference on Information Engineering and Computer Science*, pages 1–4. IEEE, 2009.

[344] Lin Zhang, Zaixi Cheng, Ying Shen, and Dongqing Wang. Palmprint and palmvein recognition based on dcnn and a new large-scale contactless palmvein dataset. *Symmetry*, 10(4):78, 2018.

[345] Renke Zhang, Di Huang, and Yunhong Wang. Textured detailed graph model for dorsal hand vein recognition: a holistic approach. In *International Conference on Biometrics (ICB)*, pages 1–7. IEEE, 2016.

[346] Yi-Bo Zhang, Qin Li, Jane You, and Prabir Bhattacharya. Palm vein extraction and matching for personal authentication. In *International Conference on Advances in Visual Information Systems*, pages 154–164. Springer, 2007.

[347] B. Zhao, J. Dang, and G. Zhang. Investigation of noun-verb dissociation based on eeg source reconstruction. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2016.

[348] Jianjun Zhao, Hogliang Tian, Weixing Xu, and Xin Li. A new approach to hand vein image enhancement. In *2nd International Conference on Intelligent Computation Technology and Automation*, volume 1, pages 499–501. IEEE, 2009.

[349] Shi Zhao, Yi-Ding Wang, and Yun-Hong Wang. Biometric identification based on low-quality hand vein pattern images. In *2008 International Conference on Machine Learning and Cybernetics (ICMLC)*, volume 2, pages 1172–1177. IEEE, 2008.

[350] Yingbo Zhou and Ajay Kumar. Human identification using palm-vein images. *IEEE Transactions on Information Forensics and Security*, 6(4):1259–1274, 2011.

[351] Yujia Zhou, Yaqin Liu, Qianjin Feng, Feng Yang, Jing Huang, and Yixiao Nie. Palm-vein classification based on principal orientation features. *PLoS One*, 9(11):e112429, 2014.

[352] Danhua Zhu, Jordi Bieger, Gary Garcia Molina, and Ronald M Aarts. A survey of stimulation methods used in ssvep-based bcis. *Computational intelligence and neuroscience*, 2010:1, 2010.

[353] Karel Zuiderveld. Contrast limited adaptive histogram equalization. In Paul S. Heckbert, editor, *Graphics Gems IV*, pages 474–485. 1994.

[354] Jinyu Zuo, N. K. Ratha, and J. H. Connell. Cancelable iris biometric. In *9th International Conference on Pattern Recognition (ICPR)*, pages 1–4, 2008.