ROYAUME DU MAROC
Université Mohammed V - RABAT -
جامعة محمد الخامس – الرباط –
Faculté des sciences
كلية العلوم
CENTRE D'ETUDES DOCTORALES - SCIENCES ET TECHNOLOGIES

N° d'ordre :3434

THESE

En vue de l'obtention du : ***DOCTORAT***

Structure de Recherche : Le Laboratoire de Recherche en Informatique et Télécommunications
Discipline : Sciences de l'ingénieur
Spécialité : Informatique et Télécommunications

Présentée et soutenue le : 27/02/2021 par :

Naima BOUSNINA

Biometric authentication systems' security: Robustness against template database and spoofing attacks

JURY

Mohammed EL HASSOUNI	*PES, Université Mohammed V - Rabat, Faculté des Lettres et des Sciences Humaines*	*Président*
Khalid MINAOUI	*PH, Université Mohammed V - Rabat Faculté des Sciences*	*Directeur de thèse*
Mounia MIKRAM	*PH, Université Mohammed V - Rabat, Ecole des Sciences de l'Information*	*Codirecteur de thèse*
Sanaa GHOUZALI	*PH, Université du Roi Saoud - Riyad, Faculté des sciences de l'informatique et de l'information*	*Codirecteur de thèse*
Chouaib MOUJAHDI	*PH, Université Mohammed V - Rabat, Institut Scientifique*	*Rapporteur/Examinateur*
Mohamed El HAZITI	*PES, Université Mohammed V - Rabat, Ecole Supérieure de Technologie*	*Rapporteur/Examinateur*
Zytoune OUADOUDI	*PH, Université Ibn Tofail – Kénitra, Ecole Nationale des Sciences Appliquées*	*Rapporteur/Examinateur*

Année Universitaire : 2020/2021

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

وَأَنْ لَيْسَ لِلْإِنْسَانِ إِلَّا مَا سَعَى* وَأَنَّ سَعْيَهُ سَوْفَ يُرَى* ثُمَّ يُجْزَاهُ الْجَزَاءَ الْأَوْفَى.

“And that man shall have nothing but what he strives for – And that his striving shall soon be seen – Then shall he be rewarded for it with the fullest reward”

Sura AN-NAJM (39-41)

To my mother, ...

Acknowledgements

The research work presented in this dissertation has been carried out at the Research Laboratory of Computer Science and Telecommunications (LRIT), Mohammed V University - Rabat, Faculty of Sciences, Morocco. The thesis is accomplished under the direction of Mr. Khalid **MINAOUI**, PH at Mohammed V University - Rabat, Faculty of Sciences and Co-direction with Mme. Mounia **MIKRAM**, PH at Mohammed V University - Rabat, School of Information Sciences and Mme. Sanaa **GHOUZALI**, PH at King Saud University, College of Computer and Information Sciences, Riyadh.

First of all, I would like to express my gratitude to my thesis director Mr. Khalid **MINAOUI**, PH at Mohammed V University - Rabat, Faculty of Sciences for the effort provided and the willingness to assist any time I need.

I would also like to thank Mme. Mounia **MIKRAM**, PH at Mohammed V University - Rabat, School of Information Sciences, for the advice, and support since my master final project. Her comments and questions in any new proposed contribution were very valuable in improving and valuating my research work.

I would like to express my recognition to Mme. Sanaa **GHOUZALI**, PH at King Saud University, College of Computer and Information Sciences, Riyadh, for her support and proficient guidance throughout all stages of the thesis. Even that the supervision has been conducted remotely, she demonstrated that distances do not matter when we have a strong intention of hard work.

I would like to document my immense gratitude for Mr. Mohammed **EL HASSOUNI**, PES at Mohammed V University - Rabat, Faculty of Letters and Human Sciences for the honor he did by approving to be President of my thesis jury.

A sincere thank you to Mr. Chouaib **MOUJAHDI**, PH at Mohammed V University - Rabat, Scientific Institute, Reviewer and Examiner of my thesis, for the interest that he carries to this work and for the interesting remarks he brings.

I am very honored to thank Mr. Mohamed **El HAZITI**, PES at Mohammed V University - Rabat, High School of Technology, Salé, Reviewer and Examiner of this thesis, for agreeing to evaluate my work and reading my manuscript thoroughly in order to bring its valuable advice.

A sincere thank you to Mr. Zytoune **OUADOUDI**, PH at Ibn Tofail University, National School of Applied Sciences, kénitra, Reviewer and Examiner of my thesis, for the interest that he carries to this work and for the interesting remarks he brings.

My special thanks go to Dr. Lilei **ZHENG** for his guidance during my last research contribution. I appreciate his discussions and interpretation of the results given in the contribution. I learned from his insight the passion for scientific research.

Finally, I would like to thank my family to whom I owe a great deal. To my mother for her unconditional love and sacrifice. To my sister Asmae, thank you for your eternal support. To my sisters Fatima Zahra and Aziza for your help and understanding. I hope this dissertation proves my love and gratitude.

Abstract

As the conventional token-based or knowledge-based personal identification techniques have become unable to satisfy higher security levels, biometric authentication systems have been suggested as an alternative solution. Despite the advantages provided by this kind of authentication mechanisms, they are not entirely protected against diverse types of attacks, namely (1) *Attacks on the user sensor*, (2) *Attacks on the interface between modules*, (3) *Attacks on software modules*, and (4) *Attacks on the system's database*.

The presented thesis points up the biometric authentication systems' resistance against the user sensor and system's database attacks. More precisely, the contributions of this dissertation are grouped into two main axes. In the first axis, we intended to design secure frameworks for biometric template protection by enhancing the preliminary presented state-of-the-art methods. Therefore, this axis consists of three contributions, including a spatial fingerprint-face-based watermarking scheme and two hybrid approaches for face and fingerprint template protection. The first hybrid approach combines a Dual-Tree Complex Wavelet Transform-Discrete Cosine Transform (DTCWT-DCT)-based watermarking algorithm and the partial Hadamard transform, while the second fuses the same watermarking algorithm, a secure sketch algorithm, and a 3D chaotic map image encryption method. The second axis presents an analytical study on a differential evolution-based adversarial attack to deep-learning-based face liveness detection models, in order to highlight practical criteria that can be used in the development of countermeasures to address face-spoofing issues.

Keywords: Biometric authentication, Biometric watermarking, Biometric template protection, Face anti-spoofing, Face spoofing, Convolutional neural networks.

Résumé

Les systèmes d'authentification biométriques ont été proposé vu que les systèmes d'authentification traditionnels sont devenus incapable de rivaliser les niveaux élevés de la sécurité dus à l'avancement technologique. En dépit des bénéfices offerts par les systèmes d'authentification biométriques, ils restent toujours faibles à l'égard de divers types d'attaques, à savoir, (1) *les attaques contre le capteur du système*, (2) *les attaques contre l'interface entre les modules du système*, (3) *les attaques contre les modules* et (4) *les attaques contre la base des données du système.*

Cette thèse se focalise sur la résistance des systèmes d'authentification biométriques aux attaques contre le capteur et la base des données du système. Par conséquent, les contributions apportées sont regroupées en deux axes principaux. Le premier axe fait le point sur la protection des modèles biométriques – visage et empreinte digitale – contre les attaques de la base des données. Dans cette contribution, trois approches de protection ont été proposé, une méthode de tatouage spatial et deux approaches hybrides. La première approche hybrid combine l'algorithme de tatouage transformée en ondelette complexe à double arbre-transformée en cosinus discret (DTCWT-DCT) et le transformée d'Hadamard partiel. Ainsi que la deuxième approche hybride fusionne la même méthode de tatouage, le secure sketch, et une méthode de cryptage de l'image basée sur la carte chaotique 3D. Le deuxième axe présente une étude analytique des attaques adverses aux mécanismes d'anti-spoofing de visage, en se basant sur l'apprentissage en profondeur. Le but de cette étude est l'évaluation du comportement de ce type de mécanisme d'anti-spoofing à l'égard des attaques adverses. De plus, ressortir les critères pratiques qui peuvent être pris en considération pour améliorer la performance des systèmes de détection de la vivacité du visage.

Mots clés: Authentification biométrique, Tatouage biométrique, Protection des modèles biométriques, Spoofing et anti-spoofing de visage, Réseaux de neurones convolutifs.

Résumé étendu

Le développement rapide d'internet et de la technologie numérique nous a fait vivre à l'ère de l'information, où des milliards d'informations numériques sont créées et livrées en une fraction de seconde. Ce phénomène a surgi de nombreux problèmes comme la protection de l'identification et la vérification de la propriété. En conséquence, les systèmes d'authentification traditionnels sont devenus incapables de rivaliser les niveaux élevés de la sécurité. Ainsi, les systèmes d'authentification biométriques ont été proposés comme une solution alternative.

Les systèmes d'authentification biométriques identifient l'individu en se basant sur ses caractéristiques physiologiques à savoir l'empreinte digitale, le visage, l'empreinte de la paume ou comportementales à savoir la parole, l'écriture manuscrite. Ce type de système a été introduit pour distinguer entre un utilisateur authentique et un utilisateur imposteur en fonction de l'identité de cet utilisateur, plutôt que de ce qu'il possède ou de ce dont il se souvient. Ce qui constitue un outil puissant de vérification de l'identité des individus. En dépit des bénéfices offerts par les systèmes d'authentification biométriques, ils restent toujours faibles à l'égard de divers types d'attaques: (1) *les attaques contre le capteur du système*, (2) *les attaques contre l'interface entre les modules du système*, (3) *les attaques contre les modules* et (4) *les attaques contre la base de données du système*. Cette thèse étudie la résistance des systèmes d'authentification biométriques par rapport aux attaques contre le capteur (spoofing) et la base de données du système.

Dans un premier lieu, et avant d'entamer les contributions apportées dans la thèse, le **premier chapitre** présente la portée de la thèse et clarifie le contexte du travail et la perspective suivie. Le **deuxième chapitre** expose l'état de l'art des systèmes d'authentification biométriques et les attaques contre ce type de systèmes, avec un accent particulier sur les attaques contre la base de données du système. Ainsi, il introduit les métriques d'évaluation de la performance des approches proposées et les bases de données utilisées.

Les **troisième**, **quatrième**, et **cinquième chapitres** font le point sur la protection des modèles biométriques - visage et empreinte digitale - contre les attaques de la base de données. Pour cette raison, la recherche s'est focalisée sur une étude de l'ensemble des techniques de l'état de l'art pour analyser leurs forces et leurs faiblesses. Par suite, une

fusion des méthodes pertinentes est faite pour unir les avantages de chacune et pallier ainsi leurs limites. Généralement, une méthode de protection des modèles biométriques est dite efficace s'elle satisfait les quatre caractéristiques: la diversité, la révocabilité, la sécurité, et la performance.

Dans le **troisième chapitre**, nous avons proposé une méthode spatiale de tatouage numérique pour sécuriser l'authenticité et améliorer la performance des systèmes d'authentification multimodaux. La méthode se base sur l'intégration des caractéristiques biométriques du visage (la marque) dans l'image de l'empreinte digitale via une clé secrète. La clé secrète est utilisée pour localiser les pixels à tatouer tout en préservant les points de minutiae de l'empreinte digitale. Ainsi que, les caractéristiques du visage sont extraites en utilisant la technique "Orthogonal Locality Preserving Projections (OLPP)". En plus de l'utilisation de l'empreinte digitale lors de l'authentification, la méthode permet aussi de récupérer et d'utiliser la marque pour établir l'authenticité de l'utilisateur.

Bien que les méthodes spatiales de tatouage numériques puissent masquer potentiellement les données biométriques dans l'image hôte, elles restent toujours vulnérables aux attaques de fréquences. En outre, aucune méthode ne satisfait la totalité des propriétés requises de la protection des modèles biométriques, à savoir révocabilité, diversité, sécurité et performances. Par conséquent, les méthodes de tatouage fréquentiel et les approches hybrides ont été proposées.

Dans le **chapitre 4**, nous avons suggéré une approche hybride pour sécuriser les modèles biométriques dans les systèmes d'authentification multimodaux. le concept de l'approche proposée est de combiner la méthode de tatouage fréquentiel "Dual-Tree Complex Wavelet Transform - Discrete Cosine Transform (DTCWT-DCT)" et la transformée d'Hadamard partielle pour satisfaire les quatre propriétés de la protection des modèles biométriques. Plus précisément, la transformée partielle d'Hadamard est utilisée pour couvrir les caractéristiques des empreintes digitales. Ensuite, les vecteurs des caractéristiques obtenus sont binarisés et intégrés dans l'image de visage via la méthode de tatouage DTCWT-DCT. Les résultats expérimentaux ont démontré la puissance de l'approche aux attaques contre la base des données du système.

Une deuxième approche hybride est proposée dans le **chapitre 5**. Cette approche consiste à incorporer la méthode de tatouage DTCWT-DCT, l'algorithme de secure sketch, et une méthode de cryptage de l'image basée sur la carte chaotique 3D. Le processus de l'approche adhère aux trois étapes: au début, le secure sketch est basé pour dissimuler les caractéristiques biométriques de l'empreinte digitale. Vu que le secure sketch est irrévocable et incapable de modéliser les variations intra-utilisateur, la méthode de tatouage DTCWT-DCT est utilisée pour intégrer les caractéristiques sécurisées de l'empreinte digitale (la marque) dans l'image du visage. Après, la méthode de cryptage est basée pour protéger l'image de visage tatouée avant d'être enregistrée dans la base de données

du système. Les résultats expérimentaux ont démontré l'efficacité de l'approche hybride en terme de satisfactions des quatre propriétés requises pour la protection des modèles biométriques.

Néanmoins, à mesure que le grand public devient familier avec la technologie biométrique, cette dernière devient fragile face aux attaques de spoofing (attaques de présentation), où l'attaqueur obtient les données biométriques de l'utilisateur authentique à partir des réseaux sociaux pour produire des modèles sophistiquées dans le but d'égarer le système d'authentification. Ceci augmente le challenge de ces systèmes non seulement pour faire face aux attaques contre la base des données, mais aussi dans la capacité de distinguer entre les vrais et les faux utilisateurs présentés devant le capteur du système.

Dans ce contexte, le **chapitre 6** expose l'état de l'art des mécanismes d'anti-spoofing des faux visages basées sur les réseaux des neurones convolutifs ainsi que les attaques adverses contre ces mécanismes. De plus, il aborde les métriques d'évaluation de performance et les bases de données utilisées.

Le **chapitre 7** introduit une étude analytique des attaques de spoofing aux mécanismes d'anti-spoofing du visage. Où nous avons étudié le compromis entre une attaque adverse basée sur l'algorithme d'évolution différentielle et un classifieur d'anti-spoofing du visage qui se base sur les réseaux des neurones convolutifs. L'étude est menée sous différents scénarios de cas d'utilisation et différentes bases de données d'anti-spoofing du visage. Le but de cette étude est d'évaluer le comportement des classifieurs d'anti-spoofing à l'égard des attaques adverses. De plus, ressortir les critères pratiques qui peuvent être pris en considération pour développer des contre-mesures et améliorer la performance des systèmes de détection de la vivacité du visage.

Pour terminer, le **chapitre 8** conclut la thèse et discute les perspectives des futurs travaux de recherche.

Contents

List of Figures

List of Tables

List of Abbreviations

FRR False Reject Rate

FAR False Acceptance Rate

EER Equal Error Rate

GAR Genuine Accept Rate

ROC Receiver Operating Characteristic

LSB Least Significant Bit

SVD Singular Value Decomposition

FFT Fast Fourier Transform

DFT Discrete Fourier Transform

DCT Discrete Cosine Transform

DWT Discrete Wavelet Transform

JPEG Acronyme de Joint Photographic Experts Group

MRICA Multi-Resolution by Independent Component Analysis

RDWT Redundant Discrete Wavelet Transform

HVS Human Visual System

BCH Bose-Chaudhuri-Hocquenghem

LDPC Low-Density Parity-Check

RSA Rivest-Shamir-Adleman

AES Advanced Encryption Standard

1D One-Dimensional

2D Two-Dimensional

3D Three-Dimensional

DPFrFT Dual Parameter Fractional Fourier Transform

DNA Deoxyribonucleic Acid

MFCC Mel Frequency Cepstral Coefficients

LPC Linear Predictive Coding

DWPT Dual Wavelet Packet Transform

DTCWPT Dual Tree Complex Wavelet Packet Transform

WPT Wavelet Packet Transform

SVM Support Vector Machine

KNN K-Nearest Neighbors

ELM Extreme Learning Machine

PSNR Peak Signal to Noise Ratio

dB decibels

FVC Fingerprint Verification Competition

ORL Olivetti Research Laboratory

OLPP Orthogonal Locality Preserving Projections

ROI Region Of Interest

PAN Performance Anchored Normalization

CPCA Column Principal Component Analysis

TRR True Rejection Rate

IBG International Biometric Group

ICAO International Civil Aviation Organization

DoG Difference Of Gaussian

LBP Local Binary Patterns

LDA Linear Discriminant Analysis

SIFT Scale-Invariant Feature Transform

SURF Speeded Up Robust Features

HOG Histogram of Oriented Gradients

OFM Optical Flow Maps

HSV Hue Saturation Value

GAN Generative Adversarial Network

LSTM Long Short-Term Memory

CVPR Conference on Computer Vision and Pattern Recognition

R-CNN Regions with Convolutional Neural Networks

AGNs Adversarial Generative Nets

DNN Deep Neural Network

LEDs Light-Emitting Diode

HTER Half Total Error Rate

UIDAI Unique Identification Authority of India

FASNet Face Anti-Spoofing Network

SGD Stochastic Gradient Descent

ReLU Rectifier Linear Unit

RGB (Red, Green, Blue)

MTCNN Multi-task Convolutional neural network

DA Data Augmentation

DTCWT Dual-Tree Complex Wavelet Transform

CNN Convolutional Neural Networks

ATM Automated Teller Machine

1 Introduction

1 Background and Motivation

With the rapid advent of the internet and digital technology, billions of digital information are created and delivered in every fraction of a second. This phenomenon reveals many challenges in the ownership verification and identification fields, such as the increase of security in the automatic personal identification and verification systems (e,g., electronic banking, border and access control). As the conventional token-based (passport, ID card ...) or knowledge-based (password, PIN code ...) personal identification techniques become unable to meet the requirements of a high level of security, biometric authentication systems have been suggested as an alternative solution. In contrast to the traditional personal authentication systems, biometric systems focus on identifying people founded on physiological (e.g., fingerprint, face, palm print) or behavioural (e.g., speech, handwriting) characteristics and binding individual at a much deeper level.

Biometric refers to the behavioural or physical human traits based to automatically identify individuals in access points. It has been deployed in a wide range of authentication applications because it puts forward various advantages over conventional authentication systems. More precisely, as ID cards and passwords can be readily shared, manipulated, stolen, or lost, biometric is presented to distinguish between a genuine and an impostor user based on who this user is, rather than what he/she owns or what he/she remembers. Which constitutes a potent tool of individuals identity verification with tremendous potential consequences.

Even though biometric systems outperform the conventional authentication systems, they are not entirely protected against diverse attacks [1, 2]. Generally, these attacks can be classified into four classes [2]: 1) *Attacks on the user interface*, where fake biometric data is presented in front of the sensor (e,g., mimicry/spoofing attacks). 2) *Attacks on the interface between modules*, where the attacker interposes the communication interfaces between modules (e,g., replay/hill-climbing attacks). 3) *Attacks on software modules*,

where the attacker modifies the executable program via Trojan-horse to get the desired information. And finally, 4) *Attack on the system's database*, herein, the attacker gets illegal access to the system's database, then replaces or usurps the biometric templates.

One of the most harmful attacks is the one against the authentication system database, where the individual biometric templates can be manipulated by the attacker for illegitimate access [2]. These templates can be tampered via four scenarios, namely, 1) Replacing the genuine template by the impostor one. 2) Removing the existing templates from the database. 3) Adding new templates to the database or even 4) Creating physical spoof from the templates to be replayed to the system sensor, as well as other systems basing on the same biometric data. Moreover, dissimilar to the forgotten or stolen passwords that can be substituted, the tampered templates are not revocable, which means even legitimate users would not be able to use the compromised models to identify themselves. To overcome the template database attacks, it becomes necessary to implement efficient solutions to deny illegitimate access to any information stored in the systems databases. To accomplish this, the perfect solution would be to secure biometric data before being stored in the database.

Unfortunately, as biometric authentication systems become familiar to the general public, their feebleness starts being known. Unlike passwords, biometric modalities, (e,g., fingerprints, retina, face or even Deoxyribonucleic Acid (DNA)) are vastly available, which makes them readily acquired and duplicated. Where attackers can get this biometric data from social networks or stolen smart-phones to produce high quality falsified models to be presented to the sensors to deceive the authentication systems. For example, in 2009, the security and vulnerability research team of Hanoi university demonstrated at Black Hat 2009 conference (the world's premier technical security conference) the mechanism of easily fooling the Lenovo's Veriface III, Asus' SmartLogon V1.0.0005, and Toshiba's Face Recognition 2.0.2.32 using fake facial images of the genuine users, to gain illegal access to the laptops. Also, two days after releasing the iPhone 5s with a Touch ID fingerprint sensor, by Apple Inc. It was hacked by a German hacker collective, chaos computer club, through a gummy finger. This type of attacks is referred to as spoofing or presentation attacks and is easy to generate with high potential to succeed. This increases the challenge of biometric authentication systems by dealing with not only systems database attacks, but also the ability to differentiate between live and fake users. Therefore, it becomes mandatory to put down typical countermeasures by building liveness detection models to uncover physiological signs of life and improve the security of biometric authentication systems.

Currently, spoofing and system database attacks are one of the main issues for organizations willing to market identity control solutions based on biometric technologies. The existing biometric authentication systems are starting to be mature enough for the real world and practical applications. However, the continuous improvement of tricky spoofing attacks and system database attacks benefiting from advanced technologies

prevents these systems from achieving their breakthrough. Thus, there is an urgently necessary to take countermeasures by finding efficient and reliable solutions for detecting and circumventing such attacks.

2 Contributions of the Thesis

First of all, it is worth mentioning that absolute security does not exist. Given the right circumstances, the appropriate technology, the funding, and the time, almost any security system can be broken. However, this does not mean that the security community of such systems succumbs to security threats. The main goal of the security community is to protect the resources needed by the attacker to compromise privacy. In this context, this dissertation addresses the biometric authentication systems' security, with a particular focus on robustness against template database attacks and presentation attacks. This choice is based on two facts: first, template database is the most critical part of biometric authentication systems, owing to, it comprises a set of salient traits that summarize the privacy and uniqueness of the genuine users. Second, the user interface is the unique point by which individuals can directly interact with the authentication mechanism, which provides more facilities to the intruder to gain unauthorized access.

The research work presented in this thesis is divided into two main objectives. The former concerns the security within the storing of face and fingerprint modalities to ensure that even in the case the attacker compromises the stored templates, it nevertheless will be possible to retrieve the original data. More precisely, the concept underlying this objective is to secure the aforementioned biometric modalities before being stored in the database. To accomplish this, the research based on extensive literature survey for the existing biometric template protection approaches to analyze their robustness and limitations. Founded on the results of the literature survey, the best and most effective set of approaches have been combined to satisfy the four requirements of biometric template protection, including security, diversity, revocability, and performance.

The latter objective concerns a sort of attacks on the user interface, that is spoofing attack. Specifically, we intend to present an analytical study of a differential evolution-based adversarial attack against CNNs-based face liveness detection systems. Herein, we analyzed the trade-off between the aforementioned spoofing and anti-spoofing algorithms under different threat models and use-case scenarios. This study aims at assessing the impact of various criteria related to both mechanisms and highlights common misconceptions to derive complementary countermeasures that aid in constructing more resistant anti-spoofing frameworks to a wide variety of adversarial attacks.

The principal contributions of the thesis are listed below:

- A spatial fingerprint-face-based watermarking scheme is presented to secure authenticity and improve the performance of multimodal authentication systems.
- A hybrid approach for face and fingerprint template protection is proposed by combining a DTCWT-DCT-based watermarking algorithm and the partial Hadamard transform to provide the four required properties of biometric template protection, including revocability, diversity, security and, performance.
- Another new hybrid approach that fuses the same DTCWT-DCT-based watermarking algorithm, a secure sketch algorithm, and a 3D chaotic map image encryption method is introduced for face and fingerprint template protection by meeting the four requirements of biometric template protection.
- A detailed analytical study of a differential evolution-based adversarial attack on CNNs-based face liveness detection frameworks is introduced, to gain insight into the behaviour of anti-spoofing models when confronted with spoofing attacks.

3 Outline of the Thesis

The thesis is organized as follows:

chapter 1 introduces the scope of the thesis, clarifies the context of the work, and the perspective followed. With a presentation of the thesis contributions, as well as a summary of the original papers.

chapter 2 reviews the stat-of-the-art of the attacks on the biometric authentication systems, with a particular focus on template database attacks. Besides, it introduces performance evaluation techniques and based databases.

chapter 3 addresses a face-fingerprint-based spatial watermarking scheme for biometric authentication systems.

chapter 4 and chapter 5 propose two hybrid approaches for face and fingerprint template protection. The former bases on a DTCWT-DCT-based watermarking algorithm and the partial Hadamard transform. While the latter combines the same watermarking method, secure sketch, and a 3D chaotic image encryption method to meet the four requirements of biometric template protection.

chapter 6 gives an overview of deep face anti-spoofing. Herein, a detailed description of CNNs-based face anti-spoofing techniques and adversarial attacks to these techniques is presented. Besides, it introduces the performance evaluation metrics and based databases.

The analytical study of the differential evolution-based adversarial attack against a CNN-based face liveness detection classifier is described in chapter 7.

Finally, chapter 8 concludes the thesis and discusses the directions and perspectives for future research works.

4 List of Publications

International Journals

- **Naima Bousnina**, Lilei Zheng, Mounia Mikram, Sanaa Ghouzali, and Khalid Minaoui. Unraveling robustness of deep face anti-spoofing models against pixel attacks. Multimedia Tools and Applications, 2020, DOI: https://doi.org/10.1007/s11042-020-10041-1

- **Naima Bousnina**, Sanaa Ghouzali, Mounia Mikram, Maryam Lafkih, Ohoud Nafea, Wadood Abdul, and Muna Al-Razgan. Hybrid Multimodal Biometric Template Protection. Intelligent Automation & Soft Computing, vol. 27, no.1, pp. 35–51, 2021.

- **Naima Bousnina**, Sanaa Ghouzali, Mounia Mikram, Abdul Wadood, and Khalid Minaoui. DTCWT-DCT Based Watermarking and Hadamard Transform for Biometric Template Protection. Journal of Computing Science and Engineering, submitted.

International Conferences

- **Naima Bousnina**, Sanaa Ghouzali, Maryam Lafkih, Ohoud Nafea, Mounia Mikram, Wadood Abdul, and Driss Aboutajdine. Watermarking for protected fingerprint authentication. 12th International Conference on Innovations in Information Technology (IIT) (1–5), 2016.

- **Naima Bousnina**, Sanaa Ghouzali, Mounia Mikram, Wadood Abdul. DTCWT-DCT watermarking method for multimodal biometric authentication. 2nd International Conference on Networking, Information Systems & Security (1–7), 2019.

2 Biometric Authentication: A State-Of-The-Art

1 Biometric Authentication Systems

Biometrics is the science of setting up the individual identity based on humans' unique physical (e,g., face, iris, fingerprint, hand geometry, retina) or behavioural (e,g., gait, voice, signature, keystroke) characteristics [3]. It has been applied in law enforcement for criminals identification during the mid of the 19th century. Since then, it has been adopted in diverse civilian applications for individuals identification and recognition, such as Automated Teller Machine (ATM) login, mobile device authentication, network access, driving license, government ID card, border control, and military [4]. Contrary to the conventional authentication systems, biometric technology has proven its ability to secure sensitive data by gaining access to only authorized persons [5].

Typically, biometric authentication systems consist of four basic modules, including *(1) Sensor module*, *(2) Feature extractor module*, *(3) Matcher module*, and *(4) Decision module*. The *sensor module* is conceived to scan the input data. Subsequently, the captured data is transmitted to the *feature extractor* module where the relevant features are extracted optimally by basically removing the noise from the input sample. Next, the obtained features are transferred to the succeeding module, which is the *matcher module*. In this stage, a comparison is carried out between the current features and the template features stored in the database to produce a matching score. Finally, the resulting matching score is transferred to the *decision module* to accept or deny the presented individual.

Generally, biometric authentication is carried out via two stages, specified as *Enrollment* and *Authentication*. During the enrollment stage, the biometric data is acquired from the user using the sensor module. Then the relevant information is extracted using the extractor module and stored in the database as a reference model. While in the authentication stage, the features of the query biometric data are extracted and compared with the reference model.

Biometric authentication can be processed using *Identification* or *Verification* modes. The purpose of the *identification* system is to answer the question: *"Who are you?"* by comparing the individual biometric data with the whole templates stored in the system database. This process is referred to as one-to-n matching. Law enforcement and border control authentication are examples of identification mode. On the other hand, the *verification* answers the question *"Are you whom you say you are?"*. Herein, the identity is proved by verifying whether the claimed user is genuine or not. Thus, the input biometric data is contrasted against a formerly collected template of the user, which is known in biometric terms as one-to-one matching. This mode is usually, based in mobile banking access management and multi-factor authentication systems for accessing a secure server.

Typically, most of the currently used biometric systems rely on single biometric modalities to establish the identity (uni-biometric). However, as these systems are deployed in sensitive real-world applications, it becomes necessary to understand their limitations for efficient and secure authentication. The uni-biometric systems are encountered by a set of challenges and problems such as [3]:

- *Noised Acquired Data:* The biometric data acquired via sensors may be compromised by noise caused by inappropriate acquisition circumstances, such as unsuitable lighting, power shortages, or physiology/health status of users. Which causes falsely classifying a genuine user as an impostor, therefore rising the False Reject Rate (FRR) of the system.
- *Lack Of Universality (limited population coverage):* All humans have biometric characteristics, but it may happen that some individuals can not physically present a standalone biometric credential due to illness or disabilities.
- *Distinctiveness Ability:* Some biometric traits are less variable than others. For example, biometric data based on hand geometry is considered less distinctive than fingerprint data.
- *Intra-Class Variations:* The biometric data earned from the legitimate user in the authentication stage, maybe very dissimilar from the biometric data based during the enrollment stage to produce the reference templates. Thereby affecting the matching process.

To overcome such limitations, one idea would be to base on multi-biometric systems that merge redoubled provenance of biometric data. The multi-biometrics can be achieved using one of the following scenarios:

- *Multi-Sensor Systems:* Herein, a single biometric trait is taken from multiple sensors, then combined to form a composite biometric feature. Owing to, using various

sensors results in gaining complementary information that improves the identification capacity of the system.

- *Multi-Algorithm Systems:* This can be fulfilled using various feature extraction or/and matching algorithms on the same biometric information to enhance the matching performance.
- *Multi-Instance Systems:* Herein, multiple instances of the same biometric data is based. For example, using the left and right irises or the left and right index fingers of the user.
- *Multi-Modal Systems:* In this case, the authentication is set up via different biometric modalities of the same individual. For instance, combining fingerprint and iris modalities to establish the identity.

Moreover, based on the type of information availability, four levels of fusion can be defined, including:

- *Fusion at Sensors Level:* Herein, the biometric data taken from multiple sensors are combined to form a composite biometric trait.
- *Fusion at Extraction Level:* Herein, the biometric characteristics are first preprocessed, and features are extracted separately via specific algorithms. Next, the obtained vectors are combined to composite a feature vector.
- *Fusion at Matching Score Level:* At this level, rather than combining the feature vectors, they are processed separately and individual matching is carried out. Then, depending on the accuracy of each matching score, the classification is performed.
- *Fusion at Decision Level:* Herein, the final recognition decision is fulfilled by combining the individual matching scores via specific methods such as majority voting [6], Bayesian decision fusion [7], and weighted majority voting [8].

2 Attacks Against Biometric Authentication Systems

Even though biometric systems offer several advantages over traditional authentication systems, they are also subject to various vulnerabilities. Authors in [9] divided biometric systems' weakness into four types, including:

- *Circumvention:* The attacker gain access to the sensitive components to ruin the protected biometric system, such as substituting templates stored in the database or revoking matcher decision.

- *Covert Acquisition:* Herein, an unauthorized person exploits the biometric data caught from the legal user to gain illegitimate access. For instance, lifting latent fingerprints to produce falsified templates or playing back a voice password or face video in front of the sensors.
- *Collusion and Coercion:* In this type of vulnerabilities, the attack is due to the legitimate user. The Collusion comes from an authorized person who has the power to change system settings by giving facilities to gain access to the hacker(perhaps by bribe). The Coercion happens when the legal user is forced (through blackmail or physical threat) to gain access to the attacker.
- *Denial of Service:* Herein, the attacker denies the legitimate user's access by degrading the performance, stopping, or slowing the system.

Based on the system's user, another category of vulnerabilities can be defined [10]. These vulnerabilities are *(1) Biometrics are not secret*, which means that our biometric traits are often all over the place and can be caught by anyone around us. *(2) Biometrics cannot be revoked.* In other words, compromising the biometric data of an individual will breach all applications that base on the same biometric template, which blocks the individual from re-enrolling. *(3) Biometrics have secondary uses*, using the same biometric information in different applications may cause user tracking by exploiting this information in multiple threat environments.

As illustrated in figure 2.1, Ratha *et al.* [11] have defined eight levels of biometric systems' attacks. Jain *et al.* summarized them into four classes [2]. The first class represents `the attacks on the user interface`, where falsified biometric data can be presented to the sensor of the systems (Level 1). Spoofing and mimicry are examples of these attacks. The second class is `the attacks on the interface between modules` where an attacker can either destroy or interpose the communication interfaces between modules. Replay and hill-climbing are examples of this type of attack (Levels 2, 4, 7). The third class indicates `the attacks on software modules` where the attacker can change the executable program using Trojan-horse to return the desired results (Levels 3 and 5). The Last category is `the attack on the system's database` where the attacker can have illegitimate access to the users' biometric templates (Level 6).

The template database is the most critical part of biometric authentication systems because it contains the set of salient traits that summarize the privacy and uniqueness of the individuals. Therefore, attacks on the template database are one of the most harmful attacks, where biometric templates can be stolen for illegitimate access. In such an attack, the biometric data can be manipulated, replaced, or stolen to produce artificial modalities. Unlike a stolen password that can be replaced, the stolen template is not revocable. This would mean that legitimate users would not be able to use compromised models to authenticate themselves [11]. Moreover, the templates may be changed to get

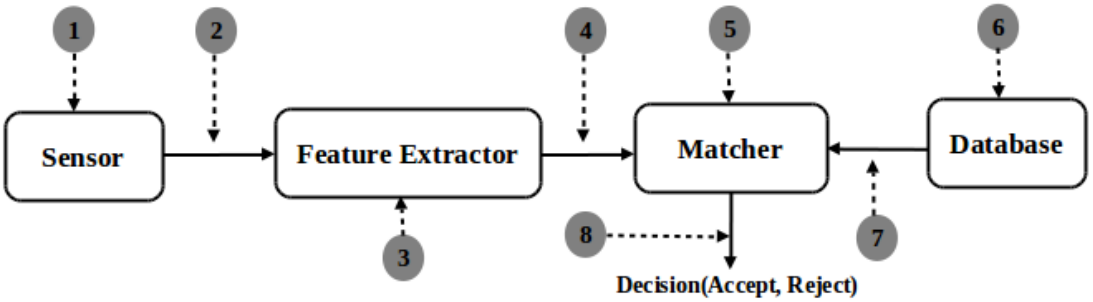


FIGURE 2.1: Biometric system attacks levels.

a high verification score, independently of the biometric data presented to the sensor of the system. This brings the system down by making the score low for legal users. Hence, to face up template database attacks, one approach would be to not store the biometric data in plain text form.

3 Biometric Template Protection

3.1 Biometric Template Protection Techniques

To provide a higher level of security against template database attacks, four requirements should be met, including [2]:

- *Diversity:* This signifies that cross-matching between databases is impossible.
- *Revocability:* Refers to the ability to cancel a template that has been breached and create a new one.
- *Security:* The extraction of an original template from a stored template should be computationally difficult.
- *Performance:* The use of a protected template must not negatively affect the system's performance.

To satisfy these requirements, two types of approaches are proposed in the literature, namely *feature transformations* and *biometric cryptosystems* [2].

3.2 Feature Transformations

Using feature transformations a transformation function F is applied during the authentication phase on the biometric data T using a specific secret key K to generate a reference template F(K,T). The reference F(K,T) is stored in the system's database

instead of the original biometric features. In the authentication phase, the same transformation F is carried out on the acquired biometric traits of the query template Q using the same key K. Next, the results of the transformation F(K,Q) is matched with the stored reference template to allow or deny the user's access.

Based on the type of the transformation function, feature transformation methods are classified into two categories, that are *bio-hashing (salting)* and *non-invertible* transformations.

3.2.1 Bio-Hashing Schemes

Bio-hashing algorithm is one of the most popular techniques that bases on biometric information salting. Herein, the function F is considered invertible. Which means, in case the attacker gains access to the transformed template and the key, the original biometric data can be recovered. Hence the security of bio-hashing techniques relies on the secrecy of the used key.

3.2.2 Non-Invertible Schemes

In the non-invertible transformation methods, a one-way transformation function is applied to the templates. Thus, it is computationally hard to recover the original biometric template, even when the secret key and/or the transformed templates are comprised.

3.3 Biometric Cryptosystems

Biometric cryptosystem techniques are designed to get the benefits of associating biometric features and cryptographic keys [12, 13]. It can also refer to generating a digital key from biometrics [14] to provide solutions to biometric-dependent key-release and secure biometric templates [15]. The biometric information obtained after the combination is named `"Helper data"` which is stored as a reference in the database. Depending on the extraction mechanism of the `helper data`, biometric cryptosystems are classified into *key binding* and *key generation* systems.

3.3.1 Key Binding Schemes

Herein, the helper data is generated by linking a secret key to the biometric template [2]. In this type of schemes, the secured templates are revocable, owing to, they are independent of the secret key. *Fuzzy Commitment* [16] and *Fuzzy Vault* [17] are examples of the key binding biometric cryptosystems schemes. In *Fuzzy Commitment*, the helper data is generated by combining the biometric features and a secret key. In the

authentication stage, the secret key is regenerated using the helper data and the biometric information of the query template. In *Fuzzy Vault* schemes, the biometric traits and the secret key are used to produce a polynomial p. Next, false points are added to the result to construct a vault V to be stored in the database. During the authentication stage, the access is gained when the secret key is built using the vault V and biometric characteristics of the query template.

3.3.2 Key Generation Schemes

Herein, the helper data is derived directly from the biometric data [18]. Besides, the storage of the helper data is not obligatory for the majority of the key generation techniques. *Fuzzy extractors* [19] and *secure sketches* [18, 20–22] are examples of the key generation schemes. In the fuzzy extractor, a uniformly random string is taken out from the biometric template, while the helper data is based in applying the reconstruction. By contrast, in the secure sketch, the helper data is applied to recover the original biometric template.

4 Alternative Information Hiding Techniques: Watermarking, Secure Sketch, Image Encryption

In [23], the authors provided a classification of existing biometric information hiding approaches where biometrics are used in combination with information hiding either to provide secure biometric authentication or to protect digital media ownership. They have concluded that when developing biometric information hiding approaches, attention should be given to feature discriminability of the biometric host data rather than the visual quality and the feature selection of biometric watermark generation for the limited capacity.

4.1 Watermarking Based Biometric Template Protection

The digital image watermarking has been proposed to solve privacy issues as well as security in many applications. Additionally, it has been classified among the most effective methods to avoid malicious and unintended data compromise. Besides, it has been used to protect copyrights and confirm ownership.

Generally speaking, the digital image watermarking process is divided into `insertion` and `extraction` stages. In the former, the algorithm embeds the watermark in a given cover image using a secret key to decide the location of the embedded watermark. While in the latter, the same key is based for extracting the embedded watermark. Based on the extraction process watermarking algorithms can be classified as `blind` and `non-blind`

techniques. In blind watermarking, only the knowledge of the secret key is enough to recover the embedded watermark, while in the non-blind one, the original cover image is required to extract the inserted watermark. Furthermore, the watermarking schemes are evaluated using three requirements, including *(1) Visibility*; the watermarked image imperceptibility should not be affected by the presence of the watermark. *(2) Robustness* refers to the resistance of the watermark so that it can withstand different image processing attacks. *(3) The capacity* refers to the amount of watermark that can be inserted in a fixed size host cover.

Watermarking techniques can be classified according to their domains into two classes: `spatial` and `frequency` domains. In the spatial domain, the watermark embedding is carried out within images by directly changing the pixels' values [24]. Modification of the Least Significant Bit (LSB) and Singular Value Decomposition (SVD) are examples of the most popular spatial domain schemes. In the frequency domain, the watermark is embedded by modulating the coefficients after applying one of the transformations, such as the Fast Fourier Transform (FFT), Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), and Discrete Wavelet Transform (DWT) [25]. However, frequency-domain methods have been demonstrated to be more robust than spatial domain methods. For instance, DCT and DWT based watermarking algorithms have shown to be robust against simple image processing operations, like low pass filtering, blurring, brightness, and contrast adjustment. Moreover, watermarking schemes can be broadly categorized into four categories, namely `robust`, `fragile`, `semi-fragile`, and `reversible` [26].

Diverse spatial domain-based watermarking approaches have been suggested in the literature for biometric template protection. For instance, authors in [27] suggested a block pyramid based adaptive quantization watermarking scheme to embed fingerprint minutiae into face images. Initially, based on the block face wise distinctions devised by Adaboost, a block pyramid is layered. Subsequently, watermark bits with higher priority are embedded into the upper pyramid level with a larger embedding strength using a first-order statics method. In [28], the authors proposed a robust watermarking method, to improve the security of multimodal biometric authentication systems, where facial features are used as a watermark and embedded into a fingerprint image with a blind second-order statistics scheme. Recently, the matter of the digital watermarking techniques was discussed in [29], by drawing on the SVD associated with multimodal biometric colour face and iris images put forth in [30].

Even though spatial domain approaches can potentially hide information from perceptibility within the host image, they remain vulnerable to several frequency attacks, such as Acronyme de Joint Photographic Experts Group (JPEG) compression, blurring, and median filtering. Thus, several contributing studies regarding the design of robust watermarking methods using frequency-domain-based schemes are proposed. As an illustration, Alkhathami *et al.* presented a DTCWT-based digital watermarking technique for

fingerprint images [31]. In this approach, the watermark is embedded in the imaginary and real parts of the DTCWT coefficients without modifying minutiae locations. The robustness of this algorithm has been evaluated under different rotation angles and various kinds of noise attacks. Another study suggested a framework for facial recognition using a fusion of DTCWT and FFT techniques [32]. The FFT and five-levels DTCWT decompositions are applied to the preprocessed facial image. The obtained DTCWT coefficients are fused with dominant absolute FFT values using arithmetic addition to generate a final set of features. In [33], a DCT-based watermarking algorithm embedded a message into the image of the fingerprint. The DCT subsequently separates the image into alternative frequency bands so that the watermarks may be easily embedded. A couple of watermarks are then embedded into the fingerprint image. The host image is then separated into an 8×8 block. Therefore, fingerprint minutiae locations need to be assessed to present the watermark within those locations. In [34], the authors proposed a watermarking approach in the DCT domain to embed the fingerprint minutiae and the iris features into an arbitrary cover image. The fusion of the two modalities is done at the decision level. Their proposed method considered only the security requirement. Another invisible watermarking approach for image protection and authentication has been proposed in [35]. This approach involves the application of the Multi-Resolution by Independent Component Analysis (MRICA) to embed two different watermarks in the independent component analysis coefficients of the cover image. Due to, using two watermarks does not only enhance the robustness of the approach but also improves the authentication and make the approach more susceptible to images of low quality attacks. In [36], the author used a DCT-based watermarking technique to fuse face and fingerprint features. To accomplish this, the watermark bits stream is generated from the fingerprint minutia points. Next, the face image is decomposed using the DCT. Then, the watermark is embedded by modifying the middle-frequency DCT coefficients. Finally, the inverse DCT is applied to form the watermarked face image. Vatsa *et al.* [37] employed a robust watermarking technique to embed voice feature descriptors in colour facial images using the Redundant Discrete Wavelet Transform (RDWT). Similarly, the DWT and RDWT form the basis of the watermarking method in [38]. The authors encode face image into a binary sequence through the use of weighted binary coding to allocate more weight to the most significant bits. Furthermore, a reverse procedure of weighted binary coding is applied to reconstruct the face image from the extracted binary sequence. The comparison of the two sets of techniques revealed that the performance of DWT-based watermarking techniques exceeds that of RDWT-based ones. Moreover, researchers in [39] proposed a DTCWT-based blind additive image watermarking method that takes into consideration the Human Visual System (HVS) characteristics at the embedding step to improve the imperceptibility. The experimental results proved that the proposed approach outperforms the existing watermarking systems in term of robustness and imperceptibility.

4.2 Secure Sketch Based Biometric Template Protection

The secure sketch is a key generation based biometric cryptosystems scheme where the original biometric information is considered as the secret key.

Typically, the process of secure sketch includes two main components [18], namely *"Encoder"* and *"Decoder"*. During the enrollment stage, the encoder (sketch generation) takes the original biometric template X as the input and outputs a sketch P. In the authentication stages, the decoder (the biometric template reconstruction) takes the query biometric template Y and the sketch P as the inputs and outputs a biometric template X'. If Y and X' are sufficiently close according to some similarity measures, we will have X = X'. An important requirement for such a scheme is that the sketch P should not reveal information about the original biometric template X and that attackers would not be able to readily forge the original data even when the sketch is compromised.

Mathematically, the secure sketch approach can be defined as follows:

Let M be a finite set of points with similarity relation S $\in M \times M$. When (X,Y) $\in$ S, we say that Y is similar to X or the pair (X, Y) is similar.

4.2.1 Secure Sketch in Discrete Domain

A secure sketch scheme can be presented in discrete domain as a tuple (M, S, Encoder, Decoder), where Enc: M $\rightarrow$ {0,1}* is the encoder and Dec: M $\times$ {0,1}* $\rightarrow$ M is the decoder, such that for all X, Y $\in$ M, if (X,Y) $\in$ S, Decoder(Y, Encoder(X)) = X. The string P = Encoder(X) is the sketch and is considered to be made public.

4.2.2 Secure Sketch in Continuous Domain

To transform the points in continuous domain to discrete domain, the quantization metric is first applied. Next, the secure sketch scheme for discrete domain is used to get the sketch P. During reconstruction stage, the quantization Q(X) is recovered instead of the original X in the continuous domain.

The quantization-based secure sketch metric is a tuple (U,S,Q, Encoder, Decoder), where Enc: M $\rightarrow$ {0,1}* is the encoder and Dec: M $\times$ {0,1}* $\rightarrow$ M is the decoder, such that for all X, Y $\in$ U, if (X,Y) $\in$ S, Decoder(Q(Y), Encoder(X)) = Q(X). The string P = Encoder(Q(X)) is the sketch.

Diverse works were proposed to develop secure sketch protection schemes. A theoretic framework based on the secure sketch to generate cryptographic keys with the use of biometrics and other noisy information is proposed by Y. Dodis *et al.* [19]. In [20], the authors proposed and analyzed a cascaded mixing approach to merge the less important

secrets with the sketch of the most important secret. They demonstrated that under certain conditions, their method provides more protection to more important secrets at the expense of increasing the risk of reduced security, compared to the less important secrets. J. Bringer *et al.* [21] presented a specific algorithm that performs well by combining several sketching techniques and cancelable transformations. A novel binary length-fixed feature generation method of the fingerprint was presented by L. Peng *et al.* [40]. This method couples the generated binary features as an input with the fuzzy commitment scheme to construct the biometric cryptosystems, by combining several error correction codes (i.e., the Bose-Chaudhuri-Hocquenghem (BCH) code, a conducted code of the BCH code, the Reed-Solomon code, and the Low-Density Parity-Check (LDPC) code). In other words, they attempted to obtain discriminating binary length-fixed features from fingerprint images, to be coupled with some error correction codes to get an efficient biometric cryptosystem based on the fuzzy commitment scheme. In [22], the authors analyzed how an error-tolerant cryptographic primitive using a secure sketch can be applied to protect biometric templates. Additionally, a general framework to design and analyze the secure sketch for face templates is presented.

4.3 Image Encryption

Cryptography is one of the best-known techniques to secure the transmission or storage of sensitive data. It is a widely used method to convert the information into a non-recognizable form. This technique includes two blocks, encryption and decryption. These blocks are accomplished using mathematical algorithms in such a way that, the original information can be decrypted only by the intended recipient.

As the biometric templates are stored in an image format, the traditional encryption algorithms, such as Rivest-Shamir-Adleman (RSA) and Advanced Encryption Standard (AES) initially designed for text data, are not suitable for biometric template protection. This fact is due to two reasons: first, the image size is much larger than the text size, thereby implying that the conventional encryption algorithms are slow to encrypt the image information. Second, the decrypted text must be equal to the original text. Conversely, this constraint is not essential for images, where it is acceptable to obtain a decrypted image containing small distortions. Subsequently, image encryption techniques are suggested to provide privacy to biometric data.

In [41], the authors suggested a hyperchaotic Rössler map based fingerprint image encryption approach to increase the robustness of the biometric systems. The authors used high pseudorandom sequences to produce efficient encryption effects, which showed better results in term of the large space of the key, the speed encryption, low correlation of pixels, and uniform distribution histograms. A modified Rubik's cube principle-based encryption scheme for fingerprint biometric is suggested in [42]. The concept underlying the Rubik's cube algorithm is to carry out the row and column rotation or changing

the position of the pixels. Additionally, the used encryption key is obtained from the fingerprint features that assist in key management and storage for further decryption and verification process. The authors in [43] analyzed the randomness, complexity, and efficiency of chaotic image encryption based on a combination of One-Dimensional (1D) logistic map, Two-Dimensional (2D) Arnold cat map, or Three-Dimensional (3D) Arnold cat map. In this work, the encryption keys are generated from the biometric image of the sender, where the fingerprint image is scanned and subsequently preprocessed to the grey-level image. Subsequently, the initial condition for the 1D logistic map and the key for the encryption are generated. The key generation is followed by the encryption process that includes randomization using a chaotic map, transformation via Dual Parameter Fractional Fourier Transform (DPFrFT), and decomposition using Hessenberg decomposition. Finally, the obtained image underwent inverse DPFrFT to obtain an encrypted image. A. M. Meligy *et al.* [44] suggested an efficient encryption algorithm to protect biometric images with fast performance speed and a high level of security. The proposed stream cypher incorporates both the logistic chaotic map and tent map to produce the cypher image. The conducted experiments revealed the high efficiency of the algorithm, its sensitivity to secret key changes, and its resistance to different types of attacks. In [45], the authors proposed a new multiple-chaos-based biometric image cryptosystem for fingerprint security. The design of the encryption algorithm is composed of four chaotic algorithms that are two 1D and two high-dimensional 3D chaotic systems. The primary idea behind the work presented in [46] is to develop a secure and reliable image encryption method based on chaotic functions, where a logistic map sequence is first generated. Subsequently, a user intrinsic key is generated from biometrics and used as the initial value of the chaotic sequence. The random sequence produced by the chaotic phenomenon is used to encrypt and secure the biometric data. X.-Y. Wang *et al.* put forward an encryption technique that combines DNA sequences, an image encryption method, and a chaotic system in [47]. The process of this approach starts by carrying out bitwise operations XOR by using the pseudorandom sequences produced by the chaotic system CML. Next, the DNA matrix is produced using a DNA coding rule. After that, a novel initial condition of CML is generated according to the DNA array and initial conditions. Then, the columns and rows of DNA matrix are permuted and the DNA matrix is confused again. Finally, after decoding the confusing DNA matrix using a kind of DNA decoding rule, the encryption image is obtained.

5 Hybrid Approaches

Generally, an efficient biometric template protection approach should satisfy the four requirements of biometric authentication systems, that are diversity, revocability, security, and performance [1]. As that no single biometric templates protection method can satisfy all the requirements at a time, several studies resort to merging the advantages

of the existing approaches while avoiding their limitations. Thus, hybrid techniques are suggested as an alternative solution, where both feature transformation approaches and biometric cryptosystems are combined at a time.

Diverse works are proposed in the literature. For instance, in [48], the authors proposed a hybrid approach based on adaptive bloom filters and Fuzzy commitment schemes for securing multi-biometric templates using iris biometric information. The experimental results showed that the proposed method satisfies the biometric template protection requirements. However, a thorough security analysis has demonstrated the vulnerability of the bloom filters based schemes to cross-matching attacks. In [49], a new hybrid encryption method is suggested to protect palm print images. The approach is proposed to fit the diffusion, confusion challenges, and supply large keyspace by carrying out the position permutations and value transformations of the image pixels using a blend of various chaos maps, including Arnold's cat map, enhanced-Lorenz map, and Ikeda map. To accomplish this, first, the Ikeda map is based for obtaining a larger keyspace to Arnold cat map. Next, the hyperchaotic Lorenz attractor is used to encrypt the scrambled template. The conducted experiments showed the strengths of the proposed cryptosystem against statistical, brute force, and differential attacks. Authors in [50] proposed a new hybrid approach based on bloom filters with additional processing steps to make the protected templates more resistant to the cross-matching attacks. The basic idea is to use the one-way functions that would reduce the recognition rate with an increasing number of users in the biometric system [51]. R. Abdullah *et al.* [52] addressed a new technique for Malaysian speaker and accent recognition. The suggested algorithm combines the Mel Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC), Dual Wavelet Packet Transform (DWPT), Dual-Tree Complex Wavelet Packet Transform (DTCWPT), and Wavelet Packet Transform (WPT) based non-linear features. During the pre-processing phase, the speech signal is first normalized and filtered to retain the pertinent characteristics. Then, MFCC, LPC, DWPT, DT-CWPT and WPT are based for extracting the features from the sampled speech signals to preserve the significant information required to distinguish accent and speaker. After that, the support vector machine, K-Nearest Neighbors (KNN), and Extreme Learning Machine (ELM) are based to measure the speaker and accent recognition efficiency. [53] introduced a feature transformation approach based on a chaos map and Hadamard matrix which was used to provide a secure biometric template protection system for access control in health-related applications.

6 Evaluation Metrics in Biometric Template Protection

6.1 Security and Quality Analysis

- **Peak Signal to Noise Ratio (PSNR)**

In the watermarking algorithms, the peak signal to noise ratio is used to measure the embedding distortion of the watermarked image. The PSNR is computed in decibels (dB) using the following formula:

$$PSNR = 10 \lg_{10}(\frac{\max(I(i,j), I_w(i,j))}{\sum_i \sum_j I(i,j) \times I_w(i,j)}) \tag{2.1}$$

Where $I(i,j)$ and $I_w(i,j)$ are the original and watermarked images respectively.

The average value of the PSNR indicates the efficiency of the watermarking approach to preserve the invisibility between the original and watermarked images. Higher the PSNR value means that the watermarked image quality is similar to the quality of the original image.

- **Correlation Coefficient**

Most of the watermarking algorithms use the correlation coefficient metric to determine the quality of the extracted watermark after attacks. The correlation coefficient is computed using the following formula:

$$C = \frac{\sum_i \sum_j W(i,j) \times W_x(i,j)}{\sqrt{(\sum_i \sum_j W(i,j)^2) \times (\sum_i \sum_j W_x(i,j)^2)}} \tag{2.2}$$

With W and W_x are the embedded and the extracted watermarks respectively.

- **Entropy Loss**

To measure the information leakage of the sketch in secure sketch algorithms, authors in [19] suggested the notion of *entropy loss* which gives the adversary the advantage to estimate the original information from a generated sketch. However, as most of the biometric templates are presented in a continuous domain, and that *entropy loss* performs the measures in a discrete domain, some sort of biometric data quantization should be utilized. To compute the *entropy loss* of a discrete variable A, the *min-entropy* defined as $\mathbf{H}_\infty(A) = -\log(\max_\alpha \Pr[A = \alpha])$ is calculated. This notion is used to know the nearly uniform random bits that can be read out from the distribution. In the case of a variable A given another random discrete variable B, this notion is recalled *average min-entropy* and defined as $\tilde{\mathbf{H}}_\infty(A|B) = -\log(\mathbb{E}_{b \leftarrow B}[2^{-\mathbf{H}_\infty(A|B=b)}])$. Thus, the *entropy loss* of the original data X given the sketch P is computed via the following equation:

$$\xi = \mathbf{H}_\infty(X) - \tilde{\mathbf{H}}_\infty(X|P) \tag{2.3}$$

- **KeySpace**

The keyspace is one of the basic metrics used to measure the robustness of encryption algorithms against brute force attacks. It refers to the set of all possible

permutations of a key. In other words, it is defined as the range of different possible values of a key. Generally, the keyspace should be sufficiently large to reduce the probability of a successful brute force attack. For instance, a password with n characters, where each of these characters can assume C different values, has a keyspace of C^n [54].

6.2 Performance Analysis

The performance evaluation in biometric authentication systems bases on the degree of similarity between two biometric characteristics referred to as similarity score. This similarity matching score includes two cases, genuine and impostor scores. The first is obtained when comparing two samples of the same biometric features originating from the same user, while the second is gained by comparing two biometric traits of different individuals.

The genuine and impostor scores are produced using the Fingerprint Verification Competition (FVC) standard protocol [55]. More specifically, the genuine scores are produced through a comparison of every image with the rest of the images of the same individual, while the impostor scores are produced through a comparison of the first image of every individual with the first image of other individuals. A biometric system decides of accepting or rejecting a user by comparing the matching scores S to a threshold η. The FRR is the measure of rejecting a genuine user rendered as an impostor and refers to the rate of genuine scores less than η, while the False Acceptance Rate (FAR) is the measure of accepting the impostors as genuine users and refers to the rate of impostor scores greater than or equal to η. After generating the genuine and impostor scores, we generate a T-set of thresholds $\eta_{j=1}^{T}$, $S_{min} \leqslant \eta_j \leqslant S_{max}$, for each $j \in \{1, 2, ..T\}$. S_{min} and S_{max} refer to the minimum and maximum scores. For each η_j, the FAR and FRR are computed via the following formula [56]:

$$FAR(\eta_j) = \frac{1}{L_0} \sum_{i=L_1+1}^{L} I(S_i \geq \eta_j) \tag{2.4}$$

$$FRR(\eta_j) = \frac{1}{L_1} \sum_{i=1}^{L_1} I(S_i < \eta_j) \tag{2.5}$$

$$I(x) = \begin{cases} 1, & If\ x\ is\ true \\ 0, & Otherwise \end{cases} \tag{2.6}$$

Where L_0 and L_1 are the set of the impostor and genuine scores. L is the total set of scores ($L = L_0 + L_1$).

The FRR and FAR values are variated by adjusting the η value. However, it is not possible to lessen both of these values at the same time.

To evaluate the reliability of the biometric authentication systems, the Equal Error Rate (EER) refers to the point at which the FAR and FRR values are equal is used. The smallest the EER value is the better, showing that the system is less likely to accept impostors as genuine or reject falsely genuine as impostors.

Additionally, the Genuine Accept Rate (GAR) is used to compute the fraction genuine scores surpassing the threshold η. The GAR is defined as $GAR = 1 - FRR$.

6.3 Diversity and Revocability Analysis

Since the user's biometric templates are linked to the same biometric traits in all the authentication systems, compromising a template from one of the systems' databases may lead to penetrating other systems. Which is known as cross-matching between databases. Thus, diversity and revocability are becoming a necessity.

A biometric template protection approach is considered to satisfy the diversity criteria if it generates multiple uncorrelated templates from the same original feature vectors. This enables the individual to enrol in different applications using the same traits without the risk of any cross-matching between their corresponding databases. The most common way of proving that a template protection scheme satisfies the diversity property is to generate multiple protected templates from the same original modality and then attempting to match the resulting sets of the protected templates. If the matching result of two secured templates of the same biometric traits is found to be lower, then, the biometric template approach is considered to meet the diversity criteria [57].

The revocability is the ability to tolerate partial compromises of data (e.g., a two-factor access control system, either the key or the stored data has been revealed to the adversary, but not both). In other words, if the stored template is broken through, it will be possible to repeal this template and reconstruct a new one. Furthermore, the renewed/ revoked template should not match the previously compromised one. Thus, revocability does not mean just to cancel the old template and issue a new one; it also means that the authentication rights of the old authenticator are revoked [58].

6.4 Databases

ORL Face Database: It is composed of 400 face images of 92×112 pixels (40 persons, 10 images per person) with 256 grey levels per pixel. The images are

captured at different times and under various facial expressions, including open or closed eyes, smiling or not smiling, and facial details (glasses/no glasses). Besides, the faces are captured in an upright position in frontal view with a slight left-right rotation. The database is built between April 1992 and April 1994 at the AT&T Laboratories Cambridge, in the context of a collaboration project between the Speech, Vision and Robotics Group of the Cambridge University Engineering Department and the AT&T Laboratories Cambridge [59].

FVC2000 DB1: It contains 880 fingerprint images of size 300×300 pixels. The fingerprint images are acquired from volunteers from 20 to 30 years-old via a low-cost optical sensor with 500 dpi resolution. Up to four fingers were collected for each volunteer (forefinger and middle finger of both the hands). The images were taken from untrained people in two different sessions and no efforts were made to assure a minimum acquisition quality [60].

FVC2002 DB1 and DB2: Each database Contains 880 fingerprint images of size 388×374 and 296×560 pixels respectively. The fingerprint images of the DB1 are taken using the optical sensor "TouchView II" by Identix, with 500 dpi resolution. While the DB2 images are captured via an optical sensor "FX2000" by Biometrika, with 569 dpi of resolution [61].

7 Conclusion

In this chapter, a detailed survey on biometric authentication systems is presented with a particular focus on biometric template protection against the system's database attacks.

To secure biometric systems against template database attacks, four requirements should be satisfied, including security, revocability, diversity, and performance. To achieve this, one solution would be to secure the biometric data before being stored in the database. To accomplish this, two sorts of techniques are suggested in the literature, namely feature transformation and biometric cryptosystems. Besides, as that no single protection method can meet the four requirements at a time, hybrid techniques are considered where both feature transformation approaches and biometric cryptosystems are combined.

3 Spatial Watermarking for Multimodal Authentication

1 Introduction

Biometric template protection is one of the most challenging tasks faced up when putting biometric authentication systems into practice. This is due to the intra-individual variability of the biometric features. Watermarking technique is one of the powerful means to secure biometric templates and add extra functionalities. The main goal of this scheme is to embed the unique pattern within the original biometric data to provide greater security. This is due to the embedded data would not be recovered only via a secret key. Besides, fusing two biometric modalities via the watermarking algorithm increases the authentication performance. Moreover, it offers the perceptual transparency between the original and watermarked images.

In this context, the current chapter presents a spatial fingerprint-face-based watermarking scheme to secure authenticity and improve the performance of multimodal authentication systems. The suggested method embeds the face features considered as a watermark into the fingerprint cover image via a generated secret key. The secret key is based for locating the pixels to be watermarked while preserving the fingerprint minutia points. The face features are extracted using the Orthogonal Locality Preserving Projections (OLPP) technique. Owing to according to [62] the OLPP features perform better since they encode more discriminating information by preserving the local structure. The proposed scheme has the advantage that, in addition to the fingerprint matching, the recovered watermark in the decoding stage is used to establish the authenticity of the user.

The suggested method is tested against the common digital image watermarking attacks such as Gaussian noise, speckle noise, poison noise, and median filter. The experiment results conducted on the ORL face database and the FVC2002 DB2 fingerprint database showed a reasonable level of robustness along with visual quality. Moreover, better

performance accuracy is found, and a better capability to distinguish between genuine and impostor users is shown.

2 Spatial Watermarking Scheme

The process of the proposed watermarking method includes two stages, watermark embedding and watermark detection.

2.1 Watermark Embedding

This phase can be summarized into three steps:

Step 1: Watermark Preparation
This step starts by extracting the face features using the OLPP method to reduce the length of the facial image by selecting the most pertinent features. The OLPP is based due to it has proven an outperform over the Eigenface, Fisherface, and Laplacianface metrics for face presentation and recognition [62]. Next, the obtained face features are converted into a bits stream and headed by two reference bits 0 and 1. The idea behind using the reference bits is to compute the adaptive threshold-based to determine the watermark bit values in the watermark detection stage.

Step 2: Minutia Area Localization
Herein, the fingerprint image minutia coordinates are located using the fingerprint minutia extraction system described in [63]. As illustrated in figure 3.1, the minutia points localization process is achieved via several preprocessing steps that are, histogram equalization, FFT enhancement, binarization, selecting the Region Of Interest (ROI) area, thinning, removing spike, and finally extracting the real minutia. By the end, a 7×7 square block is drawn around every minutia location to get the marked pixels.

Step 3: Watermark Embedding
The watermark integration into the fingerprint image is reached by changing the unmarked fingerprint pixel values according to the following equation [64]:

$$P_{WM}(i,j) = P(i,j) \times \left[1 + (2s-1)q\left(1 + \frac{SD(i,j)}{A}\right)\left(1 + \frac{GM(i,j)}{B}\right)\beta(i,j)\right] \tag{3.1}$$

Where $P(i,j)$ and $P_{WM}(i,j)$ are the original and watermarked pixels values at the location (i,j), respectively. The value of the watermark bit is denoted as s and the

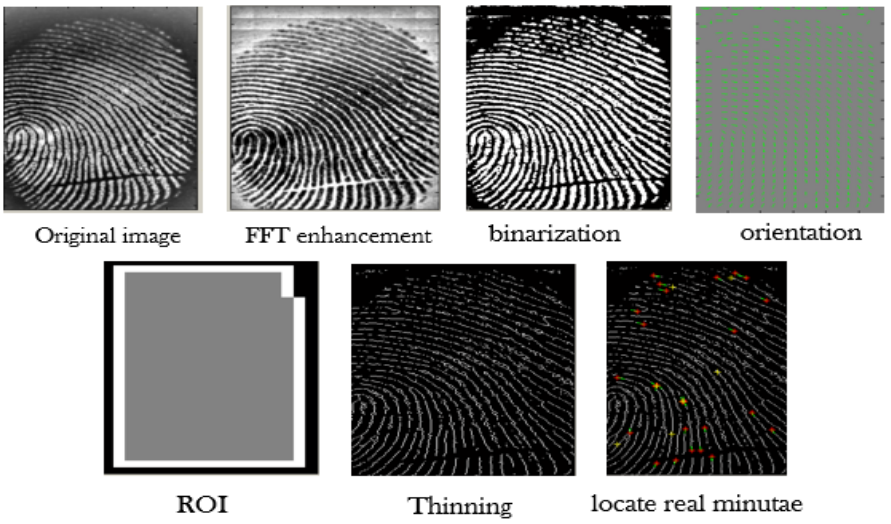


FIGURE 3.1: Fingerprint image preprocessing.

watermark embedding strength is denoted as q, where $s \in [0, 1]$ and $q > 0$. $SD(i, j)$ is the standard deviation of pixel values in a neighbourhood of location (i, j) and $GM(i, j)$ is the gradient magnitude at (i, j). A and B are the weights for the standard deviation and gradient magnitude, respectively. $\beta(i, j)$ guarantees that image pixels, whose alteration may affect the fingerprint verification performance are unchanged. It takes the value 0 if the pixel (i, j) is a marked pixel and the value 1 otherwise.

Due to the cover image large dimensions compared to the watermark ones, the watermark bits are embedded at multiple locations. Besides, a secret key is randomly generated to determine the locations of the pixels to be watermarked during the embedding process. The block diagram of the spatial watermarking algorithm is given in figure 3.2.

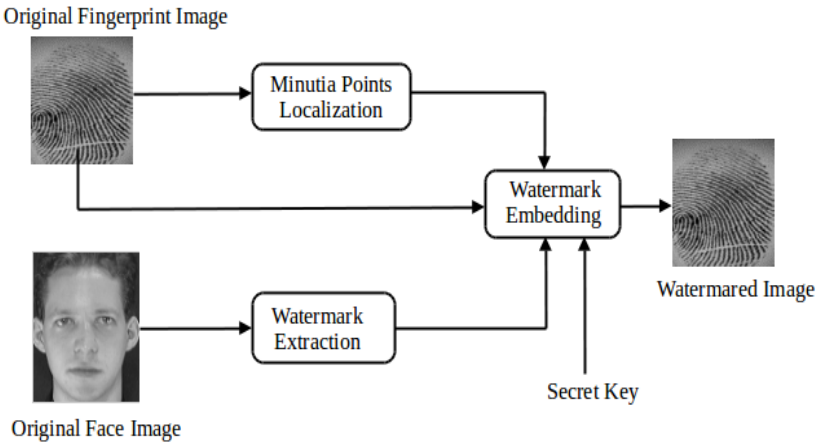


FIGURE 3.2: General watermark embedding process

2.2 Watermark Detection

To extract the embedded watermark, first, the locations of the changed pixels have recurred via the secret key employed in the embedding stage. Then the pixels values are estimated as a linear combination as presented in the following equation [64].

$$\hat{P}(i,j) = \frac{1}{4c} \times \left[\sum_{k=-c}^{c} P_{WM}(i+k,j) + \sum_{k=-c}^{c} P_{WM}(i,j+k) - 2P_{WM}(i,j) \right] \tag{3.2}$$

Where $P_{WM}(i,j)$ and $\hat{P}(i,j)$ are the watermarked and estimated pixels values at a location (i,j), respectively. $c = 2$ is the size of the neighbourhood. The difference between the watermarked and the estimated fingerprint images is computed via the following equation:

$$\delta = P_{WM}(i,j) - \hat{P}(i,j) \tag{3.3}$$

These differences are averaged over all the embedding locations associated with the same bit, to get $\bar{\delta}$. To find the adaptive threshold, the averages are computed separately for the reference bits 0 and 1 to obtain $\bar{\delta}_{R0}$ and $\bar{\delta}_{R1}$ respectively. By the end, the watermark detection is carried out using the following equation:

$$\hat{\delta} = \begin{cases} 1 & \text{If } \bar{\delta} > \frac{\bar{\delta}_{R0} + \bar{\delta}_{R1}}{2} \\ 0 & \text{Otherwise} \end{cases} \tag{3.4}$$

3 Description of the Multimodal Authentication System

As illustrated in figure 3.3, the proposed authentication system consists of two stages, enrollment and authentication. In the enrollment stage, the user presents his/her biometric data face and fingerprint. Next, the spatial watermarking algorithm is applied to fuse both modalities. The obtained watermarked fingerprint image is stored in the database as a reference image. During the authentication stage, the user presents his/her biometric data face and fingerprint. Next, the OLPP face features and the minutia points locations are extracted from the acquired images. Meanwhile, the reference image is retrieved from the database, and the watermark detection procedure is applied. Besides, the minutia points are extracted from the fingerprint cover image. After that, the reference templates and the acquired templates of both face and fingerprint modalities are compared using the Hamming distance and the minutiae matching metrics, respectively. By the end, to get the decision score, the improved min-max normalization score fusion

method [65] is employed to fuse the face and fingerprint matching scores. In case a match is found, then the user is granted permission to access, otherwise he/she is rejected.

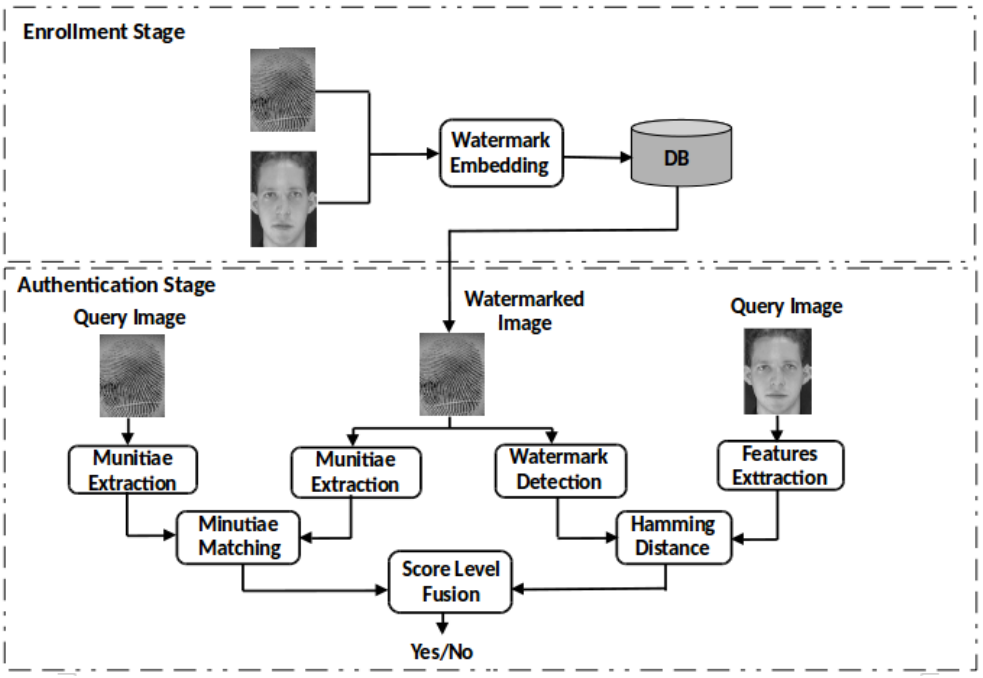


FIGURE 3.3: Overview of the enrollment and authentication stages

4 Experimental Analysis

4.1 Experimental Setup

The experiments are conducted using the ORL face and FVC2002 DB2 fingerprint databases. 40 user images, 8 images per user with a total number of 320 images are used in each dataset. Images of size 64×64 pixels and 560×296 pixels with 256 grey levels per pixel are based in the ORL face database and FVC2002 DB2 fingerprint database, respectively.

The $SD(i, j)$ value is calculated as the standard deviation of the pixel values in a cross-shaped $(5 * 5)$ neighbourhood of the embedding location (i, j). The gradient magnitude $GM(i, j)$ is computed via the $(3*3)$ Sobel operator. After applying the OLPP technique on the face images, feature coefficients vectors of size 30 are obtained. Using the binary conversion of the acquired vectors bits streams (watermark) of 1922 bits are found. In addition, the random secret keys based in the embedding process are generated with lengths similar to the watermark bits stream. Moreover, due to the large dimensions of the cover image compared to the watermark dimensions, the watermark data is embedded 20 times. The watermarking parameters are set to: $q = 0.05$, $A = 100$, and $B = 1000$.

4.2 Results and Discussions

4.2.1 Visual Quality

The imperceptibility of the watermark in the host image is assessed using the image histogram. Figure 3.4 shows the original and watermarked images along with their corresponding histograms. It can be seen from the figure that the visual quality of the fingerprint image has not been affected by the watermarking process.

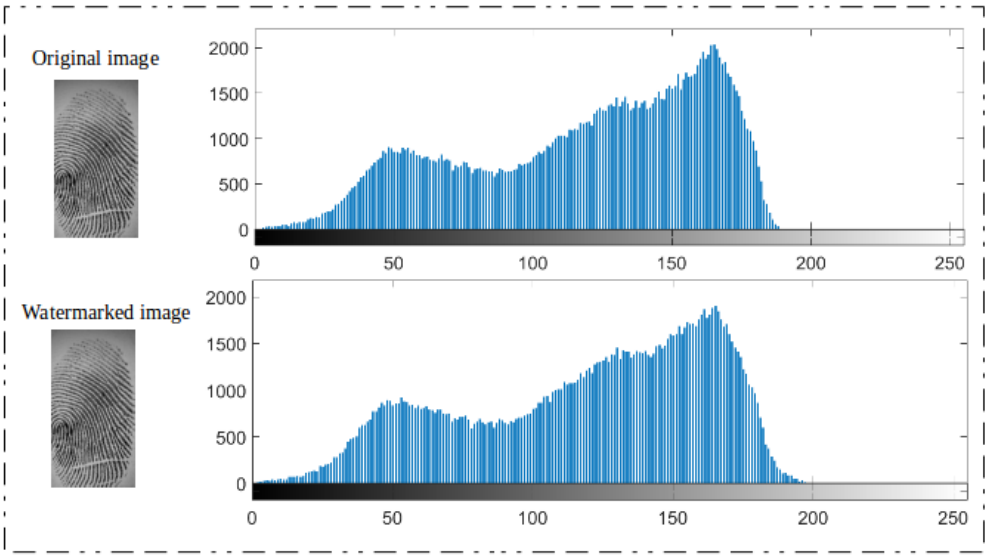


FIGURE 3.4: Original cover image (above), watermarked image (below) and their corresponding histograms.

4.2.2 Robustness Evaluation

The robustness of the suggested watermarking method is evaluated against the common digital image watermarking attacks, such as Gaussian noise, speckle noise, poison noise, and median filter. The results are reported using the PSNR values between the original fingerprint images and watermarked fingerprint images after attacks. Table 3.1 illustrates the PSNR average values of all the watermarked images before and after attacks. It can be noticed from the table that the PSNR values after attacks are slightly changed compared to images with no attack. Which proves the robustness of the watermarking method against the considered attacks, thus, providing an appreciable level of template protection.

4.2.3 Performance Evaluation

To assess the efficiency of the fingerprint-face-based multimodal authentication system, the genuine and impostor matching score distribution curves, the FAR versus FRR

TABLE 3.1: PSNR values before and after attacks.

Attacks	PSNR(dB)
No attack	33.75
Gaussian noise (Noise density=0.4)	33.65
Speckle noise (Noise density=0.001)	31.17
Poison noise ($scale = e^{10}$)	33.50
Median filter[3x3]	33.27
Median filter[5x5]	33.26

curves, and the Receiver Operating Characteristic (ROC) curves are plotted. Figure 3.5 shows the genuine and impostor matching score distribution curves of the fingerprint-based unimodal authentication system and the fingerprint-face-based multimodal authentication system respectively. We note that the overlapping between the genuine and impostor scores has been reduced, which demonstrates the ability of the fingerprint-face-based multimodal authentication system to distinguish efficiently between genuine and impostor users.

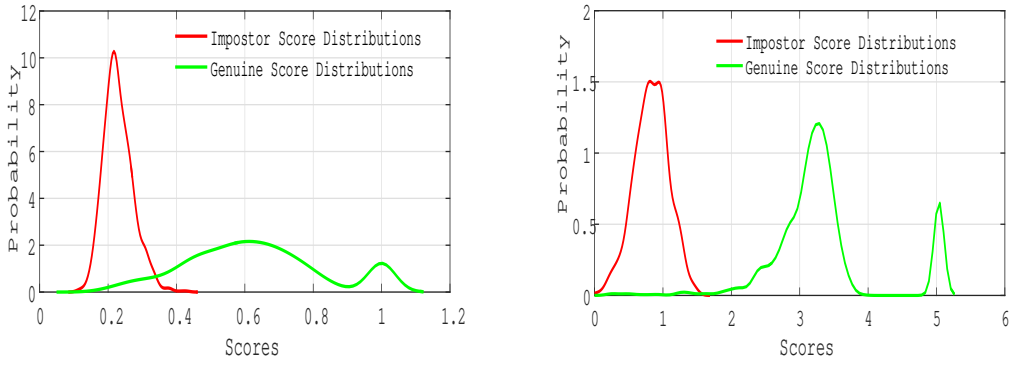


FIGURE 3.5: Genuine and impostor matching score distributions curves of the fingerprint based unimodal authentication system (left) and the fingerprint-face-based multimodal authentication system (right).

Moreover, the reliability of the authentication system is rated via the EER value, that is the point at which the FAR and FRR values are equal. The smallest the value of the ERR is the better, showing that the system is less likely to accept falsely imposters as genuine or reject falsely genuine as impostors. As illustrated in figure 3.6, an EER value of 5.40% is obtained for the fingerprint baseline authentication system, while it achieves 0.96% for the fingerprint-face-based multimodal authentication system. Henceforth, we can conclude that the proposed multimodal authentication system is more efficient compared to the unimodal authentication system by providing fewer failures in rejecting authorized users or accepting unauthorized users.

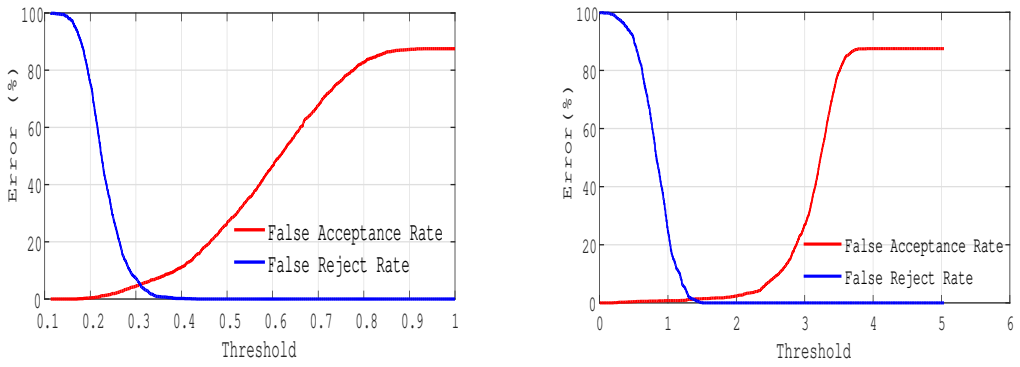


FIGURE 3.6: False acceptance rate versus false reject rate curves the fingerprint based unimodal authentication system (left) and the fingerprint-face-based multimodal authentication system (right)

The ROC curves given in figure 3.7 show the performance of (1) the fingerprint baseline unimodal authentication system (without watermarking), (2) the watermarked fingerprint-based authentication system, (3) the face based unimodal authentication system, and (4) the fingerprint-face-based multimodal authentication system. The similarity between the authentication based on original fingerprint images and the authentication based on watermarked fingerprint images indicates that the watermark embedding process does not affect the recognition performance since the minutia points are not affected. Moreover, the ROC curve of the fingerprint-face-based multimodal system shows better performance accuracy compared to the fingerprint-based unimodal authentication system.

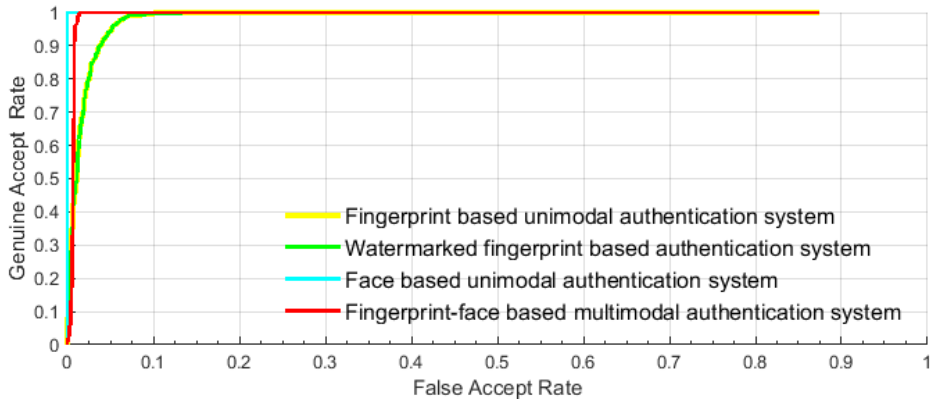


FIGURE 3.7: ROC curves of the fingerprint based unimodal authentication system, watermarked fingerprint based authentication system, the face based unimodal authentication system, and the fingerprint-face-based multimodal authentication system.

5 Conclusion

In this chapter, a spatial watermarking approach for fingerprint-face-based multimodal authentication systems is suggested. The concept underlying this method is to protect

biometric templates stored in the system database and improve the system's performance. To accomplish this, the watermark is generated by extracting the face features using the OLPP metric. The obtained face features vector is converted into a binary stream. Next, the watermark is integrated into the fingerprint image without corrupting the minutiae points using a secret key. The watermark detection is carried out using the same secret key to find the location of the manipulated pixels. The proposed method is tested against the common digital image watermarking attacks, such as Gaussian noise, speckle noise, poison noise, and median filter. The experimental results conducted on the ORL face and FVC2002 DB2 fingerprint databases showed a good level of robustness along with visual quality. Besides, the fingerprint-face-based multimodal authentication system demonstrated better performance accuracy and better capability to distinguish between genuine and impostor users.

4 DTCWT-DCT-based Watermarking and Hadamard Transform for Biometric Template Protection

1 Introduction

Generally, an ideal biometric template protection approach should satisfy four requirements of biometric authentication systems, namely diversity, revocability, security, and performance. To satisfy these requirements and ensure the template protection, several techniques are suggested in the literature [66]. Since no single biometric template protection method meets the all aforementioned requirements at a time, one idea would be to consider hybrid techniques.

In this context, the current chapter proposes a new hybrid approach for biometric template protection in multimodal authentication systems. The suggested approach combines a biometric watermarking method and the partial Hadamard transform in order to satisfy the four requirements of biometric authentication systems. The concept underlying the watermarking algorithm is to embed the watermark imperceptibly (secret information) into the cover image via a secret key to prevent illegal use. Additionally, fusing two biometric modalities via the watermarking algorithm increases the authentication performance. The partial Hadamard transform is used due to its non-invertible property. Which means, retrieving the original biometric data is a computationally arduous task even when the attacker compromises the secured template. Subsequently, meeting the diversity and revocability requirements.

The proposed hybrid approach is based to secure fingerprint-face-based authentication systems. More precisely, the partial Hadamard transform is used first to cover the fingerprint features. Next, the obtained fingerprint features vectors are binarized using a quantization metric. Then, a DTCWT-DCT-based watermarking algorithm is used to fuse the facial image and the secured fingerprint features (watermark) without significant changes in the cover image. The choice of the DTCWT is justified by the use of its high-frequency sub-bands for watermark embedding. Which enhances the security of the watermarked image [67]. Additionally, unlike the work addressed in [68],

the suggested approach uses only the high-frequency sub-bands (complex values) of the DTCWT decomposition in the embedding process. Due to, different from imaginary parts, real parts of the image contain less critical data, which helps to identify attacks in case of illegal watermark extraction. Besides, the DCT transform is employed owing to its robustness against different watermarking attacks. Moreover, the watermark is included by modifying only the middle-frequency DCT sub-bands coefficients to face compression attacks and preserve the distinctness of the watermarked image.

The approach is assessed using the ORL face database and three FVC fingerprint databases including, FVC2000 DB1 and FVC2002 DB1, DB2. Besides, The experimental results demonstrated the efficiency of the proposed approach in terms of satisfying the four biometric authentication systems' criteria.

2 Overall Hybrid Approach Description

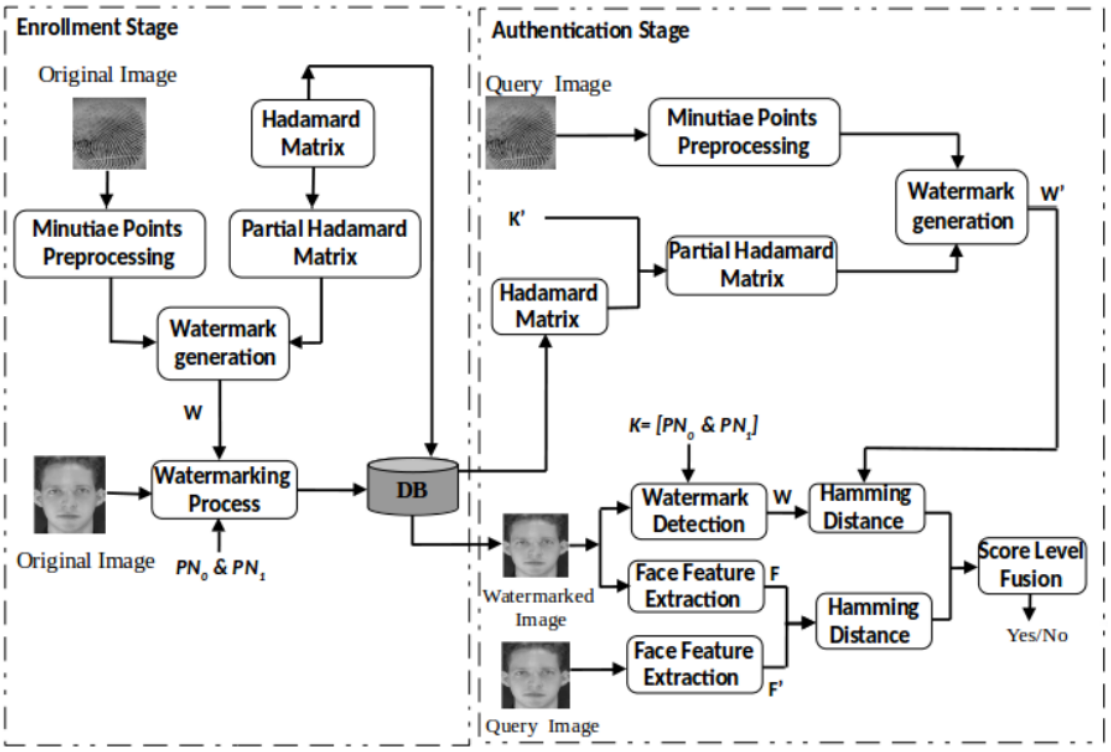


FIGURE 4.1: Overview of the enrollment and authentication stages

As illustrated in figure 4.1, the suggested hybrid approach includes two stages: *enrollment* and *authentication*.

Enrollment: During this stage, a full-order Hadamard matrix is constructed. Next, the partial Hadamard matrix is generated from the full-order Hadamard matrix. Then, the minutiae points are extracted from the fingerprint image and preprocessed. After that, the obtained fingerprint features are multiplied with the partial Hadamard matrix to get

the secured fingerprint characteristics. The acquired fingerprint features vector is then quantized to build the watermark W. Afterward, based on two pseudorandom sequences PN_0 and PN_1, the watermarking algorithm is employed to embed the watermark W into the facial image used as a reference image to be stored in the system database. Besides, the full order Hadamard matrix is stored in the same database.

Authentication: Herein, the user presents two keys K and K', and the corresponding biometric data (fingerprint, face). K is supposed to be a combination of PN_0 and PN_1, while K' contains the index of rows to be selected from the full-order Hadamard matrix to generate the same partial Hadamard matrix as in the enrollment stage. Next, the reference image (watermarked facial image) and the full-order Hadamard matrix are retrieved from the database. Then, two parallel processes are executed.

On the one hand, the same preprocessing as in the enrollment stage is applied to the fingerprint query image to obtain the fingerprint features vector. Furthermore, the partial Hadamard matrix is generated using the full-order Hadamard matrix and the key K'. Similar to the enrollment stage, the partial Hadamard matrix and the fingerprint features vector are multiplied and quantized to obtain the vector W'. Using K, the watermark W is detected from the reference image; then, the matching is conducted between W and W' via the Hamming distance method to get the fingerprint score value $S_{fingerprint}$.

On the other hand, the facial features are extracted from the reference image and the facial query image via the OLPP method [62] to obtain the two feature vectors F and F'. These are matched using the Hamming distance to get the face score value S_{face}.

By the end, the matching scores $S_{fingerprint}$ and S_{face} are then fused via the Performance Anchored Normalization (PAN) score level fusion method [65]. More precisely, the scores are fused through the application of the weighted sum rule, in which the appropriate distribution of weights is carried out based on the EER value derived in different ways from a training set. The weighted sum rule is carried out by allocating the weights ω and γ to the scores of facial and fingerprint systems, respectively via the following equation:

$$S_{final} = \omega S_{face} + \gamma S_{fingerprint} \tag{4.1}$$

The final decision is made based on the score value S_{final}. If the matching is completed, access is granted, otherwise, it is denied.

3 Hybrid Approach Components Description

In this section, we describe the main components of the proposed hybrid approach, including the watermark preparation and the DTCWT-DCT-based watermarking algorithms.

3.1 Watermark Preparation

The watermark preparation algorithm includes two stages: the preprocessing of minutiae points and the partial Hadamard transform.

3.1.1 Preprocessing of Minutiae Points

The preprocessing of the fingerprint features is accomplished using the spectral minutiae representation technique. This mechanism is used as a substitute for minutiae sets to reach high-performance speed constraints in high-dimension fingerprint databases. This process commences by taking the minutiae points out of the fingerprint image using the extraction system indicated in [63]. Next, the extracted minutiae points are represented as a spectral features vector of fixed-length based on the location and orientation of the minutiae using equation (4.2) [69].

$$\mathcal{M}_o(\omega_x, \omega_y, \sigma_o^2) = |e^{-(\frac{\omega_x^2 + \omega_y^2}{2\sigma_o^2})} \times \sum_{i=1}^{Z} \mathcal{F}(m_i(x, y, \theta))| \tag{4.2}$$

Where $\mathcal{F}$ is the Fourier transform defined as:

$$\mathcal{F}(m_i(x, y, \theta)) = j(\omega_x \cos\theta_i + \omega_y \sin\theta_i) \times e^{-j(\omega_x x_i + \omega_y y_i)} \tag{4.3}$$

Where $m_i(x, y, \theta)$ is the derivative $m_i(x, y) = \delta(x - x_i, y - y_i)$ in the direction θ_i. δ is the Dirac pulse metric. Z is the number of minutiae in the fingerprint image. (x_i, y_i, θ_i) are the locations and orientation of the i^{th} minutia in the fingerprint image. $(\omega_x, \omega_y, \sigma_o^2)$ are the frequencies and the parameters of the Gaussian kernel function, respectively.

After sampling the output of the equation (4.2) on a polar-logarithmic grid, the Column Principal Component Analysis (CPCA) method is then used to reduce the length of the spectral minutiae feature vector to decorrelate features and concentrate their power[70].

3.1.2 Partial Hadamard Transform

A Hadamard transform is a non-sinusoidal orthogonal square matrix with elements of values +1 or -1. This kind of matrix bases on Walsh functions and has no multipliers. Its mathematical core is a Sylvester-Hadamard matrix [71] or Hadamard matrix for short. With $n = 2$, an $n \times n$ Hadamard matrix is built as:

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{4.4}$$

$$H_{2n} = \begin{bmatrix} H_n & H_n \\ H_n & -H_n \end{bmatrix} \tag{4.5}$$

Hence, generally, the Hadamard matrix is defined as:

$$H_n^T H_n = nI_n \tag{4.6}$$

Where I_n is an $n \times n$ identity matrix. Owing to the invertibility of the Hadamard matrix, a column rank-deficient and non-invertible partial Hadamard transform is exploited. In other words, instead of using the $N \times N$ full-order Hadamard matrix, an $S \times N$ sub-matrix is used by randomly selecting S rows (with $S < N$).

In the case of our research work, suppose L is the $N \times 1$ fingerprint features vector acquired from the preprocessing of minutiae points section. To generate the watermark, the following steps are pursued:

- Generate a full-order Hadamard matrix H of dimensions $N \times N$; $N = 2n$ (n is the total number of bits required for quantizing pair-minutiae vectors).
- Generate the partial Hadamard matrix PH of dimension $S \times N$; $(S < N)$.
- Multiply the partial Hadamard matrix PH and the fingerprint features vector L as illustrated in equation (4.7) to engender an $S \times 1$ secured template T.

  $$T = PH \times L \tag{4.7}$$

  By applying the equation (4.7), the fingerprint features vector L is hidden among an infinite number of solutions. Which complicates recovering the original template even when the attacker compromises both T and PH.

- Quantize T to obtain a binary watermark stream. If the value of T is greater than 1, then, the watermark bit is quantized as 1, otherwise, it is considered to be 0.

3.2 DTCWT-DCT-based Watermarking

The concept of the DTCWT-DCT-based watermarking method is combining the face and fingerprint modalities to improve both the performance accuracy and security of biometric systems. This method includes watermark embedding and watermark extraction stages.

3.2.1 Watermark Embedding

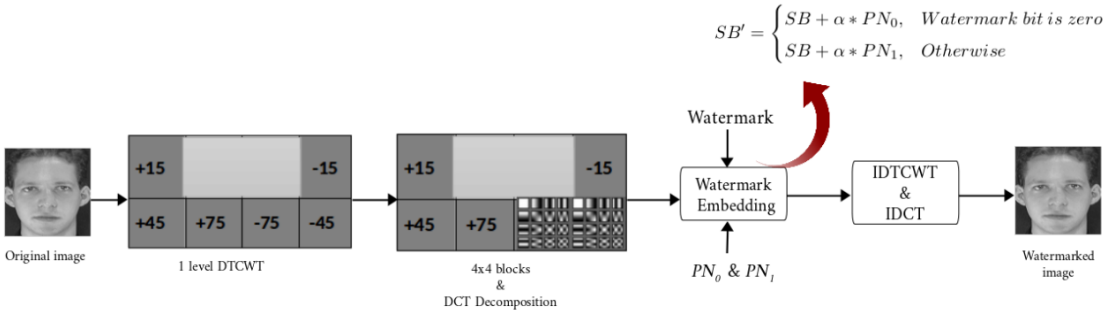


FIGURE 4.2: Watermark embedding process.

As illustrated in figure 4.2, the watermark embedding process pursues the following steps:

- Decompose the face image into six directional sub-bands (± 15 °, ± 45 °, and ± 75 °) using one DTCWT level decomposition.
- Take randomly two high-frequency sub-bands, divide each into 4×4, and apply DCT to each block.
- Generate two pseudo-random sequences PN_0 and PN_1 to embed the watermark bits 0 and watermark bits 1, respectively.
- Embed PN_0 and PN_1 into the DCT transformed 4×4 blocks with a gain factor α. The embedding is applied only to the DCT mid-band coefficients. Suppose SB is the mid-band coefficients matrix of the DCT transformed blocks. The embedding procedure is carried out as follows:

$$SB' = \begin{cases} SB + \alpha * PN_0, & Watermark\ bit\ is\ zero \\ SB + \alpha * PN_1, & Otherwise \end{cases} \tag{4.8}$$

- Apply the inverse DCT to each 4×4 blocks.
- Apply the inverse DTCWT to the transformed image, including the modified sub-bands to get the watermarked host image.

3.2.2 Watermark Extraction

As illustrated in figure 4.3, the watermark extraction algorithm does not base on the original facial image, which is referred to as blind watermark extraction. This algorithm pursues the following steps [68] :

- Apply a one level DTCWT to decompose the watermarked face image into six non-overlapping sub-bands (± 15°, ± 45°, and ± 75°).
- Take the same two high-frequency sub-bands as in the watermark embedding stage, divide each into 4×4 blocks, and apply DCT to each block.
- Extract the mid-band coefficients of the DCT transformed blocks and calculate two correlations C_0 and C_1 with respectively PN_0 and PN_1. If $C_0 > C_1$ then the extracted watermark bit is considered 0, otherwise it is considered 1.

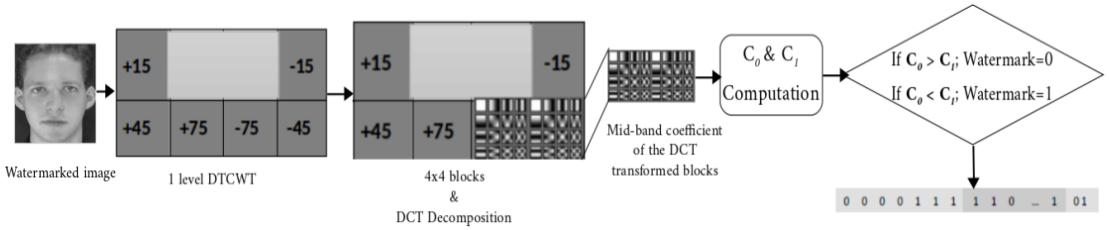


FIGURE 4.3: Watermark extraction process.

4 Experimental Analysis

4.1 Experimental Setup

To evaluate the proposed approach, our experimental study is performed using the ORL face database and three FVC fingerprint databases, including FVC2000 DB1 and FVC2002 DB1, DB2. Resulting in three datasets, *Dataset1* (ORL with FVC2000 DB1), *Dataset2* (ORL with FVC2002 DB1), and *Dataset3* (ORL with FVC2002 DB2). Facial images of size 512×512 pixels and fingerprint images of size 374×388, 560×296, and 300×300 pixels are used.

After applying the CPCA metric on the fingerprint features, a features vector of 10240 bits is obtained. Thus, to satisfy the matrices multiplication constraint, a full-order Hadamard matrix of dimensions $N \times N = 10240 \times 10240$ is generated. Next, a partial Hadamard matrix of dimensions $S \times N = 8192 \times 10240$ is produced for each user. Besides, the indexes of the selected rows from the Hadamard matrix to generate the partial Hadamard matrix are saved in the $S \times 1$ user-specific key K'.

To achieve an optimal watermarking imperceptibility and a acceptable robustness against the image watermarking attacks, the watermark embedding is performed using different values of the gain factor α among a range of 0 to 60. The conducted experiments proved that the suitable values are $\alpha = 5$ and $\alpha = 20$ for the first and second high-frequency sub-bands, respectively. Further, various values among a range of 0 to 1 are assigned to weights values ω allocated to the score of the facial system and γ allocated to the score of the fingerprint system. The values of $\omega = 0.5$ and $\gamma = 0.5$ which provide the maximum authentication performance are chosen. Moreover, as the mid-band frequencies of a DCT block have seven coefficients, pseudorandom sequences PN_0 and PN_1 of length 7 are generated.

4.2 Results and Discussions

In this section, we present and discuss the evaluation results of the conducted experiments to demonstrate the ability of the proposed approach in satisfying the four criteria of biometric template protection.

4.2.1 Security Analysis

The robustness of the watermarking algorithm is assessed experientially under the conventional digital image watermarking attacks, such as Gaussian noise, salt-and-pepper noise, speckle noise, additive white Gaussian noise, median filters, and JPEG compression. The evaluation results are reported using the PSNR, correlation coefficient and EER metrics. The PSNR average value obtained between the original facial images and the watermarked facial images is 37.98 dB. Also, the correlation coefficient average value obtained between the extracted watermark and the embedded watermark is 1. Table 4.1 illustrates the averages of the PSNR, correlation coefficient, and EER values under the early mentioned attacks. The results of the table show that the DTCWT-DCT-based watermarking algorithm can withstand these attacks in terms of the watermarked image and extracted watermark quality. Moreover, the authentication performance was not affected by the attacks, where the EER values have been slightly decreased for some attacks.

Besides, the non-invertibility of the extracted fingerprint features is gained by using the partial Hadamard transform. More precisely, selecting S random rows from the full-order Hadamard matrix produces a column rank-deficient partial Hadamard matrix with no inverse or pseudo-inverse. Which makes it computationally difficult to reconstruct the original fingerprint features, even in worst-case situations. Consequently, the original fingerprint features are safely protected.

TABLE 4.1: Correlation coefficient, EER, and PSNR values before and after attacks.

Attacks	Correlation coefficient	EER(%)	PSNR(dB)
No attack	1	0	37.98
Gaussian noise (Noise density=0.2)	0.98	0.24	37.88
Salt & pepper(Noise density=0.001)	1	0	35.08
Speckle noise (Noise density= 0.001)	0.96	0.14	36.40
JPEG compression (80%)	1	0	37.50
Median filter [3x3]	1	0.14	37.88
Median filter [5x5]	1	0	37.88
AWGNA (mean=0 & variance=10^{-4})	1	0.14	35.86
AWGNA (mean=0 & variance=3x10^{-4})	1	0.14	35.46

4.2.2 Performance Analysis

To evaluate the authentication performance of the suggested hybrid framework, the genuine and impostor matching score distribution, the FAR versus FRR, and the ROC curves are examined. Figure 4.4 illustrates the genuine and impostor matching score distribution curves obtained for (a) *Dataset1*, (b) *Dataset2*, and (c) *Dataset3*, respectively. It can be noticed from the three sub-figures the lack of overlap between the genuine and impostor matching score distributions, which demonstrates the ability of the proposed system to distinguish between authorized and unauthorized persons.

Figure 4.5 illustrates the FAR versus FRR for the three datasets. As shown in the three sub-figures, an EER value of 0.24% for the threshold 0.6541, 0% for the threshold 0.6483, and 0.26% for the threshold 0.6545 is obtained for *Dataset1*, *Dataset2*, and *Dataset3*, respectively. Which proves the efficiency of the proposed hybrid system in minimizing failures when rejecting authorized users or accepting unauthorized users.

The ROC curves presented in figure 4.6 illustrate the face-based unimodal authentication system, the fingerprint-based unimodal authentication system, and the face-fingerprint-based multimodal authentication system for (a) *Dataset1*, (b) *Dataset2*, and (c) *Dataset3*. The curve similarity between the face-based unimodal system and the multimodal system indicates that the watermarking process does not affect the recognition performance. Moreover, better performance is observed in the multimodal authentication system compared to the fingerprint-feature-based unimodal system, attaining accuracy of 100%.

4.2.3 Diversity and Revocability Analysis

Revocability and diversity are one of the essential properties required to provide biometric template protection against system database attacks. The revocability is the ability to cancel a biometric template that has been breached and create a new one. While the

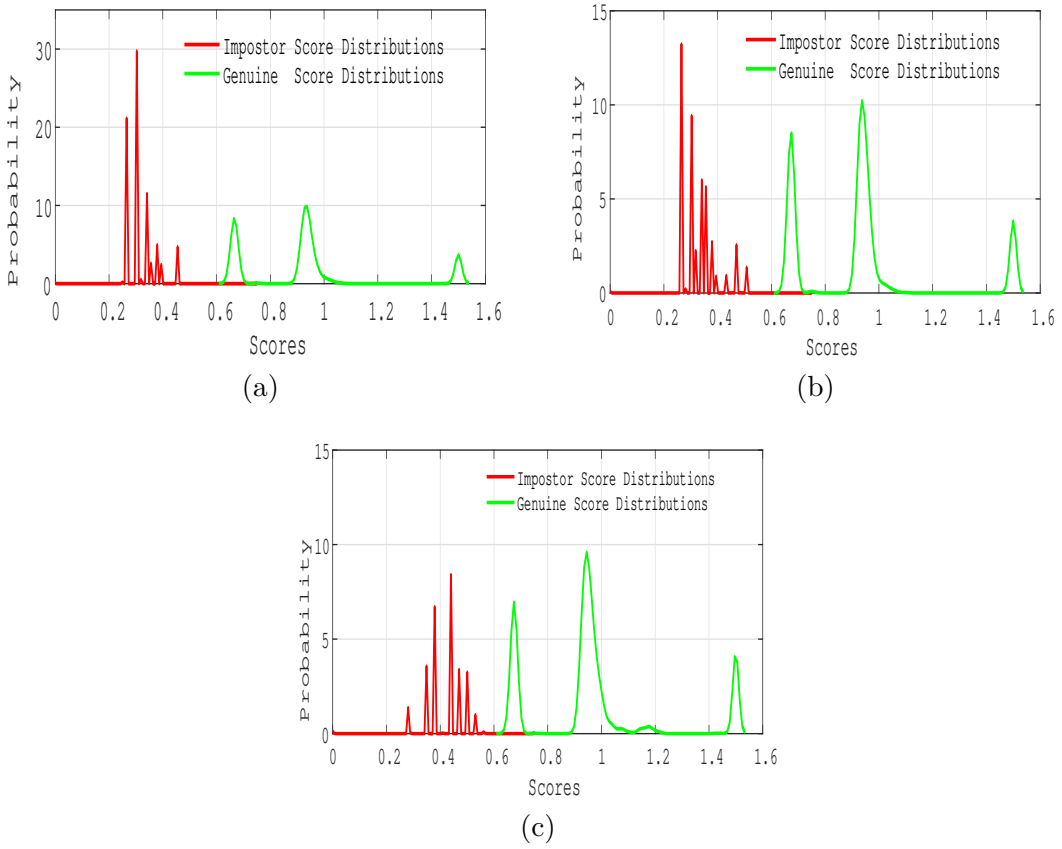


FIGURE 4.4: Genuine and imposter matching score distribution curves of face-fingerprint-based multimodal authentication system for (a) *Dataset1*, (b) *Dataset2*, and (c) *Dataset3*

diversity refers to the ability to enrol into different applications using the same biometric data without the risk of any cross-matching between their corresponding databases [57]. The typical method to provide that a template protection approach satisfies the revocability and diversity properties is to generate multiple protected templates from the same original templates and different keys.

It turned out that biometric templates produced using the partial Hadamard transform satisfy the diversity and revocability criteria [9]. Implying that applying this non-invertible transform generates complex templates that do not reveal any information about the original biometric data. To verify this, tests are conducted to investigate whether reissued templates that originate from the same fingerprint are correlated. More precisely, we generate multiple protected templates from the same original templates and different keys, then attempting to match the resulting sets of protected templates.

To measure the diversity of the proposed system, the cross-matching is derived using the True Rejection Rate (TRR). This metric computes the percentage of time the system (correctly) rejects a user with a different key. Higher the TRR value, good diversity is meet. The TRR values obtained in our experiments are 98.0%, 99.50%, and 99.35% for *Dataset1*, *Dataset2*, and *Dataset3*, respectively. Which demonstrates the ability of

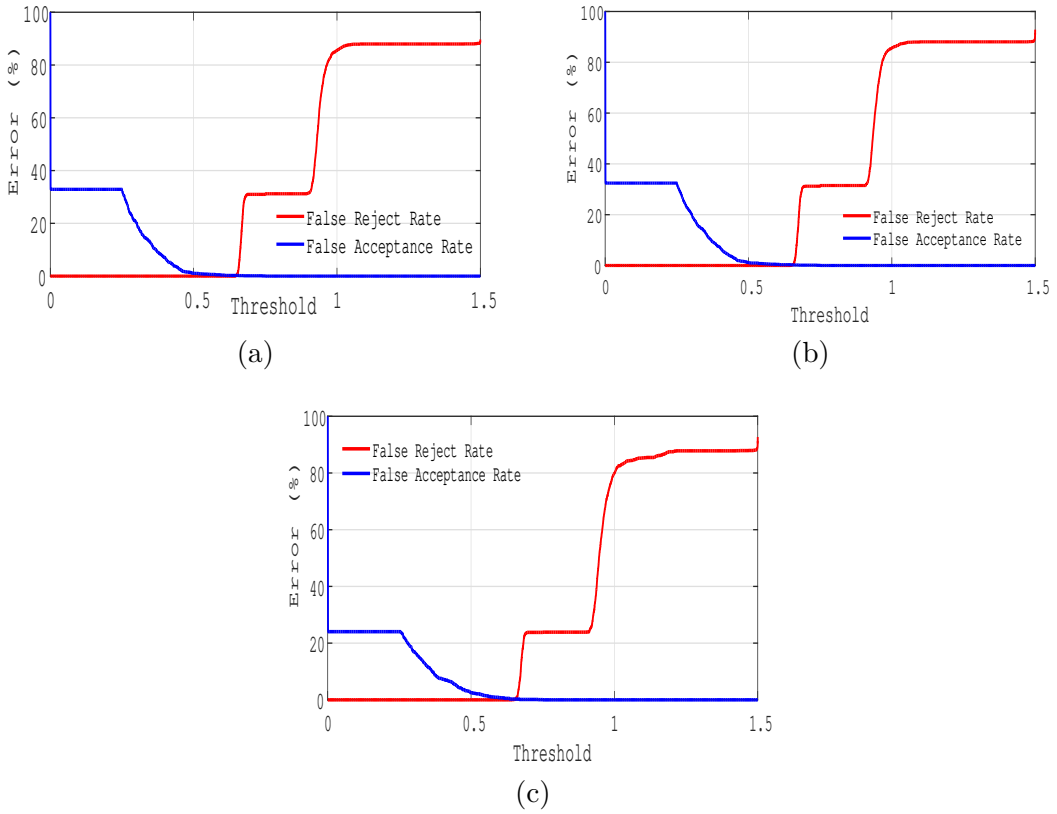


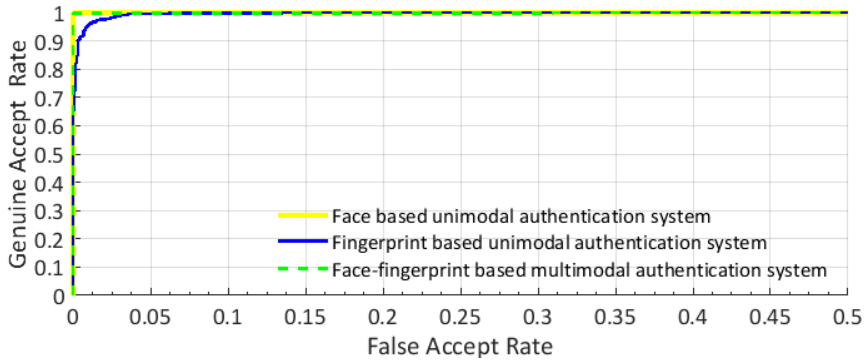
FIGURE 4.5: FAR vs. FRR curves of face-fingerprint-based multimodal authentication system for (a) *Dataset1*, (b) *Dataset2*, and (c) *Dataset3*

the proposed system to be integrated into the various applications without databases cross-matching.

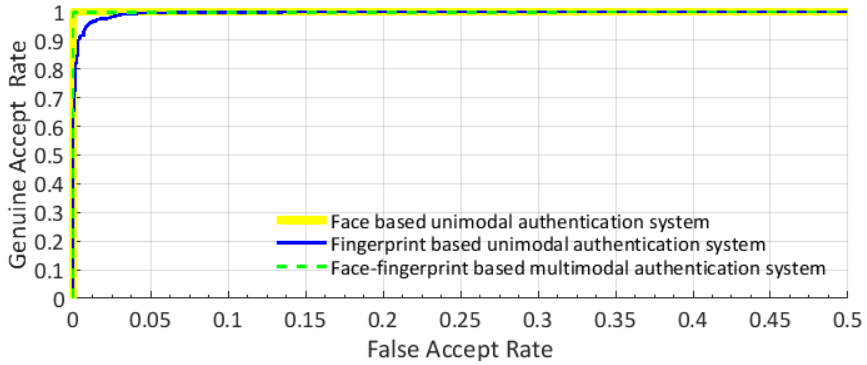
Regarding the revocability property, the matching between the protected templates of generated databases is accomplished using the EER metric. Retrieving lower EER values is referred to as achieving good revocability. Conducting the experiments tests, an EER value of 0% is obtained for each of the three datasets, indicating that the intra-class variation is preserved. Therefore, cancelling a template that has been compromised and generating a new one, does not affect the authentication system.

4.2.4 Comparison with Previous Studies

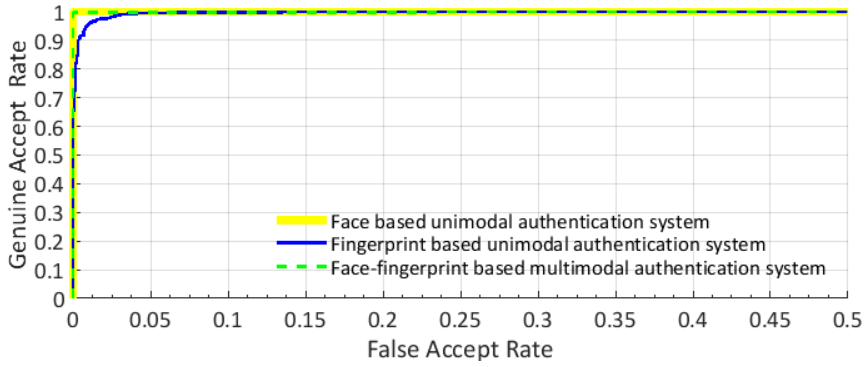
In this section, a comparison of the proposed approach with the related studies is presented. Table 4.2 illustrates the existing works related to the DTCWT-based watermarking algorithm and the Hadamard transform. The approaches are presented along with the biometric modalities, datasets, and performance results. Due to the differences in the based biometric modalities and used databases, a valid comparison cannot take place among the approaches.



(a)



(b)



(c)

FIGURE 4.6: Genuine and imposter matching score distribution curves of face-fingerprint based multimodal authentication system for *Dataset1*, *Dataset2*, and *Dataset3*

TABLE 4.2: Summary of different biometric template protection approaches.

Ref.	**Biometric Modality**	**Database**	**Technique**	**Performance (%)**	**Reqs**[1]
[31]	Fingerprint	CASIA	Dual-Tree Complex Wavelet Transform	—	S

[32]	Face	ORL JAFFE L-SPACEK CMU-PIE	DTCWT & Fast Fourier Transform	EER=4 EER=10 EER=10	P
[39]	—	—	DTCWT & Rao-test structure	EER=0.01	S
[72]	Fingerprint	FVC2002 DB2 FVC2002 DB3	Hadamard Transform	EER=3 EER=9.12	S
[73]	Fingerprint	FVC2002 DB1 FVC2002 DB2 FVC2002 DB3	Partial Hadamard Transform	EER=1 EER=2 EER=5.2	RDP
[74]	Palmprint Face	ORL Indian Face YALE PolyU CASIA	Hadamard Transform, Achlioptas matrices, one-way modulus hashing	EER=6.90 EER=11.53 EER=8.91 EER=0.56 EER=2.50	SPR
[53]	Fingerprint Iris	CASIA v1.0	Chaos & Hadamard matrices	—	SPR
[75]	Face Fingerprint	FVC2002 DB2 ORL	DTCWT-DCT watermarking	EER=0.96	SP
Our Method	Face Fingerprint	*Dataset1* *Dataset2* *Dataset3*	DTCWT-DCT watermarking, partial Hadamard Transform	EER=0.24 EER=0 EER=0.26	SPRD

We can observe from the table that the suggested approach outperforms the methods introduced in [31, 32, 39, 72, 75]. More precisely, the aforementioned existing methods meet only the security and/or performance requirements, while our approach satisfies the security, revocability, diversity, and performance requirements. Our approach performs better due to, it bases on a hybrid mechanism that fuses the biometric watermarking and the partial Hadamard transform. The watermarking is known by the ability to provide the security and good authentication performance, while the non-invertibility of the partial Hadamard transform affords the diversity and revocability requirements.

Additionally, It is noticed that the EER value of the proposed approach is lower compared to the EER values achieved in [32, 72–74]. This better performance is due to the multi-modality mechanism. More precisely, the multi-modality is gained in our biometric system via two fusion levels. The first is the fusion of face and fingerprint modalities via the watermarking algorithm. While the second is the fusion at the matching score

[1]Requirements (S: Security, R: Revocability, D: Diversity, P: Performance)

level. Herein, the facial and fingerprint feature vectors are processed separately to get individual matching scores. Then, depending on the accuracy of each biometric modality, the final matching is carried out. Which improves the authentication performance of the system.

Moreover, our method extends the work indicated in [75], by improving the approach to meet the four requirements. More precisely, the work in [75] suggested the DTCWT-DCT-based watermarking method to secure the system and improve the authentication performance. While the current approach combines the same watermarking algorithm with the partial Hadamard transform to achieve the four criteria at a time.

5 Conclusion

In this chapter, we suggested a hybrid approach to secure face-fingerprint-based authentication systems against template database attacks. To achieve this, four criteria should be satisfied, including security, performance, diversity, and revocability. Therefore, the partial Hadamard transform is first used as a cryptographic technique to cover the original fingerprint features designed as a watermark. Next, the DTCWT-DCT-based watermarking method is employed to embed the watermark into the facial image considered as a cover image. The partial Hadamard transform is based to exploit the non-invertibility of the extracted fingerprint features, which leads to satisfying diversity and revocability. While the watermarking method is employed to fuse the face and fingerprint modalities for higher performance and better security.

The experiments evaluation are conducted on the ORL face database and three FVC fingerprint databases, namely FVC2000 DB1 and FVC2002 DB1, DB2. The experimental results had proved the efficiency of the approach in term of satisfying the all biometric authentication systems' criteria at a time.

5 Hybrid Multimodal Biometric Template Protection

1 Introduction

The current chapter proposes a hybrid approach to secure biometric templates in multimodal authentication systems. The concept underlying this approach is to combine (1) a DTCWT-DCT-based watermarking algorithm, (2) the secure sketch algorithm, and (3) a 3D chaotic map image encryption method for securing face and fingerprint modalities. Fingerprints are chosen because they provide a high verification rate, while face modality is typically used in our daily recognition tasks.

The motivation behind combining these three techniques is to increase the security and efficiency of biometric authentication systems against template database attacks. This is due to, using the biometric watermarking offers a high-security level where the embedded data would not be recovered only via a secret key. Besides, fusing two biometric modalities using the watermarking algorithm increases the authentication performance. Moreover, the embedded data are kept linked to the cover image. Thus, no additional transfer resources or storage mechanisms are needed. The secure sketch algorithm is based to provide security of the original biometric data. While the 3D chaotic map image encryption method is chosen owing to its sensitivity to initial conditions, non-periodicity, non-convergence, control parameters, and robustness against brute force attacks.

The process of the approach adheres to the following steps: first, the secure sketch is used to cover the fingerprint features. Next, because of secure sketch non-revocability and incapability to model intra-user variations, the blind DTCWT-DCT-based watermarking method is employed to embed the secured fingerprint features (used as watermark) into the facial image (used as a cover image). The underlying concept in this technique is to fuse the fingerprint features and face images to relate each to other. The DTCWT domain frequency decomposition is used due to the high-frequency sub-bands that provide additional security level during the watermark integration. Further, the DCT scheme

is used given its withstanding various watermarking attacks such as noising, compressing, sharpening, and filtering. Finally, the 3D chaotic map image encryption method is employed to secure the facial watermarked image. The encryption metric used in this study is a coupling of a non-linear 3D chaos-based simple encryption technique, pixel position permutation, and pixel value transformation.

To evaluate the proposed approach, extensive experiments have been carried out on the ORL face database and three FVC fingerprint databases, including FVC2002 DB1, DB2, and FVC2000 DB1. The experimental results demonstrated a high level of security and performance. Moreover, the diversity and revocability constraints are achieved indicating the ability of the approach to be embedded into various applications without databases cross-matching. Hence, it makes a significant improvement compared with the existing related studies in terms of meeting the four requirements at a time.

2 Overall Hybrid Approach Description

Figure 5.1 illustrates an overview of the proposed hybrid template protection approach. As shown in this figure, the algorithm comprises two phases: Enrolment and Authentication.

Enrollment phase: In this stage, we applied the secure sketch encoder to the extracted and preprocessed fingerprint features to ensure their security, thus, sketch P is generated. Next, we used the watermarking algorithm to embed the sketch P considered as a watermark into the facial image used as a cover image. Further, the watermarked facial image (WI) is then encrypted using a chaos-based simple encryption technique via a user key K to enhance the security level. Furthermore, the encrypted watermarked facial image (EI) is warehoused in the database as a public reference.

Authentication phase: Herein, the genuine user presents a key K' supposed to be similar to the enrolment key K. In case the keys are not identical, it is claimed as an impostor probe. Otherwise, the user presents his/her biometric data, face, and fingerprint (query images). Next, the stored reference image EI is retrieved from the database and decrypted using the key K', to recover the watermarked facial image WI. Then, two parallel matching processes are carried out.

On the one hand, face features are extracted from both the reference and query facial images using the OLPP metric. Consequently, two feature sets F' and F are respectively obtained, binarized, and matched using the Hamming distance metric, which generates the facial modality score S_{face}.

Meanwhile, the watermark sketch P is extracted from the decrypted watermarked face image. Subsequently, the secure sketch decoder function is applied to generate a features

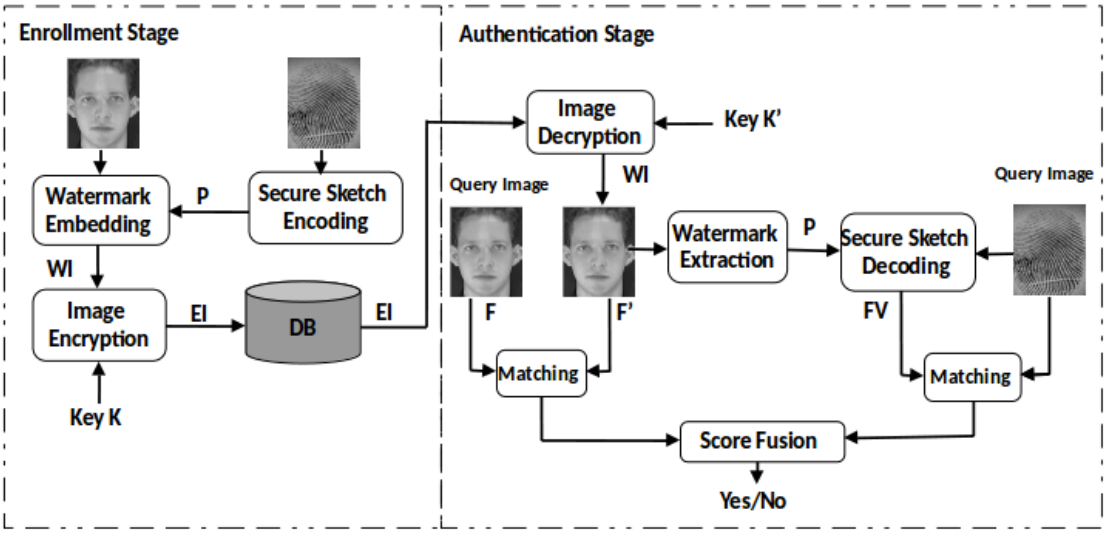


FIGURE 5.1: Proposed hybrid biometric template protection approach.

vector FV to be matched with the features vector of the fingerprint query image using the Hamming distance metric, which generates the fingerprint modality score $S_{fingerprint}$.

To determine whether the authentication is successful, the score-level fusion PAN method is used to compute the individual matching scores of face and fingerprint modalities [65]. Next, the scores are fused via the application of the weighted sum rule. The weight of υ is allocated to the score of the face modality S_{face}, while the weight of ω is allocated to the score of the fingerprint modality $S_{fingerprint}$. Finally, the matching score of the multimodal biometric authentication system is computed as:

$$S_m = \upsilon S_{face} + \omega S_{fingerprint} \tag{5.1}$$

The biometric authentication system accepts or rejects a user by comparing the matching score S_m to a threshold η.

3 Hybrid Approach Components Description

In this section, the main components of the suggested approach are described, including secure sketch, watermarking, and image encryption algorithms.

3.1 Secure Sketch Algorithm

The direct minutiae matching process is avoided in the proposed secure sketch algorithm due to two primary reasons. First, the proposed authentication system uses an extensive database, implying that, the minutiae-based matching strategy would not satisfy the

fast-performance speed requirement. Second, fingerprint minutiae representation cannot be applied directly to a template-based secure sketch, since the minutiae extraction does not provide the same number of minutiae for different fingerprint images, whereas a fixed-length feature vector is needed as an input in a secure sketch algorithm. To solve these problems, the spectral minutiae representation is based rather than the minutiae sets.

As illustrated in figure 5.2, after being elicited from the fingerprint image via the minutiae extraction system presented in [63], the minutiae points are depicted as a spectral features vector of fixed length. More precisely, suppose that Z minutiae points are extracted from the fingerprint image, with (x_i, y_i, θ_i) are the location and orientation of the i^{th} minutia and $(\omega_x, \omega_y, \sigma_o^2)$ are the frequencies and the parameters of the Gaussian kernel function, respectively. $m_i(x, y, \theta)$ is the derivative of $m_i(x, y) = \delta(x - x_i, y - y_i)$ in the direction θ_i, and δ is the Dirac Pulse metric. The minutiae exemplification is carried out via the following equation (5.2) [70]:

$$\mathcal{M}_o(\omega_x, \omega_y, \sigma_o^2) = |e^{-(\frac{\omega_x^2 + \omega_y^2}{2\sigma_o^2})} \times \sum_{i=1}^{Z} \mathcal{F}(m_i(x, y, \theta))| \tag{5.2}$$

Where $\mathcal{F}$ is the Fourier transform defined as:

$$\mathcal{F}(m_i(x, y, \theta)) = j(\omega_x \cos\theta_i + \omega_y \sin\theta_i) \times e^{-j(\omega_x x_i + \omega_y y_i)} \tag{5.3}$$

As the obtained minutia representation is of sizable dimensionality, the CPCA features reduction method [69] is utilized to lessen the spectral features vector size. This metric is based to obviate the large dimensionality problems such as template storage and computational burden requirements. Next, the minimized features vector is quantized and joined with the Reed-Solomon codeword SR via the exclusive OR (XOR) to get the sketch P.

In the secure sketch decoding phase, the same process is applied to the query fingerprint image to obtain the features vector. This vector is subsequently XORed with the sketch P retrieved from the system's database. The result is subsequently decoded using the Reed-Solomon decoder function to retrieve the codeword SR' that is supposed to be near to the codeword SR in the case of correct authentication. Finally, the features vector is restored by XORing P and SR'. To determine whether the authentication is successful, the obtained features vector is compared to the query fingerprint image features vector according to the Hamming distance.

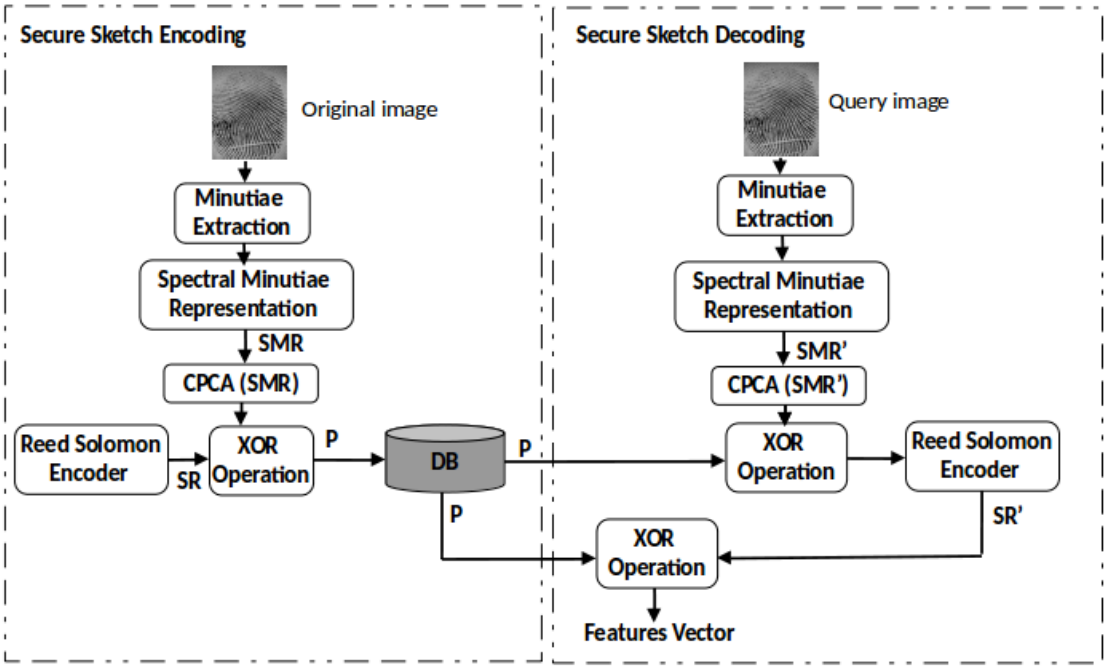


FIGURE 5.2: Block diagram of the secure sketch scheme for fingerprint template.

3.1.1 DTCWT-DCT Watermarking Algorithm

The watermarking algorithm used in this study is similar to the one based in chapter 4. It relies on DTCWT and DCT as two domain transform mechanisms. The concept based in this algorithm is to entrench the watermark with less disfigurement in the facial image while allowing blind watermark extraction via the correlation method. This algorithm comprises two stages including watermark embedding and watermark extraction.

The watermark embedding process starts by dismantling the facial image using one level DTCWT decomposition. Next, three high-frequency sub-bands are arbitrarily selected and divided into 4×4 blocks. Then, the DCT transform is applied on each 4×4 blocks. After that, the watermark bits are integrated with a gain factor α into the DCT transformed 4×4 blocks using two generated pseudorandom series PN_0 and PN_1 to integrate the watermark bits 0 and 1, respectively. Finally, to construct the watermarked facial image, the inverse DCT and inverse DTCWT are executed, respectively.

In the watermark extraction process, the same procedure as in the watermark embedding stage is carried out on the watermarked facial image to obtain the DCT transformed blocks. Then, for each DCT transformed division, the mid-band coefficients are retrieved and two correlations C_0 and C_1 are computed with PN_0 and PN_1, respectively. In case $C_0 < C_1$, the readout watermark bit is deemed 1; otherwise, it is deemed 0.

3.2 3D Chaos Image Encryption Algorithm

To encrypt the watermarked image, a 3D chaos-based encryption method with pixel position permutation and value transformation is used.

The chaos map used in this work is the logistic map. It is chosen due to its simplicity and reduced complexity. Equation (5.4) illustrates the 1D logistic map. The 3D formula is given in (5.5), where $0 < x_n < 1$ and $\mu = 4$ are the conditions to render the equation chaotic; $0 < \beta <$0.022, and 0$< \alpha <$0.015 are the growth rates, and the initial values of, x, y, and z, respectively, in between 0 and 1.

$$x_{n+1} = \mu x_n(1 - x_n) \tag{5.4}$$

$$\begin{aligned} x_{n+1} &= \gamma x_n(1 - x_n) + \beta y_n^2 x_n + \alpha z_n^3 \\ y_{n+1} &= \gamma y_n(1 - y_n) + \beta y z_n^2 y_n + \alpha x_n^3 \\ z_{n+1} &= \gamma y_n(1 - z_n) + \beta x_n^2 z_n + \alpha y_n^2 \end{aligned} \tag{5.5}$$

Suppose $M \times N$ are the dimensions of the image to be encrypted. [x_0, y_0, z_0, α, β, γ, *N1*, *N2*, *N3*, *N4*, *N5*, *N6*], is the encryption/decryption used key, where x_0, y_0, and z_0 are the initial values of the 3D logistic map population x, y, and z, respectively. *N2*, *N4*, *N6* are large random numbers used to equalize the histogram of the generated logistic map (see equation (5.6)), while *N1*, *N3*, *N5*, are used as the first index to select the chaos x, y, and z respectively.

3.2.1 Image Encryption Process

As illustrated in figure 5.3, the proposed encryption strategy comprises the following stages [76]:

- Generate a 3D logistic map based on equation (5.5) to obtain the logistic map x, y, and z.
- Equalize the histogram of the logistic map using equation (5.6) to improve the security level. The histogram equalization is performed owing to the non-uniform distribution of the logistic map histogram.

$$\begin{aligned} x &= (integer(x * N2))mod(N) \\ y &= (integer(y * N4))mod(M) \\ z &= (integer(z * N6))mod(512) \end{aligned} \tag{5.6}$$

- Permute the image pixels via a row rotation approach. This starts by generating a large random number *N1*; subsequently, select M number of chaos x starting

from index *N1*. Next, rotate the pixels based on chaos value x. If the x value is even, rotate to the left; if it is odd, rotate to the right.

- Permute the image pixels via a column rotation approach. This starts by generating a sizeable random number *N3*; subsequently, select N numbers of chaos y from index *N3*. Perform rotation according to the y value. If the y value is even, rotate up; if it is odd, rotate down.

Applying row and column rotation generates an encrypted image with an unaltered histogram, that may fail towards histogram attacks. To overcome this issue, the XOR operation is applied as an added step. The concept underlying this step is to change the image pixel value to a new value and prevent the retrieval of the original value without the chaos key. Therefore, a large random number *N5* is generated; subsequently, the $M \times N$ image is reshaped into a $1 \times MN$ vector. Finally, the XOR operation is applied between the $1 \times MN$ vector and the z chaos from the index *N5* to obtain the encrypted image.

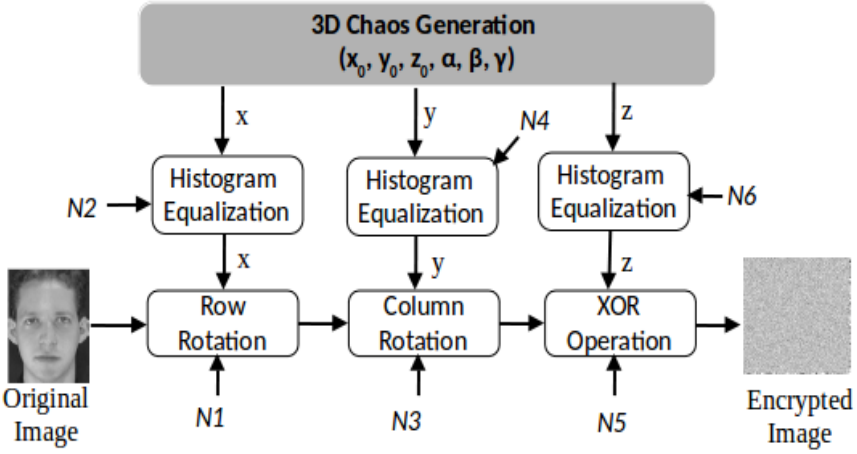


FIGURE 5.3: Encryption technique using 3D chaos.

3.2.2 Image Decryption Process

In the decryption phase, the encrypted image is first reshaped into a $1 \times MN$ vector. Then, it is XORed with the z chaos using *N5*. Next, the inverse column rotation and inverse row rotation are applied. In the inverse column rotation, the $1 \times MN$ vector is reshaped into an $M \times N$ image; subsequently, based on *N3* and the y value, the rotation process is applied. If the y value is even, we rotate down; if it is odd, we rotate up. To get the original image, the inverse row rotation is applied using *N1* and the x value. If the x value is even, we rotate to the right; if it is odd, we rotate to the left. Figure 5.4 illustrates the image decryption stages of the 3D chaos image encryption method.

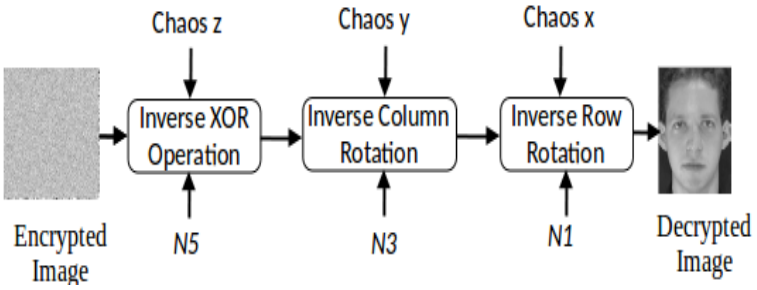


FIGURE 5.4: Decryption technique using 3D chaos

4 Experimental Analysis

4.1 Experimental Setup

Extensive tests have been carried out to assess the efficiency of the suggested hybrid approach using the ORL face database and three FVC databases: FVC2002 DB1, DB2, and FVC2000 DB1. 512×512 pixels facial images and fingerprint images of size 374×388, 560 × 296, and 300 × 300 pixels are obtained from FVC2002 DB1, DB2, and FVC2000 DB1, respectively. These images were divided into three datasets: *Dataset1* (ORL with FVC2002 DB1), *Dataset2* (ORL with FVC2002 DB2), *Dataset3* (ORL with FVC2000 DB1).

Applying the spectral minutiae representation on the extracted minutiae points, feature matrices of dimensions $128 \times 256 = 32768$ are obtained. After stratifying the CPCA features reduction metric where only the pertinent features are selected and the feature quantization method, a binary stream (sketch P) of length 10240 bits is generated.

In the watermarking stage, only three high-frequency sub-bands of size 256×256 are randomly selected from the six high-frequency sub-bands generated via the one level DTCWT decomposition. The decomposition of each selected high-frequency sub-band produces 64×64 blocks of dimension 4×4 each. Moreover, to get an optimal watermarking imperceptibility and good robustness against image watermarking attacks, the watermarking algorithm has been tested for different values of the gain factor α among a range of 0 to 60. The conducted experiments demonstrated that the convenient values are $\alpha = 20$, $\alpha = 30$, and $\alpha = 40$ for the first, second, and third high-frequency sub-bands, respectively. Besides, various values among a range of 0 to 1 are assigned to weights values υ allocated to the score of the facial system and ω allocated to the score of the fingerprint system. The values of $\upsilon = 0.5$ and $\omega = 0.5$ which provide the maximum authentication performance are chosen. Moreover, as the mid-band frequencies of a DCT block have seven coefficients, pseudorandom sequences PN_0 and PN_1 of length 7 are generated. To implement the encryption approach, *N2*= *N4*=*N6*=100000, *N1*=500, *N3*=600, and *N5*=700 are considered.

4.2 Results and Discussions

In this section, we put forward and talk over the evaluation results of the carried out experiments to prove the efficiency of the proposed approach in meeting the four requirements of biometric authentication systems.

4.2.1 Security Analysis

Three scenarios are considered below to assess the security of the suggested approach.

4.2.1.1 Compromised Encrypted Image Scenario

Suppose that the database is attacked and the encrypted facial image is stolen. It is typically computationally difficult for the attacker to estimate the key and decrypt the image due to the encryption key's large keyspace. More precisely, in the suggested 3D chaotic map encryption method, six initial conditions x_0, y_0, z_0, α, β, and γ with precision 10^{-16} are used. Consequently, the keyspace size is $(10^{16})^6$. Additionally, *N1*, *N2*, *N3*, *N4*, *N5*, and *N6* are used as random keys of precision 10^5, implying that the keyspace size is $(10^5)^6 = 10^{30}$. Consequently, the keyspace size is larger than 10^{126}, which is sufficiently immense to withstand exhaustive attacks. Moreover, the suggested image encryption method is sufficiently secure to be sensitive to slight changes in the decryption key.

4.2.1.2 Compromised Watermarked Image Scenario

Suppose that the encrypted image is compromised and decrypted to get the watermarked facial image. The conducted experiments proved the strength of the suggested watermarking algorithm against conventional digital image watermarking attacks, namely JPEG compression, Gaussian noise, speckle noise, salt and pepper noise, median filter, and additive white Gaussian noise. The evaluation is then reported using the correlation coefficient, PSNR, and EER. Table 5.1 presents the average of the EER, PSNR, and correlation coefficient values under the aforementioned attacks. From this table, it can be observed that the authentication performance of the watermarking algorithm was not affected by the attacks, where the EER value has been slightly minimized for some attacks. Besides, the PSNR and correlation coefficient values are slightly changed after applying attacks, which demonstrates the watermarking algorithm's ability to withstand these attacks. However, the DTCWT-DCT-based watermarking algorithm does not preserve higher perceptual quality between the original and watermarked face images. This is due to the trade-off between the quality of the extracted watermark after embedding and the quality of the watermarked facial image. More precisely, improving the quality

of the watermarked image decreases the quality of the extracted watermark and vice versa.

TABLE 5.1: Correlation coefficient, EER, and PSNR values before and after attacks.

Attacks	Correlation coefficient	EER(%)	PSNR(dB)
No attack	1	0	30.88
Gaussian noise (Noise density=0.2)	0.98	0.24	30.86
Salt & pepper(Noise density=0.001)	0.99	0	29.69
Speckle noise (Noise density= 0.001)	1	0.14	29.92
JPEG compression (80%)	1	0	30.77
Median filter [3x3]	1	0.14	30.46
Median filter [5x5]	1	0	30.46
AWGNA (mean = 0 and variance = 0.0001)	1	0.14	29.45
AWGNA (mean = 0 and variance = 0.0003)	1	0.14	30.32

4.2.1.3 Compromised Sketch Scenario

Assuming that the encrypted image is compromised and decrypted, and the embedded sketch is extracted. In this case, the attacker attempts to reveal the original fingerprint features vector. The robustness of the secure sketch algorithm depends on how difficult it is to esteem the original fingerprint features given that the sketch is revealed. Since sketch P is generated using the fingerprint images, the *entropy loss* is calculated only over the FVC databases. In our experiments, the informal definition of the *entropy loss* known also as the mutual information is used. It measures the quantity of information communicated on average between the sketch P and the original fingerprint features vector F. It is defined as the following formula:

$$I\left(F;P\right)=\sum_{f\in F}\sum_{p\in P}P(F,P)\log\left(\frac{P(f,p)}{P(f)P(p)}\right)$$

Where $P(F)$ and $P(P)$ are the marginal distributions of F and P.

In our experiments, first, the mutual information is calculated for each couple (F,P) of each image. Then the average of the mutual information is calculated for all images of each dataset. The average of the mutual information obtained for each dataset is 2.37×10^{-4}. The gained value shows that a tiny amount of mutual information is shared between the sketch and the original data, indicating that even when the sketch is extracted, the features vector information could not be leaked to the attacker.

4.2.2 Performance Analysis

Different curves are plotted for the three datasets to assess the performance of the proposed system, including the genuine and impostor matching score distribution curves,

the FAR against the FRR curves, and the ROC curves. The genuine and impostor scores are generated via the FVC standard protocol. Figure 5.5 presents the genuine and impostor distributions related to the suggested hybrid system for the three datasets. It is noteworthy from the sub-figures that the scores are not overlapped, demonstrating the presented approach's efficiency in discriminating between genuine and impostor users. The curves are well separated for *Dataset1* and *Dataset2* at the threshold's value 0.4517 and 0.4514, respectively, and slightly overlapped for *Dataset3* at the threshold range [0.8879, 1.0521].

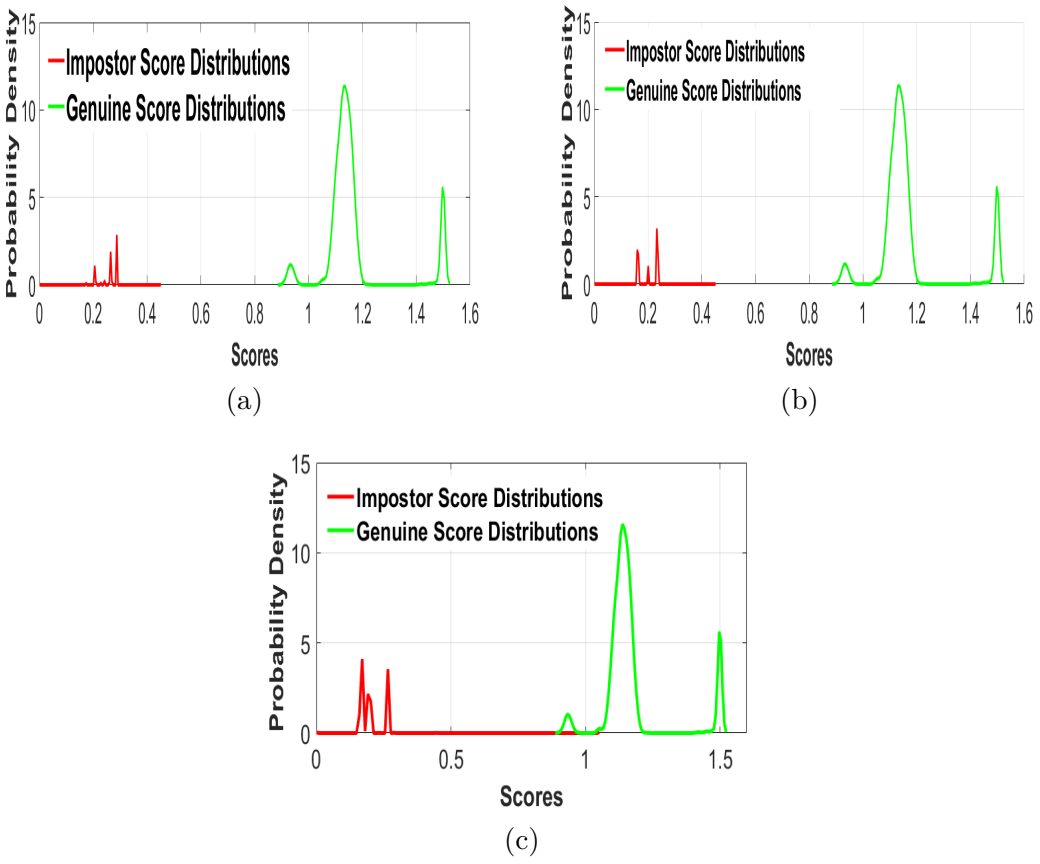


FIGURE 5.5: Genuine and imposter matching score distribution curves of face-fingerprint-based multimodal authentication system for (a) *Dataset1*, (b) *Dataset2*, and (c) *Dataset3*.

Besides, the approach's authentication performance is attested by computing the EER and the FAR as opposed to the FRR values. The smallest of the value of the EER is the better, showing that the system is less likely to falsely admit impostors as genuine or falsely reject genuine as impostors. Figure 5.6 shows the FAR as opposed to the FRR curves of the proposed system for the three different datasets. As can be observed from this figure, the proposed method reached lower EER values of 0%, 0%, and 0.12% on *Dataset1*, *Dataset2*, and *Dataset3*, respectively. Also, FRR/FAR values of 0.0083/0.2140% for the threshold 0.4512, 0.0080/0.2141% for the threshold 0.4510, and 0.0095/0.2074% for the threshold 0.9184 are obtained over the three databases,

respectively. These results demonstrated the high reliability and performance of the authentication system.

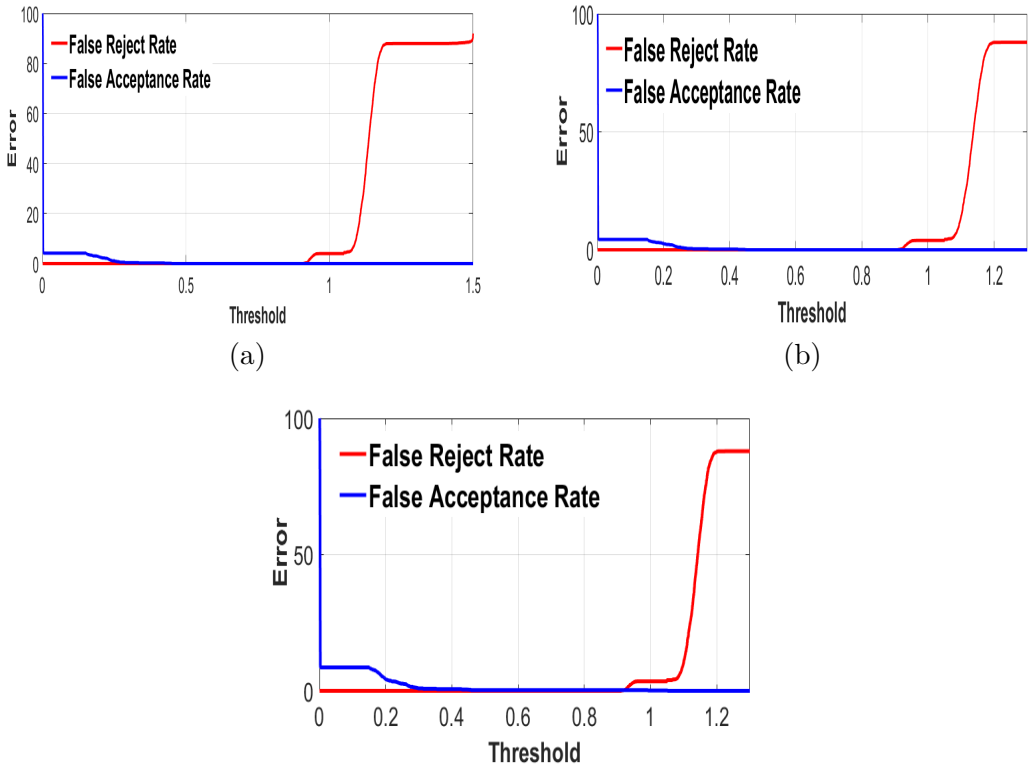


FIGURE 5.6: FAR vs. FRR curves of the face-fingerprint-based multimodal authentication system for (a) *Dataset1*, (b) *Dataset2*, and (c) *Dataset3*.

The ROC curves of the face-based unimodal authentication system, the fingerprint-based unimodal authentication system, and the proposed fingerprint-face-based multimodal authentication system for (a) *Dataset1*, (b) *Dataset2*, and (c) *Dataset3* are illustrated in figure 5.7. It can be seen from the three sub-figures that the fingerprint-based unimodal authentication system performs weaker compared to the face-based unimodal authentication system. This can be explained by the number of extracted minutiae from each fingerprint image, as the extraction process does not always provide the same number of minutiae for different fingerprint images of the same user. Additionally, significant performance improvement is shown in the fingerprint-face-based multimodal authentication system as an accuracy of 100% is achieved. The improved performance is due to two reasons: the combination of two different modalities and the effectiveness of the employed score fusion method.

4.2.3 Diversity and Revocability Analysis

A template protection scheme meets the revocability and diversity requirements if it generates multiple uncorrelated protected templates from the same original features

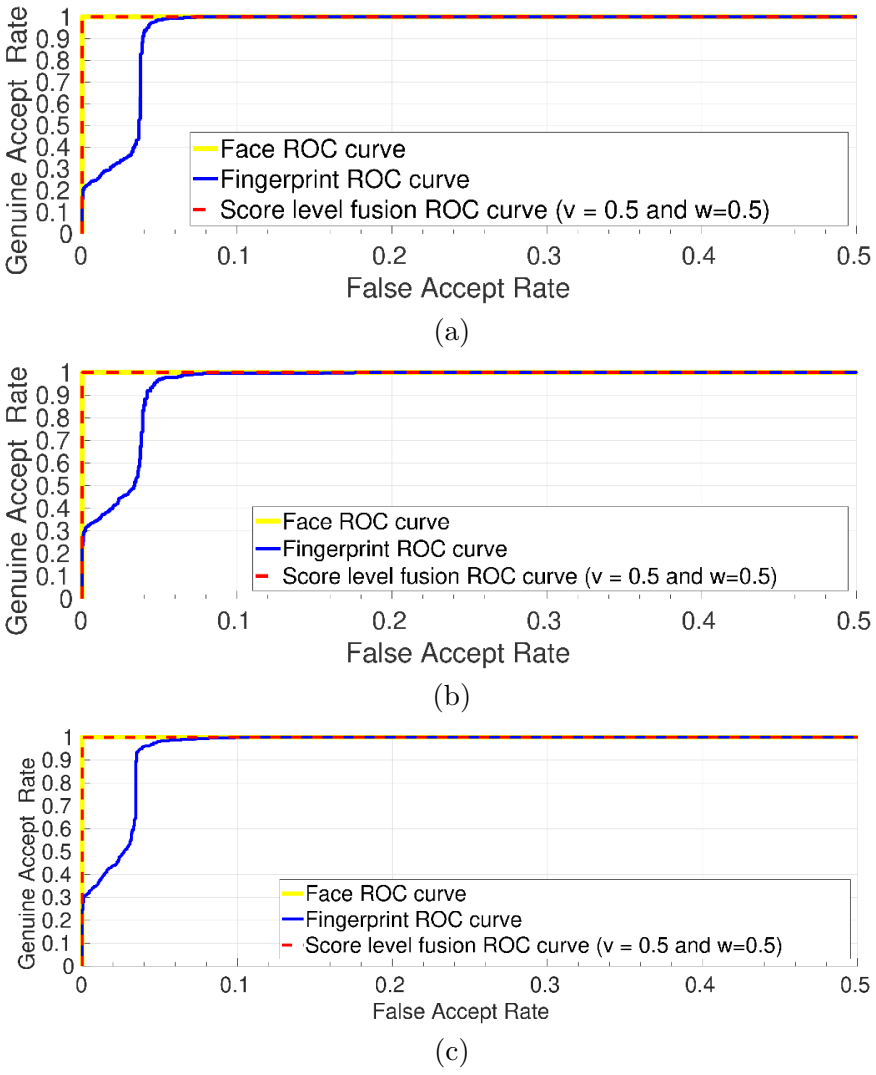


FIGURE 5.7: ROC curves of the face-based unimodal authentication system, the fingerprint-based unimodal authentication system, and the fingerprint-face-based multimodal authentication system for (a) *Dataset1*, (b) *Dataset2*, and (c) *Dataset3*.

vectors, allowing users to register in distinct applications based on the same features without the risk of cross-matching between the corresponding databases. In comparison, the revocability requirement indicates the capability to repeal a modality that has been breached and generate a novel one.

The typical means to provide that a template protection approach satisfies the revocability and diversity properties is to generate multiple protected templates from the same original templates and different keys. To simulate this scenario, we generated two databases of protected templates from the same original face and fingerprint features and different keys. Subsequently, attempting to cross-match the resulting templates of both generated databases.

Regarding the diversity criterion, the cross-matching is derived using the TRR. Higher the TRR value, good diversity is meet. In our experiments, the obtained TRR values

are 99.50%, 99.35%, and 98.94% for *Dataset1*, *Dataset2*, and *Dataset3*, respectively. Which proves the suggested approach to be embedded into various applications without databases cross-matching.

Regarding the revocability property, the matching between the protected templates of generated databases is accomplished using the EER metric. Retrieving lower EER values is referred to as achieving good revocability. An EER value of 0% is obtained for each of the three datasets, indicating that the intra-class variation is preserved. Therefore, canceling a template that has been compromised and generating a new one will not change or affect the authentication system.

4.2.4 Computational Complexity

In order to analyze the computational complexity of the proposed approach, the run time has been recorded during the authentication stage, which is the most crucial phase required to be carried out in real time. As illustrated in figure 5.1, this phase includes the image decryption, watermark extraction, and secure sketch decoding. All the algorithms were implemented in MATLAB and run on an Intel Core(TM) 2 Duo CPU P7450 at 2.13 GHz with 4 GB of memory. The run time is recorded during the experiments carried out over all the databases of 320 images. The average time obtained is 2.11s for the image decryption, 1.36s for the watermark extraction, and 3.28s for the sketch decoding process. By adding the amount of time needed to perform the matching process of 0.31s, a total average of 7.06s is attained for the overall authentication phase.

4.2.5 Comparison with Previous Studies

In the current section, a comparison of the suggested approach and the existing studies is presented. However, related studies used various databases and biometric modalities to endorse the efficiency of their suggested methods. Consequently, a useful comparison with our hybrid approach cannot be carried out. Table 5.2 outlines the biometric template protection methods suggested in the literature (based on watermarking, secure sketch, and image encryption methods), and their biometric modalities, criteria satisfaction, and performance results. It can be seen from this table that the suggested hybrid approach has evident advantages over the existing studies in terms of satisfying all the four requirements of biometric authentication systems.

The algorithms presented in [77], [78], [75] satisfied only the security and performance requirements. In [77], the authors combined cryptography and steganography to secure the storage of iris templates. The work discussed in [75] employed the same DTCWT-DCT-based watermarking algorithm used in the current work to improve the performance and security of the authentication system. Moreover, the methods presented in [20], [79],

[80], [27], [81], [82] achieved lower performance than the presented work, while ensuring the security criteria. These approaches used a single protection technique based on watermarking ([27], [81], and [82]) or secure sketch ([20] and [80]).

The methods described in [83], [84], and [85] satisfied more requirements, which demonstrated the advantage of using the hybrid approach. In [84], a DWT-SVD-based watermarking approach is employed to fuse the face and fingerprint features. Subsequently, a shuffling algorithm is carried out on the watermarked image, which is then XORed with a chaotic map and a Hadamard code to achieve the randomization and orthogonality of protected biometric features. This approach satisfied all the four requirements. However, its main drawback is that it is applicable only for biometric modality represented as an ordered set (e.g., face). This issue of using an ordered set of biometric modalities is improved in our proposed work as we used a fingerprint image as a cover image and manipulated its features using a face image. The results in [84] and our current work indicated a high recognition rate for all conditional cases, except when the noise levels are too high. In [85], the authors demonstrated how to integrate the fingerprint traits into various directional DWT sub-bands of the face image. Subsequently, each user is associated with a unique key and a hyper-chaotic map is employed to generate a keystream to encipher the watermarked image. This approach satisfied the security and diversity requirements. However, a lower level of performance (EER=3.87%) is reached compared to our method.

TABLE 5.2: Summary of different biometric template protection approaches.

Ref.	**Biometric Modality**	**Database**	**Performance (%)**	**Reqs**[1]
[21]	Fingerprint	FVC2000 University of Twente database	EER = 1.4 EER = 1.6	S
[27]	Iris & Fingerprint	FVC2004 & CASIA	EER = 1.2	S
[75]	Face & Fingerprint	ORL & FVC2002 DB1	EER = 0	S
[77]	Iris	CASIA	FAR = 0 GAR = 98.70	S
[78]	Fingerprint & Face	ORL & FVC2002 DB2	EER = 0.96	S
[79]	Face	CMU PIE FEI Extended Yale B	GAR = 78.43 GAR = 55.7 GAR = 66.5	S
[80]	Fingerprint & Face	Essex Faces94 & NIST	FRR = 1.4 FAR = 0.58	S
[81]	Face & Fingerprint	Biosecure	EER = 6.5	S
[82]	Face & Fingerprint	Biosecure	EER = 7.77	S
[84]	Face & Fingerprint	FVC2002 DB1 & ORL	EER = 0	SRD

[85]	Face & Fingerprint	FVC2002 DB1 & ORL	EER = 3.87	SD
[86]	Face	Face94	FRR < 0.2 FAR < 0.2	SR
Our method	Face Fingerprint	- Dataset1 - Dataset2 - Dataset3	FRR = 0.0083 FAR = 0.2140 GAR = 99.99 EER = 0 ——— FRR = 0.0080 FAR = 0.2141 GAR = 99.99 EER = 0 ——— FRR = 0.0095 FAR = 0.2074 GAR = 99.99 EER = 0.12	S R D

5 Conclusion

A hybrid multimodal biometric template protection approach to provide robustness against template database attacks in biometric authentication systems is proposed in this chapter. Herein, the approach combines three techniques to satisfy jointly four requirements of biometric authentication systems, including security, diversity, revocability, and performance. The concept underlying the suggested approach starts by securing the fingerprint features using the secure sketch method. Subsequently, the obtained sketch is embedded in the face image based on a DTCWT-DCT watermarking approach. Finally, the watermarked face image is secured via a 3D chaotic map-based image encryption method.

Extensive experiments were carried out on the ORL face dataset and three FVC databases, including FVC2002 DB1, DB2, and FVC2000 DB1, to evaluate the suggested system. The experimental results showed a high level of security. An excellent matching performance is also achieved with EER values of 0%, 0%, and 0.12% for the three datasets, respectively. Moreover, the diversity and revocability constraints are achieved, indicating the approach's ability to be embedded into various applications without databases cross-matching. Hence, our work makes a significant improvement compared with the existing related studies in meeting the four requirements at a time, due to its use of four

[1]Requirements (S: Security, R: Revocability, D: Diversity)

potential points: multimodality, secure sketch, biometric watermarking, and chaotic map-based image encryption.

Nevertheless, as biometric authentication systems become familiar to the general public, they become frail to spoofing attacks where an assailant can readily get the biometric data from social networks to generate sophisticated models to deceive the authentication system. This attack raises the challenge by dealing with not only systems' database attacks, but also the ability to distinguish between real and fake users.

6 Overview on Deep Face Anti-spoofing

1 Introduction

Over the past few years, face automatic secure authentication has gained an increased level of solicitude, where several studies of different backgrounds (e.g., computer vision, pattern recognition, and image processing) have been launched to introduce novel performing face authentication systems [87]. Besides, the International Biometric Group (IBG) has considered the face as one of the most widely used biometric modalities at the world level [88], where it is based in the most official identification documents, such as the national ID cards and the International Civil Aviation Organization (ICAO)-compliant biometric passport [89]. As benefits always come along with a price, the face authentication systems become a major target of arising challenges such as the resistance against the spoofing attacks.

2 Face Spoofing Attacks

A spoofing attack is defined as the ability to fool a biometric authentication system to gain illegitimate access, by showing an artificial rigged version of the original biometric data to the system's sensor. It is also known as a presentation attack defined in the first part of the ISO/IEC 30107 standard as *"Presentation of an artefact or human characteristic to the biometric capture subsystem in a fashion that could interfere with the intended policy of the biometric system"* [90]. Therefore, spoofing attacks in face authentication systems can be classified into four classes including *(1) Printed/Digital Attack-2D*, *(2) Replay Attack-2D*, *(3) Mask Attack-3D*, and *(4) Plastic Surgery Attack*.

2.1 Printed/Digital Attack-2D

In the printed/digital attack-2D, the fraudulent attempts deceiving the authentication system by presenting the genuine's photograph in front of the system's sensor. The presented photo can be a printed paper or displayed on the screen of a digital device as a cell phone or tablet. This type of attacks is considered as the most used provenance of spoofing attack, due to, a facial image can be easily captured by a camera or even recaptured from the social network [91]. Moreover, a more sophisticated sort of printed attack has been introduced to generate face motion, where a mask of high resolution is printed, face and eyes are removed, then, the attacker is placed behind the mask to show the eyes blinking and mouth movements to the system's sensor.

2.2 Replay Attack-2D

Also known as video attack. This sort of attack is executed by replaying the authorized individual's video using a digital device such as a tablet, mobile phone, or laptop in front of the system's camera. Such attacks have shown further progress due to two main reasons. First, they are hard to reveal. Second, not only the face 2D texture is used but also its dynamics.

2.3 Mask Attack-3D

Herein, the attack is carried out using the 3D mask of the authorized user face. The mask is produced with a very similar 3D shape of the target face features, which increases the difficulty of finding the attack's loopholes. The concept underlying this attack is gathering the biometric treats at a distance without direct interaction. Besides, it is considered as one of the most complex spoofing attacks, due to, manufacturing a 3D mask is not an easy mission that costs too much.

2.4 Plastic Surgery Attack

The facial plastic surgery is a part of the otolaryngology, that can be divided into two classes, including reconstructive and cosmetic. The former rectifies the face features anomalies such as cleft lip and palate birthmarks, while the second enhances the visual appearance of the facial structures and characteristics. Consequently, the original facial features transform to a new extent. Applying the cosmetic surgery to face biometric system, a person can try to spoof a face biometric authentication system, which is referred to as plastic surgery attack.

3 Face Anti-spoofing Techniques

To guarantee a high level of robustness and make face authentication systems more practical, spoofing attacks should be addressed by providing these systems with face anti-spoofing mechanisms, also referred to as liveness-detection or presentation attack detection mechanisms. The main aim of such methods is to prevent spoof attacks on biometric systems by distinguishing the feature space into living and non-living through liveness indicators.

Based on the literature, face liveness detection algorithms are categorized into various types of classifications [92–95]. In this report, two categories of classifications are presented, namely *(1) Biometric system's modules based classification* and *(2) Used technology-based classification.*

3.1 Biometric System's Modules Based Classification

This category is designed based on the biometric system's modules where the anti-spoofing algorithms can be integrated. It is divided into three classes, including *Feature-Level Techniques*, *Sensor-Level Techniques*, and *Score-Level Techniques* [96].

3.1.1 Feature-Level Techniques

Feature-level techniques are also known as `software-based techniques`, wherein the fake traits are revealed after the sample is gained from a standard sensor. More precisely, the feature data is obtained from the samples, not from the biometric feature itself. Such techniques are divided into two types that are: *feature level dynamic approaches* and *feature level static approaches* [96].

- **Feature level dynamic approach:** It is one of the most used protection techniques in 2D face recognition. It relies on revealing the motion over temporal face sequences by tracking a typical analysis of face segments. Which detects the pertinent biometric data to distinguish between live and fake faces. In such approaches, particular cues are based for performing the anti-spoofing, namely face and head gestures [97, 98] or eye blinking [99–101].

  Even the feature level dynamic techniques achieved a very competitive performance by exploiting the temporal and spatial data in face videos, they still suffering from several limitations. For instance, in some scenarios where video data has been recorded (e.g., surveillance applications), it is not rare to find that only a very few non-consecutive frames are suitable for facial analysis. Additionally, such methods can not be based in systems with only a single face image like passport related applications.

- **Feature level static approach:** In contrast to the feature level dynamic approaches, the feature level static ones focus on the analysis of single static images. Besides, to extract the pertinent biometric information, the face texture is mostly based using various image processing tools, including Fourier spectrum [102], multiple Difference Of Gaussian (DoG) filters to extract specific frequency information [103], and Local Binary Patterns (LBP) to detect photo-attacks [104].

3.1.2 Sensor-Level Techniques

Also known as `hardware-based techniques`, wherein a specific device is added to the system sensor to elicit particular characteristics of a living trait (e.g., facial thermogram, brain wave signals).

Mostly, such techniques distinguish between real and fake face modalities by evaluating one of the three following liveness traits:

- **Intrinsic properties of the living face:** Sort of visual, electrical, physical, or spectral properties.
- **Involuntary signals of the living face:** Such as perspiration, electric heart signals, and blood pressure.
- **Responses to external stimuli:** Known as challenge-response approaches. Herein, the collaboration of the individual is needed, as the analysis is based on detecting voluntary (behavioural) or involuntary (reflex reactions) to an external signal. For instance, the pupil reflex to the switched light, or head movement tracking random path determined by the system.

3.1.3 Score-Level Techniques

They are the most recently proposed liveness detection techniques that fall out of the two aforementioned techniques (sensor level and feature level classifications). These approaches concentrate on examining the biometric frameworks at the score level to fuse strategies that rise their strength against presentation attacks. Mostly, the joined scores originate from 1) Two or more unimodal biometric modules, 2) Unimodal biometric modules and anti-spoofing techniques, or 3) Only results from anti-spoofing modules. Such approaches are highly resistant to the photo, video, and mask attacks. Meanwhile, they are slower and expensive due to the additional hardware required to process the face traits.

3.2 Used Technology Based Classification

Face anti-spoofing methods vary according to the technology used. However, all of them fall into one of the two following classes, including *(1) Conventional face anti-spoofing methods* and *(2) Deep learning-based face anti-spoofing methods*.

3.2.1 Conventional Face Anti-spoofing Methods

The concept underlying the traditional face anti-spoofing approaches is adopting the hand-crafted features to extract the biometric data. Then applying the shallow learning techniques like Linear Discriminant Analysis (LDA) and SVM to build the anti-spoofing systems. This type of approaches can be classified into three classes that are:

Motion-based methods: They aim at verifying the biometric traits of the expected user by testing the spontaneous facial motions in a particular area. These facial motions include eye-blinking [105, 106], mouth movement [107], and head rotation.

Texture-based methods: The key idea of these algorithms is to emphasize the texture of the face image to single out whether the image is taken from a live individual or not. To fulfil this, several cues are used like the Scale-Invariant Feature Transform (SIFT) [105], Speeded Up Robust Features (SURF) [108], DoG [109], LBP [110, 111], and Histogram of Oriented Gradients (HOG) [112, 113]. Additionally, as these metrics can be very sensitive to camera devices, illuminations, and specific identities, new technologies as Fourier spectra [102], Optical Flow Maps (OFM) [114], Hue Saturation Value (HSV), and YCbCr [115, 116] have been introduced.

Multi-cues integration-based methods: With the development of the attacks environments and scenarios, face anti-spoofing techniques counting on single cues could not tackle all types of spoofing attacks. Thus, the integration of complementary multi-cues from diverse aspects is introduced as an alternative solution to solve several attack-specific issues simultaneously. For instance, in the 2nd competition on 2D face spoofing attacks, the CASIA and LNMIIT teams joined the motion features and the LBP-based texture features at features level, and they both obtained sound results on the REPLAY-ATTACK dataset [117]. Also, authors in [118] employed the 3D-LBP based dynamic texture characteristics to depict the local facial motions and static facial texture simultaneously. Besides, Komulainen *et al.* [119] combined the correlation between the face background regions and LBP-based texture features at the score level.

3.2.2 Deep Learning-based Face Anti-spoofing Methods

Lately, deep learning-based face anti-spoofing methods have proven their supremacy concerning the traditional anti-spoofing metrics. Different from the hand-crafted anti-spoofing methods, these algorithms extract the most pertinent end-to-end characteristics directly from the original biometric data by mapping the raw pixels of the input image to the output probability through intermediate hidden layers. Accordingly, the learned traits can catch more discriminative information that leads to better differentiate between live and spoofing faces. Moreover, such algorithms have the potential to learn more general characteristics for various spoofing sorts.

4 CNNs-based Face Liveness Detection

4.1 CNNs Architecture Overview

Convolutional neural networks [120] are deep learning algorithms that receive an image as an input, assign learnable biases and weights to the different aspects of the image, then, output a class (car, cat, plane, ...) or classe's probability that depicts the image. They are also known as CNNs or ConvNets. As long as the neural network algorithms, this architecture has been inspired by the visual cortex of the human brain, where, individual neurons respond to only specific regions of the visual field referred to as the receptive field. In 2012, the ConvNets gained their first success after the prominent achievements at the ImageNet competition by decreasing the classification error score from 26% to 15%. Since then, they have been involved in the services of a vast number of companies. For example, they have been used in Google for their photo search, Facebook for the automatic tagging, Instagram for their search infrastructure, and Amazon for their product recommendations.

Generally, convolutional neural networks are introduced owing to, ordinary nets do not scale well to images. More precisely, the regular nets take as input a single vector, then, convert it through a sequence of hidden layers. Each layer is built using a set of neurons that are fully connected to all neurons of the previous layer. Consequently, this fully-connected architecture does not fit sizable images. As an illustration, with an image of $200 \times 200 \times 3$ (200 wide, 200 high, 3 colour channels), neurons of $200 \times 200 \times 3 = 120,000$ weights are gotten in the first hidden layer. Since we are working with deep learning structures, several layers should be added. Therefore, the number of parameters will be increased exponentially, which rapidly leads to overfitting. By contrast, CNNs take in consideration that the input is an image, subsequently, they restrain the architecture more sensibly by setting up the neurons in three dimensions, including width, height, and depth. Furthermore, each layer's neurons are linked to only a small region of the previous layer instead of all of the neurons in a fully-connected way.

To build an entire convolutional neural network architecture, a sequence of three main kinds of layers are stacked, comprising Convolutional Layer, Pooling Layer, and Fully-Connected Layer.

- **Convolutional Layer:** It is the heart building of the ConvNets. To conduct the computation, a group of learnable parameters named as filters or kernels are used. As illustrated in figure 6.1, each kernel is convolved along with the height and width through the input image to calculate the dot product between the image's entries at any position and the kernel's input. This operation is carried out using two main parameters that are the stride and padding. The former is used to manage the number of cells the kernel is displaced in the input image to compute the next cell in the result. While the latter is defined as the process of appending zeros to the input matrix symmetrically to preserve more information at the border of the input image. Applying the convolution among the whole image generates a 2-dimensional activation map for each kernel, which represents the responses of that kernel at every spacial position. By the end, all the kernels' activation maps are stacked along the depth dimension to produce the output volume of the layer (figure 6.2). The convolutional layer is usually followed by a Rectifier Linear Unit (ReLU) activation function to convert the negative values of the output volume into zeros.

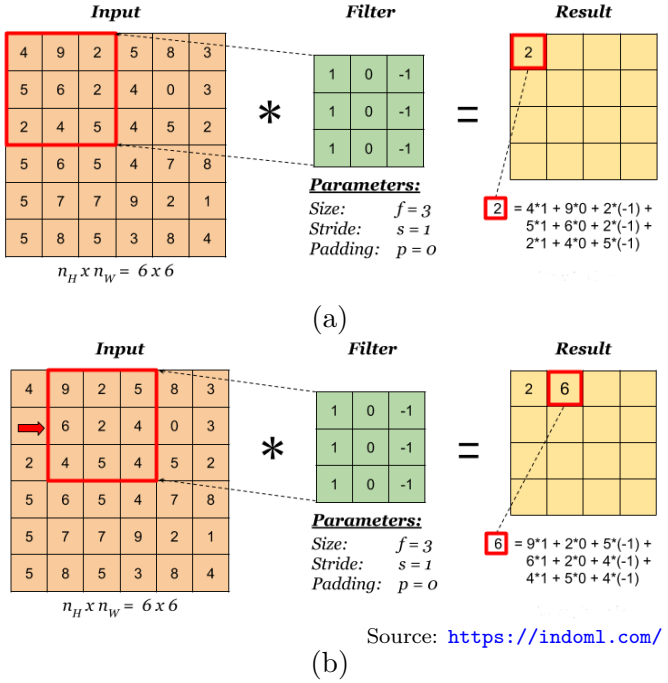
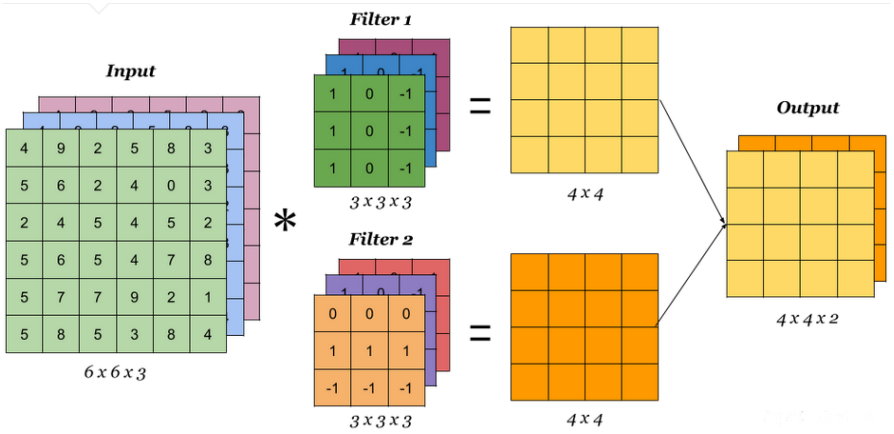


FIGURE 6.1: Visualization of a 3×3 kernel convolving around a 6×6 input image and generating the output activation map.

The main objective of the convolutional layer is to elicit the high-level traits from the input image by learning the kernels to activate when meeting specific visual

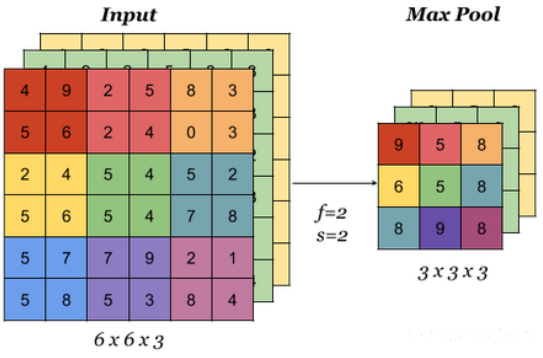


Source: https://indoml.com/

FIGURE 6.2: Overview of the convolutional operation in CNNs.

characteristics, like some orientations, blotch of some colours, and edges. Typically, the first convolution layer of a ConvNet is the main reprehensible for extracting the low-level characteristics. While joining other layers, the ConvNet architecture adapts to the high-level features as well, which leads to wholesomely understanding the images of the database.

- **Pooling Layer:** It is regularly used in-between sequential convolutional layers of a CNN structure to decrease the size of the representation by reducing the number of parameters, which speeds up the calculations and manages the overfitting. Moreover, this layer is practical for extracting dominant characteristics. There are several types of pooling layers, such as Max pooling, Average pooling, and $L2$-norm pooling. However, the most commonly used is the Max pooling layer. Figure 6.3 illustrates the Max pooling operation applied on a $6 \times 6 \times 3$ input image with a filter of 2×2 and stride of 2.

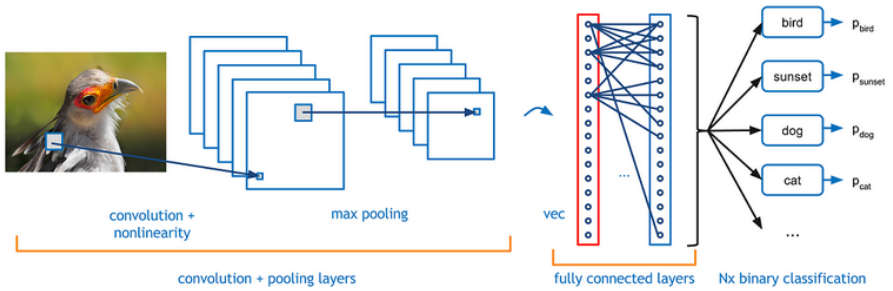


Source: https://indoml.com/

FIGURE 6.3: Max pooling layer on a $6 \times 6 \times 3$ input image, with a filter of 2×2 and stride of 2.

- **Fully-Connected Layer:** Takes as input the output volume of the previous layer (e.g., Convolution layer, Pooling layer) and produces a vector of N-dimension.

Where N is the number of classes to which the input image may belong. In contrast to the other layers, the neurons of this layer are connected to all the activations of the previous layer via a matrix multiplication followed by a bias offset. The concept underlying this layer is to learn the non-linear combinations of the high-level traits extracted by the convolutional layer to decide the characteristics that most fit a specific class. Figure 6.4 illustrates an overview of a basic CNN architecture.



Source: https://indoml.com/

FIGURE 6.4: Overview of a basic CNN architecture

There are a set of various ConvNet architectures that have been introduced to build powerful algorithms in artificial intelligence, namely LeNet, AlexNet, VGGNet, GoogLeNet, ResNet, and ZFNet.

4.2 CNNs-based Face Anti-spoofing

Convolutional neural networks have attracted numerous researchers to exploit the latent capabilities of such nets in reaching the state-of-the-art performance against presentation attack types. This potential shift towards the CNN architecture is by dint of its higher accuracy, automatic learning, and remarkable achievements in ImageNet competition [121]. As an illustration, the 13 teams that made it to the final round of the competition hosted by ChaLearn at Computer Vision and Pattern Recognition (CVPR) all adopted CNN-based solutions [122]. In [123], the authors suggested a performance evaluation of CNNs for face anti-spoofing by using ResNet and Inception architectures. The study covered different sides, such as the architecture depth, various learning rates, fine-tuning vs. training from scratch, and weight transfer vs. weight initialization. In [124], it was postulated that approaches involving treatment of the overall face either globally or in small batches diminishes algorithmic performance; based on this, a new CNN architecture to learn diverse local spoofing cues was proposed. This architecture is trained in two phases: first, each fraction of the network is trained on a specific facial area, which helps the model to catch varied spoofing cues from all parts of the face; then, the model generalization is enhanced by fine-tuning the overall model using fake and real images with the weights learned in the first step applied. In [125], a 3-D face anti-spoofing algorithm was developed using a hypergraph CNN (HGCNN). In [94], a

novel face liveness detection technique based on the fusion of two CNN architectures was introduced; in this approach, the first CNN structure was used to extract local features, while the latter was used to extract holistic features. Another two-stream CNN architecture was developed in [126] through the combination of a CNN with auxiliary supervision with a generative adversarial network (GAN)-like discriminator to achieve facial de-spoofing. The primary concept behind this approach was to reciprocally decompose spoofed faces into spoof noise and live faces and then apply classification based on the spoof noise.

A number of approaches have combined CNNs with other architectures and methods [127–130]. For example, in [129] a common CNN-LSTM network to detect face attacks across video frames was suggested. To achieve this goal, highly discriminative features of video frames were extracted via CNNs and then a long short-term memory (LSTM) architecture was used to capture the temporal dynamics of the videos. In [131], a robust feature representation scheme that combined deep texture features and eye-blink cues for facial anti-spoofing was suggested. The proposed approach involved learning deep texture features from aligned face images and unaligned video frames via CaffeNet and GoogLeNet and then employing the frame differences to detect eye-blinking.

4.3 Adversarial Attacks on CNNs-based Face Anti-spoofing

Despite the high performance, CNNs-based face liveness detection is still suffering from the adversarial attacks. The main objective of such attacks is to deceive the face anti-spoofing systems via adversarial samples that induce classification error [132]. Based on the access level to the attacked model, deep-learning-based adversarial attacks can be classified into two categories: *(1) white-box attacks* -and *(2) black-box attacks*. In the former, full access to the attacked model's parameters and architecture is granted while, in the latter, access is given only to the attacked model's inputs and outputs [133].

Since Szegedy *et al.* [134] demonstrated that even well-performing Deep Neural Networks (DNNs) are vulnerable to adversarial attacks, an extensive body of research has continued to find new adversarial attack methods. An example of this is the targeted white-box adversarial attack presented in [135] as a tool for fooling face recognition systems. The proposed attack uses the Attentional Adversarial Attack Generative Network (A^3GN) approach to produce adversarial samples identical to original face images. Unlike conventional GANs, this architecture applies facial recognition as a third component of the conflict between generator and discriminator to enable the efficient imitation of a target individual. In [136], a new approach to simulating adversarial samples by solving a constrained optimization problem using an adversarial generator network was proposed. The principle underlying the proposed strategy of generating small distortions to an input image to deceive face spoofing was tested on a faster R-CNN-based face detector. Another intriguing study proved that facial recognition systems could be fooled

by building adversarial glasses [137]. M. Sharif *et al.* proposed an adversarial generative net architecture to misclassify DNN-based face classifiers by adding physical realizations of eyeglass frames [138]. [139] proposed the concept of backdoor attack, in which poisoning samples are injected into a deep-learning-based authentication system with the goal of misleading it. In [140], 2- and 3-D face mask attack datasets constructed for the TABULA RASA project were used to prove the impact of mask attacks on 2-, 2.5-, and 3-D face recognition models. In [141], a novel adversarial attack scheme involving illumination of the attacker's face was discovered. Under this attack mode, infrared dots are projected onto specific positions on the face using minuscule InfraRed (IR) LEDs integrated into a cap, umbrella, or wig to mislead machine learning-based face recognition systems. [142] proposed a black-box adversarial attack to produce adversarial distortions based on the output of a differential evolution algorithm. The approach is designed to manipulate only a few pixels of the input face image to misclassify CNN-based classifiers. S. Moosavi-Dezfooli *et al.* [143] suggested a DeepFool approach to efficiently calculate distributions that deceive deep networks and compared the robustness of various deep network classifiers against adversarial perturbations.

5 Evaluation and Analysis

5.1 Evaluation and Analysis Metrics

To evaluate the anti-spoofing algorithms' performance, a set of common criteria are based, namely Half Total Error Rate (HTER), Recall, Precision, and Accuracy (ACC).

- **Half Total Error Rate:** A liveness detection classifier is a target of two sorts of errors that are, rejecting the live user (false rejection) or accepting the impostor (false acceptance). To statically measure the detection performance of such systems, the HTER metric that joins the FRR and FAR is employed. This metric achieves its minimum when FAR = FRR, which is referred to as EER. For a dataset $\mathbb{T}$ and a decision threshold value τ, the HTER is defined as:

$$HTER(\mathbb{T}, \tau) = \frac{FRR(\mathbb{T}, \tau) + FAR(\mathbb{T}, \tau)}{2} \tag{6.1}$$

  The lower the HTER value, the better the average performance on the anti-spoofing decision.

- **Recall:** Also known as sensitivity. It indicates the number of relevant instances that are successfully retrieved. Given the following confusion matrix:

	Actual	
Predicted	True Positive (TP)	False Positive (FP)
	False Negative (FN)	True Negative (TN)

The recall value is defined as:

$$Recall = \frac{TP}{TP + FN} \tag{6.2}$$

- **Precision:** Also known as the positive predictive value. It refers to the percentage of pertinent instances among the total retrieved instances.

  The precision value is defined as:

$$Precision = \frac{TP}{TP + FP} \tag{6.3}$$

- **Accuracy:** This metric is employed to evaluate the classification models. Specifically, it is defined as the fraction of prediction the model got correct. Mathematically, the accuracy is computed using the following equation:

$$Acuuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6.4}$$

5.2 Databases

The experiments are conducted using four publicly available face anti-spoofing datasets:

- **3DMAD:** This dataset comprises 255 videos of 17 individuals. All frames are registered via Kinect and characterized by a depth image, corresponding RGB image, and manually annotated eye position. The registrations are carried out using frontal views with neutral expressions. Each record is gathered in three different sittings: in the first two, real-access samples are registered with a time delay of two weeks between acquisitions; in the third, a 3-D mask attack is obtained.

- **CASIA:** This dataset comprises 50 genuine subjects for whom there are three genuine and nine fake videos apiece. Thus, the global dataset contains 600 video clips. Three types of attacks printed photographs with the eyes cut out, video attacks, and warped printed photographs are considered. Each subjects is recorded using three types of cameras: low, average, and high quality.

- **Replay-attack:** This dataset is produced at the Idiap research institute with the specific goal of developing anti-spoofing algorithms. Each video in this dataset is generated by placing a real individual in front of a built-in webcam or by rerecording a video or photo of the individual for at least 9-s under different lighting conditions. Depending on the type of device used, six protocols can be applied to produce a video attack: print, mobile (phone), high-definition (tablet), photo, video, or grand testing (a combination of all of the preceding).

- **ROSE-youtu:** This dataset comprises 4,225 videos of 25 subjects, with 150-200 ∼10-s video clips per subject. The front-facing cameras of five cell phones-a Hasee smart-phone, Huawei smart-phone, iPad 4, iPhone 5s, and ZTE smart-phone are used to collect the data. Genuine-face videos are recorded under five different illumination conditions, and three adversarial attacks printed paper, video replay, and masking attacks are taken into consideration.

6 Conclusion

In this chapter, an extensive overview of deep face anti-spoofing strategies is presented. Herein, four sorts of face spoofing attacks are presented, namely printed/digital attack-2D, replay attack-2D, mask attack-3D, and plastic surgery attack. Besides, two categories of face liveness detection classification are discussed, including *biometric system's modules based classification* and *the used technology-based classification*.

Due to the remarkable performance achieved by the convolutional neural networks against numerous adversarial attack types, a detailed presentation of this architecture is presented. Furthermore, CNNs-based face liveness detection algorithms and adversarial attacks on CNNs-based face anti-spoofing techniques are addressed.

7 Unraveling Robustness of Deep Face Anti-spoofing Models against Pixel Attacks

1 Introduction

Facial biometrics consistently outperform other biometric modalities (i.e., DNA, fingerprint, palm-print, iris, ...) in a wide range of daily applications in terms of their reasonable recognition cost, convenience, and high levels of performance. As examples of the applicability of the approach, Lenovo, Asus, and Toshiba laptops now come with built-in face authentication webcams [144] and the Unique Identification Authority of India (UIDAI) facial recognition system is used to identify Indian residents [145]. As the general public becomes increasingly acquainted with facial authentication systems, their loopholes are being explored. The human face can be easily acquired and duplicated by attackers who can obtain facial images or videos from social networks and use them to generate artificial models, which can then be used to deceive face authentication systems in an attack mode referred to as face spoofing. This presents a challenge to authentication mechanisms, which, in addition to delivering high recognition performance, must be able to differentiate between live and fake users.

To guarantee a strong degree of security against face-spoofing attacks, it is necessary to equip authentication frameworks with facial anti-spoofing models, an approach also referred to as presentation attack detection or liveness detection modeling. Thus, a number of approaches to solving the spoofing attack research problem have recently been developed, as witnessed by the rising number of studies in which biometric authentication system sensitivity has been evaluated and new defense techniques have been explored [146–149]. Several workshops and competitions have been organized [117, 150], dedicated datasets have been developed [103–105, 127, 151, 152], and new projects have been launched to address this issue [153]. Furthermore, as spoofing attacks' variability and complexity are growing, deep learning-based face anti-spoofing is suggested as an

alternative solution to the conventional anti-spoofing techniques. The concept underlying their idea is to extract directly pertinent end-to-end features from original data, which leads to catching variants of famed and never seen attacks.

Inspired by the significant advances in the application of deep learning to a wide range of fields, including object detection [154], recognition [155], and speech recognition [156], convolutional neural network (CNN)-based face anti-spoofing algorithms have had phenomenal success in achieving state-of-the-art results against diverse types of spoofing [157–159]. However, despite the success achieved by these algorithms, they still vulnerable to adversarial attacks. Where it was recently demonstrated that one can easily bypass face liveness detection systems via slight imperceptible perturbations by manipulating the legitimate input, which leads the system to misclassify the output.

In the context of this ongoing struggle between anti-facial-spoofing systems and spoofing attacks, in this chapter, an analytical study on adversarial attacks to deep-learning-based face liveness detection models is described. The primary goal of this study is to assess how deep face anti-spoofing mechanisms behave in the presence of spoofing attacks. The main contributions of this work are as follows:

- We present an experimental setup for emulating differential evolution-based adversarial attacks on transfer-learning-based CNNs for face liveness detection systems based on increasing the confidence with which fake faces can be classified as live faces.
- We produce an experimental assessment of more realistic circumstances in which real-world applications are emulated based on the results of tests on four different face anti-spoofing databases under different threat models and use-case scenarios.
- We assess the impact of various criteria related to both spoofing and anti-spoofing algorithms to improve the overall robustness of deep face anti-spoofing mechanisms. Based on this assessment, we attempt to outline how adversarial attacks reduce the contrast between live and fake faces, highlight common misconceptions, and derive complementary countermeasures that aid in constructing more flexible anti-spoofing frameworks and enhance their resistance to a wide variety of adversarial attacks.

2 Methodology Description

Adversarial attacks are used to minimize the resistance of liveness detection systems. In a successful case, if an adversarial sample is presented to a system senor, the authentication attempt will be viewed as a genuine identity and will be accepted. Playing the role of a neutral referee to assess the vulnerability of face anti-spoofing measures against

presentation attacks, we carried out an analytical study on adversarial attacks to deep learning-based face liveness detection systems. Our goal was to evaluate and discuss the effectiveness of face anti-spoofing models when confronted with spoofing attacks. Based on this evaluation, we attempted to highlight the various criteria for constructing more flexible anti-spoofing frameworks and improve their resistance to a wide scope of adversarial attacks. To accomplish this, we investigated the impact of differential evolution-based adversarial perturbation on transfer-learning-based convolutional neural networks for face anti-spoofing. Specifically, we experimentally evaluated how well such approaches reject adversarial samples under different use-case scenarios using four different anti-face-spoofing databases. In the following subsections, we describe the face spoofing and anti-spoofing approaches assessed in our study

3 Transfer-learning-based CNNs for Face Anti-spoofing

Following the approach used in [160], the proposed face anti-spoofing classifier was built using transfer-learning based on a pre-trained VGG16 CNN structure [161]. Within the VGG16 network, there are up to 13 convolutional (Conv) layers of 3×3 kernels and three fully connected layers of size 4,096, 1,000, and 1,000, respectively. The VGG16 architecture is built using identical padding (Zero Pad) and max-polling (Max Pool) 2×2 window and stride 2 layers. The rectifier linear unit activation function is applied to the output of each convolutional and fully connected layer.

Figure 7.1 shows the proposed face anti-spoofing network (FASNet). With the exception of the top layers, the architecture replicates the VGG16 structure; specifically, the 13 convolutional layers (highlighted in grey) are retained, while the three fully connected layers are omitted and replaced by two fully connected layers of sizes 256 and 1, respectively (highlighted in green), to enable the binary classification required to carry out the face liveness detection task. The first fully connected layer is followed by one dropout layer to avoid over-fitting. To enable finer classification, the softmax decision function is replaced by a sigmoid function. Transfer learning is carried out by freezing the weights of the first seven layers and fine-tuning the weights from the eighth up to the top layers through backpropagation.

4 Adversarial Attack Algorithm

4.1 Differential Evolution

Differential evolution is a global optimization and stochastic direct search algorithmic approach [162] in which linear combinations are applied instead of conventional crossover and mutation operators to generate new solutions.

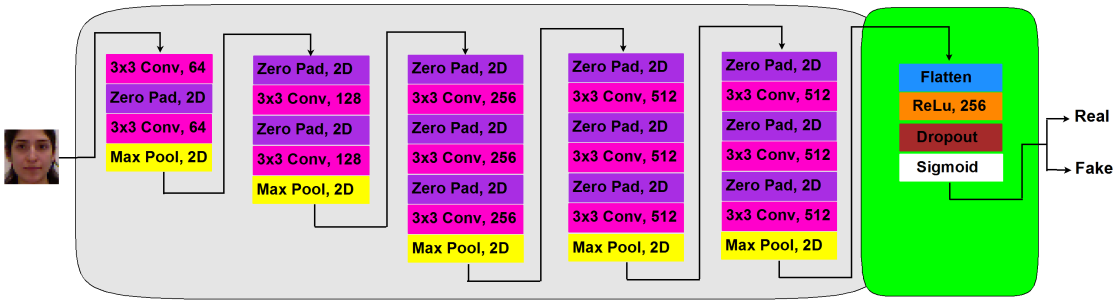


FIGURE 7.1: Face anti-spoofing network architecture.

For a candidate population $p^g_{n,i} = [p^g_{n,1}, p^g_{n,2}, p^g_{n,3}, ..., p^g_{n,D}]$ represented by an n-dimensional vector, where g is the generation and D is the number of variables, we can define f as the fitness function of a candidate solution. The differential evolution algorithm comprises four steps: initial population, mutation, recombination, and selection.

- *Step 1: Initial Population*
  To start with, the initial population is randomly generated between the upper and lower bounds as follows:

$$p_{n,i} = p^L_{n,i} + rand() \times (p^U_{n,i} - p^L_{n,i}) \tag{7.1}$$

  where $p^L_{n,i}$ and $p^U_{n,i}$ are the lower and upper bounds of the variable $p_{n,i}$, respectively, $i = 1, 2, 3, ...D$ and $n = 1, 2, 3, ...N$.

- *Step 2: Mutation*
  In the mutation stage, the current generation is perturbed by scaling the difference between randomly selected population candidates, p^g_{r2n} and p^g_{r3n}; the scaled difference is then added to a third randomly selected population candidate to generate a new donor candidate v^{g+1}_n:

$$v^{g+1}_n = p^g_{r1n} + F \times (p^g_{r2n} - p^g_{r3n}) \tag{7.2}$$

  where $F \in [0, 1]$ is the scale parameter and $r1n$, $r2n$, and $r3n$ ($r1n \neq r2n \neq r3n$) are random indices of the parent population.

- *Step 3: Recombination*
  In the recombination phase, a trial candidate $u^{g+1}_{n,i}$ is evolved using both $p^g_{n,i}$ and v^{g+1}_n as follows:

$$u_{n,i}^{g+1} = \begin{cases} v_{n,i}^{g+1} & if\ rand() \leq C_p\ or\ i = I_{rand} \quad i = 1,2,3,...,D\ and \\ p_{n,i}^{g} & if\ rand() > C_p\ and\ i \neq I_{rand} \quad n = 1,2,3,...,N \end{cases} \tag{7.3}$$

where I_{rand} is a random integer in [1,D] and C_p is the recombination probability.

- *Step 4: Selection*
  Finally, in the selection stage each generated child $u_{n,i}^{g+1}$ is compared with its corresponding parents $p_{n,i}^{g}$:

$$p_n^{g+1} = \begin{cases} u_{n,i}^{g+1} & if\ f(u_{n,i}^{g+1}) < f(p_n^g) \\ p_n^g & Otherwise \quad n = 1,2,3,...,N \end{cases} \tag{7.4}$$

  and the worse-performing parents are replaced with better children.

4.2 Differential Evolution-based Adversarial Attack

To deceive the FASNet system, the one-pixel adversarial attack introduced in [163] was utilized. The concept underlying one-pixel attacks is to generate adversarial samples by locating the pertinent pixels in the input image and their corresponding strengths of perturbation. Unlike most conventional adversarial attacks, the one-pixel attack focuses on distorting only a few pixels with varying manipulation strength, rather than distorting all pixels with an overall constraint on the strength of distortion.

Generating adversarial samples can be considered to represent an optimization problem with constraints. For an output of the image classifier. $f_t(\mathbf{z})$, where $\mathbf{z} = (z_1, z_2..., z_N)$ is the n-dimensional vector representation of an input image correctly classified as class t, the adversarial samples can be defined using the following equation:

$$\begin{aligned} \mathbf{z'} &= \mathbf{z} + e(\mathbf{z}) \\ \{\mathbf{z'} \in R^N | \arg\max_i (f_{adv}(\mathbf{z'})_i) &\neq \arg\max_j (f_t(\mathbf{z})_j)\} \end{aligned} \tag{7.5}$$

where $e(\mathbf{z}) = (e_1, e_1, ...e_n)$ is a very small additive adversarial perturbation related to $\mathbf{z}$ and the target class adv. The goal of using adversary samples is to reach an optimized perturbation $e(\mathbf{z})^*$ that raises the confidence of the target class and decreases the confidence of the original class:

$$\begin{aligned} &maximize_{e(\mathbf{z})^*} \quad f_{adv}(\mathbf{z'})) \\ &Subject\ \ to \quad ||e(\mathbf{z})||_0 \leq d \end{aligned} \tag{7.6}$$

where d is a small number that is $d = 1$ for a one-pixel attack.

To accomplish this, differential evolution as a population-based optimization algorithm is based. This algorithm follows a very limited scenario in which only the probability labels, with no inner information regarding the attacked model, are required; this categorizes the attack as a black-box adversarial attack.

In the context of the one-pixel adversarial attack, we define $\mathbf{x}$=(x,y,r,g,b) as a one-pixel perturbation in which x, y and r, g, b are the pixel coordinates and RGB channel, respectively. In a similar manner, multiple perturbations can be presented as a concatenation of several tuples as follows:

$$\begin{aligned} X &= (\mathbf{x}_1, \mathbf{x}_2, ...) \\ &= (x_1, y_1, r_1, g_1, b_1, x_2, y_2, r_2, g_2, b_2, ...) \end{aligned} \tag{7.7}$$

As illustrated in figure 7.2, the attack is generated by randomly initializing an N-dimensional population of perturbations $P = (X_1, X_2, X_3, ..., X_N)$, following which N new mutant children are created using:

$$X_i = X_{r1} + F \times (X_{r2} + X_{r3}) \tag{7.8}$$

where $r1, r2, r3 (r1 \neq r2 \neq r3)$ are random indices of the parent population P and F is the scale parameter, which is set to 0.5. Once the new generation has been created, a fitness comparison is carried out between each child and its corresponding parents according to the population index, and the best-performing individuals survive to the next iteration. This procedure is repeated for several iterations until the stopping criterion is satisfied. The perturbation corresponding to the best fitness is then taken as the best pixel for manipulation in the input image.

5 Experimental Analysis

5.1 Experimental Setup

Experiments were conducted using four publicly available face anti-spoofing datasets, namely 3D Mask Attack Database (3DMAD) [164], Replay-Attack [104], CASIA [103],

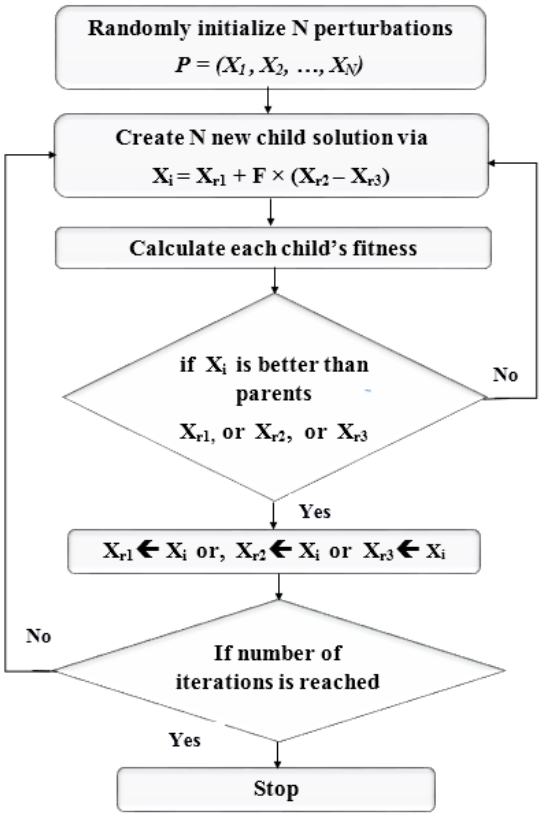


FIGURE 7.2: Flowchart of one-pixel adversarial attack algorithm

and ROSE-Youtu [165]. As the datasets comprise video clips while FASNet is a 2-D structure, a two-step pre-processing procedure, in which several frames were uniformly extracted from each video and then the Multi-Task CNN (MTCNN) was employed for face location [63], was carried out [166]. Training and test datasets comprising 32×32 pixel images containing equal numbers of live and fake images were used. Table 7.1 lists the number of images used for each dataset. The approach used in the study was to perturb the fake images so that they were identified by the face liveness detection classifier as live images and then analyze the trade-off between the spoofing and anti-spoofing algorithms.

Datasets	**Training images**		**Test images**	
	Live	**Fake**	**Live**	**Fake**
3DMAD	240	240	100	100
CASIA	360	360	540	540
Replay-Attack	1200	1200	800	800
ROSE-Youtu	2240	2240	2245	2245

TABLE 7.1: Number of images used in each dataset.

To study the effects of attack on the face anti-spoofing systems, different criteria were assessed. In particular, different training step values were used to generate the respective face liveness detection models. Eight data augmentation (DA) techniques were

applied, namely (1) *None (no DA)*, (2) *Width and Height Shift*, (3) *Shear Map*, (4) *Zooming*, (5) *Rotation*, (6) *Channel Shift*, (7) *Vertical and Horizontal Flipping*, and (8) *All DA*. In implementing differential evolution-based attacks, two additional criteria were considered the number of manipulated pixels per image and the number of attack iterations. The first parameter relates to the likelihood of a successful attack, as increasing the number of manipulated pixels will result in changes to more pixels in the image within a given time, while the second specifies the number of generations the differential evolution-based attack algorithm should run before giving up.

The initial population of the differential evolution algorithm was initialized using the Latin hypercube sampling method to minimize the number of iterations needed to achieve a reasonably accurate result. The fitness value was set as the probability (confidence) value of the original class. Once the confidence value for a fake face reached 0.5 or lower, it was classified by the anti-spoofing model as a live face. Unlike [160], in which an Adam optimizer was used, the default stochastic gradient descent (SGD)-based VGG16 optimizer was used. The FASNet model hyperparameters are listed in Table 7.2. The number of epochs used was varied experimentally based on the training steps.

TABLE 7.2: FASNet model parameters

Parameter	**FASNet Model**
Learning Rate	10^{-3}
Momentum	0.9
Dropout	0.5
Batch Size	64

The spoofing and anti-spoofing algorithms were both constructed using Keras with Tensorflow as a backend. All tests were carried out using a computer equipped with Ubuntu 18.04.2 LTS with a Tesla K80 GPU (12 GB).

5.2 Results and Discussions

In this section, we evaluate the results of the experiments against five standard metrics: accuracy (ACC), HTER, precision, recall, and recall drop.

5.2.1 Overall Results

To assess the anti-spoofing performance of the FASNet algorithm against those of other DA techniques, models for the 3DMAD, CASIA, and Replay-Attack datasets were generated using 50,000 training steps and for the ROSE-Youtu dataset using 100,000 steps. The number of steps used for the ROSE-Youtu dataset were doubled to account for the number of images used in both the training and test phases. Tables 7.3, 7.4, 7.5, and

7.6 summarize the performance of each model before and after an adversarial attack applying nine manipulated pixels over 200 attack iterations.

TABLE 7.3: Anti-spoofing performance of the FASNet model on the 3DMAD dataset; and the recall drop caused by the adversarial attack with 9 manipulated pixels.

3DMAD	Before Attack				After Attack	Recall Drop(%)
DA used	ACC (%)	HTER (%)	Precision (%)	Recall(%)	Recall (%)	
None DA	100	0	100	100	81	19.00
Shift	100	0	100	100	98	2.00
Shear	100	0	100	100	77	23.00
Zooming	100	0	100	100	41	59.00
Rotation	100	0	100	100	0	100.00
Channel shift	100	0	100	100	75	25.00
Flipping	100	0	100	100	87	13.00
All	100	0	100	100	48	52.00

TABLE 7.4: Anti-spoofing performance of the FASNet model on the CASIA dataset; and the recall drop caused by the adversarial attack with 9 manipulated pixels.

CASIA	Before Attack				After Attack	Recall Drop(%)
DA used	ACC (%)	HTER (%)	Precision (%)	Recall(%)	Recall (%)	
None DA	83.52	15.65	85.08	81.30	29.63	51.67
Shift	92.31	6.67	90.02	95.19	45.37	49.82
Shear	88.80	11.67	88.40	89.26	34.63	54.63
Zooming	92.78	7.41	92.56	93.15	29.07	64.08
Rotation	75.83	18.33	68.93	94.07	35.00	59.07
Channel shift	86.48	14.63	84.68	89.07	30.74	58.33
Flipping	91.67	8.89	89.89	93.89	37.04	56.85
All	96.02	3.80	97.15	94.81	27.41	67.40

TABLE 7.5: Anti-spoofing performance of the FASNet model on the Replay-Attack dataset; and the recall drop caused by the adversarial attack with 9 manipulated pixels.

Replay-Attack	Before Attack				After Attack	Recall Drop(%)
DA used	ACC (%)	HTER (%)	Precision (%)	Recall(%)	Recall (%)	
None DA	98.81	1.38	99.37	98.25	73.50	24.75
Shift	99.81	0.12	99.75	99.88	84.44	15.44
Shear	98.44	1.25	98.99	97.88	71.88	26.00
Zooming	99.94	0.12	99.88	100.00	63.12	36.88
Rotation	98.88	1.19	98.63	99.12	77.12	22.00
Channel shift	96.81	3.12	96.99	96.62	73.25	23.37
Flipping	98.94	1.12	98.00	99.00	82.62	16.38
All	100	0	100	100	66.38	33.62

Most importantly, the results indicate that DA significantly improved the FASNet model's anti-spoofing performance on the more complex datasets. For example, on the simple 3DMAD dataset, the model without DA reached an accuracy of 100%, whereas, on the CASIA dataset, the model without DA achieved an accuracy of 83.52%, which is much worse than the 96.02% accuracy achieved by the model applying all DA techniques.

The second important finding is that using an adversarial attack with nine manipulated pixels can significantly reduce the FASNet model's ability to recall all spoofing faces. For example, on the Replay-Attack dataset, the FASNet model with DA was able to correctly

TABLE 7.6: Anti-spoofing performance of the FASNet model on the ROSE-Youtu dataset; and the recall drop caused by the adversarial attack with 9 manipulated pixels.

ROSE-Youtu	Before Attack				After Attack	Recall Drop(%)
DA used	ACC (%)	HTER (%)	Precision (%)	Recall(%)	Recall (%)	
None DA	87.93	11.76	85.79	90.91	39.56	51.35
Shift	90.76	9.35	93.32	87.80	47.43	40.37
Shear	86.68	11.40	81.69	94.57	58.26	36.31
Zooming	91.98	8.02	92.74	91.09	39.27	51.82
Rotation	88.51	10.47	85.16	93.27	49.78	43.49
Channel shift	88.35	10.69	84.86	93.36	57.17	36.19
Flipping	89.78	9.40	78.30	93.10	54.93	38.17
All	91.92	8.57	93.52	90.07	36.35	53.72

prevent spoofing with 100% accuracy; when the attack manipulated nine pixels of a fake facial image, however, 33.62% of the fake faces were wrongly identified as actual faces. The same results were obtained on the ROSE-Youtu dataset, in which 53.72% of the fake faces were wrongly classified as genuine faces.

5.2.2 Effect of the Number of Manipulated Pixels

Next, the effect of the number of pixels manipulated by the adversarial attack on perturbation effectiveness was assessed. Specifically, different adversarial samples were generated by changing the number of manipulated pixels per image within a range from one to nine while applying a constant 200 attack iterations. Figure 7.3 shows the variations in recall drop with the number of number of changed pixels per image over the four datasets.

It is seen that increasing the number of pixels manipulated leads to a more significant recall drop, which demonstrates the impact of the number of changed pixels in increasing the likelihood of a successful attack. Given these results, we believe that even well-trained models can be relatively vulnerable to such perturbations under strict constraints.

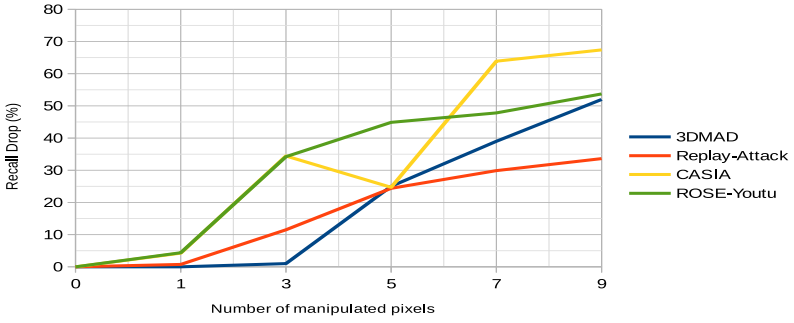


FIGURE 7.3: Number of attacked images along with number of manipulated pixels per image for the 3DMAD, CASIA, Replay-attack, and ROSE-Youtu datasets.

5.2.3 Effect of the Number of Attack Iterations

We then measured the adversarial attack performance as a function of the number of attack iterations. To do so, the pixel perturbation process was applied under a varying number of iterations to specify the number of generations that the attack algorithm should run before giving up. Figure 7.4 shows the recall drop as a function of number of attack iterations with the number of manipulated pixels held constant at nine.

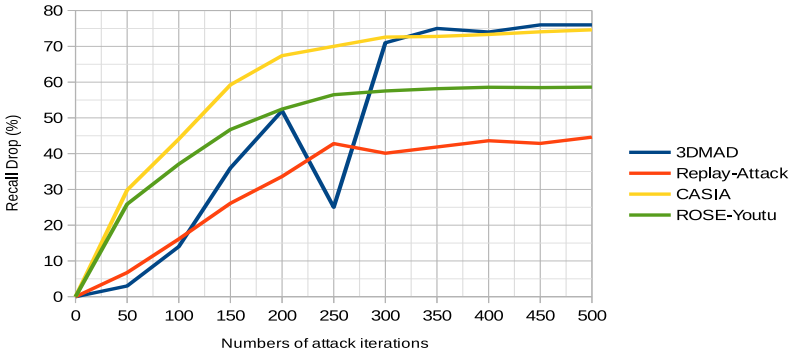


FIGURE 7.4: Recall drop with respect to different numbers of attack iterations. The number of manipulated pixels is kept as 9.

It is seen that, on all of the datasets, increasing the number of attack iterations increases the chance that the attack will deceive the anti-spoofing model, as the recall drop grows with the number of iterations. When the number of attack iterations is less than 300, increasing the number of iterations has a more significant effect on the recall drop; in general, the most significant recall drop is achieved by the time the number of attack iterations reaches 300.

5.2.4 Effect of the Number of Training Steps Used by the Anti-spoofing Model

Finally, we investigated the effect of the number of training steps used by the anti-spoofing model. The results listed in Tables 7.7 and 7.8 show the anti-spoofing performance obtained at various numbers of training steps on the CASIA and Replay-Attack datasets, respectively.

It is evident from both tables that increasing the number of training steps does not make a model more resistant to adversarial attack, as the post-attack recall variations do not decrease as the number of steps increases. For instance, the post-attack recall on the Replay-Attack dataset is 50.75% when the model is trained using 2,500 steps; this is increased to 66.38% when the number of steps is increased to 50,000. Similar results are obtained on the CASIA dataset, on which a recall value of 16.85% is obtained when the model is trained using 10,000 training steps and increases to 27.41% at 50,000 steps.

TABLE 7.7: Recall variation with respect to different numbers of training steps for the anti-spoofing model on the CASIA dataset. The number of manipulated pixels is kept as 9; the number of attack iterations is kept as 200.

CASIA	Number of training steps	10000	20000	30000	40000	50000
Before attack	Accuracy (%)	95.02	95.74	95.56	95.65	96.02
	HTER (%)	4.26	3.06	4.63	4.35	3.80
	Precision (%)	97.47	99.01	96.42	95.23	97.15
	Recall (%)	92.59	92.41	94.63	96.11	94.81
After attack	Recall (%)	16.85	22.78	20.00	28.15	27.41
	Recall drop (%)	75.74	69.63	74.63	67.96	67.40

TABLE 7.8: Recall variation with respect to different numbers of training steps for the anti-spoofing model on the Replay-Attack dataset. The number of manipulated pixels is kept as 9; the number of attack iterations is kept as 200.

Replay-Attack	Number of training steps	2500	10000	20000	30000	40000	50000
Before attack	Accuracy (%)	99.12	100.00	100.00	99.94	99.94	100.00
	HTER (%)	1.00	0.00	0.00	0.00	0.00	0.00
	Precision (%)	99.00	100.00	100.00	99.88	99.88	100.00
	Recall (%)	99.25	100.00	100.00	100.00	100.00	100.00
After attack	Recall (%)	50.75	56.88	64.75	55.00	41.75	66.38
	Recall drop (%)	48.50	43.12	35.25	45.00	58.25	33.62

5.2.5 Variation of Fitness Values

More experiments are carried out over 40 randomly selected fake images of each dataset to assess the differential evolution-based adversarial attack convergence. The adversarial attack is applied by manipulating 9 pixels for 200 attack iterations. The fitness values of the four datasets' evolution during the 200 attack iterations are illustrated in figure 7.5, where red lines highlight the average values. Since the fitness values are set to be the original classes' confidence, the attack aims to decrease these values to reach 0.5 or lower. Once this value is achieved, the attack process ends, proving the constant fitness values under the 0.5 y-axis. Additionally, it can be noted from the figure that the average fitness value regularly decreases with the number of generations, which proves that the adversarial attack works as expected for most of the attacked images.

6 Conclusion

It has been recently revealed that even the most efficient face verification and recognition systems are vulnerable to adversarial attacks, where an assailant can readily get face images or videos from social networks to generate sophisticated models to deceive the authentication systems. Therefore, numerous efforts have been done to build anti-spoofing classifiers as an additional layer to provide more protection. In this context, we presented in this chapter the results of an analytical study involving the application of

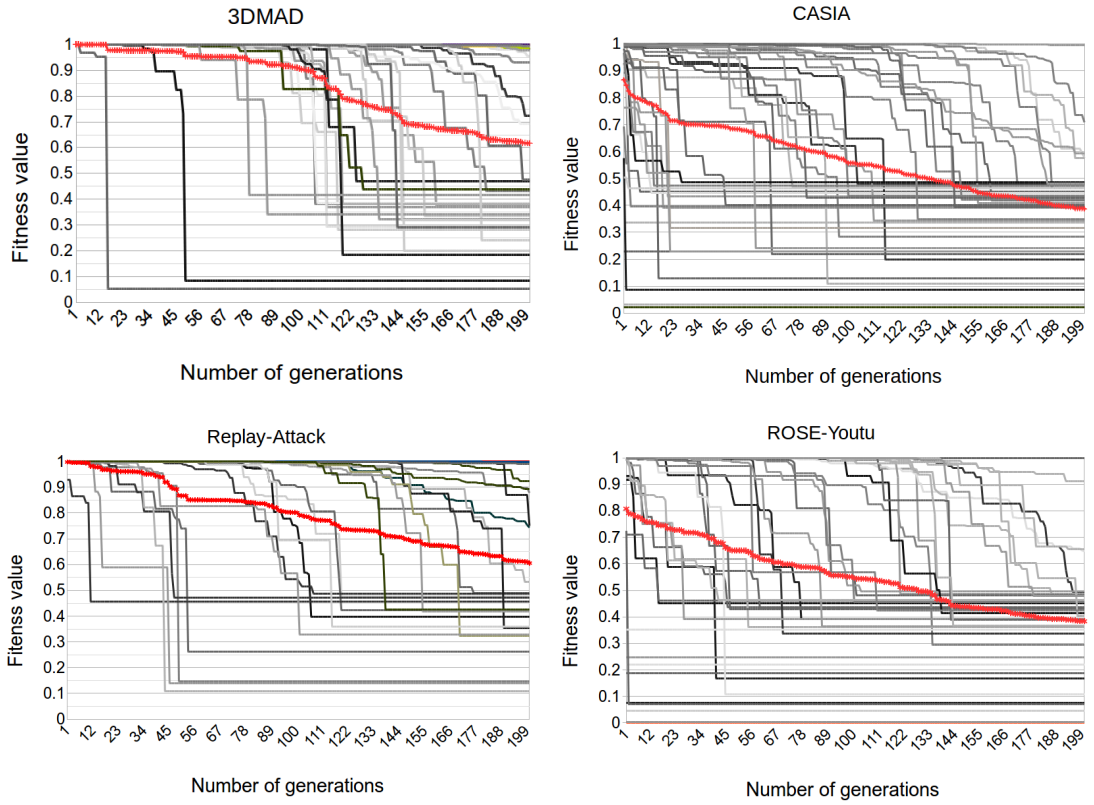


FIGURE 7.5: Fitness values changes during 200 attack iterations among the four different face anti-spoofing face datasets.

differential evolution-based adversarial attacks to a VGG16-based face liveness detection model. The main goal of this study was to highlight practical criteria that can be used to improve defense strategies for deep learning-based anti-spoofing systems against such attacks. To do this, we conducted a series of experiments under different use-case scenarios to analyze the trade-off between adversarial attack algorithms and anti-spoofing systems. The use-case scenarios applied eight data augmentation techniques and varied the number of training steps used to generate the anti-spoofing models, the number of manipulated pixels used in the attack, and the number of attack iterations. The results of experiments conducted on four face anti-spoofing databases revealed that training a model using more steps does not always reduce the recall drop. In addition, we found that increasing the number of pixels manipulated by the adversarial attack leads to a more significant recall drop and that DA could notably improve the VGG16-based anti-spoofing model's performance on the more complex datasets. However, we did not find that data augmentation can effectively protect the models from the adversarial attacks.

8 Conclusion and Perspectives

In this dissertation, we addressed the biometric authentication systems' security with a particular focus on the system's database attacks and attacks on the user sensor. Thus, the thesis is split into two main contributions: first, we highlight face-fingerprint-based authentication systems' robustness against template database attacks. Second, we tackle the spoofing attacks by analyzing deep learning face anti-spoofing classifiers resistance, when confronted with adversarial attacks.

In the first contribution, we intend to design secure approaches for biometric template protection in authentication systems by enhancing the preliminary presented state-of-the-art methods. Generally, an efficient biometric template protection approach should meet four requirements, namely diversity, revocability, security, and performance.

As a prime stage of the first contribution, we suggested a spatial watermarking scheme for multimodal authentication systems to fuse face and fingerprint modalities by embedding the face features (watermark) into the fingerprint image (cover). The experimental results showed higher authentication security and better performance accuracy.

Although watermarking-based spatial domain schemes can potentially hide information from perceptibility within the host image, they remain vulnerable to frequency attacks. Moreover, no single method provides all required properties of biometric template protection (revocability, diversity, security and performance). Thus, a hybrid approach for face and fingerprint template protection is proposed by combining a DTCWT-DCT based frequency watermarking method and the partial Hadamard transform. Another multimodal hybrid biometric template protection approach is proposed by combining three algorithms, including secure sketch, the DTCWT-DCT watermarking and a 3D chaotic map image encryption method. The experimental results demonstrated the efficiency of both hybrid approaches in meeting the four required properties at a time due to the use of potential points, including multimodality and combination of different methods.

Nonetheless, as facial authentication systems become acquainted with the general public, they become a target for spoofing attacks. Thus, the second contribution of the thesis lies in the ongoing struggle between anti-spoofing systems and adversarial attacks. Herein, we presented an analytical study emulating a differential evolution-based adversarial attack to transfer-learning-based convolutional neural networks (CNNs) face anti-spoofing classifier, in order to assess the impact of various criteria related to both spoofing and anti-spoofing algorithms in improving the overall robustness of deep face anti-spoofing mechanisms.

Perspectives

As a continuation of my PhD work, we hope to achieve the following in our future research:

- The application of the proposed template protection approaches on large-scale databases and real-world applications.
- Design and implement novel deep learning-based template protection approaches for biometric authentication systems; notably for face, fingerprint, iris, and voice modalities;
- Design and implement a novel deep-learning-based face anti-spoofing classifier based on the criteria investigated in the proposed analytical study.
- Combine data augmentation techniques and Generative Adversarial Networks (GANs) to generate adversarial attacks and new anti-spoofing datasets.

Additionally, We intend to work on face morphing attacks' detection using machine learning mechanisms.

Bibliography

[1] Umut Uludag and Anil K Jain. Attacks on biometric systems: a case study in fingerprints. In *Security, Steganography, and Watermarking of Multimedia Contents VI*, pages 622–633, 2004.

[2] Anil K Jain, Karthik Nandakumar, and Abhishek Nagar. Biometric template security. *EURASIP Journal on Advances in Signal Processing*, 2008:1–13, 2008.

[3] Anil K Jain, Patrick Flynn, and Arun A Ross. *Handbook of Biometrics*. Springer US, 2008.

[4] Anil K Jain, Arun Ross, and Salil Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14:4–20, 2004.

[5] Alejandro Enrique Flores Zuniga, Khin Than Win, and Willy Susilo. Biometrics for electronic health records. *Journal of Medical Systems*, 34:975–983, 2010.

[6] L Lam and C Y Suen. Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 27:553–568, 1997.

[7] Lei Xu, Adam Krzyzak, and Ching Y Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 22:418–435, 1992.

[8] Ludmila I Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms, 2nd Edition*. Wiley, 2004.

[9] Davide Maltoni, Dario Maio, Anil K Jain, and Salil Prabhakar. *Handbook of Fingerprint Recognition*. Springer-Verlag, 2003.

[10] Nalini K Ratha, Jonathan Connell, Ruud M Bolle, and Sharat Chikkerur. Cancelable biometrics: A case study in fingerprints. In *IEEE International Conference Pattern Recognition*, pages 370–373, 2006.

[11] Nalini K Ratha, Jonathan Connell, and Ruud M Bolle. An analysis of minutiae matching strength. In *International Conference on Audio- and Video-Based Biometric Person Authentication*, pages 223–228, 2001.

[12] Umut Uludag, Sharath Pankanti, Salil Prabhakar, and Anil K Jain. Biometric cryptosystems: issues and challenges. *Proceedings of the IEEE*, 92:948–960, 2004.

[13] Feng Hao, Ross Anderson, and John Daugman. Combining crypto with biometrics effectively. *IEEE Transactions on Computers*, 55:1081–1088, 2006.

[14] Ann Cavoukian and Alex Stoianov. *Biometric Encryption.* Springer US, 2011.

[15] Ann Cavoukian and Alex Stoianov. *Biometric Encryption: The New Breed of Untraceable Biometrics.* Biometrics: Fundamentals, Theory, and Systems Wiley, 2009.

[16] Ari Juels and Martin Wattenberg. A fuzzy commitment scheme. In *Sixth ACM Conference on Computer and Communications Security*, pages 28–36, 1999.

[17] Ari Juels and Madhu Sudan. A fuzzy vault scheme. In *IEEE International Symposium on Information Theory*, pages 237–257, 2006.

[18] Qiming Li, Yagiz Sutcu, and Nasir Memon. Secure sketch for biometric templates. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 99–113, 2006.

[19] Yevgeniy Dodis, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In *International Conference on the Theory and Applications of Cryptographic Techniques*, pages 523–540, 2004.

[20] Chengfang Fang, Qiming Li, and Ee-Chien Chang. Secure sketch for multiple secrets. In *International Conference on Applied Cryptography and Network Security*, pages 367–383, 2010.

[21] Julien Bringer, Hervé Chabanne, and Bruno Kindarji. The best of both worlds: Applying secure sketches to cancelable biometrics. *Science of Computer Programming*, 74:43–51, 2008.

[22] Yagiz Sutcu, Qiming Li, and Nasir Memon. Protecting biometric templates with sketch: Theory and practice. *IEEE Transactions on Information Forensics and Security*, 2:503–512, 2007.

[23] Bin Ma, Chunlei Li, Zhaoxiang Zhang, and Yunhong Wang. Biometric information hiding: Promoting multimedia security with content and identity authentication. In *IEEE China Summit and International Conference on Signal and Information Processing*, pages 442–446, 2013.

[24] Dong Zheng, Yan Liu, and Jiying Zhao. A survey of RST invariant image watermarking algorithms. In *Canadian Conference on Electrical and Computer Engineering*, 2006.

[25] Preeti Parashar and Rajeev Kumar Singh. A survey: Digital image watermarking techniques. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 7:111–124, 2014.

[26] S Radharani and M L Valarmathi. A study on watermarking schemes for image authentication. *International Journal of Computer Applications*, 2:24–32, 2010.

[27] Bin Ma, Chunlei Li, Yunhong Wang, Zhaoxiang Zhang, and Yiding Wang. Block pyramid based adaptive quantization watermarking for multimodal biometric authentication. In *20th International Conference on Pattern Recognition*, pages 1277–1280, 2010.

[28] Bin Ma, Chunlei Li, Zhaoxiang Zhang, and Yunhong Wang. Sparse reconstruction based watermarking for secure biometric authentication. In *Chinese Conference on Biometric Recognition*, pages 244–251, 2011.

[29] Khaled Loukhaoukha, Ahmed Refaey, and Khalil Zebbiche. Comments on "a robust color image watermarking with singular value decomposition method". *Advances in Engineering Software*, 93:44–46, 2016.

[30] Sengul Dogan, Turker Tuncer, Engin Avci, and Arif Gulten. A robust color image watermarking with singular value decomposition method. *Advances in Engineering Software*, 42:336–346, 2011.

[31] Mohammed Alkhathami, Fengling Han, and Ron Van Schyndel. Fingerprint image watermarking approach using dtcwt without corrupting minutiae. In *6th International Congress on Image and Signal Processing*, 2013.

[32] Y Rangaswamy, K B Raja, and K R Venugopal. FRDF: Face recognition using fusion of dtcwt and fft features. *Procedia Computer Science*, 54:809–817, 2015.

[33] Mohammed Alkhathami, Fengling Han, and Ron Van Schyndel. Fingerprint image protection using two watermarks without corrupting minutiae. In *IEEE 8th Conference on Industrial Electronics and Applications*, pages 1151–1155, 2013.

[34] Mita Paunwala and Suprava Patnaik. Biometric template protection with DCT-based watermarking. *Machine Vision and Applications*, 25:263–275, 2014.

[35] Wioletta Wojtowicz and Marek R Ogiela. Digital images authentication scheme based on bimodal biometric watermarking in an independent domain. *Journal of Visual Communication and Image Representation*, 38:1–10, 2016.

[36] Sanaa Ghouzali. Watermarking based multi-biometric fusion approach. In *International Conference on Codes, Cryptology, and Information Security*, pages 342–351, 2015.

[37] Mayank Vatsa, Richa Singh, and Afzel Noore. Feature based RDWT watermarking for multimodal biometric system. *Image and Vision Computing*, 27:293–304, 2009.

[38] Himanshu Agarwal, Balasubramanian Raman, and Ibrahim Venkat. Blind reliable invisible watermarking method in wavelet domain for face image watermark. *Multimedia Tools and Applications*, 74:6897–6935, 2015.

[39] Khalil Zebbiche, Fouad Khelifi, and Khaled Loukhaoukha. Robust additive watermarking in the dtcwt domain based on perceptual masking. *Multimedia Tools and Applications*, 77:21281–21304, 2018.

[40] Peng Li, Xin Yang, Hua Qiao, Kai Cao, Eryun Liu, and Jie Tian. An effective biometric cryptosystem combining fingerprints with error correction codes. *Expert Systems with Applications*, 39:6562–6574, 2012.

[41] Fausto Abundiz-Pérez, César Cruz-Hernández, Miguel Angel Murillo-Escobar, Rosa Martha López-Gutiérrez, and Adrian Arellano-Delgado. A fingerprint image encryption scheme based on hyperchaotic rössler map. *Mathematical Problems in Engineering*, 2016, 2016.

[42] Mahendra V Patil, Avinash D Gawande, and Dilendra. Biometric image encryption algorithm based on modified rubik's cube principle. In *International Conference on ISMAC in Computational Vision and Bio-Engineering*, pages 1865–1873, 2018.

[43] Jenny Joseph and Josy Elsa Varghese. Analysis of chaotic image encryption using 1D logistic map, 2D arnold cat map and 3D arnold cat map. *Interntional Journal of Scientific research and Management*, 3:3754–3761, 2015.

[44] Ali M Meligy, Hossam A Diab, and Marwa S El-Danaf. Chaos encryption algorithm using key generation from biometric image. *International Journal of Computer Applications*, 149:14–20, 2016.

[45] Hung-I Hsiao and Junghsi Lee. Fingerprint image cryptography based on multiple chaotic systems. *Signal Processing*, 113:169–181, 2015.

[46] Jincey John and Ashji S Raj. Reliable biometric data encryption using chaotic map. *International Journal of Advanced Research in Computer and Communication Engineering*, 4:52–56, 2015.

[47] Xing-Yuan Wang, Ying-Qian Zhang, and Xue-Mei Bao. A novel chaotic image encryption scheme using dna sequence operations. *Optics and Lasers in Engineering*, 73:53–61, 2015.

[48] Suzwani Ismail, Fakariah Hani Hj Mohd Ali, and Syed Ahmad Aljunid. A new hybrid approach for securing multibiometric templates based on cancelable and fuzzy commitment scheme. *Australian Journal of Basic and Applied Sciences*, 9: 72–76, 2015.

[49] Noha A Hikal and Marwa M Eid. A new approach for palmprint image encryption based on hybrid chaotic maps. *Journal of King Saud University - Computer and Information Sciences*, 2018.

[50] Marta Gomez-Barrero, Christian Rathgeb, Javier Galbally, Christoph Buschb, and Julian Fierrez. Unlinkable and irreversible biometric template protection based on bloom filters. *Information Sciences*, 370-371:18–32, 2016.

[51] Zhendong Wu, Longwei Tian, Ping Li, Ting Wu, Ming Jiang, and Chunming Wu. Generating stable biometric keys for flexible cloud computing authentication using finger vein. *Information Sciences*, 433-434:431–447, 2018.

[52] Rokiah Abdullah, Hariharan Muthusamy, Vikneswaran Vijean, Zulkapli Abdullah, and Farah Nazlia Che Kassim. Real and complex wavelet transform approaches for malaysian speaker and accent recognition. *Pertanika Journal of SCIENCE & TECHNOLOGY*, 27:737–752, 2019.

[53] Wadood Abdul, Abdullah Alzamil, Hammam Masri, Qazi Emad ul Haq, Sanaa Ghouzali, Muhammad Hussain, and Mansour AlZuair. Fingerprint and iris template protection for health information system access and security. *Journal of Medical Imaging and Health Informatics*, 7:1302–1308, 2017.

[54] L OǴorman. Comparing passwords, tokens, and biometrics for user authentication. In *Proceedings of the IEEE*, pages 2021–2040, 2004.

[55] Dario Maio, Davide Maltoni, Raffaele Cappelli, J L Wayman, and Anil K Jain. FVC2000: fingerprint verification competition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:402–412, 2002.

[56] Anil K Jain, Arun A Ross, and Karthik Nandakumar. *Introduction to Biometrics*. Springer US, 2011.

[57] Vedrana Krivokuća and Waleed Abdulla. Cancellability and diversity analysis of fingerprint template protection scheme based on compact minutiae pattern. *Information Security Journal: A Global Perspective*, 25:109–123, 2016.

[58] Sanjay Ganesh Kanade, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. Cancelable biometrics for better security and privacy in biometric systems. In *International Conference on Advances in Computing and Communications*, pages 20–34, 2011.

[59] The database of faces. http://cam-orl.co.uk/facedatabase.html, 2001. Accessed:2020-18-02.

[60] FVC2000 fingerprint verification competition. http://bias.csr.unibo.it/fvc2000/db1.asp, 2000-2002. Accessed:2020-18-02.

[61] FVC2002 fingerprint verification competition. http://bias.csr.unibo.it/fvc2002/databases.asp, 2000-2002. Accessed:2020-18-02.

[62] Deng Cai, Xiaofei He, Jiawei Han, and Hong-Jiang Zhang. Orthogonal laplacianfaces for face recognition. *IEEE Transactions on Image Processing*, 15:3608–3614, 2006.

[63] Ashraf El-Sisi. Design and implementation biometric access control system using fingerprint for restricted area based on gabor filter. *International Arab Journal of Information Technology*, 8:355–363, 2011.

[64] Anil K Jain, Umut Uludag, and Rein-Lien Hsu. Hiding a face in a fingerprint image. In *International Conference on Pattern recognition*, pages 756–759, 2002.

[65] Naser Damer, Alexander Opel, and Alexander Nouak. Performance anchored score normalization for multi-biometric fusion. In *International Symposium on Visual Computing*, pages 68–75, 2003.

[66] Abhishek Nagar. Secure biometric recognition. In *PRIP Seminar*, 2008.

[67] Nick Kingsbury. The dual-tree complex wavelet transform: A new technique for shift invariance anddirectional filters. In *Proceedings of the IEEE Digital Signal Processing Workshop*, pages 9–12, Utah.

[68] K Ramani, E V Prasad, and V Sourirajan. Protecting digital images using DTCWT-DCT. In *Proceedings of the International Conference on Advances in Information and Communication Technologies*, pages 36–44, Heidelberg, 2010.

[69] Haiyun Xu, Raymond N J Veldhuis, Tom A M Kevenaar, and Ton A H M Akkermans. A fast minutiae-based fingerprint recognition system. *IEEE Systems Journal*, 3:418–427, 2009.

[70] Haiyun Xu, Raymond N J Veldhuis, Asker M Bazen, Tom A M Kevenaar, Ton A H M Akkermans, and Berk Gokberk. Fingerprint verification using spectral minutiae representations. *IEEE Transactions on Information Forensics and Security*, 4:397–409, 2009.

[71] Rao K Yarlagadda and John E Hershey. *Hadamard Matrix Analysis and Synthesis.* Springer US, 1997.

[72] Song Wang and Jiankun Hu. A hadamard transform-based method for the design of cancellable fingerprint templates. In *Proceedings of the International Congress on Image and Signal Processing*, pages 1682–1687, Hangzhou, China, 2013.

[73] Song Wang, Guang Deng, and Jiankun Hu. A partial hadamard transform approach to the design of cancelable fingerprint templates containing binary biometric representations. *Pattern Recognition*, 61:447–458, 2017.

[74] Harkeerat Kaur and Pritee Khanna. Cancelable biometrics using hadamard transform and friendly random projections. In *Proceedings of the International Conference on Computer Vision and Image Processing*, pages 47–57, 2016.

[75] Naima Bousnina, Gouzali Sanaa, Mikram Mounia, and Wadood Abdul. DTCWT-DCT watermarking method for multimodal biometric authentication. In *International Conference on Networking, Information Systems & Security*, pages 1–7, 2019.

[76] Md Billal Hossain, Md Toufikur Rahman, A B M Saadmaan Rahman, and Sayeed Islam. A new approach of image encryption using 3D chaotic map to enhance security of multimedia component. In *International Conference on Informatics, Electronics & Vision*, pages 1–6, 2014.

[77] Oluwakemi Christiana Abikoye, Umar Abdulraheem Ojo, Joseph Bamidele Awotunde, and Roseline Oluwaseun Ogundokun. A safe and secured iris template using steganography and cryptography. *Multimedia Tools and Applications*, 79:23483–23506, 2020.

[78] Naima Bousnina, Sanaa Ghouzali, Meryem Lafkih, Ohood Nafea, Mounia Mikram, Wadood Abdul, and Driss Aboutajdine. Watermarking for protected fingerprint authentication. In *12th International Conference on Innovations in Information Technology*, pages 1–5, 2016.

[79] Sara Nazari, Mohammad-Shahram Moin, and Hamidreza Rashidy Kanan. Securing templates in a face recognition system using error-correcting output code and chaos theory. *Computers & Electrical Engineering*, 72:644–659, 2018.

[80] Yagiz Sutcu, Qiming Li, and Nasir Memon. Secure biometric templates from fingerprint-face features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6, 2007.

[81] Lamia Rzouga Haddada, Bernadette Dorizzi, and Najoua Essoukri Ben-Amaraa. A combined watermarking approach for securing biometric data. *Signal Processing: Image Communication*, 55:23–31, 2017.

[82] Lamia Rzouga Haddada and Najoua Essoukri BenAmara. Double watermarking based biometric access control for radio frequency identification card. *International Journal of RF and Microwave Computer-Aided Engineering*, 29:1–11, 2019.

[83] Jutta Hammerle-Uhl, Karl Raab, and Andreas Uhl. Watermarking as a means to enhance biometric systems: A critical survey. In *International Workshop on Information Hiding*, pages 238–254, 2011.

[84] Ohoud Nafea, Sanaa Ghouzali, Wadood Abdul, and Emad-ul-Haq Qazi. Hybrid multi-biometric template protection using watermarking. *The Computer Journal*, 59:1392–1407, 2016.

[85] Wadood Abdul, Ohoud Nafea, and Sanaa Ghouzali. Combining watermarking and hyper-chaotic map to enhance the security of stored biometric templates. *The Computer Journal*, 63:479–493, 2020.

[86] Tran Khanh Dang, Quynh Chi Truong, Thu Thi Bao Le, and Truong Hai. Cancellable fuzzy vault with periodic transformation for biometric template protection. *IET Biometrics*, 5:229–235, 2016.

[87] Anil K Jain, Arun Ross, and Sharath Pankanti. Biometrics: a tool for information security. *IEEE Transactions on Information Forensics and Security*, 1:125–143, 2006.

[88] IBG. Biometrics market and industry report 2009-2014. Technical report, 2008.

[89] Bela Gipp, Jöran Beel, and Ivo Rössling. *ePassport: The World's New Electronic Passport: A Report about the ePassport's Benefits, Risks and its Security*. CreateSpace, 2007.

[90] ISO/IEC 30107-1:2016 information technology biometric presentation attack detection part 1: Framework. https://www.iso.org/standard/53227.html, 2016. Accessed: 2019-26-08.

[91] Yan Li, Ke Xu, Qiang Yan, Yingjiu Li, and Robert H Deng. Understanding osn-based facial disclosure against face authentication systems. In *Proceedings of the 9th ACM symposium on Information, computer and communications security*, pages 413–424, 2014.

[92] Yang Xin, Yi Liu, Zhi Liu, Xuemei Zhu, Lingshuang Kong, Dongmei Wei, Wei Jiang, and Jun Chang. A survey of liveness detection methods for face biometric systems. *Sensor Review*, 37:1–25, 2017.

[93] David Menotti, Giovani Chiachia, Allan Pinto, William Robson Schwartz, Helio Pedrini, Alexandre Xavier Falcao, and Anderson Rocha. Deep representations for iris, face, and fingerprint spoofing detection. *IEEE Transactions on Information Forensics and Security*, pages 1–15, 2015.

[94] Yousef Atoum, Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Face anti-spoofing using patch and depth-based CNNs. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 319–328, 2017.

[95] Litong Feng, Lai-Man Po, Yuming Li, Xuyuan Xu, Fang Yuan, Terence Chun-HoCheung, and Kwok-Wai Cheung. Integration of image quality and motion cues for face anti-spoofing: A neural network approach. *Journal of Visual Communication and Image Representation*, 38:451–460, 2016.

[96] Javier Galbally, Sébastien Marcel, and Julian Fierrez. Biometric antispoofing methods: A survey in face recognition. *IEEE Access*, 2:1530–1552, 2014.

[97] Josef Bigun, Hartwig Fronthaler, and Klaus Kollreider. Assuring liveness in biometric identity authentication by real-time face tracking. In *Proceedings of the*

2004 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, pages 104–111, 2004.

[98] Asad Ali, Farzin Deravi, and Sanaul Hoque. Liveness detection using gaze collinearity. In *Third International Conference on Emerging Security Technologies*, pages 62–65, 2012.

[99] Klaus Kollreider, Hartwig Fronthaler, and Josef Bigun. Verifying liveness by multiple experts in face biometrics. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–6, 2008.

[100] Gang Pan, Lin Sun, Zhaohui Wu, and Shihong Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *IEEE 11th International Conference on Computer Vision*, pages 1–8, 2007.

[101] Liting Wang, Xiaoqing Ding, and Chi Fang. Face live detection method based on physiological motion analysis. *Tsinghua Science and Technology*, 14:685–690, 2009.

[102] Jiangwei Li, Yunhong Wang, Tieniu Tan, and Anil K Jain. Live face detection based on the analysis of fourier spectra. In *Proceedings Volume 5404, Biometric Technology for Human Identification*, pages 296–303, 2004.

[103] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and Stan Z Li. A face antispoofing database with diverse attacks. In *5th IAPR International Conference on Biometrics (ICB)*, pages 26–31, 2012.

[104] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *International Conference of Biometrics Special Interest Group (BIOSIG)*, pages 1–7, 2012.

[105] Keyurkumar Patel, Hu Han, and Anil K Jain. Secure face unlock: Spoof detection on smartphones. *IEEE Transactions on Information Forensics and Security*, 11: 2268–2283, 2016.

[106] Luan Tran, Xi Yin, and Xiaoming Liu. Disentangled representation learning gan for pose-invariant face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1415–1424, 2017.

[107] Avinash Kumar Singh, Piyush Joshi, and G. C. Nandi. Face recognition with liveness detection using eye and mouth movement. In *International Conference on Signal Propagation and Computer Technology*, pages 592–597, Ajmer, India, 2014.

[108] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face antispoofing using speeded-up robust features and fisher vector encoding. *IEEE Signal Processing Letters*, 24:141–145, 2016.

[109] Xiaoyang Tan, Yi Li, Jun Liu, and Lin Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *European Conference on Computer Vision*, pages 504–517, 2010.

[110] Tiago de Freitas Pereira, André Anjos, José Mario De Martino, and Sébastien Marcel. LBP-TOP based countermeasure against face spoofing attacks. In *Asian Conference on Computer Vision*, pages 121–132, 2012.

[111] Tiago de Freitas Pereira, André Anjos, José Mario De Martino, and Sébastien Marcel. Can face anti-spoofing countermeasures work in a real world scenario? In *International Conference on Biometrics*, pages 1–8, 2013.

[112] Jukka Komulainen, Abdenour Hadid, and Matti Pietikäinen. Context based face anti-spoofing. In *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems*, pages 1–8, 2013.

[113] Jianwei Yang, Zhen Lei, Shengcai Liao, and Stan Z Li. Face liveness detection with component dependent descriptor. In *International Conference on Biometrics*, pages 1–6, 2013.

[114] Wei Bao, Hong Li, Nan Li, and Wei Jiang. A liveness detection method for face recognition based on optical flow field. In *International Conference on Image Analysis and Signal Processing*, pages 233–236, 2009.

[115] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face anti-spoofing based on color texture analysis. In *IEEE International Conference on Image Processing*, pages 2636–2640, 2015.

[116] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security*, 11:1818–1830, 2016.

[117] I Chingovska, J Yang, Z Lei, D Yi, S Li, O Kahm, C Glaser, N Damer, A Kuijper, A Nouak, J Komulainen, T Pereira, S Gupta, S Khandelwal, S Bansal, A Rai, T Krishna, D Goyal, M A Waris, H Zhang, I Ahmad, S Kiranyaz, M Gabbouj, R Tronci, M Pili, N Sirena, F Roli, J Galbally, J Ficrrcz, A Pinto, H Pedrini, W Schwartz, A Rocha, A Anjos, and S Marcel. The 2nd competition on counter measures to 2d face spoofing attacks. In *IAPR Int. Conferenceon Biometrics (ICB)*, pages 1–6, 2013.

[118] Tiago de Freitas Pereira, Jukka Komulainen, André Anjos, José Mario De Martino, Abdenour Hadid, Matti Pietikäinen, and Sébastien Marcel. Face liveness detection using dynamic texture. *EURASIP Journal on Image and Video Processing*, 1, 2014.

[119] Jukka Komulainen, Abdenour Hadid, Matti Pietikäinen, André Anjos, and Sébastien Marcel. Complementary countermeasures for detecting scenic face spoofing attacks. In *International Conference on Biometrics*, pages 1–7, 2013.

[120] Cs231n: Convolutional neural networks for visual recognition. http://cs231n.stanford.edu/, 2015. Accessed:2020-13-04.

[121] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115:211–252, 2015.

[122] Ajian Liu, Jun Wan, Sergio Escalera, Hugo Jair Escalante, Zichang Tan, Qi Yuan, Kai Wang, Chi Lin, Guodong Guo, Isabelle Guyon, and Stan Z Li. Multi-modal face anti-spoofing attack detection challenge at CVPR2019. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[123] Chaitanya Nagpal and Shiv Ram Dubey. A performance evaluation of convolutional neural networks for face anti spoofing. In *International Joint Conference on Neural Networks*, pages 1–8, 2019.

[124] Gustavo Botelho de Souza, João Paulo Papa, and Aparecido Nilceu Marana. On the learning of deep local features for robust face spoofing detection. *arXiv:1806.07492*, 2018.

[125] Wei Hu, Gusi Te, Ju He, Dong Chen, and Zongming Guo. Exploring hypergraph representation on face anti-spoofing beyond 2d attacks. *arXiv:1811.11594*, 2018.

[126] Amin Jourabloo, Yaojie Liu, and Xiaoming Liu. Face de-spoofing: Anti-spoofing via noise modeling. *arXiv:1807.09968*, 2018.

[127] Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 389–398, 2018.

[128] Yao Liu, Ying Tai, Jilin Li, Shouhong Ding, Chengjie Wang, Feiyue Huang, Dongyang Li, Wenshuai Qi, and Rongrong Ji. Aurora guard: Real-time face anti-spoofing via light reflection. *arXiv:1902.10311*, 2019.

[129] Xiaoguang Tu, Hengsheng Zhang, Mei Xie, Yao Luo, Yuefei Zhang, and Zheng Ma. Enhance the motion cues for face anti-spoofing using CNN-LSTM architecture. *arXiv:1901.05635*, 2019.

[130] Yasar Abbas Ur Rehman, Lai-Man Po, Mengyang Liu, Zijie Zou, Weifeng Ou, and Yuzhi Zhao. Face liveness detection using convolutional-features fusion of real and deep network generated face images. *Journal of Visual Communication and Image Representation*, 59:574–582, 2019.

[131] Keyurkumar Patel, Hu Han, and Anil K Jain. Cross-database face antispoofing with robust feature representation. In *Chinese Conference on Biometric Recognition*, pages 611–619, 2016.

[132] Naveed Akhtar and Ajmal Mian. Threat of adversarial attacks on deep learning in computer vision: A survey. *IEEE Access*, 6:14410–14430, 2018.

[133] Nicolas Papernot, Patrick McDaniel, Ian J Goodfellow, Somesh K Jha, Z Berkay Celik, and Ananthram Swami. Practical black-box attacks against machine learning. In *ACM on Asia Conference on Computer andCommunications Security*, pages 506–519, 2017.

[134] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Fergus Rob. Intriguing properties of neural networks. *ICLR,abs/1312.6199*, 2014.

[135] Lu Yang, Qing Song, and Yingqi Wu. Attacks on state-of-the-art face recognition using attentional adversarial attack generative network. *Multimedia Tools and Applications*, pages 1–21, 2020.

[136] Avishek Joey Bose and Parham Aarabi. Adversarial attacks on face detectors using neural net based constrained optimization. *arXiv:1805.12302*, 2018.

[137] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K Reiter. Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In *ACM SIGSAC Conference on Computer and Communications Security*, pages 1528–1540, 2016.

[138] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K Reiter. A general framework for adversarial examples with objectives. *arXiv:1801.00349*, 2019.

[139] Xinyun Chen, Chang Liu, Bo Li, Kimberly Lu, and Dawn Song. Targeted backdoor attacks on deep learning systems using data poisoning. *arXiv:1712.05526*, 2017.

[140] Neslihan Kose and Jean-Luc Dugelay. On the vulnerability of face recognition systems to spoofing mask attacks. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2357–2361, Vancouver, BC, Canada, 2013.

[141] Zhe Zhou, Di Tang, Xiaofeng Wang, Weili Han, Xiangyu Liu, and Kehuan Zhang. Invisible mask: Practical attacks on face recognition with infrared. *arXiv:1803.04683*, 2018.

[142] Vargas Danilo Vasconcellos Su, Jiawei and and Kouichi Sakurai. Attacking convolutional neural network using differential evolution. *IPSJ Transactions on Computer Vision and Applications*, 11:1–12, 2019.

[143] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. Deepfool: a simple and accurate method to fool deep neural networks. *arXiv:1511.04599v3*, 2016.

[144] Minh Duc Nguyen and Quang Minh Bui. Your face is not your password. In *Black Hat DC 2009*, 2009.

[145] Unique identification authority of india (UIDAI). https://uidai.gov.in/, 2016. Accessed: 2019-26-08.

[146] Abdenour Hadid. Face biometrics under spoofing attacks: Vulnerabilities, countermeasures, open issues, and research directions. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 113–118, 2014.

[147] Luiz Souza, Luciano Oliveira, Mauricio Pamplona, and JoaO Paulo Papa. How far did we get in face spoofing detection? *Engineering Applications of Artificial Intelligence*, 72:368–381, 2018.

[148] Jukka Komulainen. *Software-based countermeasures to 2D facial spoofing attacks*. PhD thesis, University of Oulu, 2015.

[149] Lei Li, Paulo Lobato Correia, and Abdenour Hadid. Face recognition under spoofing attacks: countermeasures and research directions. *IET Biometrics*, 7:3–14, 2018.

[150] IJCB 2017 competition on generalized face presentation attack detection in mobile authentication scenarios. https://sites.google.com/site/faceantispoofing/, 2017. Accessed: 2019-26-08.

[151] Di Wen, Hu Han, and Anil K Jain. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, 10:746–761, 2015.

[152] Zinelabinde Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid. OULU-NPU: A mobile face presentation attack database with real-world variations. In *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pages 612–618, 2017.

[153] Trusted biometrics under spoofing attacks (TABULA RASA). http://www.tabularasa-euproject.org/, 2010. Accessed: 2019-26-08.

[154] Joseph Redmon and Ali Farhadi. Yolo9000: Better, faster, stronger. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017.

[155] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems (NIPS)*, pages 1097–1105, 2017.

[156] George Saon, Hong-Kwang J. Kuo, Steven Rennie, and Michael Picheny. The IBM 2015 english conversational telephone speech recognition system. *arXiv:150505899*, 2015.

[157] Junying Gan, Shanlu Li, Yikui Zhai, and Chengyun Liu. 3D convolutional neural network based on face anti-spoofing. In *2nd International Conference on Multimedia and Image Processing (ICMIP)*, pages 1–5, Wuhan, China, 2017.

[158] Lei Li, Xiaoyi Feng, Zinelabidine Boulkenafet, Zhaoqiang Xia, Mingming Li, and Abdenour Hadid. An original face anti-spoofing approach using partial convolutional neural network. In *Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6, Oulu, Finland, 2017.

[159] Jianwei Yang, Zhen Lei, and Stan Z Li. Learn convolutional neural network for face anti-spoofing. *arXiv:1408.5601v2*, 2014.

[160] Oeslle Lucena, Amadeu Junior, Vitor Moia, Roberto Souza, Eduardo Valle, and Roberto Lotufo. Transfer learning using convolutional neural networks for face anti-spoofing. In *International Conference Image Analysis and Recognition*, pages 27–34, 2017.

[161] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*, 2015.

[162] Rainer Storn and Kenneth Price. Differential evolution a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359, 1997.

[163] Jiawei Su, Danilo Vasconcellos Vargas, and Sakurai Kouichi. One pixel attack for fooling deep neural networks. *arXiv:1710.08864*, 2018.

[164] Nesli Erdogmus and Sébastien Marcel. Spoofing in 2D face recognition with 3D masks and anti-spoofing with kinect. In *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6, 2013.

[165] Haoliang Li, Wen Li, Hong Cao, Shiqi Wang, Feiyue Huang, and Alex C Kot. Unsupervised domain adaptation for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 13, 2018.

[166] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23:1499–1503, 2016.

Résumé:

Les systèmes d'authentification biométriques ont été proposé vu que les systèmes d'authentification traditionnels sont devenus incapable de rivaliser les niveaux élevés de la sécurité dus à l'avancement technologique. En dépit des bénéfices offerts par les systèmes d'authentification biométriques, ils restent toujours faibles à l'égard de divers types d'attaques, à savoir, (1) les attaques contre le capteur du système, (2) les attaques contre l'interface entre les modules du système, (3) les attaques contre les modules et (4) les attaques contre la base des données du système. Cette thèse se focalise sur la résistance des systèmes d'authentification biométriques aux attaques contre le capteur et la base des données du système. Par conséquent, les contributions apportées sont regroupées en deux axes principaux. Le premier axe fait le point sur la protection des modèles biométriques – visage et empreinte digitale – contre les attaques de la base des données. Dans cette contribution, trois approches de protection ont été proposé, une méthode de tatouage spatial et deux approaches hybrides. La première approche hybrid combine l'algorithme de tatouage transformée en ondelette complexe à double arbre-transformée en cosinus discret (DTCWT-DCT) et le transformée d'Hadamard partiel. Ainsi que la deuxième approche hybride fusionne la même méthode de tatouage, le secure sketch, et une méthode de cryptage de l'image basée sur la carte chaotique 3D. Le deuxième axe présente une étude analytique des attaques adverses aux mécanismes d'anti-spoofing de visage, en se basant sur l'apprentissage en profondeur. Le but de cette étude est l'évaluation du comportement de ce type de mécanisme d'anti-spoofing à l'égard des attaques adverses. De plus, ressortir les critères pratiques qui peuvent être pris en considération pour améliorer la performance des systèmes de détection de la vivacité du visage.

Mots clés: *Authentification biométrique, Tatouage biométrique, Protection des modèles biométriques, Spoofing et anti-spoofing de visage, Réseaux de neurones convolutifs.*

Abstract:

As the conventional token-based or knowledge-based personal identification techniques have become unable to satisfy higher security levels, biometric authentication systems have been suggested as an alternative solution. Despite the advantages provided by this kind of authentication mechanisms, they are not entirely protected against diverse types of attacks, namely (1) Attacks on the user sensor, (2) Attacks on the interface between modules, (3) Attacks on software modules, and (4) Attacks on the system's database. The presented thesis points up the biometric authentication systems' resistance against the user sensor and system's database attacks. More precisely, the contributions of this dissertation are grouped into two main axes. In the first axis, we intended to design secure frameworks for biometric template protection by enhancing the preliminary presented state-of-the-art methods. Therefore, this axis consists of three contributions, including a spatial fingerprint-face-based watermarking scheme and two hybrid approaches for face and fingerprint template protection. The first hybrid approach combines a Dual-Tree Complex Wavelet Transform-Discrete Cosine Transform (DTCWT-DCT)-based watermarking algorithm and the partial Hadamard transform, while the second fuses the same watermarking algorithm, a secure sketch algorithm, and a 3D chaotic map image encryption method. The second axis presents an analytical study on a differential evolution-based adversarial attack to deep-learning-based face liveness detection models, in order to highlight practical criteria that can be used in the development of countermeasures to address face-spoofing issues.

Keywords*: Biometric authentication, Biometric watermarking, Biometric template protection, Face anti-spoofing, Face spoofing, Convolutional neural networks.*

Année Universitaire : 2020/2021