Use of Biometrics for the Regeneration of Revocable Crypto-biometric Keys

Mohamed Amine Hmani

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2022IPPAS013

Thèse de doctorat

Use of Biometrics for the Regeneration of Revocable Crypto-biometric Keys

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à Télécom SudParis

École doctorale n°626 École doctorale de l'Institut Polytechnique de Paris (EDIPP)
Spécialité de doctorat : Informatique

Thèse présentée et soutenue à Massy, le 10 Novembre 2022, par

MOHAMED AMINE HMANI

Composition du Jury :

Hossam AFIFI Professeur, Institut Polytechnique de Paris	Président de Jury
Teddy FURON Directeur de Recherche, INRIA	Rapporteur
Chafic MOKBEL Professeur Universite de Balamand	Rapporteur
Bhiksha RAJ Professeur Carnegie Mellon University	Examinateur
Sanjay KANADE Professeur Bhivarabai Sawant College of Engineering and Research	Examinateur
Bernadette DORIZZI Professeur émérite,Institut Polytechnique de Paris	Directeur de thèse
Mme Dijana PETROVSKA DELACRETAZ Maître de conférences, Institut Polytechnique de Paris	Co-directeur de thèse

Abstract

The thesis aims to regenerate crypto-biometric keys (cryptographic keys obtained with biometric data) that are resistant to quantum cryptanalysis methods. The challenge is to obtain keys with high entropy to have a high level of security, knowing that the entropy contained in biometric references limits the entropy of the key. In the last years, mainly due to the advances of deep learning, and more concretely convolutional networks, the quality of image recognition and object detection has been progressing at a dramatic pace. With the advent of GPU computation and big datasets, neural networks saw a huge resurgence. This results in huge improvements in image recognition and consequently face recognition. Many works [Den+19a; Tai+14; Yi+14; Sun+15; PVZ15; SKP15] report near-perfect biometric performance. As such, we decided to take advantage of facial biometrics.

We started the pipeline by creating a face recognition system based on publicly available databases and models.

With the constant advancements in GPU computational power and the availability of open-source software, the reproducibility of published results should not be a problem. But, if the architectures of the systems are private and databases are proprietary, the reproducibility of published results can not be easily attained. To tackle this problem, we focus on training and evaluation of face recognition systems on publicly available data and software. We exploit the OpenFace open source system to generate a deep convolutional neural network model using publicly available datasets. We study the impact of the size of the datasets and their quality and compare the performance to a classical face recognition approach. Our focus is to have a fully reproducible model. To this end, we used publicly available datasets (FRGC, MS-celeb-1M, MOBIO, LFW), as well publicly available software (OpenFace) to train our model in order to do face recognition. We also evaluated our best model on the challenging video dataset MOBIO and report competitive results with the best-reported results on this database.

We participated in two H2020 projects using our face recognition system. For the SpeechXRays project, we provided implementations of classical and cancelable face biometrics. For the H2020 EMPATHIC project, we created a face verification REST API. We also participated in the NIST SRE19 multimedia challenge with the final version of our face recognition system which gave the best single system performance on the evaluation dataset.

To obtain cryptobiometric keys, it is necessary to have discrete biometric references. Crypto-biometric schemes, such as fuzzy commitment, require binary sources. We introduced a novel approach to binarizing biometric data using Deep Neural Networks (DNN) applied to facial biometric data. We present a data-driven template-binarization method using Deep Neural Networks, which does not degrade the performance of the baseline system. Furthermore, we seek to obtain long binary representations with high entropy to be used in crypto-biometric key regeneration schemes. The proposed binarization method has four main advantages: (i)The degradation in the recognition performance caused by the binarization is negligible compared to the baseline system. (ii)The binarization method can be applied to any type of real representation. (iii) The length of the binary representation can be controlled. The binarization method provides arbitrary length presentations that are limited only by the quality of the training database (size, noise). This allows for flexible representations that can be adapted to multiple applications, such as crypto-biometric key regeneration, fuzzy commitment, and fuzzy extraction schemes. (iv) The binarization method keeps the topology of the original space, which allows the use of binary representation in database searches and clustering.

The binary representations are evaluated on the MOBIO and Labeled Faces in the Wild (LFW) databases, where we measure their biometric recognition performance and entropy. The proposed binary embeddings give state-of-the-art performance on both databases with almost negligible degradation compared to the baseline. To get binary representations directly from face images, we proposed an original method, using auto-encoders and previously implemented classical face biometrics. We also exploited the binary representations to create a cancelable face-verification system.

Regarding our final goal, to generate crypto-biometric keys, we focused on symmetric keys. To this end, we tried to make the binary representation

longer and more discriminative. For the keys to be resistant to quantum computing, they should have double the length. The encryption keys need to have double the entropy of the keys used currently to present the same degree of security [Aug+15]. Symmetric encryption is threatened by the Grover algorithm because it reduces the complexity of a brute force attack on a symmetric key from 2^N to $2^{(N/2)}$. To mitigate the risk introduced by quantum computing, we need to increase the size of the keys. This is easy for standard symmetric keys but difficult for crypto-biometrics. Crypto-biometric keys are limited by the usable information contained in the biometric sample that they are generated from. The non-repudiation requirement is satisfied by the intrinsic properties of biometric samples. However, we must ensure that the scheme used in the key regeneration has a low False Acceptance Rate (FAR). Biometrics are unique for each user. They can not be changed without special circumstances (plastics surgery, diseases...). As such, if the regeneration scheme is not revocable, the user will be restricted to a single key across multiple applications. In addition, in the case the key is compromised, the user will not be able to create a new one. Thus, we must ensure that the regeneration scheme is revocable. Finally, the key regeneration scheme should allow for user convenience. Meaning, at the required security level, the user should not be rejected multiple times before have access to the system. The convenience of the system is shown through the FRR metric.

We succeeded in regenerating crypto-biometric keys longer than 400 bits, with low false acceptance and false rejection rates, thanks to the quality of the binary embeddings. The crypto-biometric keys have high entropy and are resistant to quantum cryptanalysis, according to the PQCrypto project, as they satisfy the length requirement. The keys are regenerated using a fuzzy commitment scheme that uses BCH codes.

The main contribution of this thesis is the binarization method based on auto-encoders, which gives long binary representations with high entropy and recognition accuracy. For future research direction, using newer face recognition systems with higher accuracy as the basis for the auto-encoder will improve the overall system performance and allow for the use of other techniques such as 'fuzzy extractor'. By implementing a fuzzy extractor scheme with high accuracy, it would be possible to generate the private key of newer quantum-resistant public encryption schemes.

Keywords

Biometrics, Face verification, Deep Learning, Cryptography.

Résumé

Dans cette thèse, nous avons abordé le problème de la régénération de clés crypto-biométriques (clés cryptographiques obtenues avec des données biométriques) résistantes aux méthodes de cryptanalyse quantique. L'enjeu est d'obtenir des clés à haute entropie pour avoir un haut niveau de sécurité, sachant que l'entropie contenue dans les données biométriques limite l'entropie de la clé. Après un chapitre d'introduction, nous présentons des travaux liés à nos travaux sur la reconnaissance faciale, la binarisation, la protection des modèles biométriques et le chiffrement dans le chapitre 2. Le chapitre 3 donne un aperçu des bases de données utilisées pour entrainer, tester et valider nos systèmes proposés.

Notre **première contribution** a été de créer un **système de reconnaissance faciale à la pointe de la technologie** basé sur des frameworks publics et des données accessibles au public. Au chapitre 4, nous présentons notre pipeline de système de reconnaissance faciale. Le système est construit sur le framework OpenFace, auquel nous avons apporté plusieurs modifications pour obtenir de meilleures performances, car il a été implémenté dans deux projets européens et utilisé dans une soumission au défi multimédia NIST SRE2019.

Nous détaillons également comment obtenir un système de reconnaissance faciale à la pointe de la technologie basé sur des logiciels accessibles au public et utilisant des ensembles de données publics. Nous essayons de donner le plus de détails possibles pour permettre la reproductibilité des résultats. Lorsque CMU a implémenté OpenFace, la reproductibilité était l'un de ses principaux objectifs. Ainsi, nous avons pu reproduire et améliorer leurs résultats. Par exemple, nous avons amélioré les performances de reconnaissance biométrique sur l'ensemble de données LFW de 92% pour le modèle CMU d'origine à 99% de précision.

D'après les résultats que nous avons obtenus, nous pouvons déduire que le goulot d'étranglement des performances se situe dans le prétraitement,

notamment la phase de détection des visages. Avec suffisamment de données, le Deep Convolutional Neural Network (DCNN) donne les meilleures performances. Néanmoins, dans les situations où les bases de données suffisamment grandes ne sont pas disponibles, les approches classiques donnent de meilleures performances.

Pour améliorer nos résultats, nous avons procédé à la suppression du bruit de mauvais étiquetage du MS-celeb-1M, qui a donné la plus grande amélioration des performances sur nos protocoles de validation.

Parmi les modifications appliquées à notre cadre, l'utilisation du détecteur de visage RetinaFace a entraîné l'amélioration la plus significative des performances. La qualité des 'landmarks' de visage détectés dépend de manière significative de la précision de la boîte englobante donnée par le détecteur de visage. L'utilisation des 'landmark' de visage corrects permet d'obtenir un meilleur alignement du visage et des modèles plus robustes. Notre choix d'utiliser DCNN pour la reconnaissance faciale a été validé lors du défi multimédia NIST SRE 2019 où notre système a obtenu la meilleure performance de système unique parmi 14 autres soumissions. Cela montre que DCNN est l'une des architectures les mieux adaptées à la reconnaissance faciale.

Enfin, l'application du filtrage des enrôlements à l'aide de certaines mesures de qualité est cruciale pour la performance du système de reconnaissance faciale. Si la référence d'enrôlement est de mauvaise qualité, une comparaison avec de bonnes références de test entraînera des scores de similarité inférieurs et de pires performances.

Les schémas crypto-biométriques, tels que l'engagement flou (fuzzy commitment), nécessitent des sources binaires. Notre **deuxième contribution**, présentée au chapitre 5, présente **une nouvelle approche de binarisation des données biométriques à l'aide de Deep Neural Network (DNN)** appliquée aux données biométriques faciales. Nous avons suivi une approche basée sur les données (data-driven) pour binariser les représentations euclidiennes basées sur l'utilisation d'auto-encodeurs sous apprentissage supervisé avec la fonction de perte "Triplet loss". Notre objectif était de créer de longues représentations binaires avec une entropie élevée pour servir dans notre schéma de régénération de clé.

Les longueurs des représentations peuvent être contrôlées. En utilisant un CNN pré-entrainé et en entrainant le modèle sur une version nettoyée de la

base de données MS-celeb-1M, nous obtenons des représentations binaires de longueur 4 096 bits et 3 300 bits d'entropie. Les représentations extraites ont une entropie élevée et sont suffisamment longues pour être utilisées dans des systèmes crypto-biométriques tels que l'engagement flou.

Nous évaluons les performances des représentations binaires sur les bases de données MOBIO et Labeled Faces in the Wild (LFW), où nous mesurons leurs performances de reconnaissance biométrique et leur entropie. Les représentations binaires proposées offrent des performances de pointe sur les deux bases de données avec une dégradation presque négligeable par rapport au système de base. L'utilisation de DNN pour extraire les représentations binaires donne des représentations avec une entropie élevée et des performances de reconnaissance élevées. Par rapport aux représentations euclidiennes de base, les projections binaires proposées offrent des performances de pointe sur les deux bases de données avec une dégradation presque négligeable. La dégradation des performances dans les deux bases de données est d'environ 0.1%.

Nous obtenons une précision de 99.12% sur la base de données LFW, en utilisant les représentations binaires, contre une précision de 99.22% en utilisant le système de base. Il en va de même pour la base de données MOBIO, où nous obtenons une précision de 98.90 % en utilisant les projections binaires par rapport à une précision de 98.93 % du système de base.

L'approche proposée au chapitre 5 peut être appliquée à n'importe quelle représentation continue, pas seulement aux représentations euclidiennes de visage. De plus, la technique de binarisation constitue un hachage préservant la localité (locality presevering hashing), où la distance relative entre les valeurs d'entrée est préservée dans la distance relative entre les valeurs de hachage de sortie. La représentation peut être utilisée pour de multiples applications, telles que la recherche de similarité, la recherche de base de données et les systèmes biométriques.

De plus, la méthode de binarisation fournit des représentations de longueur arbitraire qui ne sont limitées que par la qualité de la base de données d'apprentissage. Ainsi, la longueur des projections peut être adaptée à la sensibilité de l'application. Nous avons comparé notre approche de binarisation à certaines méthodes de binarisation classiques présentées dans [Dro+18] et

montrons que notre méthode a une meilleure performance de reconnaissance biométrique et une entropie plus élevée que les méthodes présentées.

Les représentations binaires créées sont également utilisées pour mettre en œuvre un système de reconnaissance faciale révocable basé sur une transformation de mélange utilisant un second facteur. Le système révocable est analysé selon les métriques normalisées données par la norme ISO/IEC 24745:2011. Nous montrons que le système révocable donne des modèles de haute précision et non liés lorsque le deuxième facteur n'est pas compromis. Lorsque le deuxième facteur est compromis, la sécurité du système est assurée par les performances de reconnaissance des représentations binaires, qui sont comparables au système non binarisé de base. De plus, la qualité des représentations binaires impacte le comportement du système révocable. Si le pouvoir discriminant des représentations est faible, le système révocable dépend principalement du deuxième facteur, ce qui se traduit par un FAR plus élevé.

Ces représentations sont destinées à être utilisées dans un schéma de régénération de clé crypto-biométrique basé sur un engagement flou. C'est pourquoi on cherche à obtenir des représentations binaires longues à forte entropie.

Le ***premier objectif*** de la thèse est de **créer des clés crypto-biométriques** à partir de la biométrie de l'utilisateur. Pour créer les clés crypto-biométriques, nous avons procédé en extrayant l'entropie des images de visage. Par extraction d'entropie, nous entendons extraire des informations utiles des données biométriques sous forme de format binaire.

Les représentations binaires, obtenues au chapitre 5, ne conviennent pas à une utilisation en cryptographie. La biométrie, par sa nature, n'est pas stable. Elle souffre de la variabilité introduite par de nombreux facteurs : variabilité de session, conditions d'acquisition, capteurs, etc... **Notre contribution suivante** a été d'**utiliser une cohorte pour réduire l'intra-variabilité des représentations**. L'approche que nous avons suivie pour régénérer les clés symétriques est basée sur l'engagement flou. Le schéma d'engagement flou a été implémenté à l'aide de codes de correction d'erreur Bose, Ray-Chaudhuri and Hocquenghem (BCH). Dans notre schéma d'engagement flou, une clé

aléatoire est codée à l'aide de codes de correction d'erreurs (ECC) et est ensuite XORée avec les données biométriques. Les données XORée sont cryptographiquement sécurisées, car ni la clé ni les données biométriques ne peuvent en être obtenues sans fournir l'une des deux. La clé aléatoire est récupérée au moment de la régénération de la clé en fournissant de nouvelles données biométriques. Ce système nécessite des données biométriques ordonnées sous forme binaire. Dans ce schéma, les différences de données biométriques d'une acquisition à l'autre sont traitées comme du bruit. Ce bruit provoque des erreurs dans les données transmises qui sont corrigées à l'aide des ECC. La révocabilité du schéma d'engagement flou est assurée en utilisant le même schéma de protection décrit dans la sous-section 5.4.2.

Nous rapportons le taux de fausses acceptations (FAR) et le taux de faux rejets (FRR) sur la base de données MOBIO. Nous avons réalisé 9M de tests client-client et 10M de tests client-imposteur. Pour les tests client-client, tous les échantillons biométriques sont croisés. Pour les tests client-imposteur, 21 échantillons sont sélectionnés au hasard parmi chaque utilisateur et sont ensuite croisés.

Dans toutes les expériences de régénération de clé, le FAR est de 0% car le nombre de bits erronés est supérieur à la capacité de correction de code dans le cas des tests client-imposteurs.

Le ***deuxième objectif*** de la thèse est que les clés soient post-quantiques. Par post-quantique, nous entendons que les clés doivent être résistantes aux algorithmes quantiques tels que l'algorithme de Shor [Sho94] et l'algorithme de recherche de Grover [Gro96]. Il existe deux schémas de chiffrement, symétrique et asymétrique.

L'algorithme de Grover réduit la complexité d'une attaque par force brute sur une clé symétrique de 2^N à $2^{N/2}$. Pour atténuer le risque introduit par l'informatique quantique, nous devons augmenter la taille des clés. C'est la raison pour laquelle nous avons essayé de rendre la représentation binaire plus longue et plus discriminante. Au chapitre 6, nous régénérons les clés symétriques longues pour la biométrie faciale. Les systèmes de régénération de clés à la pointe de la technologie qui utilisent la biométrie faciale souffrent d'une entropie FRR élevée et faible par rapport aux autres modalités biométriques [Wan+21]. Dans notre cas, nous avons pu régénérer des clés de

chiffrement symétriques de plus de 400 bits avec un faible FAR et un faible FRR en utilisant la biométrie faciale.

Mots clés :

Biométrie, Cryptographie, Vérification de visage, Apprentissage profond

Acknowledgements

First and foremost, I would like to express my deepest gratitude and appreciation to my PhD supervisor, DR. Dijana Petrovska-Delacrétaz. Without her support and encouragement, I would not have been able to complete this challenging journey. I am deeply thankful for her patience and perseverance, especially during the difficult times of the COVID-19 outbreak. Her guidance and valuable input have been invaluable to me throughout this process. I am also grateful to her for the numerous opportunities she has provided me, including the chance to work on two H2020 European projects, the connections she has helped me make, and the opportunity to visit multiple research institutions. It has been a great honor and pleasure to have such an exceptional supervisor as Dijana, and I am truly grateful for all she has done for me. Thank you Dijana.

I would like to extend my heartfelt thanks to my thesis director, Prof. Bernadette Dorizzi, for her invaluable guidance and wise counsel throughout the process of completing my research. Her insights and advice have been incredibly helpful, and I am deeply grateful for her support. It has been an honour to work with such a knowledgeable and skilled supervisor, and I will always be grateful for the opportunity to learn from her.

I would like to express my sincere gratitude to the members of the jury who have generously dedicated their time and expertise to the review and defense of my thesis. Their insights, critiques, and questions have been invaluable in helping me to refine my research and deepen my understanding of my topic. I am deeply grateful for their guidance, and I am truly appreciative of the opportunity to present my work to such a distinguished panel of experts. I would like to thank Dr. Teddy Furon and Prof. Chaic Mokbel, my thesis reporters, for their thorough input and corrections. Their insights and critiques were instrumental in helping me to improve my work, and I am deeply grateful for their guidance. I am also grateful to Prof. Hossam Afifi for presiding over my PhD defense, and to Prof. Bhiksha Raj for his valuable input.

I would like to express my sincere gratitude to my lab colleagues for their help and support during the period of my thesis. In particular, I would like to thank Aymen Mtibaa for his assistance and support. I am deeply grateful for his help.

I would also like to thank my family and friends for their support and encouragement during this time. The process of completing a thesis can be a challenging and demanding journey, and I am grateful to have had the support of my loved ones to help me through it. In particular, I would like to thank Ayman Ben Thabet for his unwavering support and encouragement. I am deeply grateful to have such a supportive network of loved ones.

Thank you all for your invaluable support.

Mohamed Amine HMANI, November 2022.

Contents

List of Figures

List of Tables

Acronyms

BCH	Bose, Ray-Chaudhuri and Hocquenghem.
BE	Biometric Encryption.
BRGC	Binary Reflected Gray Code.
CASIA	Chinese Academy of Sciences.
CI	Common Identifier.
CNN	Convolutional Neural Network.
CTS	Conversational Telephone Speech.
DBR	Direct Binary Represenation.
DBSCAN	Density-Based Spatial Clustering of Applications with Noise.
DCNN	Deep Convolutional Neural Network.
DLDA	Direct Linear Discriminant.
DNN	Deep Neural Network.
ECC	Error Correcting Code.
EER	Equal Error Rate.
FAR	False Acceptance Rate.
FHE	Fully Homomorphic Encryption.
FRR	False Rejection Rate.
GMM	Gaussian Mixture Model.
HTER	Half Total Error Rate.
KDF	Key Derivation Function.

LDC	Linguistic Data Consortium.
LFW	Labeled Faces in the Wild.
LPDC	Low-Density Parity-Check.
LSSC	Linearly Separable SubCode.
MDG	Minimum Distance Graphs.
MRP	Multispace Random Projections.
NIST	National Institute of Standards and Technology.
PBKDF	Password Based Key Derivation Function.
PCA	Principal Component Analysis.
PI	Pseudonymous Identifier.
PIC	Pseudonymous Identifier Comparator.
PIE	Pseudonymous Identifier Encoder.
PIR	Pseudonymous Identifier Recorder.
PRP	Probabilistic Random Projection.
RSA	Rivest–Shamir–Adleman.
SB	Shuffled Binary Embedding.
SD	Supplementary Data.
SE	Secure Element.
SK	Shuffling Key.
VGG	Visual Geometry Group.

1 Introduction

1.1 Motivation and Objectives of the Thesis

Cryptography plays an increasingly important role in society, where we have a growing need for securing data and transactions (in telecommunications, medicine, financial transactions, as well as cryptocurrency or the protection of our privacy). Cryptography is mainly based on the use of cryptographic keys to encrypt data or sign it to guarantee its authenticity and integrity.

However, cryptography has some problems. First, it is threatened by the arrival of quantum computers. The prototypes of quantum computers are increasingly efficient and advanced. It is estimated that within 20 years these computers will be able to render current cryptographic schemes obsolete. Shor's algorithm [Sho94] challenges classical asymmetric cryptography schemes, Rivest–Shamir–Adleman (RSA) encryption, and schemes based on elliptic curves in particular. Grover's algorithm [Gro96] threatens symmetric keys; it allows to divide the complexity of the search by two.

Apart from the possible problems caused by quantum computers, cryptographic keys have some inherent disadvantages. The keys are of such length that one cannot memorize them. Hence, the need to store them, which induces an additional risk of copying the keys if they are not well protected or even the loss of the key thus causing the loss of encrypted data. By using classical cryptographic keys, one cannot guarantee non-repudiation, i.e.

ensure that only the owner of the key uses it. This allows for the sharing of digital identities and the denial of liability in some cases. When encryption keys are used, there is no direct link between the keys and the true identity of the user.

Biometrics can provide answers to this problem. By using biometrics, the issue of non-repudiation is resolved. On the other hand, another problem is created; non-revocability of biometric keys. We cannot change the biometric data of an individual at will and, therefore, we always create the same biometric key for the same user.

Hence the interest of concentrating our efforts to propose methods to create crypto-biometric keys obtained from the biometric data of the user, which also have the property of revocability. By using a crypto-biometric key, the non-repudiation of the key is guaranteed. In addition, the user no longer needs to memorize the key, since it is built at the time of using the system.

The concept of crypto-biometric keys appeared towards the end of the twentieth century. The first works had the disadvantage of low entropy [Mon+01] and a non-negligible reconstruction error. Therefore, several works including [KPD09; Her+17] have sought to improve the performance of crypto-biometric keys from a security point of view by increasing the entropy and from a biometric point of view by improving the accuracy of key reconstruction.

This work aims to regenerate cancelable crypto-biometric keys that are resistant to quantum cryptanalysis methods. The challenge of this work is to obtain high entropy keys in order to obtain a high level of security. In fact, for traditional cryptographic keys, one can control the entropy during their creation. On the other hand, with crypto-biometric keys, the entropy contained

in the biometric reference limits the entropy of the key.

1.2 Research Contributions

The work is focused mainly on face biometrics as they are easily accessible. First, we try to increase the entropy extracted from biometric data. The lower the False Acceptance Rate (FAR), the higher the entropy of a given biometric system. So, to increase the entropy, we proceeded by creating a face verification system with low FAR. To this end, we improved a state-of-the-art face verification system using publicly available datasets and frameworks. The system is based on Deep Neural Network (DNN). Our contributions mainly concern face detection, landmark extraction as well as optimization of architecture and loss functions.

Cryptographic applications need the biometric features to be in binary format. The features extracted from face biometric systems are usually represented in the continuous domain. This imposes a subsequent module to transform such continuous features into a binary format without significantly deteriorating the original classification performance. To this end, we decided to follow a novel data-driven approach in which the binarization happens inside the DNN (face feature extractor). More precisely, a DNN is trained to provide binary representations with high discrimination between the users resulting in higher entropy. This approach is based on using autoencoders under supervised training with the 'Triplet loss' loss function.

The binary embeddings are first used to create a cancelable face verification system based on a shuffling transformation using a second factor. The cancelable system is analyzed according to the standardized metrics given by the ISO/IEC 24745:2011.

The binary representations in their current form are not suitable for use in cryptography. Biometric data, by their nature, are not stable. They suffer from variability introduced by many factors: session variability, acquisition conditions, sensors, etc... Thus, we proceeded to make the binary representations stable using error-correcting codes. We also used a cohort to reduce the representations intra-variability (variability of representations obtained from each user).

To obtain cancelable crypto-biometric keys, we used a key regeneration scheme based on fuzzy commitment. The keys generated by this system have a length of 512 bits, 0% FAR and 0.3 FRR on the Mobio database.

The second goal of the thesis is for the keys to be post-quantum. By post-quantum, we mean that the keys should be resistant to quantum algorithms such as Shor's algorithm and Grover search algorithm.

There are two encryption schemes, symmetric and asymmetric. We focus on symmetric keys. Symmetric encryption is threatened by the Grover algorithm because it reduces the complexity of a brute force attack on a symmetric key from 2^N to $2^{(N/2)}$. To mitigate the risk introduced by quantum computing, we need to increase the size of the keys. To this end, we tried to make the binary representation longer and more discriminative. For the keys to be resistant to quantum computing, they should have double the length. This is the reason why we tried to make the binary representation longer and more discriminative. According to the NIST SP 800-152, 256-bit security is enough for high-impact applications. As we consistently regenerate keys with length of 512 bits, they should be resistant to quantum computation.

We succeeded in regenerating crypto-biometric keys longer than 400 bits (with low false acceptance and false rejection rates) thanks to the quality of

the binary embeddings. The crypto-biometric keys have high entropy and are resistant to quantum cryptanalysis, according to the PQCrypto project[1], as they satisfy the length requirement. The keys are regenerated using a fuzzy commitment scheme that uses BCH codes.

The remainder of the thesis is structured as follows. In Chapter 2 we present related works to our work in face recognition, binarization, biometric template protection, and encryption. In Chapter 3, we give an overview of the databases used to train, test, and validate our proposed systems. Chapter 4 presents our face verification pipeline and the improvements and enhancements introduced for each of its applications. In Chapter 5, we present the DNN based binarization method that gives high entropy and long representations. Finally, before drawing the conclusions in Chapter 7, we describe our key regeneration scheme in Chapter 6.

1.3 Publications

The papers published throughout the duration of the thesis are the following:

- **Hmani, M. A.**, & Petrovska-Delacrétaz, D. (2018). State-of-the-art face recognition performance using publicly available software and datasets. 2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 1-6. IEEE.
- **Hmani, M. A.**, Mtibaa, A., Petrovska-Delacrétaz, D., Bauzou, C., Crucianu, I. (2020). Evaluation of the H2020 SpeechXRays project Cancelable Face System Under the Framework of ISO/IEC 24745:2011. 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), IEEE.

[1] https://pqcrypto.eu.org/

- **Hmani, M. A.**, Mtibaa, A., Petrovska-Delacrétaz, D.,. Joining Forces of Voice and Facial Biometrics: a Case Study in the Scope of NIST SRE'19. Chapter 9. In Voice Biometrics: Technology, trust and security. IET.

- **Hmani, M.A.**, Petrovska-Delacrétaz, D., Dorizzi, B.: Locality preserving binary face representations using auto-encoders. IET Biome.1–14 (2022). https://doi.org/10.1049/bme2.12096.

- Mtibaa, A., **Hmani, M. A.**, Petrovska-Delacrétaz, D., & Hamida, A. B. et al. (2020). Methodologies of Audio-Visual BiometricPerformance Evaluation for the H2020SpeechXRays Project. 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), IEEE.

- Nasri, M. A., **Hmani, M.A**, Mtibaa, A., Ben Hamida, A., Petrovska-Delacrétaz, D., Benslima, M.(2020). Face Emotion Recognition From Static Image Based on Convolution Neural Networks. 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), IEEE.

- Nautsch, A., Jiménez, A., Treiber, A., Kolberg, J., Jasserand, C., Kindt, E., Delgado, H., Todisco, M., **Hmani, M.A.**, Mtibaa, A. and Abdelraheem, M.A., 2019. Preserving privacy in speaker and speech characterisation. Computer Speech & Language, 58, pp.441-480

2 State of the art

2.1 Introduction

The biometric characteristics of a person are permanently associated with his identity. Although the property of permanent associatively of biometric data with the user makes biometric systems useful, it also raises some serious threats. There are two important issues related to biometric systems.

- Non-revocability: If the biometric data of a person stored in the database is somehow compromised, it cannot be cancelled or replaced. Therefore, the person cannot use the same biometric characteristic in that system and possibly in all other systems based on the same biometric characteristic. This is called non-revocability of biometrics. If it is a fingerprint-based system, the person has an opportunity to use a different finger in that system, but still this number of re- enrolments is limited. In case of face, it is not even possible.
- Privacy compromise: With an increasing use of biometric systems, the issue of protecting the privacy of a user is becoming prominent. User privacy is a complicated term. We define three types of privacy compromises:
  - Biometric data privacy compromise: The raw biometric data of

the user can be recovered from the stored templates. For example, many fingerprint-based systems use minutiae features and store minutiae extracted from a reference fingerprint image as templates. It is possible to reconstruct the original fingerprint image from the stored minutiae. In some cases, the recovered biometric data can reveal certain biological conditions (e.g., fingerprints can reveal some skin conditions). Additionally, synthesized data can be provided to the system to gain access.

- Information privacy compromise: When a person enrolls in different biometric systems with the same biometric trait, his templates in all these systems are reasonably similar (provided these systems are based on the same biometric algorithm). Therefore, templates from one database can be used to gain access to another system, and thus, the information stored in that system can be compromised.

- Identity privacy compromise: Since the templates stored in different databases of a user are reasonably similar, that person can be tracked from one system to another by cross matching his templates from the two biometric databases. Similarly, when a system operates in identification mode, it can simply reveal that a person, to which the presented biometric belongs to, is enrolled in that particular system. This can be considered as a compromise of user's privacy. For example, consider an application of biometrics to the HIV (Human Immunodeficiency Virus) patients' (or any other sensitive group) social network. This network is a closed group of HIV patients, who share information only to the members. In this scenario, if the biometric recognition system works

in identification mode, positive identification of a person based on the provided biometric data indicates that the person is a member of such a sensitive group.

Cryptography is a process employed widely in order to secure the storage and/or transmission of electronic information. The basic idea of cryptography involves two phases: encryption and decryption. During encryption, the data, denoted as plaintext, is transformed into unintelligible gibberish, denoted as ciphertext, with the help of an encryption key. The decryption process is the reverse of encryption, i.e., obtaining the plaintext from the ciphertext. The pair of algorithms that create the encryption and the reversing decryption is denoted as cipher.

According to the Kerckhoffs' principle, the security of a cryptographic system lies entirely on the secrecy of the key [Aug+83]. Additionally, for security reasons, the cryptographic keys are required to be long. For example, the possible lengths of keys required in the AES are 128, 192, or 256 bits. For public-key cryptographic systems such as RSA, the key lengths are even higher (e.g., 512, 1024, or 2048 bits). Clearly, a user cannot remember such long keys and therefore, the keys need to be stored somewhere, e.g., on a smart card or in a computer.

In order to restrict access to these keys only to legitimate users, authentication mechanisms are used. Traditionally, authentication mechanisms employed in cryptography are knowledge based (e.g., passwords) or possession based (e.g., token, smart card, etc.). These authenticators are assigned to the user identity and do not necessarily indicate the presence of the person to which they belong. Therefore, they can be (more or less easily) stolen by an attacker, and in this situation, the system cannot distinguish between the attacker and a legitimate user. Another issue related with these authentication

mechanisms is repudiation. A user can willfully share his credentials and later claim that they were stolen. Thus, such a system can be easily cheated.

Biometric systems prevent non-repudiation and can also detect whether an individual has multiple identities. Biometric systems impart higher levels of security and have seen a rapid proliferation in a wide variety of government and commercial applications around the world in the last two decades. However, various security and privacy challenges deter the public confidence in adopting biometric based authentication systems.

As described in [MB11] privacy-preserving techniques can be spread into two categories: hardware-based approaches and software-based approaches. It has to be noted that the term privacy preserving is rather vague, and term like privacy by design are also related to it. For example hardware based approaches for privacy protection like the ones proposed by FIDO[1], can be called privacy by design methods.

Hardware-based approaches: A hardware-based approach involves designing a closed recognition system. In such a system the biometric template never leaves a physically secure module such as a smart card or a hand-held device. Such a device matches the input biometric trait with the template stored in the device and releases a key in case the authentication is successful. This is the configuration adopted in the FIDO alliance proposal.

Software-based approaches: protect the biometric template by storing a modified version of it, in order to reveal as little as possible information about the original biometric trait. Software-based approaches can be spread into two

[1]The Fast Identity Online (FIDO) standard reinforces the security of online identity authentication systems on mobile devices and web applications. Its goal is to replace the exclusive use of passwords with more secure biometric authentication mechanisms that are protected by encryption systems (see https://fidoalliance.org)

main categories: template or feature transformation and biometric cryptosystems:

- Template or Feature Transformation: transform the biometric template based on parameters derived from external information such as user passwords or keys. The same transformation function is applied to the query and matched with the stored template.
- Biometric cryptosystems can be defined as the process in which a digital key (randomly generated) is bound to a biometric template (key binding) or a key is generated by a biometric template (i.e. key generation). In both modes ("key binding" and "key generation") of the Biometric Encryption (BE) methods, the key is "encrypted" with the biometric trait and the result, which is usually called biometrically encrypted key or BE template or helper data is stored either in a database or locally (i.e. smart card).

2.2 Crypto-biometrics

2.2.1 Biometric Template Protection Requirements

There are some main criteria which a cancelable biometric template should satisfy:

- **Performance**: cancelable biometric system should not degrade the verification performance of the underlying baseline biometric system;
- **Revocability**: if the stored user template is compromised it should be possible to cancel that template and reissue a new one. Additionally, the newly issued template should not match with the previously compromised template. 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. The system should be able to reject a person if he provides the authenticator linked with the old template. Note that biometrics alone cannot provide this property because biometric characteristics cannot be changed while systems using passwords and tokens have excellent revocability;

- **Diversity**: It should be possible to issue different templates for different applications related to the same user. These templates should not match with each other and should make cross-matching impossible. Password- and token-based systems are good at that, although practically, password diversity can be argued. Biometrics, by itself, cannot have template diversity;
- **Irreversibility**: It should be computationally infeasible to obtain the original biometric template from the protected template;
- **Unlinkability**: the protected biometric templates created from same biometric sample using two different secret keys should not be linkable.
- **High key entropy**: If the goal of the crypto-biometric system is to obtain crypto-bio keys, the entropy of such keys should be high.

In order to be coherent with the ISO/IEC 24745:2011 standard, we will be using the same vocabulary. The ISO/IEC 24745:2011 standard defines the architecture of biometric protection systems. The architecture is based on three important elements:

- **Pseudonymous Identifier Encoder (PIE)**: During enrolment, the PIE generates a cancelable biometric template based on the Pseudonymous Identifier (PI) and Supplementary Data (SD).

- **Pseudonymous Identifier Recorder (PIR)**: During verification, the PIR generates a pseudonymous identifier (PI*) based on the SD provided during enrolment and the biometric sample.
- **Pseudonymous Identifier Comparator (PIC)**: compares the PI created in the enrolment phase and PI* and returns a score.

2.2.2 Biometric Template Protection System Classification

Cancelable Biometrics

The transformations found in literature are of two types: reversible transformations and irreversible transformations. The reversible transformations make use of a transformation key or token which needs to be kept secret. Such transformations can be inverted to obtain the original biometric data if the transformation key is disclosed. Systems based on these reversible transformations are sometimes called salting approaches. The performance of the cancelable systems when using the reversible transformations is generally better than the classical biometric systems. The systems based on irreversible transformations, on the other hand, do not require the transformation parameters to be kept secret. Even if the transformation parameter is disclosed, it is infeasible to obtain the original biometric template from the cancelable template. However, it is observed that the performance of the cancelable biometric systems in such cases degrades compared to the classical biometric systems [PRC15; Cha+20].

In the following paragraphs, we take a brief look at some systems in transformation based cancelable biometrics category. It has to be noted that the methods are not related to a specific biometric modality. In 2001, Ratha *et*

al. [RCB01] introduced the term cancelable biometrics proposing transformation of the biometric signal (or features) using irreversible transformations. The transformation parameters are user specific. In [Rat+07] Ratha *et al.* proposed three different transformations (Cartesian, polar, and functional). These transformations provide a different amount of security to the biometric data. They tested their system on a private database with 188 pairs of fingerprint images and reported that the performance of the underlying biometric system always degrades after transformation.

Another interesting and widely used technique called BioHashing [GN03] was used by Jin *et al.* [JLG04] for cancelability. In BioHashing, a randomly generated, user-specific key (denoted as hash key) is used to generate an ortho-normal matrix. The biometric feature vector is projected onto this matrix and after thresholding; a binary vector is obtained which is denoted as BioHash. In 2007 (Lumini & Nanni) proposed an improved version of this BioHashing scheme with modifications such as binarisation threshold variation, space augmentation, feature permutation, and feature normalization. They reported that, in general, BioHashing scheme improves the performance of the underlying biometric system. But, the drawback is, in the stolen key scenario, the performance generally degrades compared to the baseline biometric system.

In 2007, Boult *et al.* [BSW07] applied the biotoken scheme to fingerprints which they earlier proposed for faces in [Bou06]. The scheme is based on robust distance matching techniques. They reported 30% improvement in the verification performance for the fingerprint biotoken system.

In 2008, Maiorana *et al.* [MCN08] proposed a different way of transformation called BioConvolving which is applied to Hidden Markov Model (HMM) based signature features. It makes use of a randomly generated sequence to

divide the features into parts on which convolution is applied. In their later papers [Mai+08; Mai+10], they showed that the number of different templates that can be generated using this technique is limited and proposed some improvements in order to increase this number. In [MCN11] they proposed a multi-biometric approach for cancelable biometrics by employing BioConvolving and using a combination of different matchers by score-level fusion.

In [TC10], a cancellable formulation for speech biometrics, which we refer as Probabilistic Random Projection (PRP) is proposed for speaker verification system. In this paper, they extend the Multispace Random Projections (MRP) by using 2D subspace techniques and Gaussian Mixture Model (GMM) for speaker modeling. PRP shows excellent performance as in MRP for the legitimate token and stolen-biometric scenario. Experiments showed that PRP does not suffer the problem of stolen-token attacks when the random subspace dimension is near to the feature dimension. Also PRP fulfilled other important properties of cancellable biometrics, i.e. diversity property and non-invertible property (no recovery of biometric template in the event of compromise).

Das *et al.*[DKG12] presented a robust alignment-free fingerprint hashing algorithm based on Minimum Distance Graphs (MDG) for secure authentication. They report that the matching performance is better than some existing fingerprint template protection schemes with an Equal Error Rate (EER) equal to 0.0227 and that this method increases the computational complexity of brute-force attack and the invertibility of the hash of the actual fingerprint is intractable.

In [WH14], Wang & Hu proposed the design of alignment-free cancelable fingerprint template via curtailed circular convolution. By quantizing and

bin-indexing pair-minutiae vectors, a binary string is generated. The transformed template fulfills the requirements of non-invertibility, revocability and diversity for cancelable fingerprint templates even when both the transformed template and parameter key are compromised. Also evaluation of the proposed scheme is reported and it shows that the new method improves the performance compared to the existing alignment-free cancelable template schemes.

In 2017, Gomez *et al.* [GB+17] presented a general framework for the evaluation of unlinkability in biometric template protection schemes, as well as an improved, unlinkable and irreversible, system based on Bloom filters. With fully reproducible experimental study, they confirm the irreversibility and unlinkability of facial Bloom filter-based protection scheme, considering an advanced adversary model, as well as a full disclosure adversary model, where a potential attacker is in possession of secret keys. They also report that the proposed scheme maintains the biometric performance of the unprotected system.

Homomorphic Encryption

Towards the end of the seventies, in a landmark paper [RAD+78], Rivest, Adleman and Dertouzos, define and investigate the applicative potential of a new notion which they call privacy homomorphisms. Indeed, building on the basic fact that the RSA cryptosystem is multiplicatively homomorphic – the product of two ciphertexts provides an encryption of the product of the two corresponding cleartexts – they end up conjecturing the existence of both secure and malleable cryptosystems, that is cryptosystems allowing to perform general calculations directly on encrypted data. This idea would remain a curiosity for a number of years, the homomorphic properties (always

limited to one operation) of several cryptosystems (most notably ElGamal and Goldwasser-Micalli) being remarked and tolerated as apparently benign is terms of security.

That situation changed towards the end of the nineties, when, mostly due to the introduction of the Paillier cryptosystem, the search for cryptosystems homomorphic at the same time for both the addition and the multiplication operations (so called Fully Homomorphic Encryption (FHE) schemes) becomes one of the grail quest of part of the cryptographic community. Indeed, on top of having reasonable performances, the Paillier cryptosystems allows to perform additions in the encrypted domain as well as multiplication by a public integer. In essence, it becomes possible to apply any (public) linear operator directly on encrypted data and to do so at reasonable computational cost. A possibility which is sufficient to give birth to a new applied research field based on homomorphic cryptography: signal processing in the encrypted domain.

In 2009, against all expectations, C. Gentry, then at Standford, proposes a first credible construction both in terms of security and of theoretical efficiency. Still, in order to properly appreciate the consequences of this initial breakthrough, it is necessary to precise what efficiency means for a theoretical computer scientist. Indeed, for a given security level ℓ (which drives the parameterization of the cryptosystem to require an order of magnitude of 2 ℓ operations for the best known attacks on the underlying mathematical problem to break the system), a homomorphic encryption system is considered theoretically efficient if the computational overhead of working in the encrypted domain is bounded by a polynomial in ℓ. Needless to say, and this was unfortunately the case for C. Gentry initial construction, that the degree of the polynomial does not need to be very large for the overhead to be

prohibitive. Furthermore, C. Gentry's initial proposal was mindbogglingly complex. Indeed, designing fully homomorphic encryption systems is a difficult task due to a noising phenomenon which amplitude grows quickly with the amount of calculations, mostly of multiplications, performed, up to the point where decryption of the results becomes impossible. To solve this issue, and this is where his main contribution lies, Gentry has introduced a denoising technique, known as bootstrapping, which loosely speaking consists in homomorphically performing a recrypt operation (i.e., equivalent to a decryption followed by an encryption without, thanks to the magic of homomorphisms, ever having the data in clear form during the operation).

Unfortunately, this first fully homomorphic cryptosystems, and more generally any bootstrapping-based cryptosystem known so far, are way too costly to have any practical relevance whatsoever.

Things subsequently got a lot better in 2012 when C. Gentry and two coauthors (Z. Brakerski and V. Vaikuntanathan) proposed a radically new homomorphic encryption schemes construction blueprint: the levelled (somewhat) homomorphic cryptosystems. In such a system, an algorithm is executed on a sequence of cryptosystems rather than a single one. When executing an algorithm over a levelled system, as a general rule, additions can be performed within the same level whereas multiplications, which are much more important noise amplifiers, require a level change. The subtlety resides in the combination of a tensorial operator which spreads the noise and a projection operator which reduces its amplitude at the cost of a level change but without bootstrapping. Additionally, levelled cryptosystems have a very interesting intrinsic parallelism property, called batching, by which independent calculations can be (quite massively) multiplexed within a unique ciphertext and thus processed in parallel at the bit level (in essence, batching parallelism is

conceptually similar to a technique known as bit-slicing, introduced during the nineties in the field of cryptanalysis of block cipher). As of 2016, the most efficient system is still due to Z. Brakerski (first published towards the end of 2013 and optimized in subsequent publications most notably by Fan and Vercauteren). It is a levelled system which intrinsic performances and memory requirements are reasonable enough to consider the first deployments of homomorphic encryption-based calculations in lightweight-enough real-world settings. Furthermore, it should be emphasized that, performance-wise, progress has been fast paced and that a kind of Moore's law seems to emerge. As originally stated by C. Aguilar, every 12 to 18 months, the performance overhead of homomorphic encryption appears to decrease by a square-root.

Additional systems, coined 2nd generation FHE, such as GSW have been proposed, but despite being conceptually simpler than leveled FHE they appear less efficient (also because they lack batching, at least at present). Very recently, a new breed of credible fast bootstrapping FHE has also started to appear.

In parallel with the above research, which has been conducted for the most part within the cryptographic community, the compilation, and parallelism community has also started to grow a fairly early (as early as end 2010) interest in homomorphic encryption techniques as a new execution environment for computer programs with a highly promising practical relevance. In particular, it should be emphasized that a homomorphic encryption system only provide bit-level operators. Thus, making the connection between an algorithm written in a high level programming language and such a low level execution environment requires a sequence of non-trivial transformations, that is, a compiler. If, furthermore, it is required that this compiler mitigates,

as much as possible, the performance overhead by means (for example) of parallelism, then techniques from the field of optimizing compilation and parallel code generation have to be brought into the picture. As of 2016, the most convincing experimental results have been obtained by combining careful optimization of the crypto-systems as well as optimized parallel code generation. In particular, still without diminishing the work remaining to address more demanding applications, it has been possible to demonstrate the execution of real yet lightweight algorithms (most notably from the field of medical diagnostic) with both acceptable performances (significantly less than a second) and security levels (128 bits) – with yet recourse only to moderate parallelism and no batching.

2.3 Face Recognition

Currently, state-of-the-art facial biometric algorithms are based on Deep Convolutional Neural Networks. Table 2.1 summarizes the most prominent published Deep Neural Network (DNN) based facial recognition systems. Most of them are either proprietary, only a description of the system is provided, or trained on private databases. To compare biometric systems objectively, it is mandatory to use the same database and the same testing protocols [PDCD09]. To this end, we will report the performance of the systems on the Labeled Faces in the Wild (LFW) [HLM14] where the comparison metric is the pair matching accuracy using the 10-fold cross validation protocol.

Some best performing systems in the face recognition ecosystem are described in the next paragraph.

The **Camvi**'s model [Eri+19] is trained on a subset of MS-celeb-1M face database, containing around 80k identities and 5 Million (M) faces. The

authors tried to remove the overlapped faces in LFW with close similarity scores. The recognition model is a single Convolutional Neural Network (CNN) with a size of 230 MB, which outputs an embedding vector with 256 float point numbers for an input image. L2 distance is used to measure the similarity between two feature vectors and compute average accuracy for each subset using the best threshold from the rest of the nine subsets of the 10-folds.

Ever.ai [Eri+19] trained their model on a private photo database with no intersection with LFW. They also trained custom face and landmark detectors for pre-processing and built their primary face recognition model on a database containing over 100k identities and 10M images. The recognition model is a single deep ResNet model [He+16], which outputs an embedding vector given an input image, and the similarity between a pair of images is evaluated via an L2-norm distance between their respective embeddings. The system is a proprietary system with no extensive description provided by the authors. We report its performance on the LFW benchmark because it is one of the best performing systems.

FaceNet [SKP15] was developed by Google. It is a unified system for face verification, identification, and clustering. It extracts Euclidean representations from images with the advantage of being general purpose. The features are also compact (with a dimension of 128) compared to traditional representations (Gabor features for example). The system was trained on a huge private database of 260 M images from 8 M subjects. It was trained for 1 000 hours.

DeepFace [Tai+14] is developed by Facebook. It processes images in two steps. First, it corrects the angle of a face so that the person in the picture becomes forward-facing, using a 3-D model of an 'average' forward-looking face. The second step is to propagate the face to the Deep Neural Network

(DNN) in order to extract its representation. The system was trained on a private database consisting of 4.4M images from 4k subjects (average of 1k per subject). It has 97.35% accuracy on LFW.

DeepID2 [Sun+15] was developed by the Department of Information Engineering of the Chinese University of Hong Kong. The features are learned using deep convolutional networks. The face identification task increases the interpersonal variations by drawing apart DeepID2 features extracted from different identities. In contrast, the face verification task reduces the intrapersonal variations by reducing the distance between DeepID2 features extracted from the same identity, both of which are essential to face recognition. It was trained on a private database consisting of 200k images from 10k subjects. Compared to other databases such as Google's or Facebook's systems, the size of the database can be considered relatively small. It gives 99.15% verification accuracy on the LFW database.

VGG-DeepFace [PVZ15] was developed by the Visual Geometry Group (VGG) from the University of Oxford. The system was trained on 2.6M images containing 2.6k identities. The published performance on LFW is 98.95%. The VGG system is essentially a very deep convolutional neural network. It leverages two distinct methods for the training: N-way classification, and triplet embedding. In the case of this system, the N-way has the advantage of faster training, while on the other hand, triplet embedding gives a better overall performance.

CASIANet [Yi+14] was developed by the Institute of Automation, Chinese Academy of Sciences (CASIA). The system is inspired by many new successful networks, including very deep architecture, low dimensional representation, and multiple loss functions. It was trained on the publicly available

CASIA database (500k images representing 10k identities). The reported performance of the system on LFW is 96.13%.

OpenFace [ALS16] is an implementation of the FaceNet system based on [SKP15]. The source code is publicly available as well as the trained model. It was trained on the publicly available CASIA-webfaces and FaceScrub databases. The system has 92.92% accuracy on LFW.

Table 2.1: Symmary of state-of-the-art Deep Neural Network based face recognition systems.

System	Size of Training database (Millions of images)	Accuracy on LFW $\pm$ Std (%)	Reproducibility
Camvi [Eri+19]	5.00	99.87 ± 0.18	No
Ever.ai [Eri+19]	10.00	99.85 ± 0.20	No
FaceNet [SKP15]	260.00	99.63 ± 0.09	No
DeepID2 [Sun+15]	0.29	99.52 ± 0.12	No
VGG-DeepFace [PVZ15]	2.60	98.95*	Yes
DeepFace [Tai+14]	4.40	97.35 ± 0.25	No
CasiaNet [Yi+14]	0.50	96.13 ± 0.30	No
OpenFace [ALS16]	0.60	92.92 ± 1.34	Yes

* Std is not reported by the authors.

Table 2.1 provides a summary of the performance of some of the face recognition systems studied. The performance are reported on the LFW benchmark. The benchmark comprises 6 000 tests, divided in 10 partitions (folds). The performance is reported in terms of average accuracy and the standard deviation over the 10 folds. When computing the accuracy for each fold, one must use the remaining nine folds in order to determine the threshold that gives the highest accuracy on the nine folds. Afterwards, the threshold is applied to the remaining fold giving the accuracy on that folds. This process is repeated for each fold separately. The accuracies are then averaged to provide the reported metric along with the standard deviation. Further details on the LFW database are provided in section 3.1.

All the mentioned methods use deep neural networks. The reproducibility of these systems hangs mainly on the availability of the training data. For example, even when the architecture is described in detail as in the case of FaceNet, OpenFace, that tried to reproduce FaceNet's perforamce, achieved only 92.92% in comparison with the 99.6% accuracy of FaceNet. This might be due to FaceNet being trained on 260 million images in contrast with the 600k of OpenFace. We suspect that another reason for the difference in performance relates to the quality of the face detector. Our face recognition system, described in the following section, is built upon the OpenFace framework. The main reason behind the choice of OpenFace is reproducibilty.

2.4 Binarisation

State-of-art face recognition systems use continuous vector embeddings to represent the users. However, the majority of biometric template protection schemes need a binary representation [LTK15] as an input. Thus, the continuous vectors need to be binarised.

Binarisation methods fall into two categories: rule-based and data-driven methods. For the rule based approaches, different schemes were proposed.
Kevenaar *et al*. [Kev+05] extract the most reliable components of facial feature vectors and binarise them for use in a template protection scheme.
Chen *et al*. [Che+09] present a detection rate optimized bit allocation principle, which is biometric characteristic-agnostic. Based on the discriminative power of the features, it assigns more or fewer bits to them during binarisation, thus improving the biometric performance of the binarised feature representation.
Bringer *et al*. [BD10] transform fingerprint minutiae set using a vicinity-based

approach, which in addition to producing a compact feature representation, also exhibits self-alignment property.

When presenting a novel fingerprint minutiae representation scheme, Cappelli *et al*. [CFM10] note that it can also operate in binarised mode, without significantly decreasing the biometric performance of the scheme.

Lee *et al*. [Lee+12] binarise facial Principal Component Analysis/Eigenfeature Regularization Extraction templates using a generalized Linnartz and Tuyl's quantization index modulation scheme for template protection.

Chen *et al*. [CV11] present a generic (for arbitrary characteristics with float-valued feature vectors) binarisation scheme using pairwise adaptive phase quantization and long-short pairing strategy.

In [LT13], Lim *et al*. propose two new encoding schemes linearly and partially linearly separable subcodes, which exhibit full-ideal and near-ideal separability capabilities, respectively.

Schlett *et al*. [SRB16] describe a simple and effective scheme for binarising multi-scale local binary pattern histograms. Almost all of the mentioned binarisation methods are either rule-based binarisation schemes or manually constructed.

In [Dro+18], Drozdowski *et al*. benchmark data-independent binarisation methods such as [Kev+05; Che+09; BD10; CFM10; CV11; Lee+12; LT13; SRB16]. These rule-based methods directly quantize the projected values with a threshold or use an orthogonal matrix to obtain the binary codes. Such methods do not preserve the locality structure in the whole learning process.

As for data-dependent approaches, recently, multiple binarisation techniques based on neural networks such as [Che+18; Sch+19; Mai+21] were introduced. These techniques focus on projecting the input on a predetermined

space. For example, in [Che+18], the authors map a Low-Density Parity-Check (LPDC) code to each identity in the training dataset. Thus, each person in the training set has their codeword, resulting in perfect discrimination between the training subjects. Nevertheless, the system's performance degrades when enrolling a user that did not belong to the training set.

Pandey *et al.* [Pan+16] use deep convolutional neural networks to learn a mapping from face images to maximum entropy binary codes. The mapping is robust enough to tackle the problem of exact matching, yielding the same code for new samples of a user as the code assigned during training. These codes are then hashed to generate protected face templates.

In [JCJ18], Jindal *et al.* generate unique binary codes with maximum entropy. In order to maximize the entropy of the binary codes, each bit of the binary code is randomly generated and has no correlation with the original biometric sample. The binary codes are used to replace the one-hot encoding used to train the VGG-Face network. The network uses binary cross-entropy as the loss function, with the last layer activation function being the sigmoid function instead of the softmax function.

Similar to our approach, Carreira *et al.* [CPR15] use auto-encoders for the binarisation of the data. The outputs of the hidden layer are passed into a step function to binarise the codes. Incorporating the step function in the learning leads to a non-smooth objective function. Optimizing this non-smooth function is NP-complete. Where the gradients do exist, they are zero nearly everywhere. They use binary SVMs to learn the model parameters to handle this difficulty. Whereas, in our case, we ignore the gradient of the binarisation layer to keep the non-zero aspect of the gradient of the loss function.

The previously mentioned binarisation methods provide binary representations with limited length. In this thesis, we aim to obtain long representations with high entropy to be used in crypto-biometric key regeneration.

As opposed to the methods which use a predefined mapping space, the approach we present aims to preserve the topology of the embeddings provided by the baseline DNN architecture. As a result, we preserve the advantages of the underlying DNN (resistance to noise, higher accuracy, robustness) while obtaining binary representations. Furthermore, persevering the topology of the data also allows for using our binarisation method in data retrieval applications.

3 Databases

In this chapter, we present the databases used throughout the thesis, as well as the experimental protocols of the databases if they were used for validating our systems. Furthermore, we explain the modifications that we introduce to the databases such as data cleaning and pruning.

3.1 Labled Faces in the Wild

The LFW dataset contains 13 233 target face images with a considerable degree of variability in facial expressions, age, race, occlusion, and illumination conditions. 1 680 of the people pictured have two or more distinct photos in the data set. The only constraint on these faces is that they were detected by the Viola-Jones face detector [PM01]. Figure 3.1 shows examples of the faces present in the LFW dataset. The protocol specifies two views of the data set. View 1 is for model selection and algorithm development. It contains two sets: 1 100 pairs per each class (matched/mismatched) for training and 500 pairs per each class for testing. View 2 is designed for performance reporting. It is divided into ten sets (folders), each with 300 matched pairs and 300 mismatched pairs. The cross-validation evaluation can be adopted among these ten folders. The final verification performance is reported as the mean recognition rate and standard error over the ten-fold cross-validation. It has to be noted that the task is to do pair matching. Given a pair of images, the

goal is to decide whether they belong to the same subject. This task is similar to face verification, except that the evaluation metrics proposed by the database collectors is the accuracy of the pair matching.

Figure 3.1: Examples from the LFW database.

3.2 WiderFace

WIDER FACE [Yan+16] dataset is a face detection benchmark dataset, of which images are selected from the publicly available WIDER dataset. 32 203 images are chosen with 393 703 labeled faces having a high degree of variability in scale, pose, and occlusion. WIDER FACE dataset is organized based

on 61 event classes (*i.e*: parade, meetings, protests).

3.3 AgeDB

AgeDB [Mos+17] is a manually collected, in-the-wild age database containing images annotated with accurate to the year, noise-free labels. As demonstrated by a series of experiments utilizing state-of-the-art algorithms, this unique property renders AgeDB suitable when performing experiments on age-invariant face verification, age estimation, and face age progression "in-the-wild". The database contains 16 488 images from 568 subjects.

3.4 MS-celeb-1M

The MS-celeb-1M is one of the largest publicly available database. It has 100K subjects, and almost 10M images. Popular search engines are used to provide about 100 images for each subject. The images are collected based on their metadata, not their content. This results in the dataset having a considerable amount of noise, as shown in Figure 3.2. The samples presented come from the same label (identity). This shows the mislabeling noise present in the MS-celeb-1M as we find males, female, non-living objects classified as the same identity in the dataset. The dataset is constructed by Microsoft and is available for noncommercial use. [Guo+16] further describes the process of assembling the images and the metric used for the choice of the 100K celebrity provided in the dataset. We used the whole dataset for training the neural network. The Ms-celeb-1M database contains a significant portion of mislabeling because it is collected automatically using web crawlers.

Figure 3.2: Examples from the MS-celeb-1M database. The samples presented come from the same label (identity).

In order to improve the performance, we leveraged clustering algorithms to clean the database. We applied Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [Est+96] to reduce the mislabelling of the database. We worked under the assumption that there is no overlap between the identities of the labels provided in the database metadata. In other words, under the same label we can find multiple identities, but there is no overlap between the identities belonging to different labels. As the number of the identities in each label is unknown, we proceed by applying DBSCAN clustering algorithm onto each label. The clustering is done on the embeddings computed using our model from [HPD18]. The cluster with the highest number of samples is kept, and the remaining clusters are discarded. In cases where the number of samples in the biggest cluster is lower than three, the label is discarded.

Furthermore, the MS-celeb-1M database has a bias towards the LFW dataset, as there is an overlap of the identities between both databases. To reduce this bias, we removed the labels (identities) that have a Euclidean distance lower than 1.2 from any sample from the LFW. Note that the threshold with which the accuracy is computed on the LFW benchmark in [HPD18] is around 1. We used a more secure threshold because the model from [HPD18] is trained on a non-cleaned version of MS-celeb-1M and presents a bias to LFW. As the comparison score is the Euclidean distance, a more secure threshold means a lower threshold.

Thus, the cleaning resulted in reducing the training database to 80k identities from the 100k users provided in the Ms-celeb-1M, and reducing the total number of images from 10M to 4.5M. The cleaning resulted in better overall performance for the baseline face recognition system. For example, in the case of the LFW database, using the same hyperparameters the accuracy

is improved from 97.53% to 98.82%. The impact of the cleaning is further shown in the case of the MOBIO database, where the Equal Error Rate (EER) of the baseline system improved from 14% to 2%.

3.5 MOBIO

The MOBIO database [McC+12] is a bi-modal (face/speaker) database recorded from 152 people. The database has a female-male ratio of nearly 1:2 (52 females 100 males). In total 12 sessions were captured for each individual. It consists of three sets; training, development, and evaluation. In our experiments, we used only the development and evaluation sets. We report the result on the protocol described in [Bou16]. The results are reported separately for males and females because for speaker recognition separating males from females gives better results. Therefore, face recognition experiments follow the same principle.

Figure 3.3 shows samples from the MOBIO database. The faces were captured in a normal setting where the subject is in front of their laptop webcam. This database presents some difficulties. Some samples have bad illumination as there is a strong light in the background. We notice that the bad illumination conditions are mainly present in the female image samples. Furthermore, some samples contain only partial face or obstructed faces. These conditions make the face detection more difficult using standard techniques such as the Viola-Jones detector [PM01] or the Dlib HoG detector [DT05; SDF11]. These faces can still be detected using newer techniques such as CNN (RetinaFace [Den+19b]), however these quality of these faces (partial, bad illumination) results in worse recognition performance.

The face detection using the Viola-Jones detector on the "Still images" of the

MOBIO database gives 84 errors from 1890 sample from the female development partition compared to 50 errors among 3600 male samples. This shows that the female partition is harder the male partition for the MOBIO database.

Figure 3.3: Examples from the MOBIO database.

3.5.1 Evaluation protocol

EER & HTER: To measure the accuracy of the presented authentication systems, we use the evaluation criteria defined in [Bou16]. These measures are the Equal Error Rate (EER) and Half Total Error Rate (HTER).

The HTER is used to represent the performance of an authentication system on the unbiased evaluation partition as a single number. To compute the HTER, a threshold θ is defined on the development partition at the intersection point of the False Acceptance Rate (FAR) and the False Rejection Rate (FRR). The corresponding FAR (or FRR) value of the development partition at this threshold θ is known as the EER. The threshold is applied to the evaluation partition to obtain the HTER:

$$HTER = \frac{FAR(\theta) + FRR(\theta)}{2} \tag{3.1}$$

θ is the threshold at the Equal Error Rate (EER) defined on the development partition. The FAR and FRR are then computed on the evaluation dataset using the threshold θ.

Accuracy: To measure the accuracy on the MOBIO database, we apply the 10-fold cross-validation pair matching protocol similar to the LFW database to have the same evaluation metric for both databases. We concatenate the development and evaluation partitions to obtain a single testing partition comprised of 100 subjects (62 males and 28 females). We use three frames from each video. Frames where the face is not present, are discarded. In order to have a balanced accuracy, we use 50 000 matched pairs and 50 000 mismatched pairs. The accuracy is computed on the 100 000 pairs using 10-fold cross-validation.

3.6 ATSIP-2018 face database

The ATSIP2018 face database is a face database recorded from 20 people, 15 male, and 5 female. This database was acquired during the ATSIP2018 workshop under the framework of the H2020 SpeechXrays European project by TPS. The subjects were instructed to acquire facial images with their mobile phones in good and bad conditions. Good meaning that the face is frontal, and bad – not frontal faces or illumination problems. Among the 282 recorded images, only in 275 images faces can be detected using the Viola-Jones algorithm. The set of 275 images, with successful face detection, is divided in 61 images for enrolment (with good conditions). For the test: 88

good images and 108 bad images were used. The number of client-client tests is 588. As for the number of client-impostor tests, it is 11 172.

The experimental protocol is the following: each subject has 3 images for enrolment: these images are used to create 3 enrolment templates for each subject.

In order to evaluate correctly biometric modules, different tests need to be done. In order to measure the False Rejection Rate (FRR), different images of the same client need to be tested against his/her enrolment data (called also templates, models). Those tests are usually denoted as client-client tests. If the verification threshold is too strict (and the system is not perfect), clients will be more annoyed, as they will be asked to repeat their verification tests more often than once.

In order to measure the False Acceptance Rate (FAR) each client's enrolment data (called also templates, models) need to be tested with verification data coming from other subjects. Those tests are usually denoted as client-impostor tests. If the verification threshold is low, more impostors can be accepted. Client-impostor tests can be done in different manners: if the impostor tests are chosen randomly from the other subjects test, they are called random impostures. If they are designed specifically to impersonate one specific user, they are called intentional impostures.

For high security applications, it is required to have as low as possible FAR. As a first step, random impostures can be done quite easily. In order to make some comparisons for different systems, the Equal Error Rate (EER) is reported where FAR=FRR.

In order to illustrate the influence of the quality of the test images, different experimental protocols need to be executed. For the ATSIP2018 database,

different experimental protocols were designed. With enrolment face images acquired in good conditions, different possibilities exist while testing with face images of various qualities. If we want to simulate tests done with only good quality images, then we need to take images with good quality for client-client tests and client-impostor tests. These protocols are detailed as follows:

- Target good/imposter good
- Target good/imposter bad
- Target bad/imposter good
- Target bad/imposter bad
- Target good/imposter good + bad
- Target bad/imposter good + bad
- Target good + bad/imposter good + bad

3.7 VAST Corpus

The VAST corpus [TS19] contains amateur video recordings (such as video blogs) collected by the LDC[1] from various online media hosting services. The videos vary in duration from a few seconds to several minutes and include speech spoken in English. Each video may contain audio-visual data from potentially multiple individuals who may or may not be visible in the recording. Manually produced diarization labels (i.e., speaker time marks), as well as key face frames and bounding boxes (that mark an individual's

[1]The Linguistic Data Consortium (LDC) is an open consortium of universities, libraries, corporations and government research laboratories (see https://www.ldc.upenn.edu).

face in the video) were provided for both the DEV set and TEST set enrollment videos (but not for the test videos in either set). The audio is sampled at 16kHz. This corpus is used in the NIST SRE19 multimedia challenge.

3.7.1 Evaluation Metric

In the NIST SRE challenges, three metrics are considered: EER, min_C, and act_C. The EER is the Equal Error Rate, where the False Acceptance Rate (FAR) is equal to the False Rejection Rate (FRR)[2]. act_C is the primary metric, and min_C is the secondary metric.

act_C, the actual detection cost, is computed according to a basic cost model. This model is used to measure the detection performance of the submitted systems in SRE'19, which is defined as a weighted sum of false-rejection and false-acceptance error probabilities for some decision threshold θ. The cost function is normalized to give $C_{Norm}(\theta)$ which is defined as follows:

$$C_{norm}(\theta) = P_{fr}(\theta) + \beta \times P_{fa}(\theta) \tag{3.2}$$

where β is defined as:

$$\beta = \frac{C_{fa}}{C_{fr}} \times \frac{1 - P_{target}}{P_{target}} \tag{3.3}$$

The actual detection cost is computed from the trial scores by applying a detection threshold of $\log(\beta)$, where log denotes the natural logarithm.

[2]NIST SREs adopt the terminology "false alarm" and "miss" instead of "false acceptance" and "false rejection".

For the CTS challenge, the primary cost function is computed using two thresholds. The thresholds are computed for two values of β, β_1 for $P_{Target_1} = 0.01$ and β_2 for $P_{Target_2} = 0.005$, where P_{target} is the a-priori probability of the specified target speaker. We note that the prior P_{target} is a synthetic parameter used in order to reduce the multi-class problem into a binary classification problem [Bru10, Chapter 8]. The values are fixed by NIST.

In this case, the actual detection cost is the following:

$$C_{Primary} = \frac{C_{Norm}(\log(\beta_1)) + C_{Norm}(\log(\beta_2))}{2} \tag{3.4}$$

As for the multimedia challenge, the primary cost function is defined using a single threshold. This threshold is computed for $P_{Target} = 0.005$.

$$C_{Primary} = C_{Norm}(\log(\beta)) \tag{3.5}$$

4 Proposed Face Recognition System

4.1 Introduction

The goal of this thesis is to generate crypto-biometric keys from facial biometric data. As such, we started the pipeline by creating a face recognition system based on publicly available databases and models. With the constant advancements in GPU computational power and the availability of open-source software, the reproducibility of published results should not be a problem. But, if the architectures of the systems are private and databases are proprietary, the reproducibility of published results can not be easily attained. To tackle this problem, we focus on the training and evaluation of face recognition systems on publicly available data and software. In this chapter we exploit the OpenFace open-source system to generate a deep convolutional neural network model using publicly available datasets. We study the impact of the size of the datasets and their quality and compare the performance to a classical face recognition approach. Our focus is to have a fully reproducible model. To this end, we used publicly available datasets (FRGC, MS-celeb-1M, MOBIO, LFW), and publicly available software (OpenFace) to train our model in order to do face recognition. We also evaluate our best model on the challenging video dataset MOBIO and report competitive results with the

best-reported results on this database.

In the last years, mainly due to the advances of deep learning, and more concretely convolutional networks, the quality of image recognition and object detection has been progressing at a dramatic pace. With the advent of GPU computation and big datasets, neural networks saw a huge resurgence. This results in huge improvements in image recognition and consequently face recognition. Many works [Den+19a; Tai+14; Yi+14; Sun+15; PVZ15; SKP15] report near-perfect biometric performance. But in most cases, all systems are either proprietary or trained on private datasets. This raises the problem of the difficulty of reproducing published results [PDCD09].

In this chapter, we try to reach the best-reported results on the Labeled Faces in the Wild (LFW) [Hua+07] database, by using the open-source OpenFace [ALS16] software. This software is based on Google's FaceNet architecture [SKP15] that achieves the best results on LFW, but is fully proprietary. CMU has already worked in this direction, but their published results of 92.92% are far from the 99.6% that Google got on LFW. We have chosen to exploit the publicly available MS-celeb-1M [Guo+16] dataset. We evaluate the performance of our newly trained system on the (LFW), as well as the MOBIO [McC+12] dataset (a very challenging audio-visual dataset). We also provide the improvements that we introduced to the face recognition system in order to improve the performance.

The face recognition system described in this chapter was also implemented in two H2020 European projects and used in a submission to National Institute of Standards and Technology (NIST) SRE2019 competition.

4.2 Proposed Face Recognition System Pipeline

In this section, we describe our face recognition system pipeline. Figure 4.1 shows the block diagram of the face recognition system, which is built using open-source implementations of 1) a face detector named RetinaFace [Den+19b] and 2) a face embedding extractor based on FaceNet [SKP15] (built using the OpenFace implementation [ALS16]).

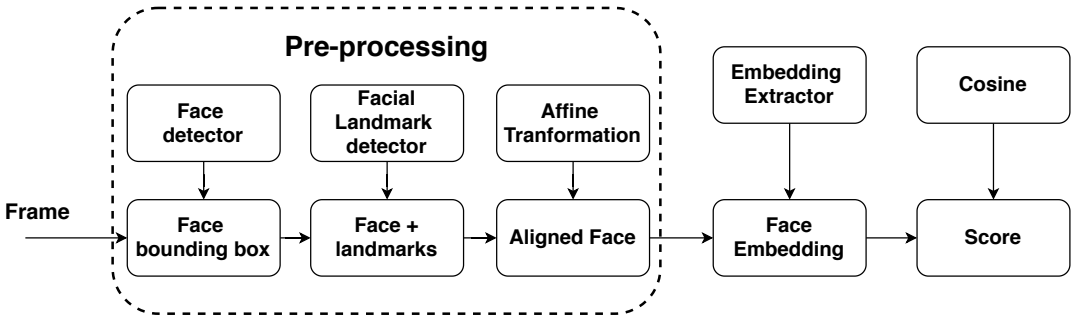


Figure 4.1: Block diagram of our face recognition system.

In the following subsections, we detail the system's components and explain the improvement provided by each modification to the OpenFace framework.

4.2.1 Face Pre-processing

Before using the DNN to construct the face embedding, the image containing the face should be pre-processed. The pre-processing consists of a geometric alignment of the face. The alignment contains two steps. The first step is to detect the bounding box of the face. Once the face is detected, we need to detect the facial landmarks, which in our case, are the 68 facial points defined by the Multi-PIE 68 points mark-up shown in Figure 4.2. The landmarks that are used for normalization are the eyes and the nose. Using these landmarks, the face is rotated, scaled, and cropped. The resulting image has 96x96 pixels. Figure 4.3 and Figure 4.4 show the effects of pre-processing on one image.

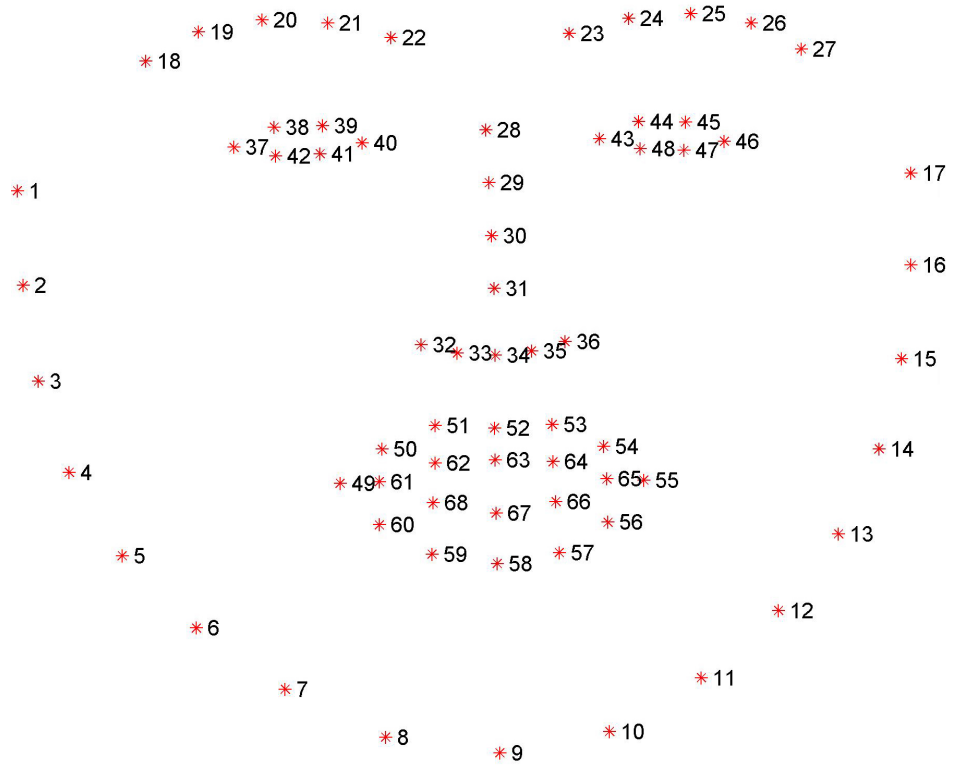


Figure 4.2: The Multi-PIE 68 points mark-up [Gro+10] used for face landmark annotation.

The alignment used in the pre-processing of the training set should be applied in the enrollment and verification phase in the same manner. To achieve the best performance, the same face detector and landmark detector used in pre-processing the training data should be used when exploiting the DNN.

Figure 4.5 and Figure 4.6 show examples of the alignment using the affine transformation on samples from the ATSIP2018[1] database of good quality as well as of bad quality. When the subject's face is not facing forward and presents a high degree of rotation the alignment results in a stretched face.

The first challenge in the face alignment phase is face detection. This step constitutes a high impact on the overall performance of the face recognition system. This impact is especially visible when the acquisition conditions are

[1] ATSIP2018 is a database acquired in the H2020 SpeechXRays project by TSP.

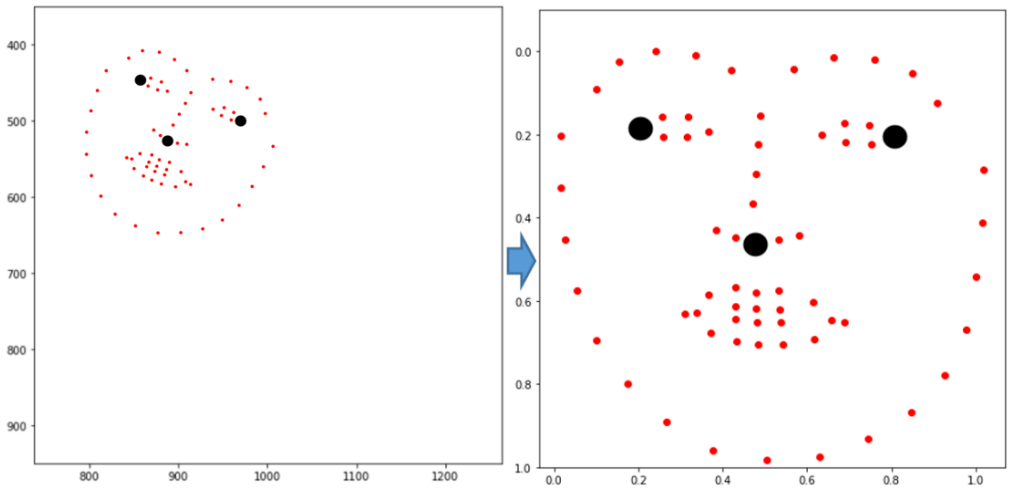


Figure 4.3: Face alignment by applying an affine transformation computed using the outer eyes and nose landmarks. Points of interest (outer eyes and nose) are shown using black dots.

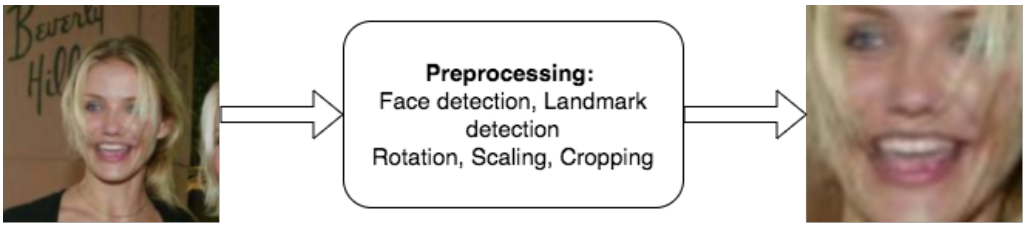


Figure 4.4: Example of the pre-processing of an image from LFW using eyes and nose positions.

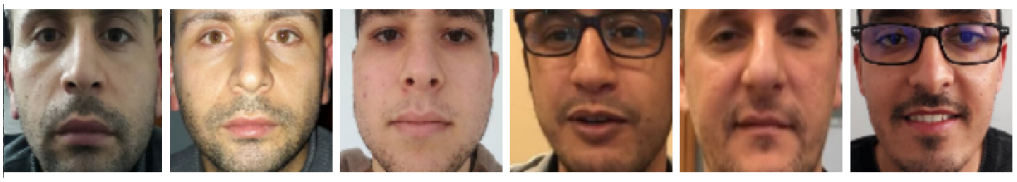


Figure 4.5: Examples from the alignment of images from the ATSIP2018 database with good acquisition conditions, i.e.: frontal face, good illumination.

Figure 4.6: Examples from the alignment of images from the ATSIP2018 database with bad acquisition conditions, i.e.: face turned to a great degree.

adverse (shadows, partial face, occlusion)

To study this impact, we tried different face detection methods ranging from classical methods such as the Viola-Jones [PM01] algorithm to newer ones based on Deep Neural Networks such as Single Shot MultiBox Detector (SSD) [Liu+16] and RetinaFace [Den+19b]. As shown in Table 4.1, the choice of the face detector has a noticeable impact on the performance of the whole system.

Table 4.1: Biometric performance of the face recognition system using different face detectors. Face landmark detection is done using DLIB implementation of ERT. Face Embeddings are extracted using the FaceNet architecture trained on the cleaned version of MS-celeb-1M.

Face detection method	Accuracy on LFW (%)	EER on SRE'19 multimedia DEV (%)
Viola-Jones	97.53	17.00
SSD	98.82	14.36
RetinaFace	99.32	11.20

In order to apply the affine transformation using the position of the outer points eyes and nose, we need to detect the facial landmarks. In order to understand the impact of the quality of the landmark detector we used three landmark detectors: Ensemble of Regression Trees (ERT) proposed by [KS14a] and implemented in the DLIB [Kin09a] toolbox; a 2DFAN DNN based solution was introduced by [BT17]; and a CNN that we trained on ibug 300W [Sag+16]. From Table 4.2, we can conclude that the impact of the face landmark detector is negligible on the overall performance of the face recognition system. This might be due to the embedding extractor neural network being trained using a training set with the same face landmark detector. In fact, the DNN gets used to the errors induced by the landmark detector. What is important, is to use the same landmark detector used for training when exploiting the system for face verification.

Table 4.2: Biometric recognition performance of the face recognition system using different face landmark detection methods. Face detection is done using the SSD model. Face Embeddings are extracted using the FaceNet architecture trained on the cleaned version of MS-celeb-1M.

Landmark detection method	Accuracy on LFW (%)	EER on SRE'19 multimedia DEV (%)
ERT DLIB implementation [KS14a]	98.82	14.36
2DFAN [BT17]	98.90	14.20
CNN (trained on ibug 300W)	98.68	14.56

4.2.2 Embedding Extractor

The embedding extractor used in the face recognition system is a deep convolution neural network. The DNN architecture used in OpenFace is an implementation of the FaceNet model based on [SKP15]. It was inspired from the inception network [Sze+16].

Initial version of the DNN architecture: The initial architecture that we used, consists of an input layer, an output layer and 24 hidden layers among which there are 7 inception layers. The initial version of the network counts 3 733 968 parameters. The DNN extracts feature vectors that give the best possible separation between subjects. It uses triplet embedding to optimize the representations. [SKP15] details the process of the triplet selection and optimization. The loss function defined in eq. 4.1 is based on the triplet loss optimization scheme which consists of choosing two samples from the same class (the anchor and the positive) and a sample for a different class (the negative). The goal of the training is to separate the Anchor-Positive pairs from the Anchor-Negative pairs by at least the margin α. The triplet mining is done online by selecting the triplets that do not follow the rule given

by eq. 4.2. We select the hard negative triplets where the anchor negative-distance is less than the anchor-positive distance.

$$L(\theta) = \sum_{i}^{N} max(\left\|f(x_i^a) - f(x_i^p)\right\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha, 0) \tag{4.1}$$

In eq. 4.1, θ represents the network parameters, x_i^a is the anchor sample, x_i^p the positive sample, and x_i^n the negative sample for subject i. $f(x)$ is the DNN representation of the image x. In order for the training to be efficient (to save computing time), only the triplets that verify eq. 4.2 rule are selected, as other triplets will not improve the network performance.

$$\left\|f(x_i^a) - f(x_i^p)\right\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha > 0 \tag{4.2}$$

This selection process allows for the training to run faster and be more efficient because we will not need to back-propagate triplets that have little effect. If a triplet does not verify the inequality from eq. 4.2 then the considered samples contain too little intra-class variance and a high inter-class variance. As a result of such training, the network outputs a low-dimensional representation of an input image, which consists of a normalized feature vector of size 128. This representation can be leveraged to do either verification (one to one comparisons) or identification (one to many).

The main target of this study is to understand the impact of the training dataset on the performance. In order to be able to study the effect of the database we first made a baseline system based on recommended parameters from [SKP15]. We set the parameters as follows. The embedding size, meaning the length of the representation, was set to 128. We decided to

stop the training based on two criteria, either we reach 1 000 epochs or after 170 hours with the condition that results are stagnant. Each epoch consisted of 250 batches. 20 subjects were uniformly sampled in each batch from the dataset and 18 images per subject were also uniformly sampled form the available images for each subject. If less than 18 images are available, we take all available images. Because we are using the triplet loss function we need at least 2 images per subject. Before the training we removed all subjects from the dataset who have less than 2 images where DLIB successfully detected a face. α is a margin used in the process of triplet selection and serves also in separating the anchor form the negatives. α's impact is further explained in [SKP15]. It is set to 0.2 which constitutes a compromise between the complexity of the triplet mining and the separation between the triplets. The hardware configuration is as follows: an Intel core i7 7700k, 64 Go of DDR4 RAM, 1 TB SSD for storage and a NVIDIA Geforce GTX 1080Ti with 12 GB of VRAM.

Each epoch of the training consists in optimizing the loss function 250 times (once every batch). The batch training is done as follows:

1. Generate a batch by random sampling from the database.
2. Represent every image in the batch (forward propagation).
3. Select triplet verifying eq. 4.2. If no triplets are found, return to step 1. Else compute the loss function.
4. Optimize the network parameters (backward propagation).

For the specified training parameters, the batch generation takes 0.02 seconds. The forward propagation takes 0.4 seconds. The triplet selection, if

enough triplets are found, takes 0.001 seconds and the backward propagation takes 0.3 seconds. Thus, the batch lasts for almost 0.7 seconds on average. However, if no triplets are found (for example due to not enough variability in the training dataset) the processing time for the batch increases considerably.

Figure 4.7 illustrates the evolution of the epoch time (250 x average batch time) where there are not enough triplets. This training was done to study the limits of the triplet selection process. We used a small dataset with 50 subjects with 4 000 images taken from the MOBIO database. In the beginning, the model can not separate the dataset correctly, thus we find enough triplets to optimize the network. As the network performance improves, it becomes able to discern the identities. This results in less triplets verifying eq. 4.2. The training process is stacked at step 1, trying different samples in order to find the triplets it needs to compute the loss function. The process may try thousands of configurations before finding hard-negative triplets. This results in exponential increase of the training time. This made us decide to add another condition to stop the training: if the training period exceeds one week (170h) and the results are stagnant.

Experimental Results

In this section, we detail the performance of the DNN models we obtained by training the architecture using different datasets on the LFW dataset and the MOBIO dataset. We also compare the result of the DNN with a "classic" (not DNN based) approach. For this purpose we have chosen the Direct Linear Discriminant (DLDA) [YY01] based system, because it has a similar strategy of building a compact image (face) representation model (at the training phase) that we can use to project the new incoming faces in order to be able

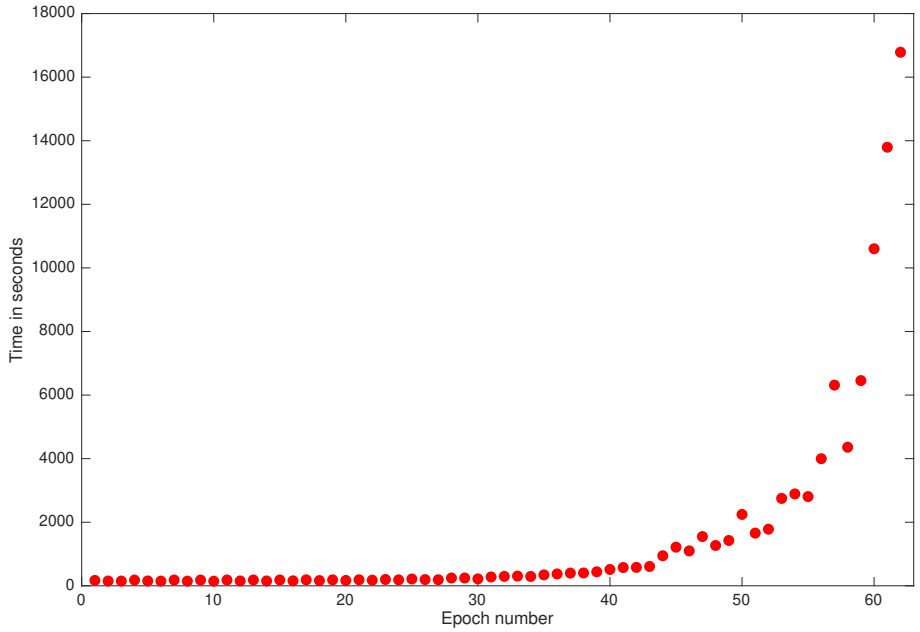


Figure 4.7: Illustration of the evolution of the epoch training time using a low variability dataset originating from the MOBIO dataset.

to compare two face images.

Table 4.3: Our results on the LFW dataset reporting the influence of the training images compared with Google and CMU results

Exps	Preprocessing	Training Dataset	subjects	Number of images	Epochs	Loss	Accuracy
Google	FaceNet	private	8 M	260 M	-	-	99.96%
CMU	OpenFace	FaceScrub CASIA-WebFace	11k	600k	-	-	92.92
Exp.1	OpenFace	FRGC	568	39328	700	0.06	77%
Exp.2	OpenFace	FRGC	568	39328	1000	0.03	80%
Exp.3	Microsoft	MS-celeb-1M	100k	8 M	1000	0.19	86%
Exp.4	OpenFace	MS-celeb-1M	100k	4 M	1000	0.19	96.82%
Exp.5	OpenFace	MS-celeb-1M	100k	8 M	1000	0.18	97.52%

Our goal is to obtain the best performance with the available datasets. We achieve a pair matching accuracy of 97.6% on LFW using all of the available images from the MS-Celeb-1M for training the DNN.

We use the preprocessing of OpenFace. As the preprocessing is based on DLIB face detector, it is not able to detect faces in 58 images from LFW. As a

fallback, we use images from the deep funnelled set of LFW in order to do the verification tests.

We follow the 10-folds cross-validation protocol provided by LFW on the view two. In which, 6 000 pair matching tests are split into 10 partitions. The accuracy is defined as the mean value of correctly matched pairs divided by the number of pairs in each of the 10 folds. The accuracy is defined in eq. 4.3

$$accuracy = mean(\frac{Nbr\ of\ correctly\ matched\ pairs\ for\ fold\ k}{Total\ Nbr\ of\ pairs\ in\ fold\ k}) \tag{4.3}$$

Table 4.3 summarizes the most interesting experiments we have done using OpenFace. In **Exp.1**, we used the FRGC dataset, we stopped the training process at 700 epochs because the training time became too long due to not finding enough triplets satisfying the constraint defined in eq. 4.2. In **Exp.2**, we tried to get better results using the same dataset by pushing the training further. The loss on the training partition and the accuracy on the LFW were both improved by 3 percentile. Nevertheless, the results were not convincing. This made evident the need for bigger datasets. The biggest public dataset that we found was MS-celeb-1M. This dataset was the core of the remaining experiments. Microsoft provides a pre-aligned version as well as a raw version of the dataset. In **Exp.3** we used the pre-aligned version by Microsoft. However, the preprocessing was not adequate to the input of the DNN. The images were of varying sizes. After 1 000 epochs we obtained 86% accuracy on LFW. The results are better than when using only FRGC as training data, but still not at the level of the reported results in the literature. Thus we decided to apply OpenFace alignment on the raw data. This resulted in better overall performance, as shown in experiments 4 and 5. In **Exp.4**, only half of the images were used, and at 1 000 epochs we obtained 96.82% accuracy on LFW. When we used the whole dataset in **Exp.5** we got 97.52% accuracy

on LFW after 1 000 epochs. The performance only improved by less than 1 percentile, even when doubling the number of images used. We deduced from both these experiments that the most important aspect is the variability in the dataset. It is more beneficial to have more identities than to have more samples per person as the limit for the intra-class variability is achieved fast. We retained the model created in Exp 5 for the remaining tests. Further on we will refer to it as **OpenFace_best**.

Performance on the MOBIO dataset

The MOBIO dataset is divided into 3 partitions: training, development and evaluation. For the purpose of this work we did not use the training partition as we wanted to validate the model obtained from training on the MS-celeb-1M. Table 4.4 details the results on MOBIO of our model with the best performance on LFW (**OpenFace_best**). In the table we report the verification performance on both still and automatic protocols. Both these protocols are described further in [Bou16]. For the still protocol we used the still images provided in the framework of the ICB2013 challenge. For the automatic protocol we used 3 and 10 frames from the videos. The frames were selected uniformly from the videos, ie: for 3 frames we took the first, the middle, and the last frame. The results that we obtained on MOBIO are equivalent if not better than the commercial system studied in [Bou16]. To measure the performance on MOBIO we used the HTER metric which is defined in section 3.5.

Table 4.4: Results of our **OpenFace_best** model on MOBIO.

Openface_best	Eval Female (HTER)	Eval Male (HTER)
Still	14.57%	6.43%
3 frames	10.04%	4.79%
10 frames	8.84%	3.99%

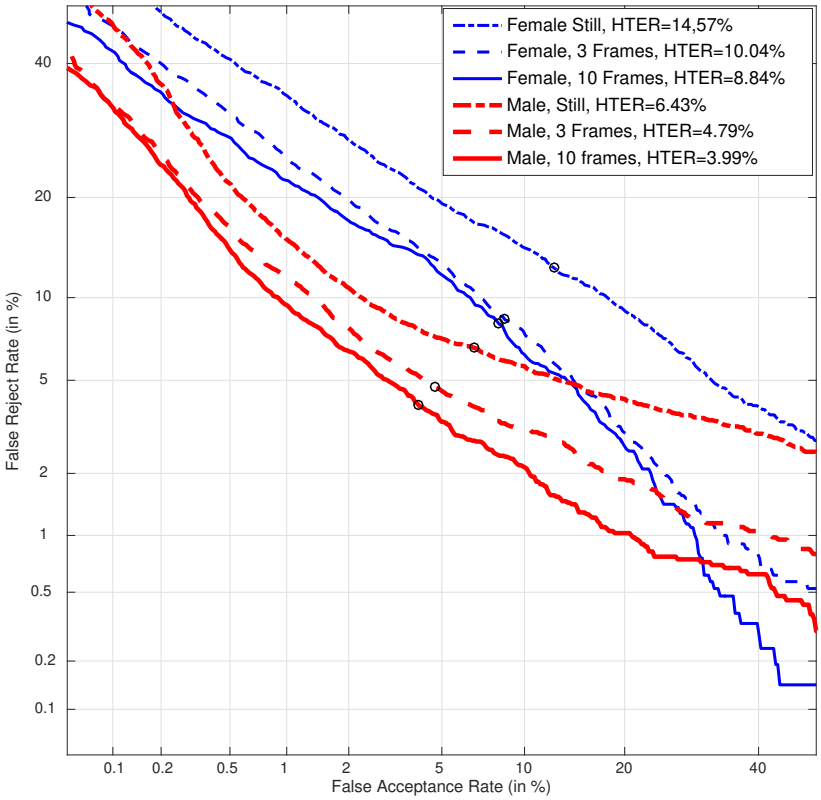


Figure 4.8: DET curves of OpenFace on MOBIO.

Table 4.5: Comparison of our results of the DNN and the DLDA on MOBIO still images and LFW

System	Training Dataset	Subjects	Images	MOBIO		LFW
				Eval Female (HTER)	Eval Male (HTER)	
SudFrog_1	FRGC	568	39328	17.43%	10.9%	79.94%
OpenFace_1	FRGC	568	39328	21.87%	18.97%	80%
SudFrog_best	Mobio train set + FRGC	100	4000	12.64%	7.68%	86%
OpenFace_best	MS-celeb-1M	100K	10M	14.57%	6.43%	97.52%

The MOBIO dataset is biased towards males with females representing about 30%. We trained OpenFace on a gender independent database. However we find relatively different results when comparing the performance between males and females. The same tendency appear in the systems studied in [Bou16]. The best reported results are 9% on the eval female partition and 5.5% on the eval male partition when using 10 frames, whereas we got 8.8% on the eval female partition and 4% HTER on the eval male. We can attribute the difference in the performance to the poor performance of the face detector on the female images. OpenCV fails to detect the face in 80 female images and only 19 in male images. This may be explained either by a bias in the pretrained face detector module or by bad illumination in the female images.

We studied the impact of the size of the training data on the performance in both cases of traditional DLDA approach using the SudFrog software and the deep neural network architecture provided by OpenFace. We decided to compare OpenFace to the DLDA approach because of fundamental similarities. Both, triplet embeddings and DLDA try to reduce the intraclass distance and enlarge the interclass distance. SudFrog is a face recognition system that was developed in Institut Mines Telecom, Telecom SudParis[2]. It is based on space reduction techniques. SudFrog does not do neither face detection nor landmark detection. Moreover, SudFrog aims to construct an Euclidean projection space, similar to OpenFace. It must be provided with the eyes, nose and mouth positions for it to do face recognition. For face detection and

[2]https://github.com/sudfrog/sudfrog

landmark detection, we use a combination of OpenCV and DLIB. OpenCV was used for face detection. DLIB was used for landmark detection. We used the default detectors provided by the software (front_face.xml for face detection and shape_predictor_68_face_landmarks.dat for landmark detection). In comparison, OpenFace uses DLIB both for face as well as landmark detection. OpenCV is slower, but detects more faces than DLIB on the somehow difficult MOBIO dataset. Using the same amount of data, SudFrog outshines the DNN. However, once we use the huge MS-celeb-1M dataset, the positions are reversed. We can not train SudFrog with MS-celeb-1M dataset as it is technically infeasible. The feature space becomes too huge for the memory.

Final version of the DNN architecture: The final version used in the SRE'19 multimedia submission follows the same approach as the initial DNN model. However, we introduced the following modifications.

- The inception-v3 layers were replaced by inception-resnet-v1[Sze+17] layers. Inception-resnet-v1 has a computational cost that is similar to that of inception-v3 and allows for faster training.
- The triplet loss function was modified according to the equation eq. 4.4. The modified version takes into account the cosine similarity between the embeddings. To compensate for the added terms, α is changed from 0.2 to 0.6. The value 0.6 was chosen empirically.
- The size of the embeddings was changed from 128 real components to 512 by trial. The choice of the embedding size was carried out using the LFW benchmark. We chose the size that gave the best accuracy on LFW.

$$L(\theta) = \sum_{i}^{N} max(\ \|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 \\ - f(x_i^a) \cdot f(x_i^p) + f(x_i^a) \cdot f(x_i^n) + f(x_i^p) \cdot f(x_i^n) \\ + \alpha \ , 0) \tag{4.4}$$

where "." is the dot product. The Euclidean distance component and the cos-distance have an overlap. We do not remove the overlapping components of the loss function in order to give more importance to the distance between the anchors and the negatives samples than the distance between the positives and the negatives sampled. If we develop eq. 4.4 we obtain:

$$L(\theta) = \sum_{i}^{N} max(\|f(x_i^n)\|_2^2 + f(x_i^n) \cdot f(x_i^p) + 3f(x_i^a) \cdot f(x_i^n) - 3f(x_i^a) \cdot f(x_i^p) \\ - \|f(x_i^p)\|_2^2 + \alpha, 0) \tag{4.5}$$

The impact of the improvements to the DNN is not evident on the LFW dataset, however on the AgeDB-30 benchmark and on the SRE'19 multimedia development partition, the improvement in the performance is much more visible. On the AgeDB-30 benchmark, the accuracy improved from 89% to 97% as shown in Table 4.6. This shows that the improved architecture is more robust to age variance than the initial version provided by the OpenFace framework.

4.3 NIST SRE2019 submission

We used our face recognition pipeline to participate in the NIST SRE2019 multimedia challenge.

Table 4.6: Biometric recognition performance of the studied DNN architectures. The face detection is done using the RetinaFace face detector. Face landmark detection is carried out using the DLIB implementation of ERT. The DNN is trained on the cleaned version of MS-celeb-1M.

	Accuracy on LFW (%)	Accuracy on AgeDB-30 (%)	EER on SRE'19 multimedia DEV (%)
Initial version	99.32	89.00	11.20
Final version	99.80	97.00	4.36

The 2019 speaker recognition evaluation (SRE'19)[3] [Sad+20] is part of an ongoing series of speaker recognition evaluations conducted by the US National Institute of Standards and Technology (NIST) since 1996. They provide a common test bed that enables the research community to explore promising new ideas in speaker recognition, and have a valuable impact to support the community in their development of advanced technology incorporating these ideas.

SRE'19 consisted of two separate activities. The first one was a leaderboard-style challenge using Conversational Telephone Speech (CTS), for text-independent speaker detection. Moreover, in addition to the regular audio-only track, the SRE'19 introduced, for the first time, an audio-visual and visual-only tracks, denoted as Multimedia track.

The audio-visual data for the multimedia challenge were extracted from the unexposed portions of the Video Annotation for Speech Technology (VAST) corpus, collected by the Linguistic Data Consortium (LDC).

We trained the face detector using the WIDER FACE dataset with the default configuration described in [Den+19b]. As for the face embedding extractor,

[3] https://sre.nist.gov/

we used the MS-celeb-1M dataset for training. After removing label noise using the DBSCAN clustering algorithm, the training dataset comprises 80 000 subjects with a total of 4 000 000 images.

For the enrolment, we select the frames provided by the manual annotations given by National Institute of Standards and Technology (NIST). We then crop the frame to the bounding box specified in the metadata. The face is then detected inside the provided bounding box. DLIB [Kin09a] landmark detector is used to obtain the face landmarks. We use the outer eyes points and the nose tip in order to align the face to a predefined layout. The aligned image is resized to 96*96 pixel rectangle and fed to the face embedding extractor. The result is a 512 component face embedding per frame.

As for the test videos, we begin the processing by extracting one frame per 0.5 seconds from each video. Then we apply the RetinaFace face detector to each extracted frame to get all the faces present in the frame. We then run the landmark detector on each bounding box found by the face detector. After aligning and resizing the faces, we extract the embedding of each face. In order to compute a single score for each trial involving an enrollment video and a test video, we compute the maximum of the cosine similarity scores obtained by comparing all of the enrollment embeddings and the embeddings of the faces found in the video. The scores are not post-processed.

Figure 4.9 reports the performance of the submitted face recognition system on the dev and test partitions of the multimedia challenge. The curves show that the submitted system performs better on the test set than on the dev set. This might be due to the higher number of tests in the test partition in comparison with the dev partition. In fact, the number of tests in the test partition is 12 times higher. The difference in the number of tests results also in higher error margin for the dev set. For example, the error margin at the

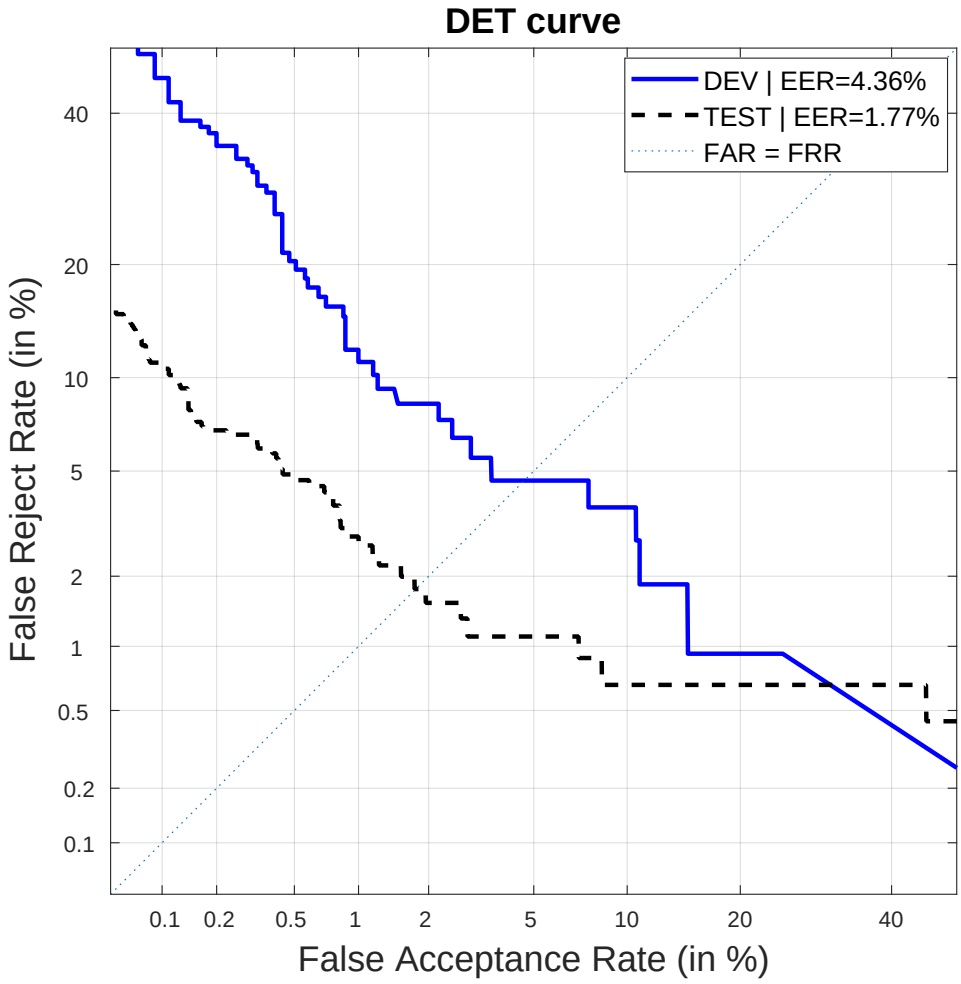


Figure 4.9: DET curve performance of the submitted face recognition system on the DEV and TEST partitions of the multimedia challenge.

EER is 1.89% for dev set and 0.55% for the test set.

The scores provided by our face recognition system are computed using the cosine distance. These scores are suitable for the EER metric. However, they cannot be used to compute the actual cost (act_C) which is the primary metric for the SRE'19 challenge. The act_C metric is based on the log likelihood ratio (LLR) scores. The scores were not post-processed (no Z-norm, T-norm etc...), but were calibrated using the bosaris toolkit [BV11]. The score calibration is

done in the same step as the fusion. The multimedia challenge is an audio-visual task. The face recognition scores need to be fused with the speaker recognition scores.

Among the modifications applied to the framework, the use of the RetinaFace face detector resulted in the most significant improvement in performance on the development set. The quality of the detected face landmarks depends significantly on the correctness of the bounding box given by the face detector. Using the correct face landmark results in better face alignment and more robust templates. It is worth noting that this detector, RetinaFace, was also used by other submissions, which shows its success in the task of face detection.

Although the majority of the submissions to the challenge used CNN models for face recognition, a system submitted by one team used TDNN for face recognition, which gave 19.84% EER on the development set. This shows that CNN is one of the better-suited architectures for face recognition.

We also note that applying whitening and using PLDA for scoring does not improve the recognition performance, at least in our submission.

Finally, applying enrollment filtering using some quality measures is crucial to the performance of the face recognition system. If the enrollment reference is of bad quality, the comparison against good test references will result in lower similarity scores and, as a result, impacts the decision threshold. Some participants implemented enrollment provisioning procedures where they locate the target face in frames other than the keyframes given by the metadata to obtain a better enrollment reference. However, in our submission, we only used the keyframes of the enrollment videos. This helped us

obtain the best performance in the EER metric between the 14 submitted systems to the SRE'19 multimedia challenge.

4.4 Implementation in H2020 European Projects

4.4.1 The SpeechXRays project

The H2020 SpeechXRays project is a European project that took place between 2015 and 2019. Its goal is to develop and test in real-life environments a user recognition platform based on audio-visual identity verification. Under the framework of the SpeechXRays project, a cancelable face recognition prototype was developed and implemented. The project also provides classical biometrics using state-of-the-art Deep Convolutional Neural Network for face recognition coupled with well-established techniques for speaker recognition. The platform includes anti-spoofing by doing liveness detection and prompting the user a different sentence at each access attempt. The system also provides cancelable face biometrics based on binary embedding shuffling and using a second factor in the form of a password or a Secure Element[4]. The SpeechXRays platform provides two types of face recognition systems; the first method is based on linear space reduction technique DLDA [YY01], the second method is a DNN based face recognition system [HPD18] inspired by the FaceNet architecture [SKP15].

The DNN module (which constitutes the default component of the project face recognition system) outputs a low dimensional representation of an input image, which consists of a normalized Euclidean feature vector of size 128. The resultant embeddings are compared using Euclidean distance. As

[4]A Secure Element (SE) is a tamper-resistant combination of hardware, software, and protocols capable of embedding smart card-grade applications. Typical implementations include UICC, embedded Secure Element, and removable memory cards.

for the cancelable version, it creates binary templates of 2048 bits and uses the hamming distance for comparison.

The face biometric module was delivered by Télécom SudParis (TSP) and came in different versions. The very first version, delivered on June 2017 and denoted as SudFrog, was based on Gabor features, Linear Discriminant Analysis, and Direct Linear Discriminant Analysis, which improved cross-class discrimination. We denote them in the SpeechXRays documents as TSP-DNN face biometric module. Different versions of the initial system were proposed. The main releases are SpeechXRays-DNN 4.0 delivered to SIVECO on August, 2018, and a new improved version of this system SpeechXRaysDNN 4.1 delivered to SIVECO on October, 2018. The TSP-DNN face biometric modules are based on an OpenFace implementation of the FaceNet system. It is inspired by the inception network.

The latest version of TSP-DNN system achieves 98.9%. This is a huge improvement over the previous SudFrog system that gave 80% on this same LFW database.

To facilitate face biometrics solution RealEyes[5] has provided different versions of a Face Detection and Face Landmark Alignment SDK. The SDK is provided for OSX, Windows, Android, and iOS operating systems, in both 32 and 64 bit versions. The first version of the SDK is based on a variation of a Viola-Jones type Haar features based face detector. It is a state-of-the-art face detector that balances well between speed, accuracy and algorithmic complexity. Its main drawback is inability to detect non-frontal faces. For landmark tracking we have used an implementation of an ERT (Ensemble of Regression Trees) based landmark tracking method. It trades off accuracy

[5]One of the consortium members

over speed and has been shown to have a very good performance on processing power constrained hardware, such as mobile devices. The second version of the SDK that was integrated in the final version of the SpeechXRays solution, has had two major upgrades, both aimed at improvements in accuracy of face detection and landmark tracking. Its face detector is replaced with a FastCNN method based on deep convolutional neural networks and trained on a much larger training dataset, consisting of a mix of RealEyes proprietary and third-party datasets. Landmark tracking algorithm is upgraded to an SDM (Supervised Descent Method) based algorithm, which is capable of achieving higher landmark fitting accuracy.

4.4.2 Empathic Project

We denote the Empathic Project implementation of our face recognition system as IMT-DNN. The IMT-DNN face biometric module is based on the OpenFace [ALS16] implementation of the FaceNet system. The initial implementation of this module was trained on the the MS-celeb-1M dataset as it was provided. The Ms-celeb-1M dataset contains a big portion of mislabelling because it was generated automatically using web crawlers. In order to improve the performance, we cleaned the database using DBSCAN clustering algorithm.

4.4.3 Version 1

The module leverages the following technologies:

- Java as the programming language.
- SQLite database for session management.
- Libvlc for decoding the h264 video stream.

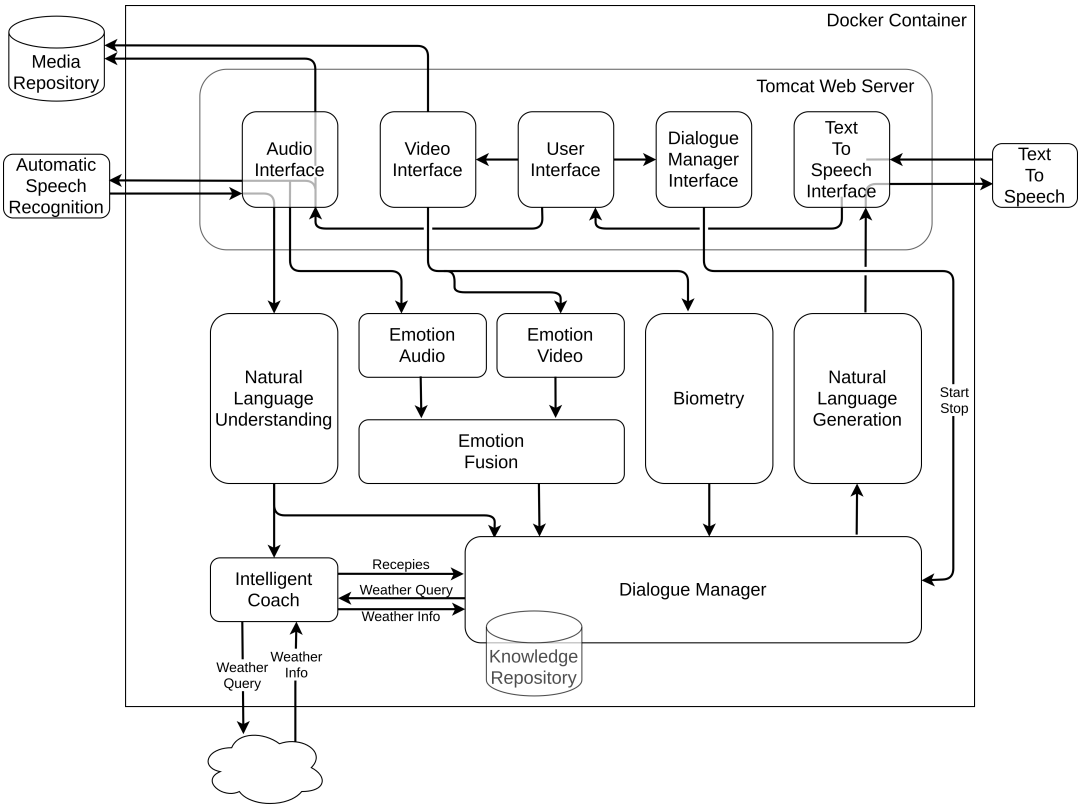


Figure 4.10: Final EMPATHIC system architecture.

- OpenCV for image processing and executing the neural networks (face detection, landmark detection, Euclidean representation extraction).
- Apache webserver to serve the endpoints of the Rest APi.

The biometric authentication module implements a REST API which the Web UI connects to in order to authenticate the users. Due to difficulties in implementing the media server which manages the video streams we implemented a second version of our face recognition system.

4.4.4 Version 2

The module has two input channels:

- ActiveMQ Message Broker queue.

- Socket: Through these connections the module receives the image frames.

As for the output, for every input image frame processed, the module sends a JSON message to the AMQ message broker (2DM queue). It is the role of the Dialogue Manager to decide how to act based on the information provided by the biometric authentication module. Finally, it should be said that the module uses an enrollment protocol in order to distinguish between users and to verify them. The protocol is as follows:

1. When receiving a start message from messaging queue, get user ID.
2. Check if there is a template enrolled using the user ID. If False then go to step 3 (we are in the first session of the user) else: Go to step 4 (we are not in the first session).
3. If there is no template enrolled, enroll the user using the first 5 frames that contains a face. Following this step the templates will be created, stored and able to be used for further verifications.
4. Check the presence of the enrolled user and send a message for every image frame processed.

Due to privacy issues and regulations as the Empathic project may contain sensitve medical data, it was decided to use the face recognition system as a means to check for the presence of the user in video feed and to check that it is the same user connecting using the same account across sessions.

During the testing of the final version of the system, the biometric module successfully enrolled and verified 78 elderly users across multiple sessions. There were 31 Spanish, 29 French and 18 Norwegian users. According to the

feedback from the users, they did not have any complaints about the face biometric module.

4.5 Conclusions

In this chapter, we presented our face recognition system pipeline. The system is built upon the OpenFace framework, to which we introduced several modifications to obtain better performance as it was implemented in two European projects and used in a submission to the NIST SRE2019 multimedia challenge.

We also detail how to obtain a state-of-the-art face recognition system based on publicly available software and using public datasets. We try to give the most possible details to allow for the reproducibility of the results. When CMU implemented OpenFace, reproduciblity was one of their main goals. Thus we were able to reproduce their results and improve upon them. However, we couldn't get the same results as Google who used huge proprietary datasets and a proprietary face alignment system. Our **OpenFace_best** DNN model gives good verification results on both evaluation datasets, MOBIO and LFW. From the results that we obtained we can infer that the performance bottleneck lays in the preprocessing, notably the face detection phase. Given enough data, the DNN is unmatched. Nevertheless, in situations where the databases are not available classical approaches give better performance.

In order to improve our results, we proceeded to remove the misalbeling noise from the MS-celeb-1M which gave the biggest improvement in performance on our validation protocols.

5 Binarisation

5.1 Introduction

Face is one of the most widely used biometric characteristics. With the availability of huge face recognition datasets [Guo+16; HLM14] and growing computational power, face recognition performance keeps improving [Eri+19; SKP15; ALS16; Liu+17; Den+19a]. Face recognition has seen vast adoption thanks to its accuracy and ease of use. From Smartphones and computers to CCTV cameras and surveillance, face recognition is present everywhere. This widespread presence raises privacy and security concerns. A solution to these concerns is to employ biometric template protection schemes such as crypto-systems and cancelable biometrics. However, to protect the face templates, most of the techniques employed need a binary representation of the face. In addition, most face verification systems employ continuous representations, which are less suitable for template protection schemes.

Crypto-biometric schemes such as fuzzy commitment require binary sources. This chapter introduces a novel approach to binarising biometric data using Deep Neural Networks (DNN) applied to facial biometric data. The binary representations are evaluated on the MOBIO and the Labeled Faces in the Wild (LFW) databases, where we measure their biometric recognition performance and entropy. The proposed binary embeddings give a state-of-the-art performance on both databases with almost negligible degradation

compared to the baseline.

The representations' length can be controlled. Using a pre-trained CNN and training the model on a cleaned version of the MS-celeb-1M database, we obtain binary representations of length 4 096 bits and 3 300 bits of entropy. The extracted representations have high entropy and are long enough to be used in crypto-biometric systems such as fuzzy commitment.

Furthermore, the proposed approach is data-driven and constitutes a locality preserving hashing that can be leveraged for data clustering and similarity searches. As a use-case of the binary representations, we create a cancelable system based on the binary embeddings using a shuffling transformation with a randomization key as a second factor.

The major contribution presented in this chapter is introducing a data-driven template binarisation method using Deep Neural Networks, which does not degrade the performance of the baseline system. Furthermore, we seek to obtain long binary representations with high entropy to be used in crypto-biometric key regeneration schemes.

The proposed binarisation method has four main advantages:

- The degradation in the recognition performance caused by the binarisation is negligible compared to the baseline system.
- The binarisation method can be applied to any type of real representation.
- The length of the binary representation can be controlled. The binarisation method provides arbitrary length presentations that are limited only by the quality of the training database (size, noise). This allows

for flexible representations that can be adapted to multiple applications, such as crypto-biometric key regeneration, fuzzy commitment, and fuzzy extraction schemes.

- The binarisation method keeps the topology of the original space, which allows for the use of the binary representation in database searches and clustering.

The chapter is organized as follows: In Section 5.2, we explain the different approaches we followed to extract binary embeddings directly using Deep Neural Networks (DNN). In Section 5.3, we analyze the performance of the binary representations in terms of biometric recognition and entropy. In Section 5.4, we study a use-case for the binary embeddings consisting in creating a cancelable biometric system using a shuffling transformation as a protection scheme. In Section 5.5, we describe a cancelable system that we implemented in the H2020 SpeechXRays project based on the same template protection scheme. In Section 5.6, we compare both implementation to show the importance of the performance of the binary representations on the security of the cancelable system before concluding in Section 5.7.

5.2 Proposed Face Binarisation Method

This study uses deep neural networks to extract binary biometric representations from face images. This way, we take advantage of data-driven approaches to generate an optimized binary representation.

Our binarisation method consists of training an end-to-end binary embedding extractor directly from aligned face images. Thus, the binarisation layer considers the loss function and is optimized for the task. We aim to obtain locality-preserving binary representations. The locality preserving property

is defined by eq. 5.1 where a, p and n are three random points from the original space and $f(.)$ is the projection function. The triplet loss function (shown in eq. 5.2 is suitable for this task as the optimization criterion is equivalent to eq. 5.1. To this end, we based our DNN on the FaceNet [SKP15] architecture which uses the triplet loss function for the training.

$$d(a,p) < d(a,n) \Rightarrow \|f(a) - f(p)\| < \|f(a) - f(n)\| \tag{5.1}$$

In the following subsections, we present the baseline face recognition system and describe the approaches taken to binarise the biometric data.

5.2.1 Baseline Face Recognition System

Our goal is to obtain discriminating binary representations from faces that do not degrade the performance of the baseline system. The proposed binarisation method transforms Euclidean face embeddings into binary embeddings of different lengths. The Euclidean embeddings are constructed using a Deep Neural Network based on FaceNet [SKP15]. In [HPD18] we describe in detail the methodology we followed to create the face recognition system based on the OpenFace implementation [ALS16]. We trained a convolutional DNN using the triplet loss function. The triplet loss function, given by eq. 5.2, takes a triplet comprised of an anchor x_i^a and a positive sample x_i^p from the same subject, and a negative sample x_i^n selected randomly from the rest of the dataset. The training goal is to bring closer the anchor and positive samples and distance the negative sample using the margin α.

$$L = \sum_{i}^{N} \max\left(0, \left\|f(x_i^a) - f(x_i^p)\right\|_2^2 - \left\|f(x_i^a) - f(x_i^n)\right\|_2^2 + \alpha\right) \tag{5.2}$$

The DNN architecture for training the face projection space is composed of 24 layers and 3 733 968 parameters. The training phase aims to obtain the best representation that separates the positive identities from negative ones using the triplet loss function. After the training phase, the network outputs a low dimensional representation of an input image consisting of a normalized Euclidean feature vector of size 128. Further details about the DNN architecture are provided by Table 5.1.

Table 5.1: Details of the nn4.small2 Inception architecture which is a version of the nn4 model from FaceNet [SKP15] hand-tuned by [ALS16] to have less parameters. Each row is a layer in the neural network and the last six columns indicate the parameters of pooling or the inception layers from [Sze+15]. This model is almost identical to the one described in [Sze+15].The pooling is always 3×3 (aside from the final average pooling) and in parallel to the convolutional modules inside each Inception module. If there is a dimensionality reduction after the pooling it is denoted with p. 1×1, 3×3, and 5×5 pooling are then concatenated to get the final output.

type	output size	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj
conv1 (7 × 7 × 3, 2)	48 × 48 × 64						
max pool + norm	24 × 24 × 64						m 3 × 3, 2
inception (2)	24 × 24 × 192		64	192			
norm + max pool	12 × 12 × 192						m 3 × 3, 2
inception (3a)	12 × 12 × 256	64	96	128	16	32	m, 32p
inception (3b)	12 × 12 × 320	64	96	128	32	64	l_2 , 64p
inception (3c)	6 × 6 × 640		128	256,2	32	64,2	m 3 × 3, 2
inception (4a)	6 × 6 × 640	256	96	192	32	64	l_2, 128p
inception (4e)	3 × 3 × 1024		160	256,2	64	128,2	m 3 × 3, 2
inception (5a)	3 × 3 × 736	256	96	384			l_2, 96p
inception (5b)	3 × 3 × 736	256	96	384			m, 96p
avg pool	736						
linear (fc)	128						
l_2 normalization	128						
linear	N						
binarisation	N						
linear	128						
l_2 normalization	128						

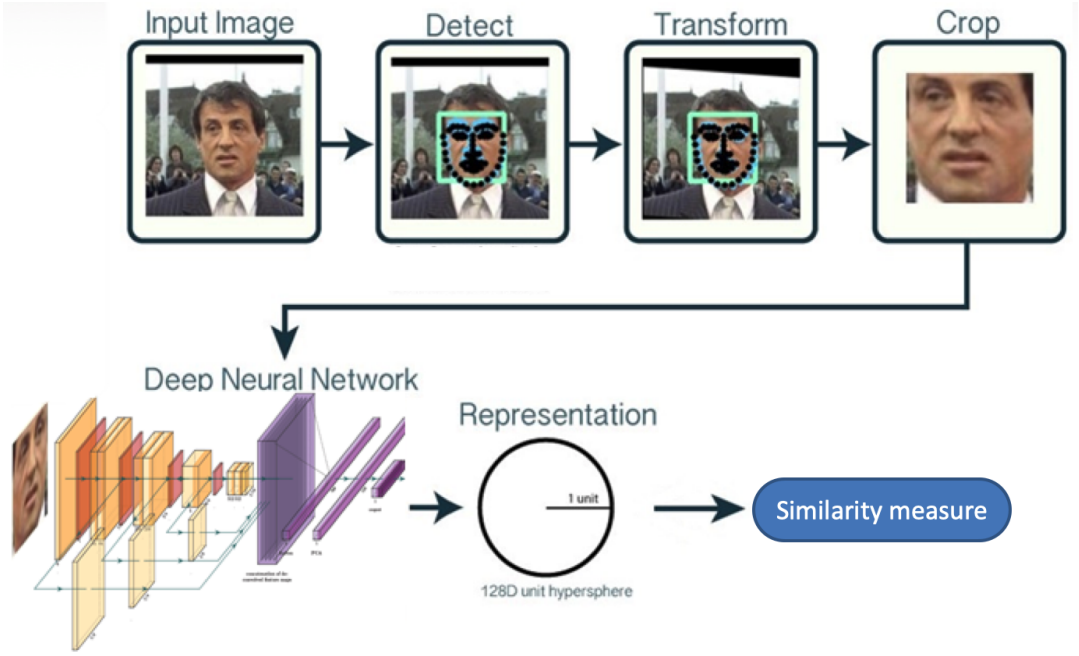


Figure 5.1: Pipeline of the baseline face recognition system.

Figure 5.1 shows the pipeline of the face recognition system. First, the face is detected and aligned according to a predefined template. Afterward, the aligned face is processed by the DNN in order to extract a Euclidean representation. This Euclidean representation constitutes the template that defines the user either for enrolment or verification.

Face alignment consists of three steps: face detection, landmark detection, affine transformation, and face cropping. The face detection is carried out using a deep convolutional neural network provided by OpenCV based on Single-Shot-Multibox Detector (SSD) [Liu+16] and uses ResNet-10 architecture as a backbone. This model gives state-of-the-art performance with a low computational overhead. The image needs to be resized to 300x300 pixels to use the face detector. The image is provided in RGB format after subtracting the mean from each value. The output of the SSD detector is a bounding box. Given the face-bounding box, we use an implementation of [KS14b] provided by DLIB [Kin09b] to detect the facial landmarks. Further details on

the face alignment are provided in [HMD21].

Finally, the Euclidean embedding extracted from the aligned face using DNN can be used for face recognition either in identification or verification scenarios. This work aims to binarise the Euclidean embeddings with the least amount of degradation, which we explain in the next section.

5.2.2 Locality Preserving Binary Face Representations using Auto-encoders

Figure 5.2 shows the architecture of the proposed approaches. Both approaches (a) and (b) follow the same architecture. The difference lies in how the training data is used. In approaches (a) and (b), we opted to use an auto-encoder on top of the deep convolutional neural network (FaceNet based) to obtain the binary code.

The idea was to use an encoder to project the Euclidean representation that we get from the DNN onto another vector. This vector has the same size as the intended binary representation. Afterward, we apply a custom binarisation layer on the vector and finally use a decoder to get back to the Euclidean representation.

The binarisation layer is defined as follows. In this layer, we apply a threshold to each input component. The output of this layer is defined in eq. 5.3. The input is compared to a threshold that is specified beforehand. The choice of the threshold is based on the type of the previous layer activation function. In our case, we chose a threshold of "0" as the previous activation function is the hyperbolic tangent. This layer does not have trainable parameters. In the back-propagation phase of the training, this layer is treated as the identity

function, and its gradient is equal to 1.

$$F(input) = \begin{cases} 0, & \text{if } input \leq threshold \\ 1, & \text{otherwise} \end{cases} \tag{5.3}$$

This idea has two benefits. First, we get more control over the length of the binary representation (we only need to modify the auto-encoder). The second benefit is that we get a continuous output from the auto-encoder, allowing us to use standard optimization methods in conjunction with the triplet loss criteria. Figure 5.2 illustrates the example where we use a code length of P. First, the image is fed to the DNN, and we extract a Euclidean representation of size 128. Next, encode it on a P-component real vector, which is, in turn, binarised. Then we reconstruct the initial Euclidean representation. The architecture described in Figure 5.2 is only used during the training phase. After training, we remove the decoder, and obtain a binary code given a face image.

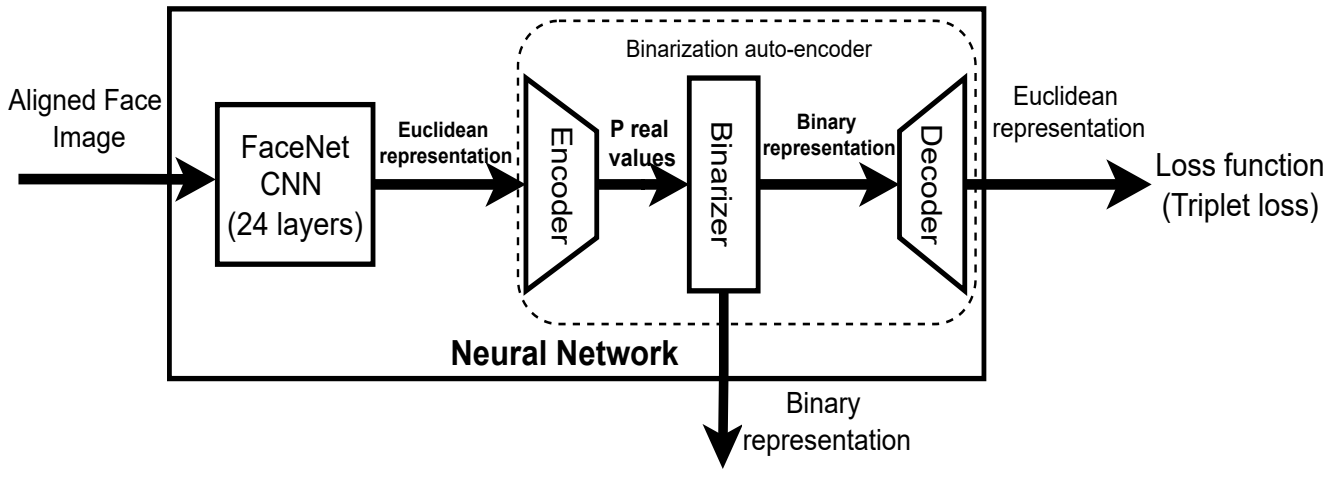


Figure 5.2: Block diagram of the binarisation method used in approaches approach (a) and (b). In approach (a), the whole model is trained from scratch. In approach (b) the FaceNet CNN is pre-trained using the MS-celeb-1M.

To put this idea into practice, we first needed to find an auto-encoder architecture suitable for the output. In other words, we sought to find the hyper-parameters of the auto-encoder (number of hidden layers, width of the layers, and the activation functions) that result in the least degradation in the recognition performance compared to the original Euclidean representation. In this step, we did not use the binarisation layer. As the binarisation step generally degrades the performance, we would not be able to say whether the performance was degraded due to the auto-encoder or the binarisation step. The architecture that resulted in the least degradation was constructed using three layers. The encoder consisted of two linear layers with a hyperbolic tangent as an activation function. We used a single layer with a ReLU activation function for the decoder. The auto-encoder choice was based on the difference between the auto-encoder performance and the baseline performance of the original DNN architecture, which is 97.52% on the LFW. Once the auto-encoder architecture is set, we introduce the binarisation layer between the encoder and the decoder.

The difference between approach (a) and approach (b) is that in (a), we train the whole architecture from scratch, while in (b), we use a pre-trained model on MS-celeb-1m. This model is described in [HPD18]. Compared to the approach (a), where the training is done from scratch, it is much faster for the DNN to converge towards good results. The loss of the models constructed using approach (a) stabilizes around 1 000 epochs compared to 100 epochs for models constructed using approach (b).

5.3 Experimental Performance of the Binary Representations

In this section, we present the biometric performance of the binary representations. We evaluate the performance on the LFW and the MOBIO databases using the accuracy, as a common evaluation metric, computed using the 10-fold cross-validation protocol.

We evaluate the recognition performance and the entropy of the models. As the goal of the work is to binarise the biometric samples to be suitable for biometric crypto-systems and biometric protection schemes, the binary representations should have high entropy and good recognition performance.

In approach (a), we train the network, shown in Figure 5.2, from scratch on the MS-celeb-1M dataset using the triplet loss function. We report in Table 5.2 the performance of this approach for various lengths of the binary representations. The training was carried out for 1 000 epochs. We note that the best performance on LFW is obtained with 512-bit representations. On the other hand, 512-bit representations provide the best performance on the MOBIO dataset. We attribute that to the overlap of the original MS-celeb-1M dataset with the LFW dataset. As the representation length grows, the model overfits to MS-celeb-1M resulting in worse performance on MOBIO.

When the length of the embeddings reaches 4 096 bits, the recognition performance decreases dramatically. On LFW, the accuracy plummets from 93% to 81% compared to representation with a length of 2 048. The performance degradation is more accentuated on the MOBIO dataset, where the error reaches almost 50%. We attribute the cause of the degradation when using 4 096-bit embeddings to the loss of information in the training phase of the neural network due to the thresholding process. The information propagated

backward is not enough to optimize the system's parameters. The number of trainable parameters in the auto-encoder evolves exponentially from 33 024 parameters for representations with a length of 128 bits to 1 056 768 parameters for the 4 096-bit representations.

Table 5.2: Impact of the length of the binary representations on the biometric performance of approach (a): Training the Auto-encoder using Triplet Loss from scratch. The baseline system is the system used in [HPD18]. The results in the second row (row '128*') are obtained by applying a median binarisation on the output of the CNN used in [HPD18]. The maximum standard deviation (std) on LFW is around 1%. The maximum std on MOBIO is around 0.1%.

Length	Accuracy on LFW %	Accuracy on Mobio %
baseline system	97.52	90.58
128* (median)	89.32	79.74
128	91.73	82.50
256	93.18	83.50
512	**94.12**	**84.23**
1024	93.62	81.46
2048	93.07	79.46
4096	81.13	53.60

Table 5.3: Entropy of the representations created using approach (a). The entropy was measured using 5M samples from MS-celeb-1M. $p(x = 1)$ is the probability of a bit is equal to 1.

Length	$p(x = 1)$	Entropy
128	0.487	98.26
256	0.532	113.87
512	0.514	**163.4**
1024	0.496	116.65
2048	0.511	143.92
4096	0.503	49.87

Studying the biometric performance of the binary representation alone is not enough, especially where we are trying to have long representations. Appending a fixed portion to all the representations will not degrade the recognition performance of the system. However, as our primary goal is to obtain

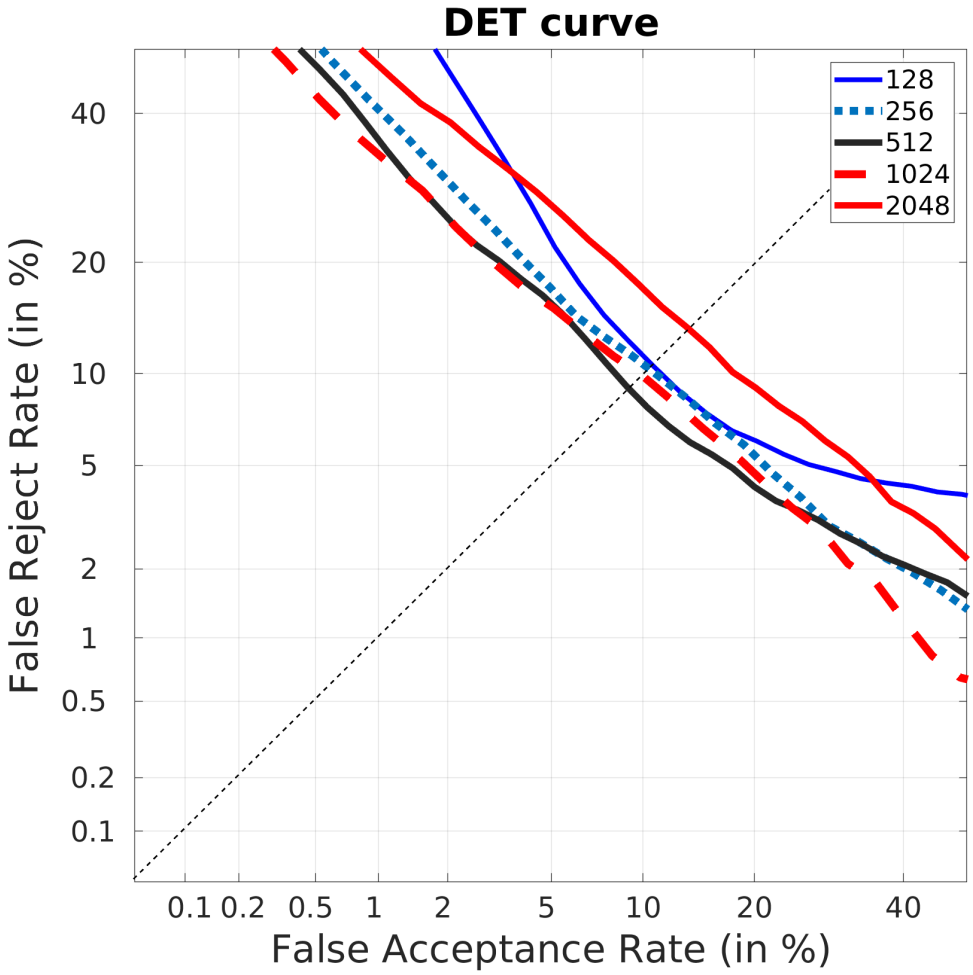


Figure 5.3: DET curves of the Eval male partition of the MOBIO database using the standard protocol [Bou16]. The training of the models is done using the MS-celeb-1M. The training is done from scratch on the original version of MS-celeb-1M following approach (a).

a long binary representation, we need to study the entropy of the representations.

In information theory, entropy is a measure of the amount of uncertainty or randomness in a system. It is a key concept in the field of information theory, which was developed in the 1940s by Claude Shannon in order to understand the limits of communication and information storage.

It is defined as:

$$H = -\sum(p(x) \times \log(p(x))) \tag{5.4}$$

where H is the Shannon entropy, $p(x)$ is the probability of outcome x, and the sum is taken over all possible outcomes x. One of the most important applications of entropy in information theory is in the field of security and cryptography. In these fields, entropy is used to measure the strength of a password, the randomness of a key, or the security of a communication channel.

For example, when choosing a password, it is important to select a password that is difficult for an attacker to guess. One way to measure the difficulty of a password is to calculate its entropy, or the amount of uncertainty or randomness it contains. A password with high entropy is more secure, because it is less likely to be guessed by an attacker.

Similarly, when generating a key for encryption, it is important to select a key that is random and has high entropy. A key with high entropy is more secure, because it is less likely to be predicted by an attacker.

Thus, we decided Shannon entropy to measure the performance of the binary representations, as the main goal of the thesis is crypto-biometric key regeneration.

The entropy is measured on 5 million samples from MS-celeb-1M. We use Monte Carlo random sampling in order to compute the entropy. From the 5M samples, we select 500k samples randomly and measure the entropy based on those 500k samples. This step is repeated for 1 000 iterations. The entropy provided in the tables is the average of the 1 000 iterations.

We report in Table 5.3 the entropy of the binary representations according to their length. Presentations of length 512 provide the highest entropy with 163 bits. On the other hand, embeddings of length 4 096 give the lowest value for entropy which is consistent with their biometric recognition performance.

The results of the approach (a), especially the low entropy, led us to use the pre-trained face recognition models instead of training from scratch. Table 5.4 reports the performance of the system when we use a pre-trained CNN. Using a pre-trained CNN significantly reduces the degradation, especially for embeddings with 4 096 bits. In addition, the pre-training reduces the loss of information introduced by the auto-encoder. Using a pre-trained model on the cleaned version of MS-celeb-1M and adding the auto-encoder previously discussed resulted in better biometric verification performance compared to training the model from scratch. The pre-trained CNN is the FaceNet model trained on the same dataset as the auto-encoder. So, when we present the performance of the models trained on the original/cleaned version of MS-celeb-1M, the pre-trained CNN is trained separately on the same set as the whole module.

We report the entropy of the binary representations obtained using an auto-encoder with a pre-trained CNN in Table 5.5. For the version trained on the original MS-celeb-1M, we see that the entropy of the keys reaches its maximum of 260 for representations of size 1 024. Besides, the $p(x = 1)$ is around

0.5 (except for length 4 096), which shows that many bits of the representations are constant. In addition, when we use the cleaned training database for training the system, we see that entropy improves significantly, in particular when the length of the representation exceeds 512 bits.

To estimate the degradation of the biometric performance introduced by the binarisation, we compare the performance of the approach (a) to the system presented in [HPD18]. For approach (b), we compare the performance of the binarised embeddings to the pre-trained CNN that was used. Approach (a) shows higher degradation in the performance, from 97.53% to 94.12% accuracy on LFW and from 90.58% accuracy on MOBIO to 84.23%. The degradation is more pronounced on the MOBIO database due to bias in the original version of MS-celeb-1M towards the LFW dataset.

Table 5.4: Impact of the length of the binary representation on the biometric recognition performance of approach (b) (Using a pre-trained CNN with an auto-encoder). Values in **bold** are given by models trained using the cleaned version MS-celeb-1M. The first row is provided to show the degradation in recognition performance between the initial system (Euclidean embeddings) and the binarised embeddings. By 'pre-trained CNN', we denote the initial OpenFace DNN. The results in the second row (row '128*') are obtained by applying a median binarisation on the output of the pre-trained CNN.

Length	Accuracy on LFW %		Accuracy on MOBIO %	
Pretrained CNN	97.52	**99.22**	90.58	**98.93**
128*(median)	89.32	**93.22**	79.74	**90.15**
128	94.88	**97.30**	81.31	**95.27**
256	95.37	**97.50**	87.62	**97.84**
512	95.85	**98.80**	87.11	**98.28**
1024	96.32	**99.12**	89.35	**98.87**
2048	95.06	**99.00**	85.60	**98.58**
4096	95.15	**99.00**	80.12	**98.90**

As for approach (b), we present two cases. The first case is when the pre-trained CNN and the auto-encoder are trained on the original MS-celeb-1M.

Table 5.5: Entropy of the representations created using the approach (b). The entropy was measured using 5M samples from MS-celeb-1M. $p(x = 1)$ is the probability of a bit being equal to 1. Values in **bold** are given by models trained using the cleaned version MS-celeb-1M.

Length	$p(x = 1)$	Entropy
128	0.497 | **0.489**	112.22 | **116.20**
256	0.486 | **0.481**	205.67 | **233.59**
512	0.493 | **0.482**	252.01 | **473.74**
1024	0.506 | **0.454**	261.29 | **944.24**
2048	0.498 | **0.315**	223.99 | **1679.25**
4096	0.826 | **0.308**	179.08 | **3349.47**

In this case, as shown in Table 5.4, the accuracy on LFW is decreased by about 1% to 2% compared to the baseline. On the other hand, the accuracy on the MOBIO dataset improved compared to the performance of approach (a). We attribute the difference of behaviour of the system to the overlap between the training and LFW databases. However, when the training is carried out on the cleaned database, the degradation on both datasets is lower than 1%. On the LFW database, we obtain 99.12% accuracy using the binary representations, whereas we get 99.22% accuracy using the baseline system. The same applies to the MOBIO database, where we get 98.9% accuracy using the binary representations, compared to an accuracy of 98.93% with the baseline system. This shows that our binarisation methods are highly dependent on the quality of the training data. By the quality of the training data, we refer to the level of the noise, the mislabelling, the quality of the images, and the size of the database. If we have little data, it will result in low entropy of the representations. The mislabelling and the noise will also reduce the system's accuracy and lower the entropy of the representations at the same time.

As shown in Table 5.5, the entropy of the representations depends on the quality of the training dataset (non-cleaned/cleaned). For the non-cleaned version, representations with lengths longer than 256 bits have no further

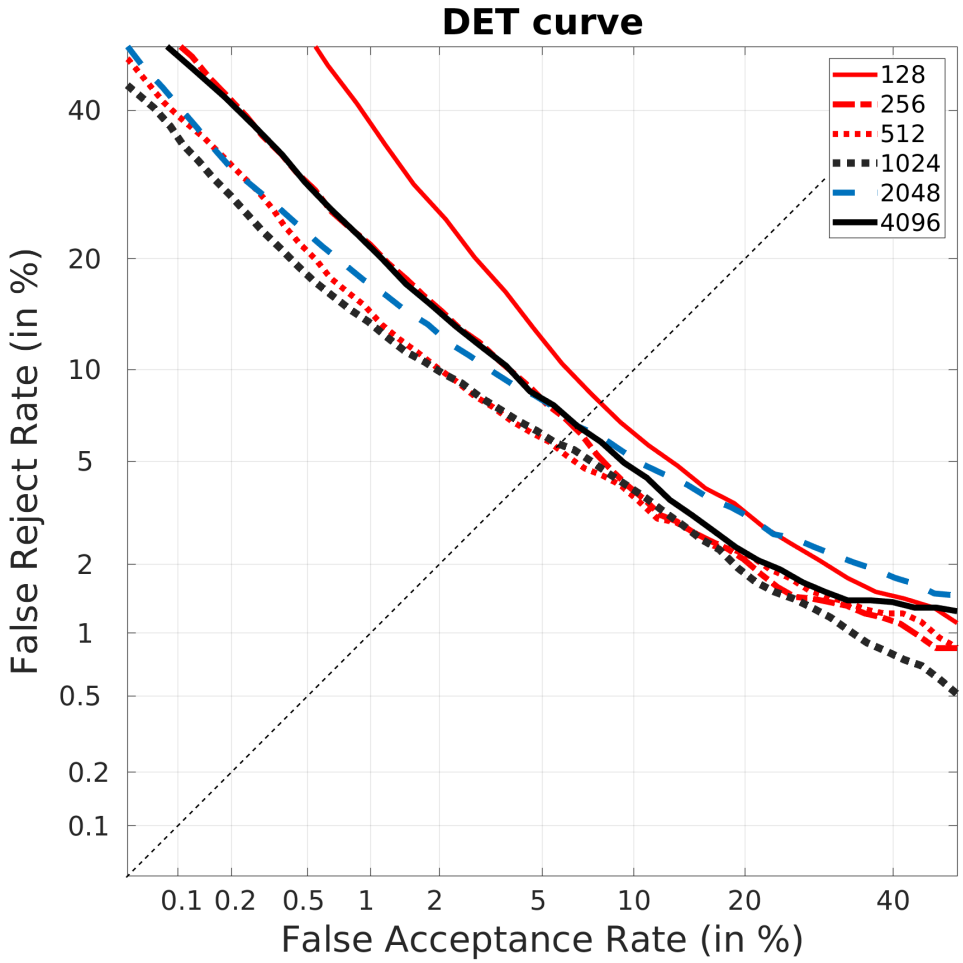


Figure 5.4: DET curves of the Eval male partition of the MOBIO database using the standard protocol [Bou16]. The training of the models is done using the MS-celeb-1M. The training is done using a pre-trained CNN on the original version of MS-celeb-1M following approach (b).

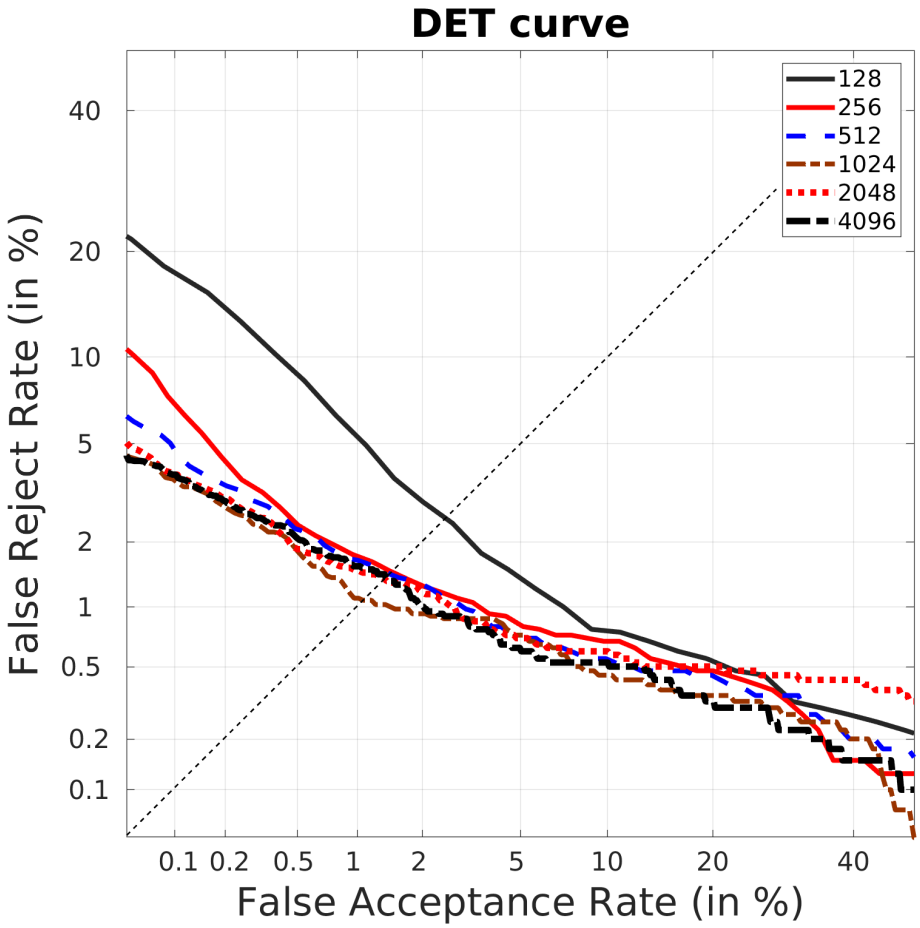


Figure 5.5: DET curves of the Eval male partition of the MOBIO database using the standard protocol [Bou16]. The training of the models is done using the MS-celeb-1M. The training is done using a pre-trained CNN on the cleaned version of MS-celeb-1M following approach (b).

useful information. As for the cleaned version, this behaviour appears when we exceed the length of 4 096 bits. The proposed auto-encoder can provide longer representations, but their real length, which is shown through their entropy, is limited.

Finally, in both approaches (a) and (b), the performance is better than binarising simply using the median as described in [Hma+20]. Moreover, our method has the advantage of providing arbitrary length presentations, limited only by the quality of the training dataset. The representation length can

thus be adapted to the sensitivity of the application. We present in Table 5.6 a comparison between our proposed approach and some classical binarisation methods. These classical methods were presented in [Dro+18; LT13] and benchmarked on the AR and FERET datasets. We followed the proposed approach presented in [Dro+18] for binarising the output of the CNN by quantizing the feature space and applying an encoding to the codebook obtained in the quantization step. We follow the same processing chain presented in [Dro+18] but we use our DNN features as input for the binarisation methods. The binarisation methods that we re-implemented are :

- Direct Binary Represenation (DBR), where the decimal values from the quantization are directly convected into their binary representations.
- Binary Reflected Gray Code (BRGC), similar to the DBR method, the decimal values are encoded directly to binary format using their BRGC representations.
- Linearly Separable SubCode (LSSC) [LT13], an encoding method that aims to keep the distance from the decimal space to the binary space.
- Sparse, in this scheme which is similar to one-hot encoding, the number of encoded bits per real value is equal to the number of quantization intervals, and only one bit is set to 1 per encoding.

We followed an equal-width quantization approach where the feature space is divided into intervals of the same size.

In our comparison, we used the same output lengths for each of the systems as in [Dro+18]. Furthermore, we also adapted the schemes to obtain 1 024 bits for all the methods, mainly by changing the number of quantization intervals. For example, for DBR to obtain representations with 1 024 bits, we used 256 quantization intervals to obtain a DBR representation on 8 bits for

each real value. BRGC, LSSC, and Sparse were quantized over 256, 9, and 8 intervals, respectively.

As shown in Table 5.6, our approach gives better recognition performance than the classical methods. Furthermore, the entropy of our approach is higher than the classical approaches presented. For example, LSSC shows the best performance among the studied classical binarisation approaches with 98.62% accuracy on LFW compared to an original baseline of 99.2%. Thus, the recognition degradation of this approach is minor. However, it provides less than half the entropy provided by our binarisation approach. In addition, some methods show significant degradation in the performance when using longer representations (such as DBR) and, as such, limiting the length of the representation. BRGC and Sparse, and especially DBR suffer from degradation in performance when increasing the length of the representations. We attribute the degradation in performance for DBR to two factors, first, the high number of quantization intervals; second, the fact the DBR code does not conserve distances as opposed to LSSC and BGRC. On the other hand, our method keeps the system's performance even with much longer representations as we do not need to change the number of quantization intervals by increasing the number of neurons in the bottleneck layer in the auto-encoder; we can increase the length of the binary representation.

In the following section, we provide a use-case of the binary representations consisting of a cancelable face verification system.

5.4 Application to Cancelable Biometrics

Biometrics systems are strongly associated with identity, and therefore, biometric recognition creates a strong link between the user's identity and the

Table 5.6: Performance of the classical binarisation methods on the LFW dataset. The binarisation methods are applied to the output of our version of OpenFace CNN trained on the **cleaned** version of MS-celeb-1M. The entropy of the methods is computed using the same approach presented previously.

Encoding	**Length (bits)**	**Accuracy on LFW (%)**	**Entropy**
Euclidean representation (OpenFace)	128 floats	99.22	$\sim$
DBR	256	97.28	253.23
	1024	84.25	650.50
BRGC	256	97.37	146.04
	1024	96.17	561.74
LSSC	348	97.38	148.60
	1024	98.62	409.03
Sparse	512	96.93	275.31
	1024	94.35	418.67
Ours	1024	**99.12**	**944.24**

authenticator. However, many privacy concerns are being raised about biometrics. Since biometric characteristics are permanently associated with the person, they cannot be replaced in case of compromise. This lack of revocability is a serious issue for user authentication systems. Moreover, biometric templates originating from the same biometric characteristics stored in different databases are similar. Therefore, biometrics lack diversity, and two biometric databases can be cross-linked, compromising the user's privacy. Recovery of biometric data from the biometric references and possibly revealing physical conditions bring additional privacy issues with biometric systems.

Cancelable biometrics is proposed in order to address these problems. It consists of transforming the original biometric template to obtain a cancelable biometric reference that can be revoked. Therefore, when a biometric template is compromised, it can be canceled and replaced.

5.4.1 Cancelable System Requirements

There are some main criteria which a cancelable biometric template should satisfy:

- **Performance**: the cancelable biometric system should not degrade the verification performance of the underlying baseline biometric system;
- **Revocability**: if the protected biometric template is stolen, it should be possible to revoke that template and re-issue a new one;
- **Diversity**: is the maximum number of independently protected templates that can be created from one biometric sample;
- **Irreversibility**: it should be computationally infeasible to obtain the original biometric template from the protected template;
- **Unlinkability**: the protected biometric templates created from the same biometric sample using two different secret keys should not be linkable.

In the following subsection, we present and evaluate the performance of the biometric protection scheme applied to the binary representations created using our binarisation method. In the following evaluation, we use the terminology of the ISO/IEC 24745:2011 [ISO11]. We use **PI** to denote the Pseudonymous Identifier and **SD** for Supplementary Data. According to the ISO/IEC 24745:2011, SD is data intended for security amplification of renewable biometric references by means of possession, knowledge or application-based secrets that are both required during enrolment and verification and are not stored with biometric references nor dependent on biometric characteristics, that are either provided by the data subject or the identity management system.

In order to be coherent with the SO/IEC 24745:2011 standard, we will be using the same vocabulary. The ISO/IEC 24745:2011 standard defines the architecture of biometric protection systems. The architecture is based on three important elements:

- **Pseudonymous Identifier Encoder (PIE)**: During enrolment, the PIE generates a cancelable biometric template based on the pseudonymous identifier (PI) and supplementary data (SD).
- **Pseudonymous Identifier Recorder (PIR)**: During verification, the PIR generates a pseudonymous identifier (PI*) based on the SD provided during enrolment and the biometric sample.
- **Pseudonymous Identifier Comparator (PIC)**: compares the PI created in the enrolment phase and PI* and returns a score.

5.4.2 Proposed Cancelable System

To protect the template, we apply the shuffling scheme proposed by Kanade *et al.* in [KPDD12]. The shuffling scheme (shown in Figure 5.6) uses a binary shuffling key. Since this key is a long bit-string, it is stored on a secure token, or it can be derived from a password. The binary embedding is divided into blocks of the same length. Two distinct parts are created: the first part contains all the blocks corresponding to the positions where the shuffling key bit value is "1". All the remaining blocks are taken into the second part. These two parts are concatenated to form the shuffled binary embedding, treated as the protected template. The original and shuffled templates have a one-to-one correspondence. A block from the original vector is placed at a different position in the shuffled embedding. When two binary embeddings are shuffled using the same shuffling key, the absolute positions of the blocks change,

but this change occurs in the same way for both of the representations. As a result, the Hamming distance between them does not change. On the other hand, if they are shuffled using two different keys, the result is a randomization of the representations, and the Hamming distance increases.

For this use-case, we chose a block size of "1" compared to "7" in [KPDD12]. This has two main advantages. First, the size of the shuffling key will be longer, thus harder to brute-force. Secondly, the permutation space becomes bigger, allowing for a higher number of possible templates. The shuffled binary embedding, which is the cancelable template, is the result of combining the biometric sample and the Supplementary Data (**SD**) (the shuffling key in our case). Therefore, it can be revoked in case of compromise, and a new template can be generated by changing the shuffling key. In our case, we chose a block size of "1" with a shuffling key of size 1 024. The shuffling keys can be either generated and stored in the Secure Element or derived from the password using, for example, a Password Based Key Derivation Function (PBKDF) such as PBKDF2[1].

A unique shuffling key is assigned to each user during enrolment, and he/she has to provide that same key during every subsequent verification. This means, in an ideal case, that the genuine users always provide the correct shuffling key.

Figure 5.7 illustrates the architecture of the cancelable system. According to the ISO/IEC 24745:2011, the system falls under **Model G** where "Store distributed on token and server, compare on server". The token in our case being

[1] PBKDF2 is part of RSA Laboratories' Public-Key Cryptography Standards (PKCS) series, specifically PKCS #5 v2.0, also published as Internet Engineering Task Force's RFC 2898

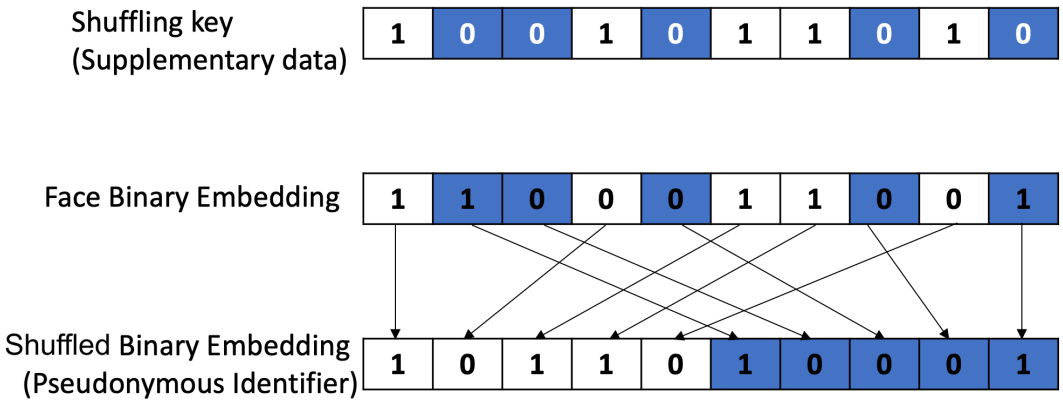


Figure 5.6: Shuffling scheme with block size of "1" bit.

the Secure Element (SE), contains the shuffling key and the Common Identifier (CI) to allow checking for the integrity of the data. The CI is an identifier for correlating identity references and biometric references in physically or logically separated databases. The communication between the SE and the client as well as the communication between the SE and the server is done using asymmetric cryptography. To be precise, RSA 2 048 keys are used to secure the communication. The data storage is distributed between the token and the server. At each verification attempt, the client asks the token, the SE, for the shuffling key. Afterwards, it generates the **PI*** and sends it to the server. Then, the server measures the distance between the stored **PI** and the **PI*** and decides based on predefined threshold the outcome of the verification. In the case where the user opts not to use the SE, instead, he/she chooses to provide a password at each access attempt. Then as the storage and the comparisons will be done on the server side, the system follows the **Model A** architecture of the ISO 24745 "store on server and compare on server using RBRs" .

According to the results reported in Table 5.7, a binary embedding of 1 024 bits gives the best trade-off between size and performance. As such, all subsequent evaluation analyses are carried out using 1 024-bit representations.

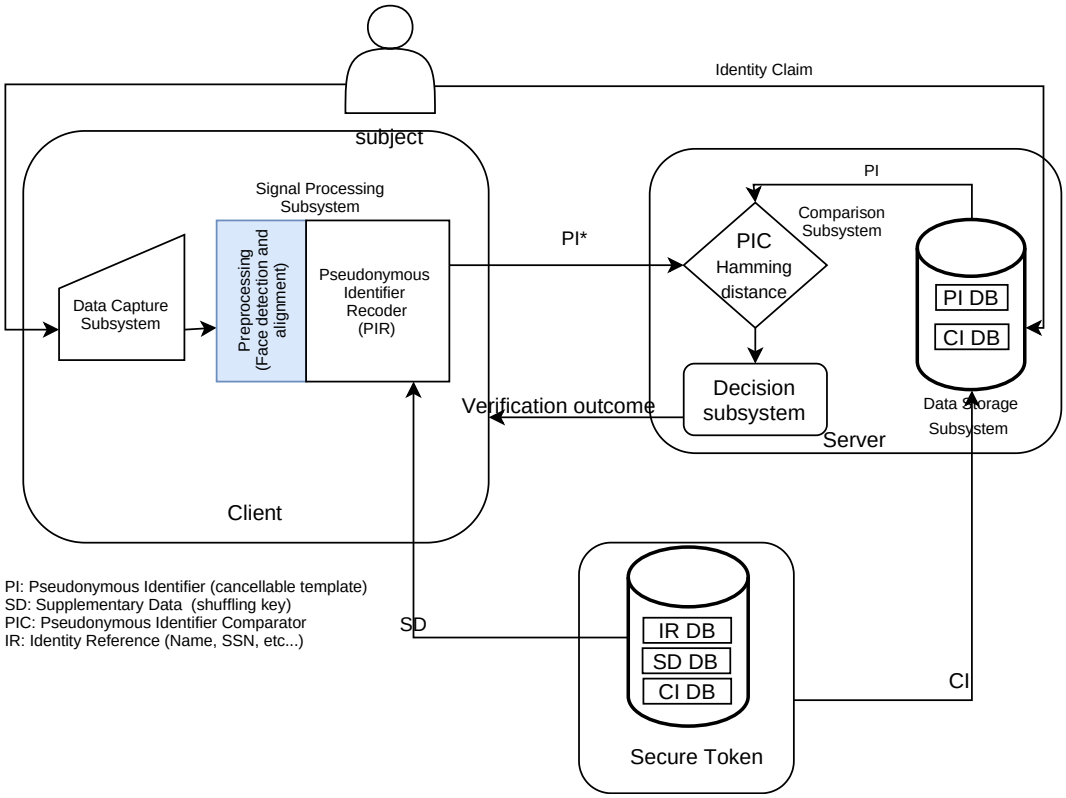


Figure 5.7: Proposed cancelable biometric scheme.

Biometric Recognition Performance

The performance of the verification system is an important point that must not be degraded by the transformation scheme. Therefore, for a fair comparison, first, the biometric verification performance of the baseline biometric system should be evaluated, then the performance of the proposed cancelable biometric system. It is necessary to evaluate the system performance when one of the two factors is compromised. Hence, two impostor scenarios are considered:

- Stolen biometric data: when the biometric data for the user is compromised. Here an impostor will try to provide the stolen biometric data with the wrong **SD**;

Length	Accuracy on LFW %		Accuracy on MOBIO %	
128*	98.32	**98.00**	99.72	**99.67**
128	98.27	**98.82**	99.88	**99.67**
256	99.68	**99.22**	99.91	**99.88**
512	100	**99.80**	100	**100**
1024	100	**99.88**	100	**100**
2048	100	**99.88**	100	**100**
4096	100	**99.77**	100	**100**

Table 5.7: Impact of the length of the shuffled binary representations obtained following approach (b) (using a pre-trained CNN with an auto-encoder) on the recognition performance. Values in **bold** are given by DNN models trained using the cleaned version MS-celeb-1M. The results in the second row (row '128*') are obtained by applying a median binarisation on the output of the initial OpenFace DNN.

- Stolen Supplementary data: when the **SD** of the user is compromised. Here an impostor will try to provide erroneous biometric data with the stolen **SD**.

The biometric recognition performance of the system is reported in Table 5.7. The performance of the system is improved compared to using non-shuffled representations. Moreover, we obtain better overall performance for the shuffling when using our proposed binarisation method compared to using median threshold as shown in the first and second row of Table 5.7. We also note that thanks to the fact that we can control the length of the generated binary representation, we can improve the recognition performance by using longer representations.

For the stolen biometric scenario, the system has a False Acceptance Rate (FAR) of 0%. This point is further developed in the unlinkability analysis. Therefore, the protected biometric templates created from the same biometric sample using two different secret keys should not be linkable, which is the same as using a compromised biometric sample with a different key.

As for the stolen **SD** scenario, the performance of the system reverts to the case of non-shuffled representations shown in Table 5.4. In fact, the attacker can revert to the original binary representation if they have access to the PI and SD. As such, this scenario is reduced to the case of regular biometric verification task using the binary representations. As a result, the performance of the system in this scenario is the performance of the system presented in Table 5.4.

Diversity

It is necessary to calculate the maximum number of pseudonymous identifiers (a pseudonymous identifier (**PI**) is a part of a renewable biometric reference that represents an individual or data subject) that can be generated. After that, unlinkability and irreversibility analysis should be done as a function of **PI** issued. In the case of the previously described shuffling scheme, the maximum number of **PI** is given using the number of possible permutations. Moreover, because the decision-making is based on a threshold comparison, we should not account for templates falling in the same neighbourhood. We estimate the maximum number of templates using the hamming-packing bound. Using a threshold $t = 0.2$, for binary representations of length 1 024 we get around 2^{659} possible **PI** for each user.

$$\begin{aligned} Number\ \ Of\ \ PI &= \frac{number\,of\ possible\,permutation}{volume\,of\ Hamming\,spheres} \\ &= \frac{2^n - n}{\sum_{k=0}^{t\times n/2} \binom{n/2}{k}^2} \\ &= \frac{2^{1024} - 1024}{\sum_{k=0}^{0.2\times 512} \binom{512}{k}^2} \approx 2^{659} \end{aligned} \tag{5.5}$$

Irreversibility

There are two types of irreversibility analysis. The first type is to analyse whether we can revert to the original template given the **SD**. The second analysis is the analysis of the protected template without having the **SD**. As the applied transformation is a shuffling of the bits of the embedding without a loss of information, given the second factor, the scheme is fully reversible. However, without access to the second factor and prior knowledge about the distribution of the non-shuffled templates, it is computationally not feasible to revert to the original binary embedding as the number of permutations to be tested which is equal to $2^{1024} - 1024 \approx 2^{1018}$ is too big to be brute-forced.

Unlinkability

For this metric, we follow the methodology defined in [GB+17]. Two types of score distributions will be analysed for the assessment of the unlinkability provided by the protected templates:

- **Mated instances:** scores computed from templates extracted from different samples of the same subject using different keys.
- **Non-mated instances:** scores obtained from templates generated from samples of different subjects using different keys.

As described in [GB+17] two measures are computed, $D_{\leftrightarrow}(s) \in$ [0,1] gives an estimation of the linkability of a system for a specific score s, and $D_{\leftrightarrow}^{sys} \in$[0,1] gives an estimation of the linkability of a system as a whole, independently of the score. If for a specific score s_0 $D_{\leftrightarrow}(s_0) = 0$, this means that the system is fully unlinkable for this particular score. Also, if $D_{\leftrightarrow}^{sys} = 0$ where both score distributions (mated and non-mated) are overlapping means that the system is fully unlinkable for the whole score range. The computation of $D_{\leftrightarrow}(s)$ and

$D_{\leftrightarrow}^{sys}$ depends on the prior probability ratio ω of the mated and non-mated distributions, which may result in $D_{\leftrightarrow}^{sys} = 0$ even if the distribution are not perfectly overlapping. In this analysis, we specify $\omega = 0.2$.

As observed in Figure 5.8, the distribution of mated and non-mated scores overlap, thus making the function $D_{\leftrightarrow}(s)$ identically-zero over the range of the possible scores. In addition $D_{\leftrightarrow}^{sys}$ is equal to *0* rendering the system fully unlinkable. The scores used to estimate the probabilities are computed using the whole LFW dataset of 5 749 users. For each user, we generate 50 different shuffling keys and thus 50 protected templates. By considering the whole population of the LFW dataset, we get around 14M mated scores and 80 000M non-mated scores. To have the same number of samples from each population, we sample uniformly 10M mated scores and non-mated scores. Hence, $D_{\leftrightarrow}(s)$ and $D_{\leftrightarrow}^{sys}$ are estimated in the mean case and do not take account of user-specific distributions. Based on $D_{\leftrightarrow}(s)$ and $D_{\leftrightarrow}^{sys}$ we conclude that the proposed system is fully unlinkable for the whole score range.

The diversity, irreversibility, and unlinkability metrics are tightly correlated. If the system can not satisfy the diversity requirement, and as such, can not create different **PI**s using the same biometric data with different **SD**s, then the identities will be linkable. Furthermore, if the irreversibility requirement is not satisfied, the templates can be linked. Finally, if the system is fully linkable, then it does not satisfy the diversity requirement as all the generated **PI**s are equal. Furthermore, even if the system is not fully linkable and only partially unlinkable, it will result in easier attacks on the original templates.

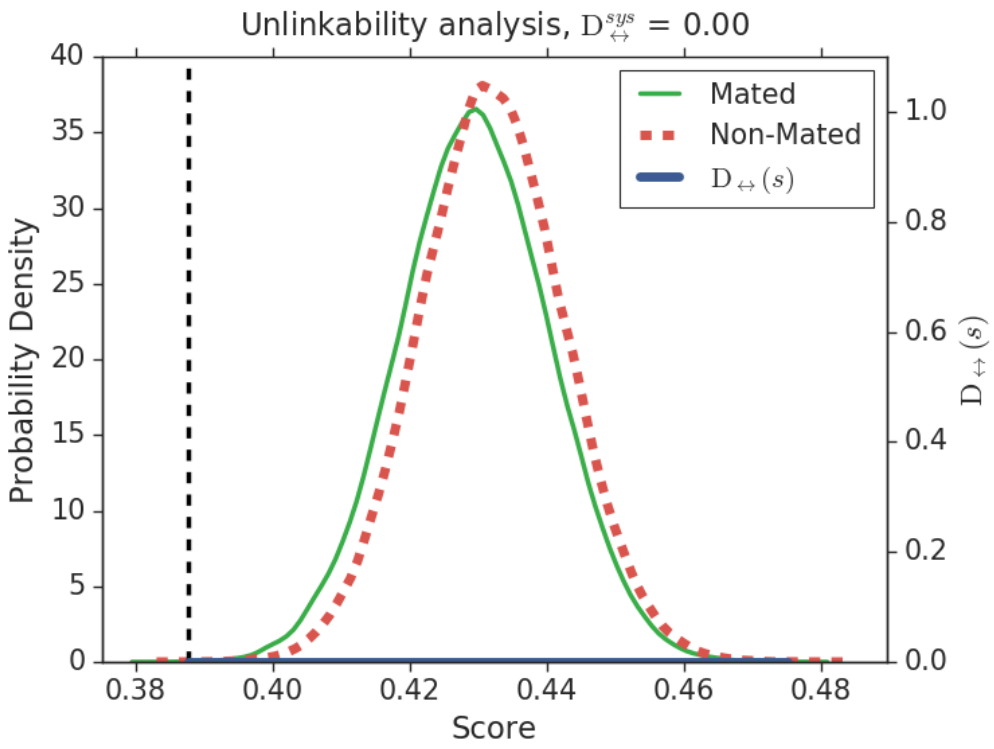


Figure 5.8: Unlinkability analysis of the system based on scores computed on the LFW dataset. Templates used are of length 1 024. The templates are obtained using DNN created corresponding to approach (b) (using a pre-trained CNN with an auto-encoder) and trained on the **cleaned** version of MS-celeb-1M.

5.5 Implementation in the SpeechXRays Project

The H2020 SpeechXRays project aims to achieve this privacy requirement by implementing a cancelable biometric system. Using a shuffling transformation on the binary embeddings extracted from face images combined with a shuffling key, the users templates are made cancelable and unlinkable to the users in the same time. We explain how the system follows the ISO/IEC 24745:2011 compliance recommendation, and we report its performance and evaluate its properties following the ISO standardized metrics, notably the system irreversibility and its unlinkability. When working under ideal circumstances (the second factor is not stolen), the system gives 100% accuracy

on the MOBIO dataset. Moreover, it is fully unlinkable and it is computationally infeasible to recover the original template without the second factor.

5.5.1 The Cancelable Face System Prototype

In the prototype, we are using Gabor feature vectors of size 40 960. The dimension of the feature vectors is reduced from 40 960 to 2 048 components using Principal Component Analysis (PCA). To binarise the vectors, we apply thresholding. The thresholding is done based on the median value of the vectors. From a feature vector $X = (x_1, ..., x_{2048})$, we obtain a binary embedding $X_{bin} = (b_1, ..., b_{2048})$ by comparing each component to the median value of the vector.

$$b_i = \begin{cases} 0, & \text{if } x_i \leq median(X) \\ 1, & \text{otherwise} \end{cases} \qquad \text{for } i \text{ in } (1, ..., 2048)$$

The median value of vector X is computed for each feature vector separately. The result is a binary representation with an equal number of *ones* and *zeros* for each embedding.

To protect the template, we apply the shuffling scheme described in [KAN10].

5.5.2 Evaluation of the System

A workforce use case is one of the main use cases of the SpeechXRays project. In this use case, the workers need to access workstations, computers and handheld devices to carry out their tasks. We chose the MOBIO dataset to evaluate the performance of the system as it illustrates this use case by providing samples from mobiles and computers with realistic conditions. The

MOBIO [McC+12] database contains audio visual data recorded from 152 subjects. The database comprises 52 females and 100 males. In total, there are 12 sessions for each individual. The database is divided into three parts: training, development and evaluation. We report the results on the protocol described in [Bou16]. The results are reported separately for males and females because for speaker recognition separating males from females is a common practice. Therefore, the face recognition experiments follow the same protocol. The development partition consists of 1 890 true and 32 130 false access scores for female trials and 2 520 true and 57 960 false access male trials. The evaluation partition consists of 2 100 true and 39 900 false attempts for female trials and 3 990 true and 147 630 false access scores for male trials.

Biometric Performance

In Table 5.8, we report the performance on the MOBIO dataset. The DLDA approach constitutes the baseline of the comparison. In fact, the protected templates are created from the Gabor features used in the DLDA representation, hence to be able to measure the degradation of the performance, we compare the cancelable system to the DLDA system. Certainly, the DNN approach for face recognition give better performance than the DLDA approach. Nevertheless, the cancelable biometrics scheme takes advantage of the second factor providing better recognition performance. It is only in the case of stolen shuffling key that the performance degrades compared to the baseline (and the DNN). Figure 5.11 shows the impact of the shuffling on the scores. The true access attempts remain unchanged by the shuffling, whereas the mean of imposter scores is moved to around 0.5. As a consequence, the client and imposter distributions are separated, allowing for better decision

Table 5.8: Performance on MOBIO dataset, EER is computed using 42 000 tests for the female partition and 151 620 for the male partition.

	EER on Eval Female partition (%)	EER on Eval Male partition (%)
DNN [HPD18]	5.26	2.97
DLDA (Baseline)	12.43	7.86
Gabor Binary Representation	17.18	13.48
Shuffled Gabor Binary Representation	0	0
Stolen Biometrics	0	0
Stolen Second Factor	17.18	13.48

making. The separation of the two distributions allows for perfect verification performance for the system, as shown in Figure 5.9. However, in the case of a stolen shuffling key, the performance of the system regresses as if the shuffling was not applied. Thus the performance in case of the stolen second factor is the same as the non-shuffled binary representations. On the other hand, if only the biometric sample is stolen, the system will not grant access. As will be explained when measuring the unlinkability of the system, the system is fully unlinkable, hence, with a different second factor, the biometric data will be considered as if coming from another user.

Diversity

It is necessary to calculate the maximum number of PI that can be generated. After that, unlinkability and irreversibility analysis should be done as a function of PI issued. In the case of the previously described shuffling scheme, the maximum number of PI is given using the number of possible permutations. Moreover, because the decision making is based on a threshold comparison, we should not account for templates falling in the same neighborhood. We estimate the maximum number of templates using the hamming-packing bound. Using a threshold $t = 0.05$, for binary representations of length 2 048

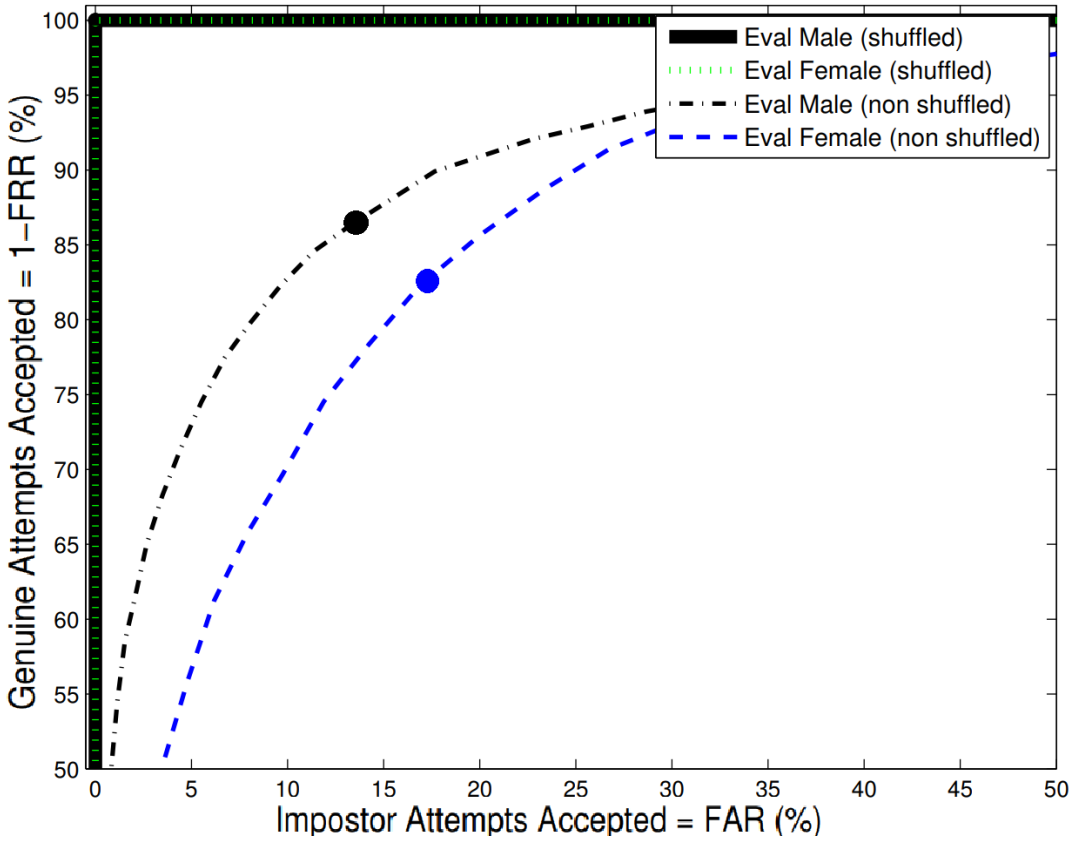


Figure 5.9: ROC curve for performance of the protected and non-protected systems on the MOBIO dataset.

we get almost 2^{1759} possible **PI** for each user.

$$Number\ \ Of\ \ PI = \frac{2048! - 2048}{\sum_{k=0}^{0.05*1024} \binom{1024}{k}^2} \approx 2^{1759} \tag{5.6}$$

Irreversibility

Without access to the second factor and prior knowledge about the distribution of the non-shuffled templates, it is computationally not feasible to revert to the original binary embedding as the number of permutations to be tested which is equal to $2^{2048} - 2048 \approx 2^{2048}$. The NIST SP 800-152 standard recommends 192 bits of entropy for applications with high impact [BSB14]. Thus recovering 2048 bits is computationally infeasible. We also note that due to the thresholding step in the binarisation process, it is not computationally feasible to recover original Gabor feature vectors.

Unlinkability

For this metric, we follow the methodology defined in [GB+17]. Two types of score distributions will be analysed for the assessment of the unlinkability provided by the protected templates:

- **Mated instances:** scores computed from templates extracted from different samples of the same subject using different keys. It represents the probabilities $p(s|Hm)$.
- **Non-mated instances:** scores yielded by templates generated from samples of different subjects using different keys. It represents $p(s|Hnm)$.

As described in [GB+17] two measures are computed, $D_{\leftrightarrow}(s) \in [0,1]$ gives an estimation of the linkability of a system for a specific score s and $D_{\leftrightarrow}^{sys} \in [0,1]$ gives an estimation of the linkability of a system as a whole, independently of the score.

If for a specific score s_0 $D_{\leftrightarrow}(s_0) = 0$, this means that the system is fully unlinkable for this particular score. Also, if $D_{\leftrightarrow}^{sys} = 0$ this means that the system is fully unlinkable for the whole score range. As observed in Figure 5.10, the distribution of mated and non-mated scores overlap, thus making the function $D_{\leftrightarrow}(s)$ identically-zero over the range of the possible scores. In addition $D_{\leftrightarrow}^{sys}$ is equal to *0* rendering the system fully unlinkable. The scores used to estimate the probabilities $p(s|Hm)$ and $p(s|Hnm)$ are computed using the whole MOBIO dataset of 150 users. For each user, we generate 1 000 different shuffling keys and thus 1 000 protected templates. By taking into account the whole population of the MOBIO dataset, we get 150 000 000 mated score and 1 500 000 000 non-mated score. Hence, $D_{\leftrightarrow}(s)$ and $D_{\leftrightarrow}^{sys}$ are estimated in the mean case and do not take account of user-specific distributions. However, thanks to the construction scheme of the binary representations making sure that independently of the user, each embedding contains exactly 1 024 *zeros* and 1 024 *ones* the user-specific distribution is not different from the whole population distribution. Based on $D_{\leftrightarrow}(s)$ and $D_{\leftrightarrow}^{sys}$ we conclude that the proposed system is fully unlinkable for the whole score range.

We report the cancelable biometric system implemented in the SpeechXRays based on binary representations and shuffling keys for face verification. The proposed system improves the verification performance and satisfies the evaluation criteria of the template protection, mainly the revocability and unlinkability. Due to the shuffling scheme, the protected face template satisfies the requirement of revocability. The proposed system can generate different

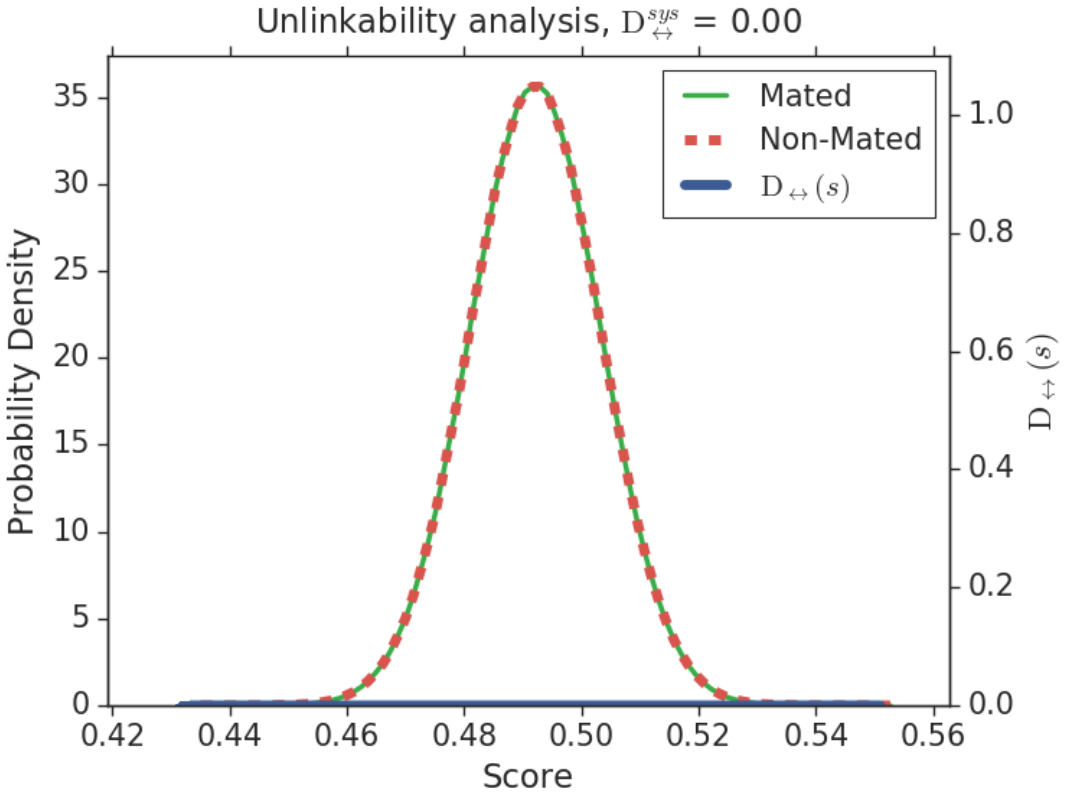


Figure 5.10: Unlinkability analysis of the system based on scores computed on the MOBIO dataset.

templates for various applications using the same biometric sample, which conserves privacy. If the stored shuffled template is stolen, the administrator can cancel the old enrolled template and issue a new one by changing the shuffling key. Besides, the proposed system achieves excellent recognition performance with an EER of 0% on the MOBIO dataset. The binary embedding being extracted directly from the Gabor feature vectors suffer from low accuracy without the additional security provided by the shuffling key. This results in having degraded performance when the second factor is stolen compared to the state of the art systems in face verification. A data-driven binary embedding extraction approach should improve the performance of the system, especially in the event of a stolen shuffling key.

5.6 Impact of the Performance of the Binary Representations on the Cancelable System

In addition to the evaluation criteria proposed by the ISO/IEC 24745:2011 standard [ISO11], in the case of cancelable biometrics, one should check if the security of the system is only based on the second factor. Cancelable systems tend to rely on the second factor ignoring the biometric component, which is one of the shortcomings of cancelable biometrics as shown in [RU11; Kon+06]. In fact, for the used shuffling scheme, if all the users have the same initial representation, after shuffling, we obtain 100% verification accuracy. Thus, the protection scheme based on shuffling benefits greatly from the security of the second factor. However, the combination of the binarisation method we propose with the shuffling scheme constitutes a system that relies on biometrics as well as on the second factor. This is especially shown in the difference between the systems trained on the original and cleaned

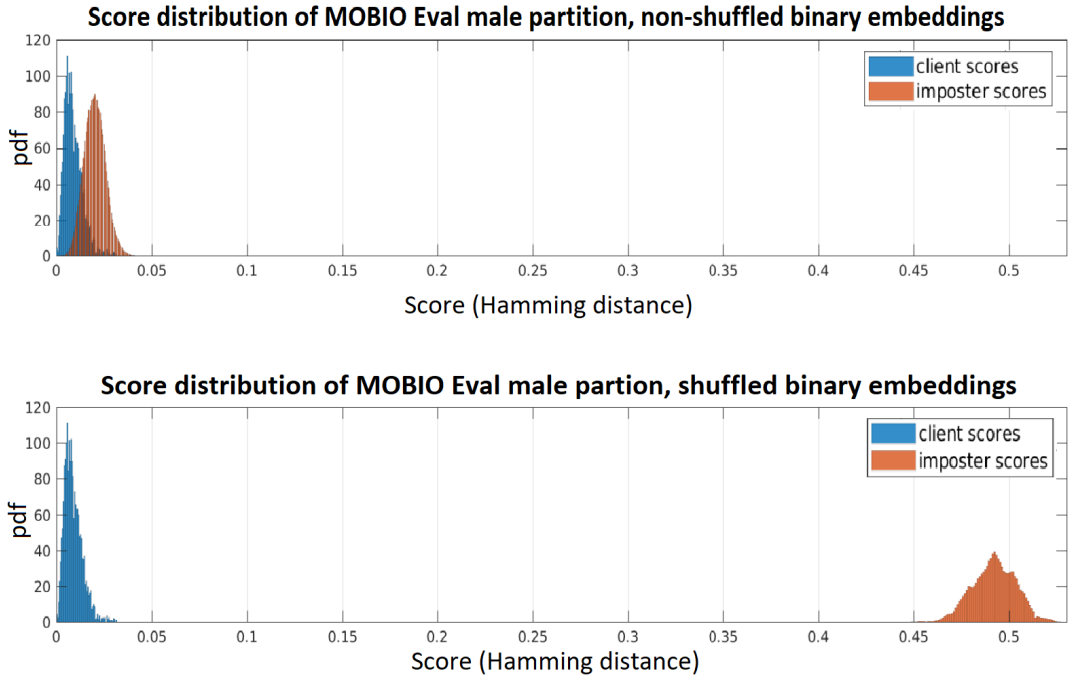


Figure 5.11: Score distribution for shuffled and non shuffled representations of MOBIO Eval male partition.

version of MS-celeb-1M. The degradation in performance of the cancelable system shown in Table 5.7 when the training is done on the cleaned version of the MS-celeb-1M is, in fact, due to the bad quality of the images used in the tests. The system should not accept these images because the face is obstructed, distorted, or not present. When the binary embedding extractor is trained on the cleaned data set, the system rejects client-client tests where either the enrolment or probe samples are of low quality. On the other hand, the version trained on the original version of MS-celeb-1M (non-cleaned) accepts these images because the verification is done using the second factor, not the biometric reference. Examples of the images with bad quality are presented in Figure 5.13b. The test scores from these images are circled in red in Figure 5.12 (Hamming Distance $>$ 0.4). The face image samples are taken from the MOBIO dataset. The images of bad quality, such as those presented in the figure, are not accepted by the cancelable system based on binarised

DNN embeddings trained on the **cleaned** version of MS-celeb-1M. The system is intended to work with images such as those in Figure 5.13a.

This shows that the system considers the biometric information and does not only focus on the second factor. As the system trained on the cleaned version of MS-celeb-1M rejects images of the same user of low quality, it does not rely solely on the second factor.

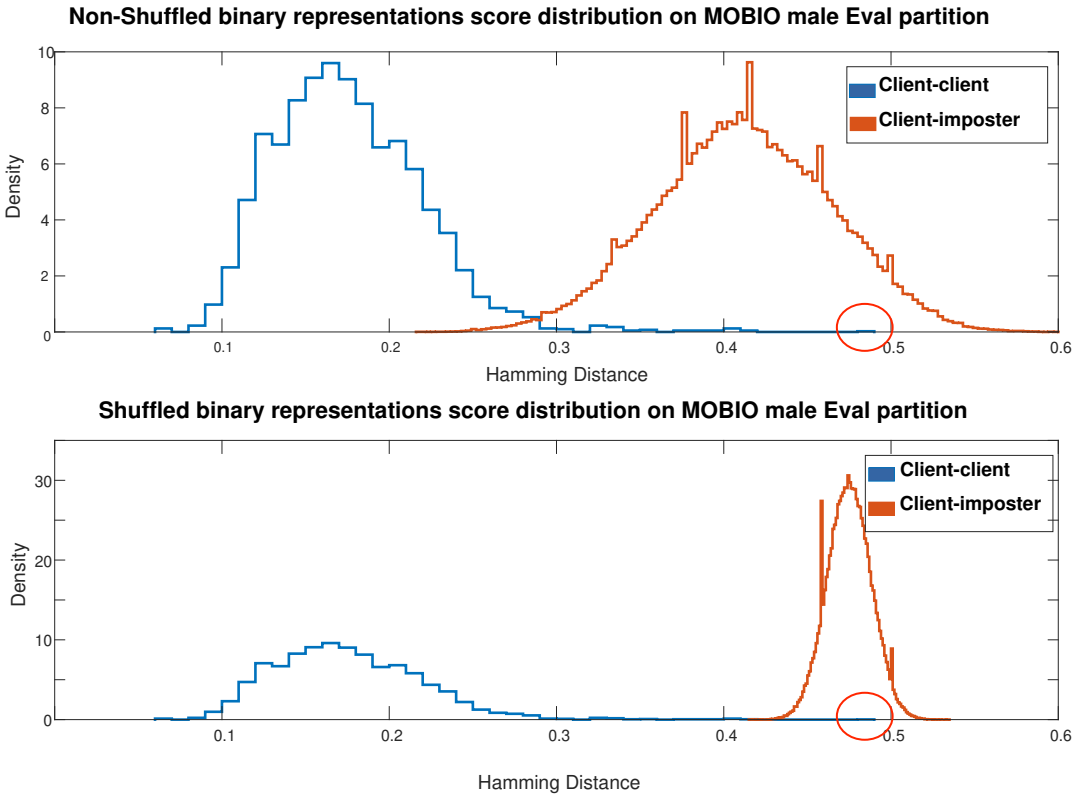
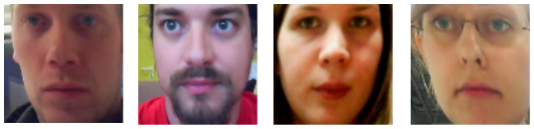
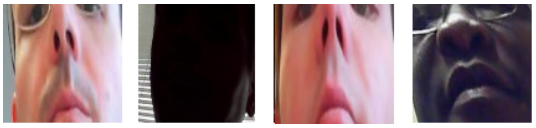


Figure 5.12: Impact of the shuffling on the score distribution of the data. Score distribution from templates of length 1 024. The templates are obtained using the DNN corresponding to approach (b) (using a pre-trained CNN with an auto-encoder) and trained on the **cleaned** version of MS-celeb-1M.



(a) Examples of face images with good quality



(b) Examples of face images with bad quality

Figure 5.13: Face image samples taken from the MOBIO database. Face detection is done using the OpenCV SSD face detector. Alignment is done using DLIB 68 points landmark detector.

5.7 Conclusion

In this chapter we present a novel approach to extract binary embeddings directly from face images using a Deep Neural Network. We followed a data-driven approach to binarise the embeddings based on using auto-encoders under supervised training with the 'Triplet loss' loss function.

The binary embeddings are analyzed in terms of biometric recognition performance and entropy. The performance is evaluated on the LFW and MOBIO databases. The degradation in performance on both databases is around 0.1%.

We obtain 99.12% accuracy on the LFW database, using the binary representation, compared to 99.22% accuracy using the baseline system. The same applies to the MOBIO database, where we get 98.90% accuracy using the binary embeddings compared to an accuracy of 98.93% of the baseline system. Using DNN to extract the binary embeddings results in representations with high entropy and high recognition performance. Compared to the baseline Euclidean representations, the proposed binary embeddings give a state-of-the-art performance on both databases with almost negligible degradation.

The approach proposed in this chapter can be applied to any continuous representation, not only Euclidean face representations. Moreover, the binarisation technique constitutes a locality-preserving hash where the relative distance between the input values is preserved in the relative distance between the output hash values. The representation can be used for multiple applications such as similarity search, database search, and biometric systems.

Furthermore, the binarisation method provides arbitrary length presentations that are limited only by the quality of the training database. The embedding length can thus be adapted to the sensitivity of the application. In addition, we compared our binarisation approach to some classical binarisation methods presented in [Dro+18] and show that our method has better biometric recognition performance and higher entropy than the presented methods.

The binary embeddings are also used to create a cancelable face recognition system based on a shuffling transformation using a second factor. The cancelable system is analyzed according to the standardized metrics given by the ISO/IEC 24745:2011. We show that the cancelable system gives high accuracy and unlinkable templates when the second factor is not compromised. When the second factor is compromised, the system's security is assured by the recognition performance of the binary representations, which is comparable to the baseline non-binarised system. Furthermore, the quality of the binary representations impacts the behavior of the cancelable system. If the discriminative power of the representations is low, the cancelable system depends mainly on the second factor, which results in higher FAR.

These representations are meant to be used in a crypto-biometric key regeneration scheme based on fuzzy commitment. This is why we seek to obtain long binary representations with high entropy.

6 Crypto-biometric Key Regeneration

6.1 Introduction

Obtaining cryptographic keys using biometrics is a remarkable concept because it offers a distinct advantage over classical methods of generating cryptographic keys. Classical cryptographic systems rely on identifiers such as passwords or tokens, that are assigned to the users by system administrators in order to authenticate the user and generate secure keys for that user. However, these assigned secrets have their own disadvantages as they can be stolen or shared, and hence, are insufficient to prove the user's identity. Using biometrics to obtain crypto-biometric keys can provide a better solution as far as identity verification is concerned.

Biometrics can be employed for obtaining crypto-bio keys in different ways, such as cryptographic key release, key generation, and key regeneration. In this chapter, we present the work carried out on key regeneration. We focus on the regeneration of symmetric keys. The schemes used can be extended to asymmetric encryption by regenerating the private key of the key pair.

Encryption systems are threatened by quantum computing algorithms. For symmetric keys, Grover's algorithm reduces the entropy of brute force attacks by half. Meaning the complexity of a brute force attack on a symmetric

key is reduced from 2^N to $2^{N/2}$. As for asymmetric encryption, Shor's algorithm is able to break the most common encryption schemes used currently, such as RSA and elliptic curve, instantly.

The Grover algorithm is a quantum algorithm that can search an unsorted database more efficiently than is possible with classical algorithms. It can search a database of N items in $O(\sqrt{N})$ time, which is a significant improvement over the $O(N)$ time required by classical algorithms.

While the Grover algorithm is not specifically designed to break symmetric encryption algorithms, it could potentially be used to break certain types of symmetric key cryptography if the key size is small enough. This is because the Grover algorithm can be used to perform a "brute force" search, in which all possible keys are tried one by one until the correct key is found.

If the key size of a symmetric encryption algorithm is small enough, the Grover algorithm could potentially be used to perform a brute force search in a shorter amount of time than is possible with classical algorithms. For example, if the key size is 128 bits, the Grover algorithm could potentially find the correct key in $O(2^{64})$ time, which is significantly faster than the $O(2^{128})$ time required by classical algorithms.

As we stand right now, quantum computers still lack the power to run these algorithms to their full potential. But, the threat is there nevertheless. The day that quantum supremacy for these algorithms is achieved, the majority of encryption systems currently used will be in danger. As for how long before quantum supremacy is reached, it is uncertain (quantum supremacy is the goal of demonstrating that a programmable quantum device can solve a problem that no classical computer can solve in a feasible amount of time). However, the state of the art in quantum computing is advancing at a rapid

pace.

Vast progress toward quantum supremacy was made in the 2000s, especially in the last decade. In 2011, D-Wave Systems of Burnaby in British Columbia became the first company to sell a quantum computer commercially [Mer11]. In 2017, IBM demonstrated the simulation of 56 qubits on a classical supercomputer, thereby increasing the computational power needed to establish quantum supremacy. In December 2020, a group based in the University of Science and Technology of China reached quantum supremacy by implementing a type of Boson sampling on 76 photons with their photonic quantum computer Jiuzhang [Bal20]. The paper states that to generate the number of samples the quantum computer generates in 20 seconds, a classical supercomputer would require 600 million years of computation.

The goal of this work is to obtain post-quantum crypto-biometric keys. This goal has multiple requirements:

- To be resistant to quantum computing algorithms;
- Non-repudiation: the user cannot share his key and claim that he was not the one using it;
- To be cancelable;
- Convenience, meaning the regenerated keys has to have low False Rejection Rate (FRR) at the required security level.

As we focus on symmetric key regeneration, the keys need to have double the entropy of the keys used currently to present the same degree of security [Aug+15]. This is easy for standard symmetric keys but difficult for crypto-biometrics. Crypto-biometric keys are limited by the usable information contained in the biometric sample that they are generated from. The

non-repudiation requirement is satisfied by the intrinsic properties of biometric samples. However, we must ensure that the scheme used in the key regeneration has a low False Acceptance Rate (FAR). Biometrics are unique for each user. They can not be changed without special circumstances (plastics surgery, diseases...). As such, if the regeneration scheme is not revocable, the user will be restricted to a single key across multiple applications. In addition, in the case the key is compromised, the user will not be able to create a new one. Thus, we must ensure that the regeneration scheme is revocable. Finally, the key regeneration scheme should allow for user convenience. Meaning, at the required security level, the user should not be rejected multiple times before having access to the system. The convenience of the system is shown through the FRR metric.

This chapter is structured as follows. First, we introduce the key regeneration scheme used in the proposed system. Afterward, we provide the performance of the system. Before concluding, we present the security analysis of the system.

6.2 Key Regeneration Scheme

6.2.1 Fuzzy Commitment

The system proposed in this section is based on the fuzzy commitment scheme presented in Figure 6.1 and Figure 6.2. The fuzzy commitment scheme was first introduced in 1999 by Juels and Wattenberg [JW99]. A random key is encoded using Error Correcting Codes (ECC) and is then XORed with the biometric data. The XORed data is cryptographically secure because neither the key nor the biometric data can be obtained from it without

providing one of the two. The random key is retrieved at the time of key regeneration by providing fresh biometric data. This system requires ordered biometric data in binary form. In this scheme, the differences in the biometric data from one acquisition to another are treated as noise. This noise causes errors in the data being transmitted which are corrected using ECC.

The system is comprised of two phases. The first phase, shown in Figure 6.1 is the enrollment phase where the user generates a symmetric key and links it to his/her identity. The second phase, shown in Figure 6.2, is the verification phase where the user regenerates his/her key from his biometric data, stored helper data, and a secret second factor used to ensure the revocability of the scheme.

The revocability of the fuzzy commitment scheme is assured using the same shuffling scheme described in Subsection 5.4.2.

In the **enrolment phase**, the user provides an image containing his/her face $\boldsymbol{I}$ and a secret second factor $\boldsymbol{S}$. The second factor can be a shuffling key stored on a secure token or a password that is used to derive the shuffling key. The image $\boldsymbol{I}$ is processed using the DNN described in Section 5.2.2 to provide a binary embedding $\boldsymbol{B}$.

The binary embedding is then shuffled using the shuffling key $\boldsymbol{S}$ provided by the user to achieve the revocability requirement. The shuffled binary embedding is denoted $\boldsymbol{SB}$. The $\boldsymbol{SB}$ is used to protect the encryption key $\boldsymbol{K}$ generated by the system at the beginning of the process. The encryption key $\boldsymbol{K}$ is encoded using the error-correcting code described in following section creating an encoded encryption key $\boldsymbol{E(K)}$. The encryption key $\boldsymbol{K}$ is also hashed using a hash function to provide a hashstring that will be used to verify the successful regeneration of the encryption key $\boldsymbol{K}$ in the regeneration phase.

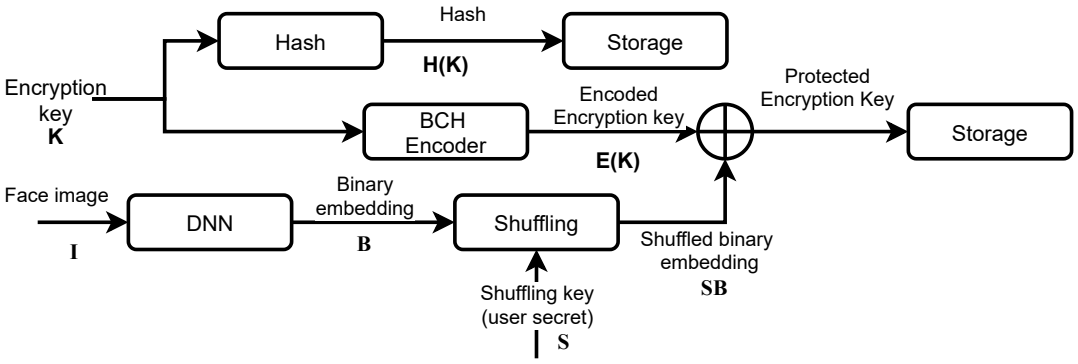


Figure 6.1: Enrolment phase of the fuzzy commitment scheme used in the key regeneration.

The ***SB*** and the encoded key ***E(K)*** are XORed to provide the protected encryption key. The security of the scheme is mainly provided by the ***SB***. In fact, the entropy of the system is the minimum between the entropy of ***SB*** and ***K***.

In the enrolment phase, the system stores the hash of the encryption key ***H(K)*** and the protected encryption key. The two pieces of data do not need to be protected and can be stored in a non-encrypted database. On the other hand, the Shuffling Key ***(SK)*** must be protected. Should the shuffling key be compromised, the user's enrolment must be revoked and a new enrolment using a different shuffling key ***SK*** need to be generated.

In the **verification phase**, the user provides a new face image sample ***I'***. This image is processed following the same method as in the enrolment phase to provide a binary embedding ***B'***. The user also provides the same second factor, either in the form of a password or a shuffling key ***S***, used in the enrolment phase. The binary embedding ***B'*** is shuffled using this second factor resulting in a shuffled binary embedding ***SB'***. This shuffled binary embdding ***SB'*** will have some differences from ***SB*** due to the variability in the face image sample provided at the beginning of the verification step. ***SB'*** is then XORed with the protected encryption key recovered from storage. The result

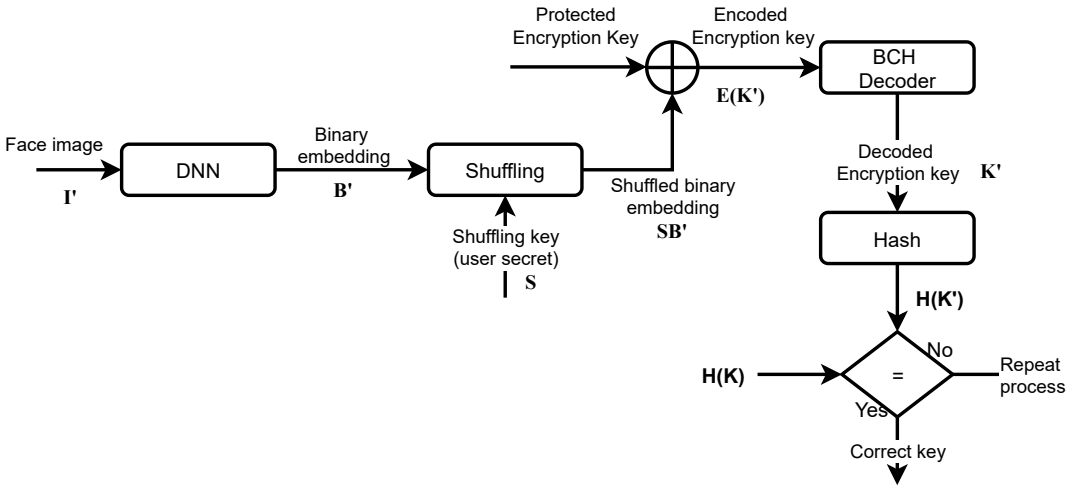


Figure 6.2: Regeneration phase of the fuzzy commitment scheme used in the key regeneration.

of the XOR operation is ***E(K')***. The difference between ***E(K)*** and ***E(K')*** is considered the noise introduced by the communication channel. ***E(K')*** is then decoded using the same error correcting code used in the enrolment phase. The decoded encryption key is denoted ***K'***. To check if the regeneration of the key is successful or not, we compare the hash of the decoded key ***H(K')*** with the hash stored in the enrolment phase ***H(K)***. If ***H(K)*** is equal to ***H(K')*** the regeneration is successful and the system provides the user with the encryption key ***K'*** which is identical to the key ***K***. If ***H(K)*** and ***H(K')*** are different the regeneration fails and the user is asked to provide a new face image.

In a real use case of the system, in order to reduce the failures, quality measures of the face image should be employed. These quality measures are discussed in Chapter 4.

The description of the fuzzy commitment scheme provided above explains only the generic data flow in the system. In our system, the binary embedding extractor (DNN) provides a fixed output of 4096 bits. To adapt the length of the binary embedding to the need of the security requirement of

the system, we added a process of bit selection. In fact, according to the parameters of the error-correcting code and the length of the encryption key ***K***, the length of the encoded encryption key ***E(K)*** will be different. Therefore, the length of ***SB*** will vary as it hides the encoded encryption key ***E(K)***. As such, the length of the binary embedding ***B*** and shuffling key ***S*** has to vary. The details about the error-correcting code used as well as the process of bit selection are provided in the following sections.

6.2.2 Bit Selection

According to the length of the encryption key, the block size used and the codeword size of the ECC, the output encoded key ***E(K)*** will have varying sizes.

For example, if the length of the encryption key ***K*** is 512, we divide it into 16 blocks of 32 bits. Each block is encoded on a codeword of 63 bits. The encoded encryption key will have 16 words resulting in a total length of 1 008 bits. If the length of the encryption key is 516 bits (a 512-bit key padded with 4 zeros), the key can be divided into 86 blocks of 6 bits. Each block will be encoded on a 31-bit codeword. The final output will have a length of 2635 bits.

However, the binary embedding extractor provides binary representations of a fixed length of 4096 bits. As the histogram in Figure 6.3 shows, not all the bits of the binary representation have high entropy. As such, we proceed to select the bits with the most entropy of the binary representations. Furthermore, we need to preserve the recognition performance of the binary representations when selecting the bits.

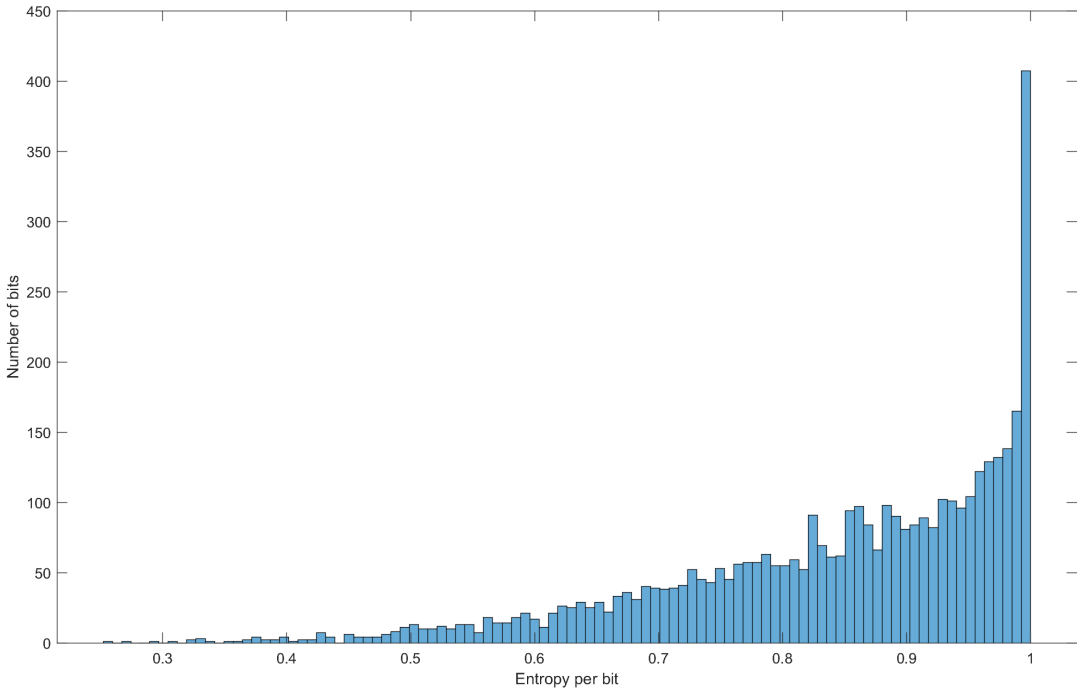


Figure 6.3: Entropy per bit of the 4096-bit binary representations.

The bit selection is made by first reordering the bits of the binary representation following the inter-class variance, intra-class variance, or both. Then, we take the first N bits needed in the fuzzy commitment scheme. The computation of the inter-class variance and intra-class variance is done on a cohort database that has no overlap with the validation.

We selected the FRGC database [Phi+05] as the cohort database because of two reasons. First, it was not acquired in the wild like LFW [LM+16] or MS-celeb-1m [Guo+16]. As such, there is a low risk of mislabeling or overlap with other databases. Secondly, it provides images of high quality in controlled conditions, which are useful for computing inter-class variance without introducing ambiguities due to the acquisition conditions. Figure 6.4 shows examples of the images used to compute the inter-class variance. The images are frontal facing with good lighting and uniform background. We also

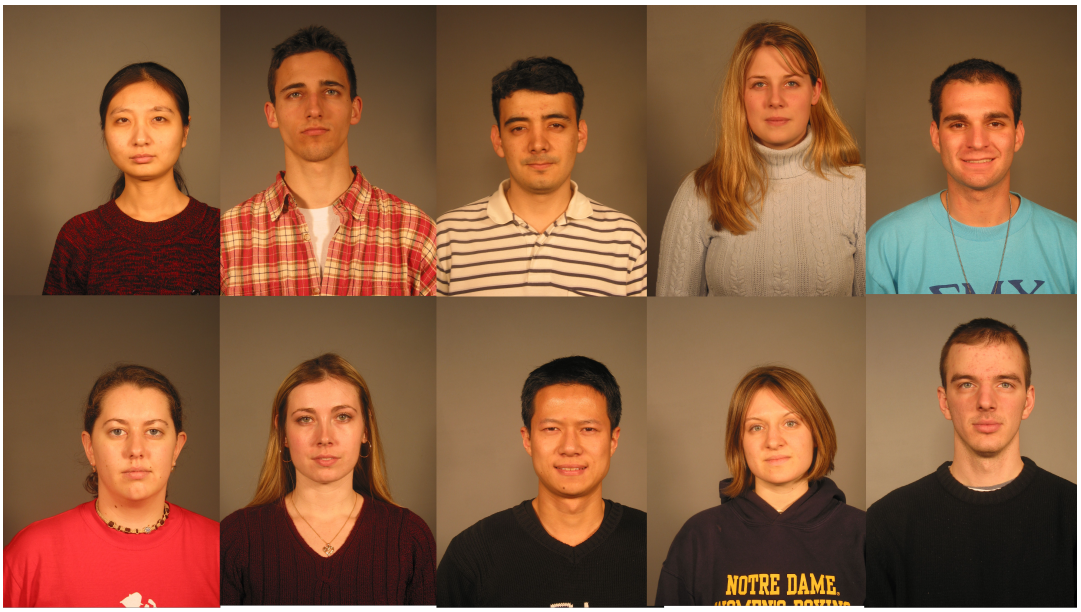


Figure 6.4: Example of the images used to compute the inter-class variance. The images are taken from the controlled partition of the FRGC database.

computed the intra-class variance using the FRGC dataset, but using the uncontrolled partition as shown in Figure 6.5.

Figure 6.6 shows the process of reordering the bits using the inter-class variance computed using FRGC. We take a binary representation of a controlled sample from each subject in the database. Then we compute the variance for each component (column). We then reorder variance using the descending order. Following this process, we store the indices following the new order for future use.

In the key regeneration system, each time the binary representation is extracted, we use the stored indices to select the suitable number of bits for the parameters of the system. Using these indices, we can select the N bits needed for the scheme. The selected bits will be the bits with the highest entropy from the binary face representations. This is under the assumption that there is no session noise when the data was acquired. That is why we only computed the inter-class variance using samples only from the controlled

Figure 6.5: Example of the images used to compute the intra-class variance. The images are taken from the uncontrolled partition of the FRGC database.

partition.

As for the computation of the intra-class variance, we used the uncontrolled partition of the FRGC database. In this case, we take all the samples pertaining to each user. Then, we compute the variance for each user independently using all the samples. As a result, we obtain a variance vector for each user as illustrated in Figure 6.7. The variance vectors are then averaged to provide the mean variance vector. In this mean variance vector, the bits with the highest variance are the bits that represent the session noise contained in the input samples. In this case, the variance vector is ordered using an ascending order. And same as the case of the inter variance, we store the indices of the new order.

As the goal of this type of bit selection is to remove the noisy bits, the need for using the FRGC database is further emphasized as we are sure that the

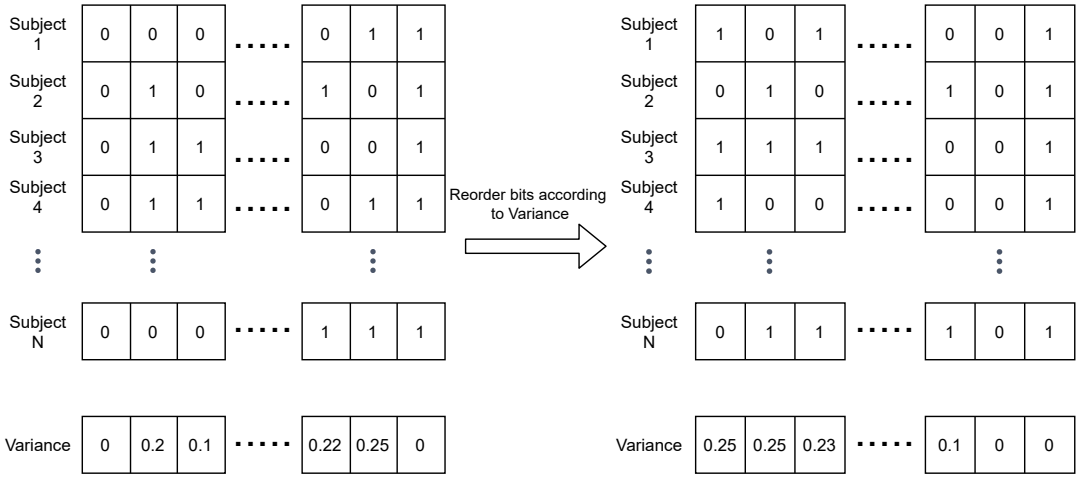


Figure 6.6: Bit reordering of the binary representations created by the DNN according to the inter-class variance.

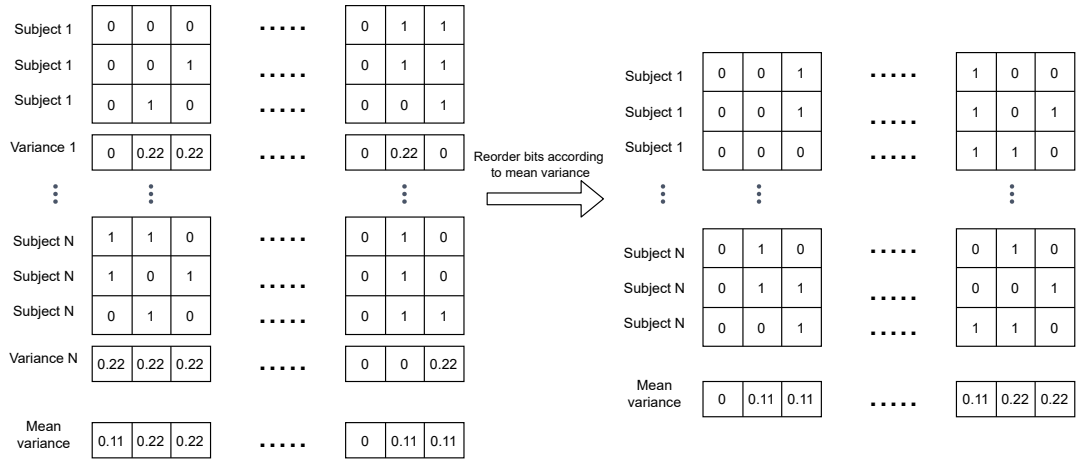


Figure 6.7: Bit reordering of the binary representations created by the DNN according to the intra-class variance.

data does not contain mislabeling.

Selecting the bits using the inter-class variance gives importance to the security of the system at the cost of the convenience of the user. By taking the bits with the highest inter-class variance, we reduce the false acceptance rate, which increases the false rejection rate of the system. On the other hand, if we focus only on the bits with the lowest average intra-class variance, the user will have an easier time regenerating his/her key, but this will increase the risk of false acceptance. As such, we tried to combine both approaches

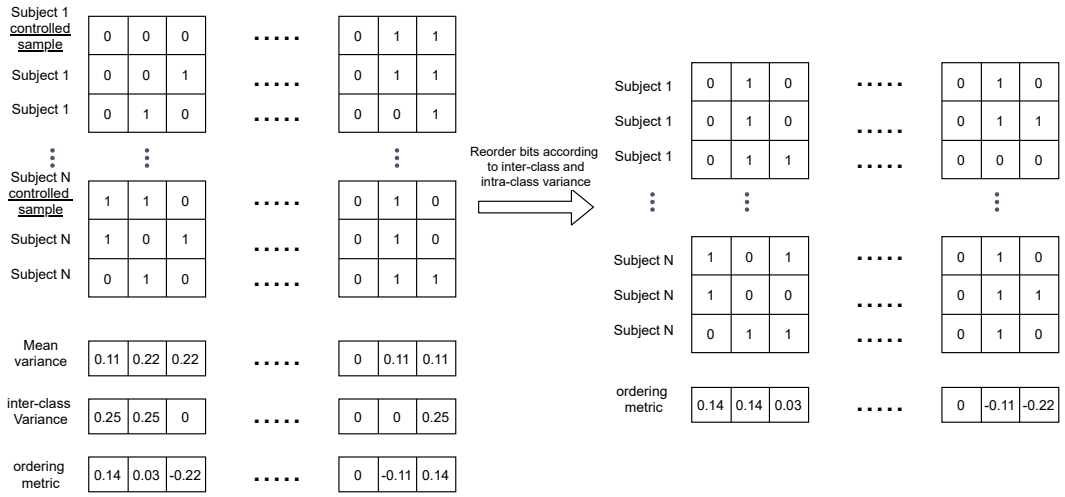


Figure 6.8: Bit reordering of the binary representations created by the DNN using the inter-class variance and intra-class variance.

by selecting the bits using both the intra-class variance and the inter-class variance.

Figure 6.8 illustrates the process of this selection. We subtract the intra-class mean variance from the inter-class variance vector. If the bit has high inter-class variance and low intra-class variance, in the resulting vector, it will obtain a high weight. If the bit has low inter-class variance and high intra-class variance, it will get a low weight in the new vector. Finally, this new vector is ordered in descending order, and we use the same selection process previously described.

To validate the selection process, we used the accuracy on the LFW database as the benchmark. The tests carried out on the LFW database are divided into 3 000 client-client tests and 3 000 client-imposter tests. As such, the accuracy metric can give a sensible measure of the usability of the system. If the system has high accuracy, then it achieves both a low false acceptance rate and a false rejection rate.

Figure 6.9 shows the performance of the inter-class, intra-class and inter-class

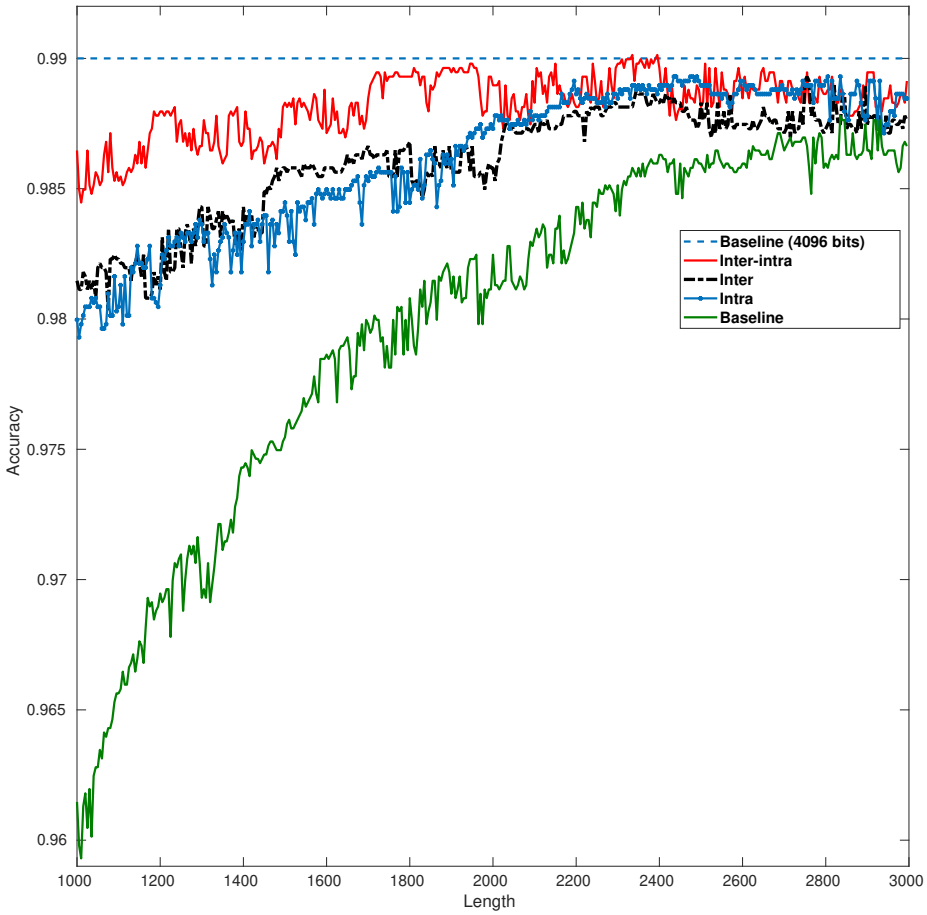


Figure 6.9: Impact of the bit selection strategy on the accuracy of the binary representations on LFW. The accuracy is represented as a function of the length of the representation.

+ intra-class bit selection strategies. The figure shows the accuracy on LFW as a function of the length of the binary representation. All the curves are drawn using the same initial binary representation; we only changed the selection strategy. As the lowest length of the encoded message (encryption key) is above 1 000 bits, the curves start at length 1 000, and each data point is computed using an increment of 10 bits.

The average size of the encoded keys in our experiments is between 2 000 and 3 000 bits. Thus, we chose the bit selection based on the third strategy, which is to use both the inter-class and intra-class variance to select the bits as this strategy has the best performance in this range as shown in Figure 6.9.

6.2.3 Error Correcting Code

Error correcting codes are techniques used to detect and correct errors that may occur during the transmission or storage of digital data. These errors can be caused by various factors such as noise, interference, or hardware malfunctions. By adding redundant information to the data being transmitted or stored, error correcting codes can help ensure that the data is received or retrieved accurately.

The error-correcting code that we used for the fuzzy commitment scheme is the Bose, Ray-Chaudhuri and Hocquenghem (BCH) code. We chose the BCH code because it is a robust code capable of correcting random errors. BCH codes are a class of cyclic error correcting codes that are based on algebraic concepts. They are widely used in communication systems and storage devices to detect and correct errors.

The basic idea behind BCH codes is to add redundant bits to the data being transmitted or stored in such a way that the receiver or reader can use these

bits to detect and correct errors. These redundant bits are called check bits. The number of check bits added to the data is determined by the length of the BCH code and the desired error correction capability.

The BCH code is constructed using a generating polynomial, which is a polynomial equation with coefficients that are used to generate the check bits. The generating polynomial is chosen such that it has a certain number of roots, which are values that make the polynomial equation equal to zero. These roots are used to generate the check bits, and the number of roots determines the error correction capability of the BCH code. Each coefficient in the polynomial represents a bit of data, and the polynomial as a whole represents the entire data set. The polynomial is then used to generate a set of check bits, which are added to the data set to form the coded message.

To encode data using a BCH code, the data is first divided into blocks of a certain size, and the check bits are generated for each block using the generating polynomial. The check bits are then appended to the data block to form the encoded data block. When the encoded data block is transmitted or stored, errors may occur due to noise or other factors.

To detect and correct errors in the received or retrieved data, the receiver or reader uses the generating polynomial to calculate the check bits for the received or retrieved data block. If the calculated check bits match the check bits in the received or retrieved data block, it is assumed that the data is error-free. If the calculated check bits do not match the check bits in the received or retrieved data block, it is assumed that errors have occurred and the receiver or reader uses the generating polynomial to determine the locations of the errors and correct them.

One of the advantages of BCH codes is that they have a high error correction

capability. For example, a BCH code with a length of 127 bits and a desired error correction capability of t bits can correct up to t errors in the data. This means that the BCH code can detect and correct errors in the data even if up to t errors have occurred.

Another advantage of BCH codes is their robustness against errors. These codes are able to correct a certain number of errors based on their designed parameters, and are therefore able to tolerate a certain level of noise or interference during the transmission or storage of data. This makes them particularly useful for applications where the transmission or storage of data may be subject to noise or interference.

In summary, BCH codes are a type of error correcting code that are used to detect and correct errors in digital data. They are based on algebraic concepts and use a generating polynomial to generate check bits that are appended to the data being transmitted or stored. The receiver or reader uses the generating polynomial to detect and correct errors in the received or retrieved data. BCH codes have a high error correction capability and low overhead, making them efficient for use in communication systems and storage devices.

The BCH code takes a block of size $\boldsymbol{k}$ and encodes it on a code word of length $\boldsymbol{n}$. The code has correction capacity of $\boldsymbol{t}$, meaning in each codeword we can correct at most $\boldsymbol{t}$ errors. The parameters $\boldsymbol{n}$, $\boldsymbol{k}$ and $\boldsymbol{t}$ are defined by eq. 6.1

For any positive integers $m \geq 3$ and $t < 2^m - 1$, there exists a binary BCH code with the following parameters:

$$\begin{aligned} &\text{codeword length} && n = 2^m - 1 \\ &\text{number of partiy check bits} && n - k \leq m.t \\ &\text{minimum distance} && d_{min} \geq 2t + 1 \end{aligned} \tag{6.1}$$

With the minimum distance being the minimum number of positions in which any two distinct codewords differ.

The encryption key of length L is divided in N blocks of k bits. Each block is encoded into a new block of n bits using the BCH error correcting code. The total length of the encoded key is $(N * n)$. The scheme can correct up to $T = (N * t)$ errors where t is the correction capacity of the code.

6.3 Results of the Proposed Key Regeneration Scheme on the MOBIO Database

In this section, we report the performance of the key regeneration scheme on the MOBIO database. The experimental protocol of the MOBIO [Kho+13] databases was originally developed for identity verification not for key regeneration. As such, we introduced some changes to the protocol. We combined the Male and Female partitions of the MOBIO database to obtain more samples. Furthermore, before carrying the experiments, we removed the samples that were of low quality where either the face is not present or obstructed as shown in Figure 6.10. After the pruning we get a total of 53k biometric samples from 152 users.

Table 6.1 reports the regeneration performance of the fuzzy commitment scheme. We report the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) on the MOBIO database. We carried out 9 M client-client tests

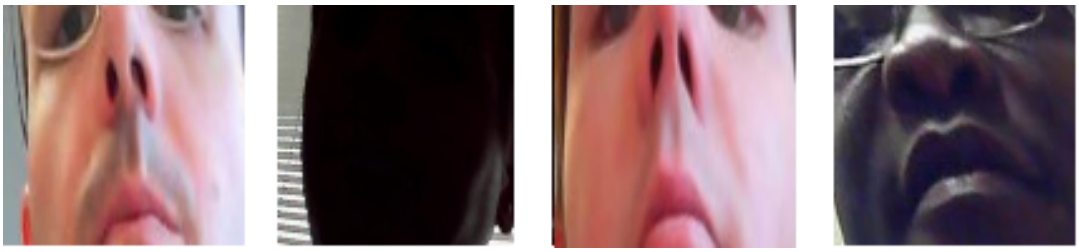


Figure 6.10: Examples of bad face samples of MOBIO database. These images were removed from the testing dataset.

and 10 M client-imposter tests. For the client-client tests, all the biometric samples are cross-matched. As for the client-imposter tests, 21 samples are randomly selected from each user and are then cross-matched.

In all experiments, the FAR is 0% as the number of erroneous bits is bigger than the code correction capacity. As shown in Figure 6.11, the minimum imposter distance is 0.196 for representations with 3 000 bits. This means we need to correct 588 bits for the imposter to be accepted as the correct user. However, from Table 6.1 we see that we can correct at most 552 bits.

The goal of the experiments was to regenerate keys with more than 400 bits to be resistant to quantum computing. As such, we focused on encryption keys with a length between 400 and 512.

Table 6.1: Performance of the key regeneration scheme. BCH codes are presented in (n, k, t) format where n is the length of the encoded block, k is length of the message block and t is the number of bits that can be corrected in the encoded block. The FAR and FRR are computed on the MOBIO database using 9 M client-client tests and 10 M client-imposter tests.

Key length	Encoded key length	BCH code	FRR on MOBIO	FAR on MOBIO
516	2666	(31,6,7)	0.7 %	0%
512	1008	(63,32,11)	0.3 %	0%
510	3213	(63,10,13)	0.3 %	0 %
528	3084	(127,22,23)	0.8 %	0%
420	3556	(127,15,27)	0.3%	0%
430	2047	(2047,430, 214)	1.6%	0%
430	4095	(4095, 495, 430)	1.3 %	0%

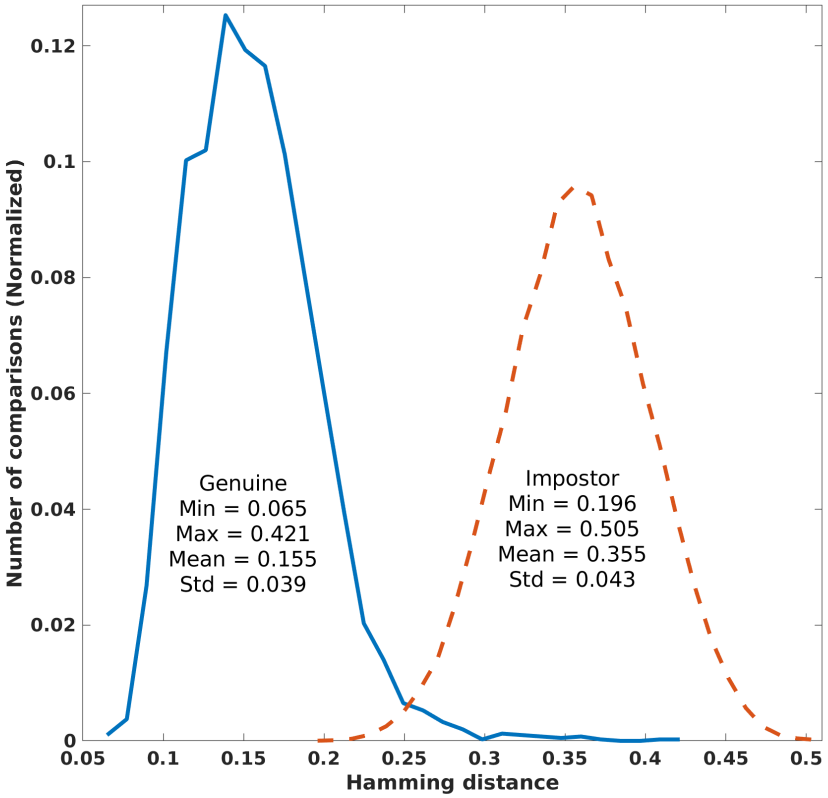


Figure 6.11: Normalized Hamming distance distribution for genuine and impostor comparisons on the MOBIO Eval male partition. The template used in the comparisons are binary templates of length 3 000 bits created using bit selection process described in the previous subsection.

6.4 Security Analysis

Due to the sensitivity of the application as it is intended to be used in cryptographic systems, we need to check the security of the scheme. In this section, we evaluate the security of the proposed key regeneration system based on fuzzy commitment against different scenarios of attacks:

- Stolen second factor,
- Stolen biometrics,
- Stolen database,
- Brute force attacks.

The security analysis was carried out on the MOBIO database using the same protected templates generated in the previous section.

6.4.1 Stolen Second Factor

In this scenario, we study the impact of the theft of the second factor on the security of the system. The attacker will use the second factor, in this case, the shuffling key, to regenerate the crypto-biometric key. We assume that the attacker also has access to the hash of the encryption key and the protected encryption key of the target user but does not have access to the target biometric data.

As such, the attacker tries using a facial dataset to access the system. We simulated this type of attack using the MOBIO database and comparing each user against the rest of the database. We report in the Table 6.2 the FAR obtained using the different system configurations.

In fact, as in our protocol we removed the enrollment faces with low quality, and as the minimum client-imposter distance is 0.19 which is higher than the

Table 6.2: FAR on the MOBIO database in the scenario of stolen second factor.
BCH codes are presented in (n, k, t) format where n is the length of the encoded block, k is length of the message block and t is the number of bits that can be corrected in the encoded block.

Encryption key length	BCH code	FAR
516	(31,6,7)	0
512	(63,32,11)	0
510	(63,10,13)	0
528	(127,22,23)	0
420	(127,15,27)	0
430	(2047,430, 214)	0
430	(4095, 495, 430)	0

error correction capacity of the used ECCs the FAR is 0% in all the experiments.

6.4.2 Stolen Biometrics

In this scenario, we study the impact of the theft of biometric data of the user on the security of the system. The biometric data, in this case, is the face of the user. As biometric data are easily accessible, especially using social networks. This type of attack is also known as spoofing, where the attacker introduces previously captured data of the user to try and access the system. The best mitigation is to implement an anti-spoofing measure such as liveness detection to assure that the user is present in front of the sensor and that it is not a picture or a video. In our system, the use of the second factor helps mitigating the spoofing attack.

We simulate this scenario by using the MOBIO protocol and using wrong shuffling keys for the client client tests of the protocol. In all the experiments, the attacker is rejected with 100% accuracy.

6.4.3 Stolen Database

In this scenario, the attacker has access to the system and to the storage of the application, where he/she recovers the protected encryption keys **PK** and the encryption key hashes **H(K)** (signature).

This allows for two types of attacks. The first type of attack is to try to generate an encryption key with the same length. In this case, the complexity of the attack and the possibility of recovering the original encryption key **K** depends only on the security of the hash function used to generate the signature.

The second type of attack is to try to use protected encryption keys to recover the original encryption key. In this case, the attacker tries to XOR the protected encryption key with binary strings generated using the metadata of the system. We suppose that attackers know the statistical distribution of the binary face representations and how the shuffling keys are generated. The shuffling keys can either be randomly generated and stored on a special medium for the user to use (a smart card) or derived using a "Password-Based Key Derivation Function"[1] from a password provided by the user. Using this information, the attacker can reduce the complexity of his attack by knowing the average number of activated bits in the shuffled binary embeddings **SB** used to protect the encoded encryption keys **E(K)**. In this attack, the goal of the adversary is mainly to generate a "pseudo" shuffled binary embedding to reverse the **XOR** operation applied to the encoded encryption

[1] A Key Derivation Function (KDF) is simply any mechanism for taking a password (something a user remembers or stores in a password manager) and turning it into a symmetric key suitable for cryptographic operations (i.e., AES).

key **E(K)**. It can be considered that the attacker is trying to find the Pseudonymous Identifier **PI** of the user of the cancelable face verification system described in Section 5.4.2. The complexity of the attack is equivalent to the diversity of the cancelable system computed using the parameters of the fuzzy commitment scheme. The maximum number of Shuffled Binary Embedding ***(SB)*** is given using the number of possible permutations. Moreover, because the decision-making is based on a threshold comparison, we should not account for templates falling in the same neighborhood. We estimate the maximum number of templates using the hamming-packing bound.

$$\begin{aligned} Number\ \ Of\ \ SB &= \frac{number\,of\ permutation}{volume\,of\ Hamming\,spheres} \\ &= \frac{L!}{L_0!L_1!\sum_{k=0}^{t}\binom{L}{k}} \end{aligned} \tag{6.2}$$

Where :

L : is the length of the encoded encryption key/shuffled binary representations.

L_0 : is the average number of zeros in the shuffled binary representations.

L_1 : is the average number of ones in the shuffled binary representations.

t : the maximum of number of bits that can be corrected using the ECC used in the fuzzy commitment scheme. Where $2t+1$ is the minimum distance of the ECC code.

For example, in the case where we use encryption keys of length 528 bits with a BCH(127,22,23) code with a correction capacity of 23 bits in a block of 127 bits of the encoded message, the encryption key is divided into 24 blocks of 22 bits. Each block is encoded onto a 127-bit block. Thus, the maximum number of bits that can be corrected is 23×24. However, this does not mean that

we can correct 552 random errors in the encoded message. We can correct only 23 random bits in each 127-bit block of the encoded message. As such, the number of possible ***SB*** in this case is higher than the lower bound provided by eq. 6.2. We obtain more than 2^{992} possible ***SB*** for this configuration as shown in eq. 6.3.

$$Number\ \ Of\ \ SB = \frac{3084!}{1542!1542! \sum_{k=0}^{23\times 24} \binom{3084}{k}} \approx 2^{992} \tag{6.3}$$

Table 6.3: Number of possible ***SB*** for each system configuration. The number of ***SB*** is provided in $\log_2$ format. BCH codes are presented in (n, k, t) format where n is the length of the encoded block, k is length of the message block and t is the number of bits that can be corrected in the encoded block.

Encryption key length	**BCH code**	**Number of SB ($\log_2$)**
516	(31,6,7)	610
512	(63,32,11)	333
510	(63,10,13)	852
528	(127,22,23)	992
420	(127,15,27)	901
430	(2047,430, 214)	1056
430	(4095, 495, 430)	1916

We show in Table 6.3 the minimum number of possible ***SB*** for each configuration of the system. We study the number of ***SB*** as it is equivalent to the entropy of the shuffled binary representations used to protect the encryption keys. For the case of BCH(63,32,11), the complexity of the brute-force attack is reduced from 512 to 333, which is still not brute-forceable with current technology but lower than the security requirement we established in Section 6.1. As for the rest of the configurations presented in Table 6.3, except for BCH(63,32,11) , it is computationally infeasible to attack the system using the information recovered from the protected encryption key. It is easier to try to brute-force the encryption key directly using its signature.

6.4.4 Brute Force Attacks

In this paragraph, we study the brute-force attacks possible on the system. The attacker can either try to directly brute-force the encryption key or brute force the system by presenting faces and random second factors. The first attack is computationally infeasible for all the system configurations presented in Table 6.1. The encryption key lengths are longer than 400 bits which are not brute-froceable even using quantum algorithms such as the Grover algorithm.

As for brute-forcing the system using its inputs (face and shuffling key), this attack is more complex than brute-forcing the encryption key as the shuffling key is longer than the original encryption key. Furthermore, if we consider the best case for the attacker when they have biometric samples of the target, the attack reverts to the stolen biometric scenario, which is not brute-forceable.

6.5 Conclusion

This work has two main goals. The first goal is to create crypto-biometric keys from the users' biometrics. To create crypto-biometric keys, we proceeded by extracting entropy from face images. By extracting entropy, we mean extracting the useful information from the biometric data in the form of binary format. In sections 4 and 5, we explained how to get entropy from face templates using a neural network in the form of binary representations. The entropy of the representations can be controlled using the neural network hyper-parameters.

The binary representations in their current form are not suitable for use in cryptography. Biometrics, by their nature, are not stable. They suffer from

variability introduced by many factors: session variability, acquisition conditions, sensors, etc... As such, we used a cohort to reduce the representations intra-variability (variability of representations obtained from each user). The approach that we followed for regenerating symmetric keys is based on 'fuzzy commitment'.

The second goal of the thesis is for the keys to be post-quantum. By post-quantum, we mean that the keys should be resistant to quantum algorithms such as Shor's algorithm [Sho94] and Grover search algorithm [Gro96]. There are two encryption schemes, symmetric and asymmetric. Grover algorithm reduces the complexity of a brute force attack on a symmetric key from 2^N to $2^{N/2}$. To mitigate the risk introduced by quantum computing, we need to increase the size of the keys. This is the reason why we tried to make the binary representation longer and more discriminative. In this chapter, we regenerate long symmetric keys for face biometrics. State-of-the-art key regeneration systems that use face biometrics suffer from high FRR and low entropy compared to other biometric modalities [Wan+21]. In our case, we succeeded in regenerating symmetric encryption keys longer than 400 bits with low FAR and low FRR using face biometrics.

On the other hand, current public-key encryption is mainly based on schemes that are vulnerable to Shor's algorithm. RSA encryption, which relies on the factorization problem, and Elliptic-Curve Cryptography (ECC), which is based on the Discrete Logarithm Problem, are easily broken by Shor's algorithm. As such, to create asymmetric crypto-biometric keys, we need to use schemes that are based on other mathematical problems that have no known vulnerabilities to quantum computing. There are multiple post-quantum encryption schemes [Bas+19]. However, the binary representation that we created are not stable enough to be used in fuzzy extractor schemes to reliably

generate a private key for the proposed post-quantum asymmetric schemes. On the other hand, if the private key of the asymmetric encryption scheme is shorter than 600 bits, we can use the same fuzzy commitment scheme to regenerate a post-quantum asymmetric crypto-biometric private key.

7 Conclusions and Perspectives

7.1 Summary

In this thesis, we addressed the problem of regenerating crypto-biometric keys (cryptographic keys obtained with biometric data) that are resistant to quantum cryptanalysis methods. The challenge is to obtain keys with high entropy to have a high level of security, knowing that the entropy contained in biometric references limits the entropy of the key.

After an introductory chapter, we present related works to our work in face recognition, binarization, biometric template protection, and encryption in Chapter 2. Chapter 3 gives an overview of the databases used to train, test, and validate our proposed systems.

Our **first contribution** was to create a **state-of-the-art face recognition system** based on public frameworks and publicly available data. In Chapter 4, we present our face recognition system pipeline. The system is built on the OpenFace framework, to which we introduced several modifications to obtain better performance, as it was implemented in two European projects and used in a submission to the NIST SRE2019 multimedia challenge.

We also detail how to obtain a state-of-the-art face recognition system based on publicly available software and using public datasets. We try to give the most possible details to allow for the reproducibility of the results. When

CMU implemented OpenFace, reproducibility was one of their main goals. Thus, we were able to reproduce and improve upon their results. For example, we improved the biometric recognition performance on the LFW dataset from 92% for the original CMU model to 99% accuracy.

From the results that we obtained, we can infer that the performance bottleneck is in the preprocessing, notably the face detection phase. Given enough data, the Deep Convolutional Neural Network (DCNN) gives the best performance. Nevertheless, in situations where the large enough databases are not available, classical approaches give better performance.

To improve our results, we proceeded to remove the misalbeling noise from the MS-celeb-1M, that gave the greatest improvement in performance on our validation protocols.

Among the modifications applied to our framework, the use of the RetinaFace face detector resulted in the most significant improvement in performance. The quality of the detected face landmarks is significantly dependent on the accuracy of the bounding box given by the face detector. Using the correct face landmarks results in better face alignment and more robust templates.

Our choice to use DCNN for face recognition was further validated during the NIST SRE 2019 multimedia challenge where our system obtained the best single system performance among 14 other submissions. This shows that DCNN is one of the better suited architectures for face recognition.

Finally, the application of enrollment filtering using some quality measures is crucial to the performance of the face recognition system. If the enrollment reference is poor quality, a comparison with good test references will result in lower similarity scores and worse performance.

Crypto-biometric schemes, such as fuzzy commitment, require binary sources. Our **second contribution**, presented in Chapter 5, is introducing **a novel approach to binarizing biometric data using Deep Neural Network (DNN)** applied to facial biometric data. We followed a data-driven approach to binarize the embeddings based on using auto-encoders under supervised training with the 'Triplet loss' loss function. Our goal was to create long binary representations with high entropy to serve in our key regeneration scheme.

The lengths of the representations can be controlled. Using a pre-trained CNN and training the model on a cleaned version of the MS-celeb-1M database, we obtain binary representations of length 4 096 bits and 3 300 bits of entropy. The extracted representations have high entropy and are long enough to be used in crypto-biometric systems such as fuzzy commitment.

We evaluate the performance of the binary representations on the MOBIO and Labeled Faces in the Wild (LFW) databases, where we measure their biometric recognition performance and entropy. The proposed binary embeddings provide state-of-the-art performance on both databases with almost negligible degradation compared to the baseline. Using DNN to extract binary embeddings results in representations with high entropy and high recognition performance. Compared to the baseline Euclidean representations, the proposed binary embeddings give state-of-the-art performance on both databases with almost negligible degradation. The performance degradation in both databases is around 0.1%.

We obtain 99.12% accuracy on the LFW database, using the binary representations, compared to 99.22% accuracy using the baseline system. The same applies to the MOBIO database, where we obtain 98.90% accuracy using the

binary embeddings compared to an accuracy of 98.93 % of the baseline system.

The approach proposed in Chapter 5 can be applied to any continuous representation, not only Euclidean face representations. Moreover, the binarization technique constitutes a locality-preserving hash, where the relative distance between the input values is preserved in the relative distance between the output hash values. The representation can be used for multiple applications, such as similarity search, database search, and biometric systems.

Furthermore, the binarization method provides representations of arbitrary length that are limited only by the quality of the training database. Therefore, the embedding length can be adapted to the sensitivity of the application. We compared our binarization approach to some classical binarization methods presented in [Dro+18] and show that our method has a better biometric recognition performance and higher entropy than the presented methods.

The created binary embeddings are also used to implement a cancelable face recognition system based on a shuffling transformation using a second factor. The cancelable system is analyzed according to the standardized metrics given by ISO/IEC 24745:2011. We show that the cancelable system gives high accuracy and unlinkable templates when the second factor is not compromised. When the second factor is compromised, the system's security is ensured by the recognition performance of the binary representations, which is comparable to the baseline non-binarized system. Furthermore, the quality of the binary representations impacts the behavior of the cancelable system. If the discriminative power of the representations is low, the cancelable system depends mainly on the second factor, which results in a higher FAR.

These representations are meant to be used in a crypto-biometric key regeneration scheme based on fuzzy commitment. This is why we seek to obtain long binary representations with high entropy.

The ***first goal*** of the thesis is to **create crypto-biometric keys** from the user biometrics. To create the crypto-biometric keys, we proceeded by extracting entropy from face images. By extracting entropy, we mean extracting useful information from the biometric data in the form of binary format.

The binary representations, obtained in Chapter 5, are not suitable for use in cryptography. Biometrics, by their nature, are not stable. They suffer from variability introduced by many factors: session variability, acquisition conditions, sensors, etc... **Our next contribution** was to **use a cohort to reduce the representations intra-variability** (variability of representations obtained from each user). The approach that we followed for regenerating symmetric keys is based on 'fuzzy commitment'. The fuzzy commitment scheme was implemented using Bose, Ray-Chaudhuri and Hocquenghem (BCH) error correcting codes. In our fuzzy commitment scheme, a random key is encoded using Error Correcting Codes (ECC) and is then XORed with the biometric data. The XORed data is cryptographically secure because neither the key nor the biometric data can be obtained from it without providing one of the two. The random key is retrieved at the time of key regeneration by providing fresh biometric data. This system requires ordered biometric data in binary form. In this scheme, the differences in the biometric data from one acquisition to another are treated as noise. This noise causes errors in the transmitted data that are corrected using ECC. The revocability of the fuzzy commitment scheme is assured using the same shuffling scheme described in subsection 5.4.2.

We report the False Acceptance Rate (FAR) and the False Rejection Rate (FRR)

on the MOBIO database. We carried out 9M client-client tests and 10M client-imposter tests. For the client-client tests, all the biometric samples are cross-matched. As for the client-imposter tests, 21 samples are randomly selected from each user and are then cross-matched.

In all the key regeneration experiments, the FAR is 0% as the number of erroneous bits is bigger than the code correction capacity.

The ***second goal*** of the thesis is for the keys to be post-quantum. By post-quantum, we mean that the keys should be resistant to quantum algorithms such as Shor's algorithm [Sho94] and Grover's search algorithm [Gro96]. There are two encryption schemes, symmetric and asymmetric.

Grover algorithm reduces the complexity of a brute-force attack on a symmetric key from 2^N to $2^{N/2}$. To mitigate the risk introduced by quantum computing, we need to increase the size of the keys. This is the reason why we tried to make the binary representation longer and more discriminative. In Chapter 6, we regenerate long symmetric keys for face biometrics. State-of-the-art key regeneration systems that use face biometrics suffer from high FRR and low entropy compared to other biometric modalities [Wan+21]. In our case, we were able to regenerate symmetric encryption keys of more than 400 bits with low FAR and low FRR using face biometrics.

7.2 Future Research Directions

Suggested future research works resulting from this thesis can be summarized as follows.

- Using newer face recognition systems with higher accuracy as the basis

for the auto-encoder. This allows for more stable and longer representations. As such, the fuzzy commitment scheme described in Chapter 6, will provide better security and convenience.

- Current public-key (asymmetric) encryption is mainly based on schemes that are vulnerable to Shor's algorithm. RSA encryption, which relies on the factorization problem, and Elliptic-Curve Cryptography (ECC), which is based on the Discrete Logarithm Problem, are easily broken by Shor's algorithm. To create asymmetric crypto-biometric keys, we need to use schemes that are based on other mathematical problems that have no known vulnerabilities to quantum computing. As such, a possible research direction is to implement a fuzzy extractor scheme based on binary embeddings to generate the private key of newer quantum-resistant public encryption schemes.

Bibliography

[ALS16] Brandon Amos, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. "Openface: A general-purpose face recognition library with mobile applications". In: *CMU School of Computer Science* 6 (2016).

[Aug+15] Daniel Augot et al. "Initial recommendations of long-term secure post-quantum systems". In: (2015).

[Aug+83] Kerckhoffs Auguste et al. "La cryptographie militaire". In: *Journal des sciences militaires* 9.538 (1883), p. 5.

[Bal20] Philip Ball. "Physicists in China challenge Google's 'quantum advantage'". In: *Nature* 588.7838 (2020), p. 380. ISSN: 14764687. DOI: 10.1038/d41586-020-03434-7.

[Bas+19] Kanad Basu et al. "NIST Post-Quantum Cryptography-A Hardware Evaluation Study." In: *IACR Cryptol. ePrint Arch.* 2019 (2019), p. 47.

[BD10] Julien Bringer and Vincent Despiegel. "Binary feature vector fingerprint representation from minutiae vicinities". In: *IEEE 4th International Conference on Biometrics: Theory, Applications and Systems, BTAS 2010* (2010), pp. 1–6. DOI: 10.1109/BTAS.2010.5634488.

[Bou06] T Boult. "Robust distance measures for face-recognition supporting revocable biometric tokens". In: *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*. IEEE. 2006, pp. 560–566.

[Bou16] Thirimachos Bourlai. "Face Recognition in Challenging Environments: An Experimental and Reproducible Research Survey". In: *Face recognition across the imaging spectrum*. Springer, 2016, pp. 269–270. ISBN: 3319285017.

[Bru10] Niko Brummer. "Measuring, refining and calibrating speaker and language information extracted from speech". PhD thesis. Stellenbosch: University of Stellenbosch, 2010.

[BSB14] Elaine Barker, Miles Smid, and Dennis Branstad. "A Profile for US Federal Cryptographic Key Management Systems". In: *NIST Special Publication* 800 (2014), p. 152.

[BSW07] Terrance E Boult, Walter J Scheirer, and Robert Woodworth. "Revocable fingerprint biotokens: Accuracy and security analysis". In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2007, pp. 1–8.

[BT17] Adrian Bulat and Georgios Tzimiropoulos. "How Far are We from Solving the 2D & 3D Face Alignment Problem? (and a Dataset of 230,000 3D Facial Landmarks)". In: *Proceedings of the IEEE International Conference on Computer Vision* 2017-Octob (2017), pp. 1021–1030. ISSN: 15505499. DOI: 10.1109/ICCV.2017.116. arXiv: 1703.07332.

[BV11] Niko Brummer and Edward de Villiers. "The BOSARIS Toolkit User Guide: Theory, Algorithms and Code for Binary Classifier Score Processing". In: i (2011), pp. 1–24. URL: https://sites.google.com/site/bosaristoolkit/.

[CFM10] Raffaele Cappelli, Matteo Ferrara, and Davide Maltoni. "Minutia Cylinder-Code: A new representation and matching technique for fingerprint recognition". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.12 (2010), pp. 2128–2141. ISSN: 01628828. DOI: 10.1109/TPAMI.2010.52.

[Cha+20] Donghoon Chang et al. "Cancelable multi-biometric approach using fuzzy extractor and novel bit-wise encryption". In: *IEEE Transactions on Information Forensics and Security* 15 (2020), pp. 3152–3167.

[Che+09] C. Chen et al. "Biometric quantization through detection rate optimized bit allocation". In: *Eurasip Journal on Advances in Signal Processing* 2009 (2009). ISSN: 16876172. DOI: 10.1155/2009/784834.

[Che+18] Lingying Chen et al. "Face template protection using deep LDPC codes learning". In: *IET Biometrics* 8.3 (2018), pp. 190–197. ISSN: 2047-4946. DOI: 10.1049/iet-bmt.2018.5156.

[CPR15] Miguel A Carreira-Perpinán and Ramin Raziperchikolaei. "Hashing with binary autoencoders". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 557–566.

[CV11] Chun Chen and Raymond Veldhuis. "Binary biometric representation through pairwise adaptive phase quantization". In: *EURASIP Journal on Information Security* 2011.1 (2011), p. 543106. ISSN: 1687-417X.

[Den+19a] Jiankang Deng et al. "ArcFace: Additive Angular Margin Loss for Deep Face Recognition". In: *CVPR*. 2019.

[Den+19b] Jiankang Deng et al. "RetinaFace: Single-stage Dense Face Localisation in the Wild". In: (2019). arXiv: 1905.00641. URL: http://arxiv.org/abs/1905.00641.

[DKG12] Priyanka Das, Kannan Karthik, and Boul Chandra Garai. "A robust alignment-free fingerprint hashing algorithm based on minimum distance graphs". In: *Pattern Recognition* 45.9 (2012), pp. 3373–3388.

[Dro+18] P. Drozdowski et al. "Benchmarking Binarisation Schemes for Deep Face Templates". In: *Proceedings - International Conference on Image Processing, ICIP* (2018), pp. 191–195. ISSN: 15224880. DOI: 10.1109/ICIP.2018.8451291.

[DT05] Navneet Dalal and Bill Triggs. "Histograms of oriented gradients for human detection". In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. Vol. 1. Ieee. 2005, pp. 886–893.

[Eri+19] Learned-Miller Erik et al. *LFW : Results*. 2019. URL: http://vis-www.cs.umass.edu/lfw/results.html (visited on 04/29/2019).

[Est+96] Martin Ester et al. "A density-based algorithm for discovering clusters in large spatial databases with noise." In: *Kdd*. Vol. 96. 34. 1996, pp. 226–231.

[GB+17] Marta Gomez-Barrero et al. "General Framework to Evaluate Unlinkability in Biometric Template Protection Systems". In: *IEEE Transactions on Information Forensics and Security* 13.6 (2017), pp. 1406–1420. ISSN: 15566013. DOI: 10.1109/TIFS.2017.2788000.

[GN03] Alwyn Goh and David CL Ngo. "Computation of cryptographic keys from face biometrics". In: *IFIP International Conference on*

Communications and Multimedia Security. Springer. 2003, pp. 1–13.

[Gro+10] Ralph Gross et al. "Multi-pie". In: *Image and Vision Computing* 28.5 (2010), pp. 807–813.

[Gro96] Lov K Grover. "A fast quantum mechanical algorithm for database search". In: *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. 1996, pp. 212–219.

[Guo+16] Yandong Guo et al. "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition". In: *European Conference on Computer Vision*. Springer, 2016, pp. 87–102.

[He+16] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[Her+17] Charles Herder et al. "Public Key Cryptosystems with Noisy Secret Keys." In: *IACR Cryptology ePrint Archive* 2017 (2017), p. 210.

[HLM14] Gary B Huang and Erik Learned-Miller. "Labeled faces in the wild: Updates and new reporting procedures". In: *Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep* (2014), pp. 3–14.

[Hma+20] Mohamed Hmani et al. "Evaluation of the H2020 SpeechXRays project Cancelable Face System Under the Framework of ISO/IEC 24745:2011". In: *2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. IEEE. 2020, pp. 1–6.

[HMD21] Mohamed Amine Hmani, Aymen Mtibaa, and Dijana Petrovska Delacretaz. "Voice Biometrics: Technology, trust and security". In: Security. Institution of Engineering and Technology, 2021. Chap. Joining forces of voice and facial biometrics: a case

study in the scope of NIST SRE19, pp. 187–217. DOI: 10.1049/PBSE012E_ch9.

[HPD18] Mohamed Amine Hmani and Dijana Petrovska-Delacrétaz. "State-of-the-art face recognition performance using publicly available software and datasets". In: *2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. IEEE, 2018, pp. 1–6. ISBN: 1538652390.

[Hua+07] Gary B Huang et al. *Labeled faces in the wild: A database for studying face recognition in unconstrained environments*. Tech. rep. Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[ISO11] ISO/IEC JTC1 SC27 Security Techniques. *ISO/IEC 24745:2011. Information Technology - Security Techniques - Biometric Information Protection*. International Organization for Standardization. 2011.

[JCJ18] Arun Kumar Jindal, Srinivas Chalamala, and Santosh Kumar Jami. "Face template protection using deep convolutional neural network". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* 2018-June (2018), pp. 575–583. ISSN: 21607516. DOI: 10.1109/CVPRW.2018.00087.

[JLG04] Andrew Teoh Beng Jin, David Ngo Chek Ling, and Alwyn Goh. "Biohashing: two factor authentication featuring fingerprint data and tokenised random number". In: *Pattern recognition* 37.11 (2004), pp. 2245–2255.

[JW99] Ari Juels and Martin Wattenberg. "A fuzzy commitment scheme". In: *Proceedings of the 6th ACM conference on Computer and communications security*. 1999, pp. 28–36.

[KAN10] Sanjay Ganesh KANADE. "Enhancing information security and privacy by combining biometrics with cryptography". Theses.

Institut National des Télécommunications, 2010. URL: https://tel.archives-ouvertes.fr/tel-01057728.

[Kev+05] T. A.M. Kevenaar et al. "Face recognition with renewable and privacy preserving binary templates". In: *Proceedings - Fourth IEEE Workshop on Automatic Identification Advanced Technologies, AUTO ID 2005* 2005 (2005), pp. 21–26. DOI: 10.1109/AUTOID.2005.24.

[Kho+13] Elie Khoury et al. "The 2013 speaker recognition evaluation in mobile environment". In: *2013 International Conference on Biometrics (ICB)*. IEEE. 2013, pp. 1–8.

[Kin09a] Davis E. King. "Dlib-ml: A machine learning toolkit". In: *Journal of Machine Learning Research* 10 (2009), pp. 1755–1758. ISSN: 15324435.

[Kin09b] Davis E. King. "Dlib-ml: A Machine Learning Toolkit". In: *Journal of Machine Learning Research* 10 (2009), pp. 1755–1758.

[Kon+06] Adams Kong et al. "An analysis of BioHashing and its variants". In: *Pattern recognition* 39.7 (2006), pp. 1359–1368.

[KPD09] Sanjay Kanade, Dijana Petrovska, and Bernadette Dorizzi. "Multi-biometrics based cryptographic key regeneration scheme". In: *IEEE 3rd International Conference on Biometrics: Theory, Applications and Systems, BTAS 2009* (2009). DOI: 10.1109/BTAS.2009.5339034.

[KPDD12] Sanjay G Kanade, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. "Enhancing information security and privacy by combining biometrics with cryptography". In: *Synthesis Lectures on Information Security, Privacy, and Trust* 3.1 (2012), pp. 1–140. ISSN: 1945-9742.

[KS14a] Vahid Kazemi and Josephine Sullivan. "One millisecond face alignment with an ensemble of regression trees". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2014), pp. 1867–1874. ISSN: 10636919. DOI: 10.1109/CVPR.2014.241.

[KS14b] Vahid Kazemi and Josephine Sullivan. "One millisecond face alignment with an ensemble of regression trees". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 1867–1874.

[Lee+12] Hyunggu Lee et al. "A secure biometric discretization scheme for face template protection". In: *Future Generation Computer Systems* 28.1 (2012), pp. 218–231. ISSN: 0167-739X.

[Liu+16] Wei Liu et al. "Ssd: Single shot multibox detector". In: *European conference on computer vision*. Springer, 2016, pp. 21–37.

[Liu+17] Weiyang Liu et al. "SphereFace: Deep hypersphere embedding for face recognition". In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* 2017-Janua (2017), pp. 6738–6746. DOI: 10.1109/CVPR.2017.713.

[LM+16] Erik Learned-Miller et al. "Labeled faces in the wild: A survey". In: *Advances in Face Detection and Facial Image Analysis*. 2016, pp. 189–248. ISBN: 9783319259581. DOI: 10.1007/978-3-319-25958-1_8.

[LT13] Meng Hui Lim and Andrew Beng Jin Teoh. "A novel encoding scheme for effective biometric discretization: Linearly separable subcode". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.2 (2013), pp. 300–313. ISSN: 01628828. DOI: 10.1109/TPAMI.2012.122.

[LTK15] Menghui Lim, Andrew Beng Jin Teoh, and Jaihie Kim. "Biometric feature-type transformation: Making templates compatible for secret protection". In: *IEEE Signal Processing Magazine* 32.5 (2015), pp. 77–87. ISSN: 10535888. DOI: 10.1109/MSP.2015.2423693.

[Mai+08] Emanuele Maiorana et al. "Cancelable biometrics for hmm-based signature recognition". In: *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems*. IEEE. 2008, pp. 1–6.

[Mai+10] Emanuele Maiorana et al. "Cancelable templates for sequence-based biometrics with application to on-line signature recognition". In: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 40.3 (2010), pp. 525–538.

[Mai+21] Guangcan Mai et al. "SecureFace: Face Template Protection". In: *IEEE Transactions on Information Forensics and Security* 16 (2021), pp. 262–277. ISSN: 15566021. DOI: 10.1109/TIFS.2020.3009590.

[MB11] A Mitrokotsa and J Bringer. "D5. 1: Privacy preservation techniques". In: *Evaluation* 1 (2011).

[McC+12] Christopher McCool et al. "Bi-modal person recognition on a mobile phone: using mobile phone data". In: *2012 IEEE International Conference on Multimedia and Expo Workshops*. IEEE, 2012, pp. 635–640. ISBN: 1467320277.

[MCN08] Emanuele Maiorana, Patrizio Campisi, and Alessandro Neri. "On-line signature authentication: user adaptive template protection and renewability". In: *Mobile Multimedia/Image Processing, Security, and Applications 2008*. Vol. 6982. SPIE. 2008, pp. 263–274.

[MCN11] Emanuele Maiorana, Patrizio Campisi, and Alessandro Neri. "Bioconvolving: Cancelable templates for a multi-biometrics signature recognition system". In: *2011 IEEE International Systems Conference*. IEEE. 2011, pp. 495–500.

[Mer11] Zeeya Merali. "First sale for quantum computing". In: *Nature* 474.7349 (2011), p. 18. ISSN: 00280836. DOI: 10.1038/474018a.

[Mon+01] Fabian Monrose et al. "Using voice to generate cryptographic keys". In: *2001: A Speaker Odyssey-The Speaker Recognition Workshop*. 2001.

[Mos+17] Stylianos Moschoglou et al. "AgeDB: The First Manually Collected, In-the-Wild Age Database". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* 2017-July (2017), pp. 1997–2005. ISSN: 21607516. DOI: 10.1109/CVPRW.2017.250.

[Pan+16] Rohit Kumar Pandey et al. "Deep Secure Encoding for Face Template Protection". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (2016), pp. 77–83. ISSN: 21607516. DOI: 10.1109/CVPRW.2016.17.

[PDCD09] Dijana Petrovska-Delacrétaz, Gérard Chollet, and Bernadette Dorizzi. *Guide to biometric reference systems and performance evaluation*. Springer, 2009. ISBN: 1848002912.

[Phi+05] P Jonathon Phillips et al. "Overview of the face recognition grand challenge". In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. Vol. 1. IEEE. 2005, pp. 947–954.

[PM01] Viola Paul and Jones Michael. "Prefacio Prólogo". In: February (2001). ISSN: 1063-6919. DOI: 10.1109/CVPR.2001.990517. arXiv: arXiv:1011.1669v3.

[PRC15] Vishal M Patel, Nalini K Ratha, and Rama Chellappa. "Cancelable biometrics: A review". In: *IEEE signal processing magazine* 32.5 (2015), pp. 54–65.

[PVZ15] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. "Deep face recognition." In: *bmvc*. Vol. 1. 3. 2015, p. 6.

[RAD+78] Ronald L Rivest, Len Adleman, Michael L Dertouzos, et al. "On data banks and privacy homomorphisms". In: *Foundations of secure computation* 4.11 (1978), pp. 169–180.

[Rat+07] Nalini K Ratha et al. "Generating cancelable fingerprint templates". In: *IEEE Transactions on pattern analysis and machine intelligence* 29.4 (2007), pp. 561–572.

[RCB01] Nalini K. Ratha, Jonathan H. Connell, and Ruud M. Bolle. "Enhancing security and privacy in biometrics-based authentication systems". In: *IBM systems Journal* 40.3 (2001), pp. 614–634.

[RU11] Christian Rathgeb and Andreas Uhl. "A survey on biometric cryptosystems and cancelable biometrics". In: *EURASIP Journal on Information Security* 2011.1 (2011), p. 3. ISSN: 1687-417X.

[Sad+20] Seyed Omid Sadjadi et al. "The 2019 NIST audio-visual speaker recognition evaluation". In: *Proc. Speaker Odyssey (submitted), Tokyo, Japan* (2020).

[Sag+16] Christos Sagonas et al. "300 Faces In-The-Wild Challenge : database and results". In: *Image and Vision Computing* 47 (2016), pp. 3–18. ISSN: 02628856. DOI: 10.1016/j.imavis.2016.01.002.

[Sch+19] Jo Schlemper et al. "Deep Hashing using Entropy Regularised Product Quantisation Network". In: (2019), pp. 1–11. arXiv: 1902.03876. URL: http://arxiv.org/abs/1902.03876.

[SDF11] Chang Shu, Xiaoqing Ding, and Chi Fang. "Histogram of the oriented gradient for face recognition". In: *Tsinghua Science and Technology* 16.2 (2011), pp. 216–224. ISSN: 1007-0214.

[Sho94] Peter W Shor. "Algorithms for quantum computation: discrete logarithms and factoring". In: *Proceedings 35th annual symposium on foundations of computer science*. Ieee. 1994, pp. 124–134.

[SKP15] Florian Schroff, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 815–823.

[SRB16] Torsten Schlett, Christian Rathgeb, and Christoph Busch. "A binarization scheme for recognition based on multi-scale block local binary patterns". In: *Biosig 2016* (2016). ISSN: 3885796546.

[Sun+15] Yi Sun et al. "Deepid3: Face recognition with very deep neural networks". In: *arXiv preprint arXiv:1502.00873* (2015).

[Sze+15] Christian Szegedy et al. "Going deeper with convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

[Sze+16] Christian Szegedy et al. "Rethinking the inception architecture for computer vision". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826.

[Sze+17] Christian Szegedy et al. "Inception-v4, inception-ResNet and the impact of residual connections on learning". In: *31st AAAI Conference on Artificial Intelligence, AAAI 2017* (2017), pp. 4278–4284. arXiv: 1602.07261.

[Tai+14] Yaniv Taigman et al. "Deepface: Closing the gap to human-level performance in face verification". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 1701–1708.

[TC10] Andrew Beng Jin Teoh and Lee-Ying Chong. "Secure speech template protection in speaker verification system". In: *Speech communication* 52.2 (2010), pp. 150–163.

[TS19] Jennifer Tracey and Stephanie Strassel. "VAST: A corpus of video annotation for speech technologies". In: *LREC 2018 - 11th International Conference on Language Resources and Evaluation* (2019), pp. 4318–4321.

[Wan+21] Peiyi Wang et al. "Biometric key generation based on generated intervals and two-layer error correcting technique". In: *Pattern Recognition* 111 (2021), p. 107733. ISSN: 00313203. DOI: 10.1016/j.patcog.2020.107733. URL: https://doi.org/10.1016/j.patcog.2020.107733.

[WH14] Song Wang and Jiankun Hu. "Design of alignment-free cancelable fingerprint templates via curtailed circular convolution". In: *Pattern Recognition* 47.3 (2014), pp. 1321–1329.

[Yan+16] Shuo Yang et al. "WIDER FACE: A Face Detection Benchmark". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.

[Yi+14] Dong Yi et al. "Learning face representation from scratch". In: *arXiv preprint arXiv:1411.7923* (2014).

[YY01] Hua Yu and Jie Yang. "A direct LDA algorithm for high dimensional data with application to face recognition". In: *Pattern recognition* 34.10 (2001), pp. 2067–2070.

ECOLE DOCTORALE

Titre : Utilisation des données biométriques pour la régénération des clés cryptobiométriques révocables

Mots clés : biometrie, cryptographie, verification de visage, Apprentissage profond

Résumé : Ce travail de thèse vise à régénérer des clés crypto-biométriques (clés cryptographiques obtenues avec des données biométriques) résistantes aux méthodes de cryptanalyse quantique. Le défi est d'obtenir des clés avec une haute entropie pour avoir un haut niveau de sécurité, sachant que le l'entropie contenue dans les références biométriques limite l'entropie de la clé. Notre choix a été d'exploiter la biométrie faciale.

Nous avons d'abord créé un système de reconnaissance faciale de pointe basé en utilisant des bases de données publiques. Notre architecture utilise des réseaux de neurones profonds avec une fonction de perte'Triplet loss'. Nous avons participé à deux Projets européens H2020 pour lesquelles nous avons fournit des adapations de notres systeme de reconnaise de visage. Nous avons également participé au challenge multimédia NIST SRE19 avec la version finale de notre système classique de reconnaissance faciale qui a donnée d'excellents résultats.

Pour obtenir des clés crypto-biométriques, il est nécessaire de disposer de références biométriques binaires. Pour obtenir les représentations binaires directement à partir d'images de visage, nous avons proposé une méthode novatrice tirant parti des auto-encodeurs et la biométrie faciale classique précédemment mise en œuvre. Nous avons également exploité les représentations binaires pour créer un système de vérification de visage cancelable.

Concernant notre objectif final, générer des clés crypto-biométriques, nous nous sommes concentrés sur les clés symétriques. Le chiffrement symétrique est menacé par l'algorithme Groover parce qu'il réduit la complexité d'une attaque par force brute de 2^N à $2^{(N/2)}$. Pour atténuer le risque introduit par l'informatique quantique, nous devons augmenter la taille des clés. Pour cela, nous avons essayé de faire la représentation binaire plus longue et plus discriminante.

Nous avons réussi à régénérer des clés crypto-biométriques de plus de 400 bits grâce à la qualité des plongements binaires. Les clés crypto-biométriques ont une haute entropie et résistent à la cryptanalyse quantique selon le PQCrypto projet car ils satisfont à l'exigence de longueur. Les clés sont régénérées à l'aide d'un schéma de "fuzzy commitment" en utilisant les codes BCH.

Title : Use of Biometrics for the Regeneration of Revocable Crypto-biometric Keys

Keywords : biometrics, cryptography, face verification, Deep Learning

Abstract : This thesis aims to regenerate crypto-biometric keys (cryptographic keys obtained with biometric data) that are resistant to quantum cryptanalysis methods. The challenge is to obtain keys with high entropy to have a high level of security, knowing that the entropy contained in biometric references limits the entropy of the key. Our choice was to exploit facial biometrics.

We first created a state-of-the-art face recognition system based on public frameworks and publicly available data based on DNN embedding extractor architecture and triplet loss function. We participated in two H2020 projects. For the SpeechXRays project, we provided implementations of classical and cancelable face biometrics. For the H2020 EMPATHIC project, we created a face verification REST API. We also participated in the NIST SRE19 multimedia challenge with the final version of our classical face recognition system.

In order to obtain crypto-biometric keys, it is necessary to have binary biometric references. To obtain the binary representations directly from face images, we proposed an original method, leveraging autoencoders and the previously implemented classical face biometrics. We also exploited the binary representations to create a cancelable face verification system.

Regarding our final goal, to generate crypto-biometric keys, we focused on symmetric keys. Symmetric encryption is threatened by the Groover algorithm because it reduces the complexity of a brute force attack on a symmetric key from 2^N to $2^{(N/2)}$. To mitigate the risk introduced by quantum computing, we need to increase the size of the keys. To this end, we tried to make the binary representation longer and more discriminative. For the keys to be resistant to quantum computing, they should have double the length.

We succeeded in regenerating crypto-biometric keys longer than 400bits (with low false acceptance and false rejection rates) thanks to the quality of the binary embeddings. The crypto-biometric keys have high entropy and are resistant to quantum cryptanalysis, according to the PQCrypto project, as they satisfy the length requirement. The keys are regenerated using a fuzzy commitment scheme leveraging BCH codes.

Institut Polytechnique de Paris
91120 Palaiseau, France