MONASH University

Towards a machine learning based identifier-free face cryptosystem

Xingbo Dong
Doctor of Philosophy

A Thesis Submitted for the Degree of Doctor of Philosophy at
Monash University in 2021
Faculty of Information Technology

Copyright notice

Abstract

Among different biometric modalities, face is becoming a hot research topic due to its distinctive advantages, such as non-intrusive, friendliness, contact-less, and convenience. Nowadays, face recognition system has been widely deployed by banks, stores, and transportation services providers. The prevalence of face recognition technology brings convenience to the public. However, security issues have drawn extensive attention recently, such as the security risk of the biometric templates stored in the system directly. It is reported that the face images can be reconstructed easily from the raw face features. Biometric template protection (BTP) techniques have been proposed to alleviate those issues by transforming the conventional biometric features into a protected template. However, the introduction of BTP transformations usually leads to performance degradation.

Besides the trade-off between performance and security, BTP approaches such as biometric cryptosystems also lead to a user-unfriendly experience. Popular biometric cryptosystem approaches such as fuzzy vaults can be considered as an instance of a 1-to-1 match or verification system which returns correct secret (yes) or null (no). In such approaches, the user must input both identity credentials and biometrics, hence compromises the convenience of biometrics.

On the other hand, most face recognition systems' accuracy is evaluated on a closed-set protocol, assuming that all probe samples are registered in the gallery. While in practice, the accuracy usually drops significantly in open-set settings when probe samples may not be enrolled. Currently, it is still a big challenge to have an accurate open-set face recognition system.

In this thesis, in order to protect the face features extracted from deep models, a learning-based Index-of-Max (LIoM) hashing has been proposed. LIoM is utilized to hash the face features into compact hash codes in integer/binary form; hence matching can be easily performed by the Hamming distance in a highly efficient manner. Furthermore, since LIoM hashing transforms the original facial features non-invertibly, the privacy of users can also be preserved. In order to achieve better performance under open-set settings, several techniques are explored, such as Extreme-Value-Machine, feature fusion, and score-level fusion. A large-scale 1-to-N face searching system has been built based on LIoM, and its performance has been evaluated.

To build face cryptosystems that only need biometric input, a sole-input face cryptosystem for identification (FCI) is proposed. The FCI composes an open-set 1-to-N search subsystem and a 1-to-1 match chaff-less fuzzy vault (CFV) subsystem. The first subsystem stores N facial features protected by LIoM hashing and enhanced by a fusion module for searching accuracy. When a face image of the user is presented, the subsystem returns top k match scores, and thus the corresponding vaults in the CFV subsystem will be activated. The 1-to-1 matching occurs among k vaults alongside the query face, and an identifier associated with the user will be retrieved from the correct matched vault. We demonstrated that the coupling of the LIoM hashing and the CFV avoids the risks of Chaff-points in conventional fuzzy vault. Meanwhile, user privacy, unlinkability, and cancellability have been achieved. Finally, the FCI system is evaluated under three large-scale public unconstrained face datasets, namely LFW, VGG2, and IJB-C, with respect to its accuracy, computation cost, template protection criteria, and security.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature: ____________________

Print Name: Xingbo Dong

Date: ____________________

Publications during enrolment

Publications included in this thesis:

[1] **Xingbo Dong**, Jung-Yeon Hwang, Zhe Jin, Soohyong Kim, Sangrae Cho and Andrew Beng Jin Teoh. Open-set Face Identification with Index-of-Max Hashing by Learning. ***Pattern Recognition*** (2020). https://doi.org/10.1016/j.patcog.2020.107277

[2] **Xingbo Dong**, Jung-Yeon Hwang, Zhe Jin, Soohyong Kim, Sangrae Cho and Andrew Beng Jin Teoh. A Secure Chaff-less Fuzzy Vault for Face Identification System. ***ACM Transactions on Multimedia Computing Communications and Applications*** (2021).

Other publications not included in this thesis:

1. **Xingbo Dong**, Zhe Jin, Andrew Beng Jin Teoh, M Tistarelli, KS Wong. On the Security Risk of Cancelable Biometrics. ***Pattern Recognition*** (under review).
2. **Xingbo Dong**,Sangrae Cho, Soohyong Kim and Andrew Beng Jin Teoh. Deep Rank Hashing Network for Cancellable Face Identification. ***Pattern Recognition*** (under review).
3. **Xingbo Dong**, Zhe Jin,and Andrew Beng Jin Teoh. A Genetic Algorithm Enabled Similarity-Based Attack on Cancellable Biometrics.***10th IEEE International Conference on Biometrics: Theory, Applications and Systems*** (BTAS2019).
4. **Xingbo Dong**, KokSheik Wong, Zhe Jin, Jean-luc Dugelay.A Secure Visual-Thermal Fused Face Recognition System. ***IEEE 21st International Workshop on Multimedia Signal Processing***(MMSP 2019).
5. **Xingbo Dong**, KokSheik Wong, Zhe Jin, Jean-luc Dugelay. A Cancellable Face Template Scheme Based on Nonlinear Multi-Dimension Spectral Hashing. ***The 7th IAPR/IEEE International Workshop on Biometrics and Forensics*** (IWBF 2019).
6. **Xingbo Dong**, Zhe Jin, and Wong, K. A Generalized Approach for Cancellable Template and Its Realization for Minutia Cylinder-Code. ***In Proceedings, APSIPA Annual Summit and Conference*** (Vol. 2018, pp. 12-15).

Acknowledgements

This research was supported by Institute for Information and Communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No.2016-0-00097, Development of Biometrics-based Key Infrastructure Technology for On-line Identification and No.2018-0-01369, Developing blockchain identity management system with implicit augmented authentication and privacy protection for O2O services, and No. 2018-0-00189, Security Technology for Portal Device to Connect Human-Infrastructure-Service in Highly Trusted Intelligent Information Service).

I would like to express my gratitude to my main supervisor Dr. Jin Zhe. Dr. Jin offered tremendous help and support not only on academic and career but also on daily life hardness. He was the one who always guides and encourages me with patience, which directed me to a mature researcher. I would also like to thank my associate supervisor A. Prof. Dr. Wong KokSheik, for his scientific input and support, such as reviewing and proof-reading my manuscripts. Dr. Jin and Dr. Wong also provided a lot of research collaboration opportunities, such as secondment at EU countries, which have broadened my mind. Besides, I am very grateful to Dr. Soon Lay Ki for all the support on my Ph.D. studies, including the arrangement of my secondment at Tsinghua University during the pandemic time.

I would like to give a special thanks to Prof. Andrew Teoh from Yousei University for his guidance on my research experiment and manuscript writing. That guidance and discussions on my research topic help me to have an in-depth understanding and interest in academia.

I would like to thank my milestone review panels, Dr. Pan Shirui and Dr. Lim Chern Hong, for their invaluable feedback on my research. Their discussions and feedback drive my research work into a more solid one. Thanks for their guidance and suggestions.

Besides, I would like to acknowledge Prof. Massimo Tistarelli, Prof. Jean-Luc DUGELAY, and Prof. Andreas Uhl, for their' hosting in my secondment period at EU countries. I would also like to thank Prof. Guo Zhenhua from Tsinghua. I appreciate the weekly individual meeting with him and the group meeting with Tsinghua's peers, which broaden my horizons widely.

Finally, I want to thank my parents, my sister and my girlfriend. Without their understanding and support, I could not achieve this. They always show their trust and support to me for every hard decision I made. Especially when I decide to withdraw from life science research career, they tried their best to help me get out of the hardship and encourage me to explore new possibilities.

Contents

Copyright notice i

Abstract ii

Declaration iv

Publications during enrolment v

Acknowledgements vi

List of Figures x

List of Tables xi

Abbreviations xiii

1 Introduction 1

1.1 Problem Statement . . . 6

1.2 Research Question . . . 7

1.3 Research Objectives . . . 7

1.4 Contribution Overview . . . 8

1.5 Practical Usage . . . 9

1.6 Organization . . . 10

2 Literature Review and Background Study 11

2.1 Deep Learning based Face Feature Extraction . . . 11

2.2 Cancellable Biometrics . . . 13

2.3 Biometric Cryptosystem . . . 16

2.4 Closed-set and Open-set Face Identification . . . 21

3 Face Identification System with Learning-based Index-of-Max Hashing 25

3.1 Introduction . . . 25

3.2 Motivations and Contributions . . . 27

3.3 Preliminary . . . 28

3.3.1 Locality Sensitive Hashing (LSH) . . . 28

3.3.2 Winner-Take-All Hashing . . . 29

3.3.3 Index-of-Max (IoM) Hashing . . . 30

3.4 Face Identification Framework . . . 32

3.4.1 Deep Face Feature Extraction . . . 32

3.4.2 Learning based Index-of-Max hashing 34
3.4.3 Fusion Strategy 39
3.5 Chapter Conclusion 42

4 Evaluation and Analysis of the Learning-based Index-of-Max Hashing 43
4.1 Performance Evaluation 43
4.1.1 Learning-based IoM Hashing 44
4.1.2 Open-set and Closed-set Identification 49
4.1.2.1 Database Configuration 49
4.1.2.2 Performance Metrics 53
4.1.2.3 Unprotected Identification System Performance 53
4.1.2.4 Parameters Tuning 54
4.1.2.5 Protected System Performance without Fusion Strategy 55
4.1.2.6 Protected System Performance with Fusion Strategy 58
4.1.2.7 Visual Evaluation of the system 63
4.1.2.8 Time Cost of the System 64
4.1.2.9 Summary 65
4.1.3 State-of-the-Art Comparison 67
4.2 Unlinkability and Revocability Analysis of LIoM Hashing 67
4.2.1 Unlinkability Evaluation 68
4.2.2 Revocability Evaluation 70
4.3 Chapter Conclusion 71

5 Secure Chaff-less Fuzzy Vault for Face Identification System 72
5.1 Introduction 72
5.2 Motivations and Contributions 73
5.3 Overview of the Facial Cryptosystem for Identification 75
5.4 Chaff-less Fuzzy Vault 76
5.5 Performance Evaluation 79
5.5.1 Performance Metrics 80
5.5.2 Parameter Tuning for 1-to-N Searching 81
5.5.3 Secret Retrieval Performance 82
5.5.4 Computation Efficiency 86
5.6 Security Analysis for the CFV 87
5.6.1 Security Model 87
5.6.2 Sneak Key Inversion (SKI) Attack 88
5.6.3 Surreptitious Key Inversion (SuKI) Attack 88
5.6.4 Brute-force Attack 89
5.6.5 MisIdentify Attack 89
5.7 Chapter Conclusion 90

6 Conclusions and Futures 91
6.1 Conclusion 91
6.2 Future work 93
6.2.1 Discriminative Features 93
6.2.2 Integrate LIoM to Deep Neural Network 94
6.2.3 Apply FCI to Other Biometric Features 94

6.3 A Cat-and-Mouse Game 94

Bibliography **95**

List of Figures

1.1 Two face recognition working modes. 3

1.2 The differences of closed-set and open-set face identification. 4

1.3 Two biometric template protection categories. 5

2.1 Diagram of fuzzy vault. 20

2.2 Close and Open Set Face Identification (1-to-N Match) System. 21

3.1 Locality sensitive hashing (From [3]). 29

3.2 Winner-Take-All Hashing. 30

3.3 Face identification module. 32

3.4 Different fusion strategies. 40

4.1 Performance comparisons under identity-dependent scenario. 48

4.2 Performance comparisons under identity-independent scenario. 48

4.3 Performance comparisons under database-independent scenario. 48

4.4 LFW configuration for face identification evaluation. 50

4.5 mAP vs k and mAP vs LIoM length m on fusion feature of LFW (closed-set). . 55

4.6 Protected identification system performance on LFW with InsightFace where (a) is for closed-set evaluation and (b), (c), (d) are for open-set evaluation. 56

4.7 Protected identification system performance on LFW with FaceNet where (a) is for closed-set evaluation and (b), (c), (d) are for open-set evaluation. 57

4.8 Accuracy of performance on LFW score with IoM hashing (protection+fusion). 59

4.9 Accuracy of performance on VGG2 score with IoM hashing). 60

4.10 Accuracy of performance on IJB-C score with IoM hashing). 61

4.11 Similarity score distribution on LFW. 62

4.12 Illustration of face search in open-set 1 (LFW, O1). 63

4.13 Genuine and imposter score distribution on LFW. 64

4.14 Unlinkability of LIoM and random IoM hashing under VGG2. 70

4.15 Revocability of LIoM and random IoM Hashing under VGG2. 71

5.1 Overview of the face cryptosystem for identification system architecture. 76

5.2 Chaff-less Facial Fuzzy Vault Illustration. 77

5.3 IR (%) vs k of LIoM hashing on the LFW: (a) FaceNet; (b) InsightFace. 81

5.4 Key retrieval performance of LFW by fusion. 83

5.5 Key retrieval performance of VGG2 by fusion. 84

5.6 Key retrieval performance of IJB-C by fusion. 85

5.7 Identifier Retrieval Time Cost for One Probe against p on LFW. 87

List of Tables

1.1 Examples of FR system. 3

2.1 Latest deep models of three loss types. 13
2.2 Summary of selected cancellable biometrics on face. 16
2.3 Summary of fuzzy commitment and fuzzy vault. 21
2.4 The list of works on Face Search and Face Retrieval. 24

3.1 Deep features used in this study. 34

4.1 Databases used in this project. 44
4.2 Dataset configuration for identity-dependent evaluation. 45
4.3 Datasets configuration for identity-independent and dataset-independent evaluation. 46
4.4 Number of matching scores under different scenarios. 47
4.5 Risk of different situations. 49
4.6 Division of subjects in LFW according to knowns, known unknowns and unknown unknowns sets. 51
4.7 Division of subjects in LFW according to training, gallery and probes sets. . . . 51
4.8 Division of subjects in VGG2 according to knowns, known unknowns and unknown unknowns. 51
4.9 Division of subjects in VGG2 according to training, gallery and probes. 51
4.10 Division of subjects in IJB-C according to knowns, known unknowns and unknown unknowns. 51
4.11 Division of subjects in IJB-C according to training, gallery and probes. 52
4.12 Unprotected closed-set & open-set system performance in LFW, VGG2 and IJB-C. 54
4.13 Protected closed-set & open-set system performance without fusion strategy in LFW. 55
4.14 Protected closed-set & open-set system performance with fusion strategy in LFW. 58
4.15 Protected closed-set & open-set system performance with fusion strategy in VGG2. 58
4.16 Protected closed-set & open-set system performance with fusion strategy in IJB-C. 62
4.17 Average time of processing in training, enrollment and matching on FaceNet face vector. 65
4.18 Comparison of unprotected, protected without and with fusion strategies closed-set identification systems in terms of IR(%)@1 performance. 66
4.19 Comparison of unprotected, protected without and with fusion strategies open-set identification system in terms of DIR(%)@FAR=1%. 66
4.20 Comparison of unprotected, protected without and with fusion strategies open-set identification system in terms of DIR(%)@FAR=0.1%. 67
4.21 Accuracy comparison with state of the art. 68

4.22 Summary of the proposed method performance on FaceNet. 68

5.1 An example of secret retrieval matching with gallery. 80
5.2 TIR@0.1%MIR of FCI key retrieval (%). 82
5.3 Summary of Identifier Retrieval Time Cost of one probe with $p = 10$ and $p = 15$. 86

Abbreviations

AD	**A**uxiliary **D**ata
BTP	**B**iometric **T**emplate **P**rotection
BC	**B**iometric **C**ryptosystem
CFV	**C**haff-less **F**uzzy **V**ault
CMC	**C**umulative **M**atching **C**haracteristic
DIR	**D**etection and **I**dentification **R**ate
FCI	**F**ace **C**ryptosystem for **I**dentification
FR	**F**ace **R**ecognition
FIR	**F**ailure **I**dentify **R**ate
HD	**H**elp **D**ata
IR	**I**dentification **R**ate
LSH	**L**ocality **S**ensitive **H**ashing
LIoM	**L**earing based **I**ndex-of-**M**ax hashing
MIR	**M**is**i**dentify **R**ate
PI	**P**seudonymous **I**dentifier
TIR	**T**rue **I**dentify **R**ate

Chapter 1

Introduction

Can you prove who you are? This is a typical and critical question asked in so many situations. Personal identification is required by a wide variety of applications such as bank counter, customs check, smartphone unlocking, or email login. There are a lot of ways you can prove your identity. You can generally prove your identity by 1) *what you have*, such as the badge, credit card, ID card, or passport. 2) *what you know*, such as PIN, passwords. However, the aforementioned methods require you to remember some numbers/words in your mind or take a physical card with you all the time, which can cause inconvenience in today's rapid life. The traditional password or ID card is far from meeting the needs of reliability and safety because a password is easy to be lapsed in memory [4] and a card is easy to be lost, stolen, guessed, or forgery.

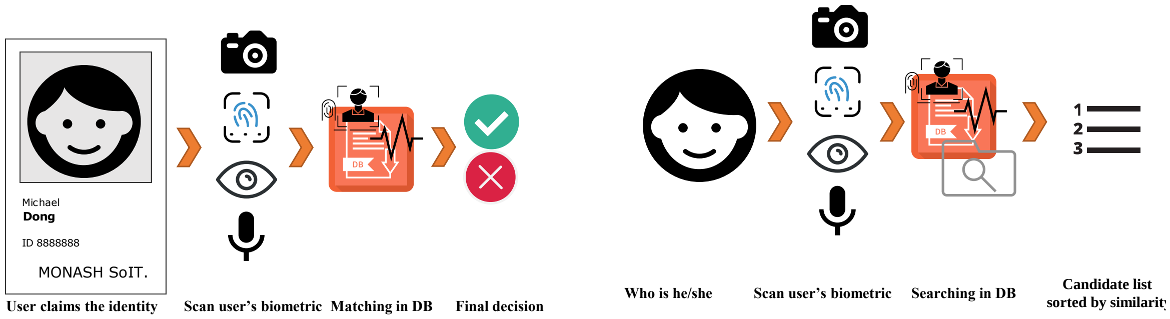
One alternative way to alleviate the above issues is biometrics, or in other words, identification based on *who you are*. According to Morris's definition in 1875, biometrics refers to the combination of ancient Greek "bios" (life) and "metron" (measure), which indicates the biological measurements that are bind strongly to the individual identity. Biometrics can be classified as physical and behavioral biometric according to the measurement and statistical analysis characteristics. The biometrics used for recognition systems are normally unique (i.e., the characteristics of any two people should be different), robust (i.e., the features do not change with time), collectible (i.e., the feature can be quantitatively collected), natural(i.e., all people comes with such traits, instead of artificially generated) and reliable (e.g., high credibility and the high accuracy).

Among various biometric modalities, face is one of the most popular biometric traits (along with fingerprint and iris) used in biometric recognition. This is because that face recognition (FR) has its special advantages:

- **Non-intrusive**. FR can be easily deployed without disturbing people's normal behavior. Users do not need to press the sensor like fingerprints or align their eyes with the iris scanner like iris recognition.
- **Convenience**. In general, a common camera can be used to collect facial images without the need for special complicated equipment. Image acquisition can be made in seconds.
- **Friendliness**. FR is human compliant, International Civil Aviation Organization (ICAO) compliant. The method of recognizing a person's face is consistent with human minds, and both humans and machines can use face images to identify a person. Fingerprints, iris, or finger-vein can not be recognized by a human being without participating in special training.
- **Contact-less**. To collect fingerprints, the system generally requires a finger to contact the collection device, which is neither hygienic nor easy. While there is no direct contact with the device in the face image acquisition process.
- **Scalability**. FR's employment can lead to many practical applications, such as application control, face image search, credit card, terrorist identification, etc.
- **Accuracy**. FR can achieve high accuracy while preserving extra information such as age, gender, emotion, etc.

Face recognition system has been deployed by many companies and organizations due to its advantages. Chinese e-commerce firm Alibaba and affiliate payment software Alipay demonstrated the “Smile to Pay” system in 2015. In 2016, a mobile application named “MasterCard Identity Check” is released by MasterCard to let customers confirm a payment using their face. In 2018, AirAsia unveiled an FR system, namely “FACES”, Malaysia's first airport facial recognition system, and deployed in the airport to help passengers to board by the self-boarding gate. Currently, FR has been widely deployed by banks, stores, and transportation services provider [5].

A conventional face recognition system is a typical pattern recognition system that is composed of preprocessing, feature extractor, and matcher modules. From an operation point of



(a) Verification (a) Identification

Figure 1.1: Two face recognition working modes.
(adopted from [1])

view, face recognition technology can be used either to verify who they claim to be (1-to-1 match or verification/authentication) or to automatically identify the N subjects enrolled in a system (1-to-N match or identification), as shown in Figure 1.1 (a) and (b), respectively. For verification mode, an identifier[1] such as ID, name, etc., is required alongside face biometric for identity claim, whereas identification demands solely face input. Some of the practical usage examples are listed in Table 1.1. In reality, border control, criminal watching, and surveillance systems are usually designed as identification systems, while passport checking, access control, and attendance checking use authentication.

Table 1.1: Examples of FR system.

Scenario	Application case
Identification	Border control
	Criminal watch list
	Conference ID admin
	Surveillance camera
Authentication	Passport checking
	Building access control
	Attendence admin
	Bank customer authentication

The identification can be further divided into open-set and closed-set identification [6]. Closed-set assumes that all query subjects are enrolled or registered in the gallery, while open-set contains unknown identities not enrolled in the gallery. Currently, a lot of FR researches focus on the closed-set protocol, however, the closed-set assumption cannot be true in the real world for most FR systems. Instead, unknown identities can be included in the probes, and

[1] identifier and secret are used interchangeably in this thesis.

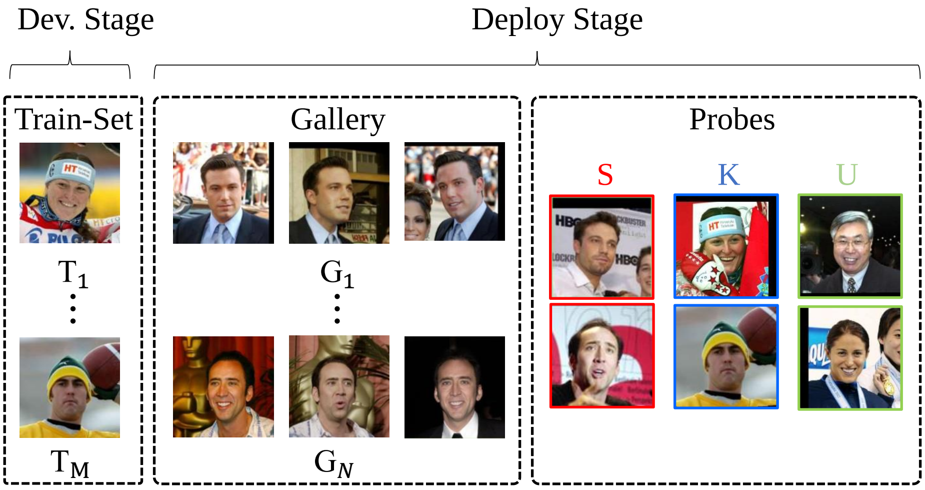
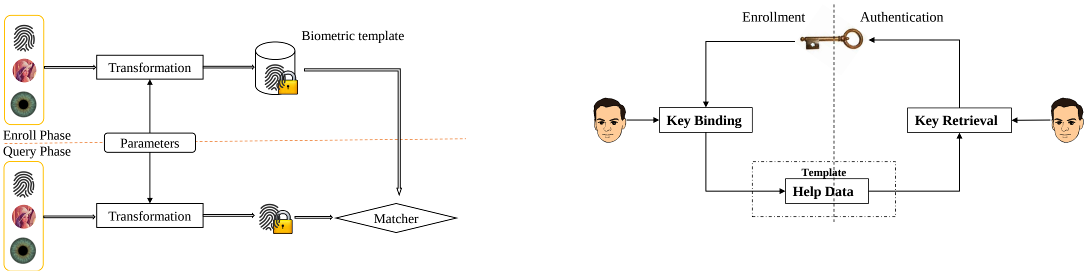


Figure 1.2: The differences of closed-set and open-set face identification. Three sub-types of probes can appear in the probe set, including 1) Known probe (S), subjects include in the gallery; 2) Known unknown probe (K), subjects used for training but not include in the gallery; 3) Unknown unknown probe (U), subjects neither in training set nor gallery. Samples are from LFW dataset.

thus the FR system should be able to reject/ignore those queries. Figure 1.2 illustrates the difference between closed-set and open-set face recognition.

Employing the subject's biometrics as an identity credential typically results in improved comfort and convenience. However, except for answering the question "Can you prove who you are?", another problem is also critical to raise out: "Can others masquerade you?' Suppose a person's biometric template stored in the database is intercepted or stolen by an attacker. In that case, the attacker might reconstruct the biometric traits based on the compromised template, then masquerade or monitor that person. This will lead to severe security and privacy risks. Considering that biometrics are strongly associated with the owner, they cannot be changed when compromised, thus aggravating security and privacy concerns.

It is highly demanding to have a secure and privacy-preserving biometric system. Biometric template protection (BTP) schemes are invented to tackle the above issues [7–10]. BTP schemes permanently keep the enrolled biometric data as secret by performing a non-invertible transformation to protect the original data. The emerging BTP schemes can be broadly divided into feature transformation (also known as cancellable biometrics) and biometric cryptosystems. Cancellable biometrics can be sub-categorized as salting and non-invertible transform according to the transformation function's characteristics. The salting scheme is similar to the traditional salting algorithm in cryptography, which concatenates a random salt r with a secret k, then



(a) Cancellable Biometrics

(a) Biometric Cryptosystem

Figure 1.3: Two biometric template protection categories.

stores the hash $H(r+k)$ in the database, where $H(\cdot)$ is a hashing function. A protected biometric template (pseudonymous identifier (PI)/auxiliary data (AD)) is extracted in a biometric salting scheme by integrating a user-specific key such as a password or random numbers with the biometric data. Non-invertible transformation schemes employ a one-way transformation function to generate the protected template or PI[2], thus grants the irreversibility of a biometric system even if any parameter of the transformation function is revealed.

On the other hand, Biometric Cryptosystem (BC) mainly includes two approaches: (1) key generation, which generates a key from the biometric feature data; (2) key binding, which secures a key (e.g., PIN, private keys, public keys, a hash of message for digital signature, etc.) using the biometric feature, and the key can only be released if the biometric query sample is the right original person. Besides template protection purposes, BC is also used to serve for the secret (keys) management purpose. The BC technique guarantees the biometric applications can be deployed under privacy-preserving and secure policies, such as the EU General Data Protection Regulation (GDPR) [11].

A typical biometric cryptosystem accepts both identifier (ID, for instance) and biometrics. The ID is meant to retrieve the encrypted data of a specific subject, and then biometric is used to reclaim the secret embedded in the encrypted data. Therefore, the biometric cryptosystem can be considered as an instance of a 1-to-1 match or verification system; in a sense, the system returns correct secret (yes) or null (no).

Even though there are many BTP schemes, it is still unsatisfactory to achieve a balance between performance and security. Several critical criteria are defined in the ISO/IEC 24745

[2] The terms PI and AD and "protected template" are used interchangeably in this thesis, as they are the secure form of plaintext biometric templates.

standard [12] on BTP: irreversibility, revocability, and unlinkability. Revocability requires the system can issue new protected templates to replace the compromised ones. Unlinkability requires that infer any information by matching two protected templates from two different applications is computationally impossible. Irreversibility means that the retrieving of the original biometric data from stored biometric templates should be computationally infeasible. Meanwhile, the employing of BTP should preserve the matching accuracy performance compared to that of before-transformed counterparts.

1.1 Problem Statement

Thanks to the technological advancements in imaging devices, FR applications have been deployed in many places. However, due to the privacy and security issues discussed above, the FR application is still far from secure and reliable.

Face template security: Deep face feature is invertible. Despite the deep face feature is showing excellent accuracy in the face recognition task, its vulnerability in terms of privacy and security is of great concern to the public. In [13], a neighborly de-convolutional neural network (NbNet) is designed to reconstruct face images from their deep face features successfully. When a face template is stolen, permanent compromise is inevitable since biometric characteristics are largely immutable. Furthermore, the same unprotected biometric source enrolled in multiple databases for different applications is completely correlated. An adversary can perform cross-match to track and potentially monitor personal activities if one biometric template is compromised [14, 15].

Limitation of identifier-based biometric cryptosystem: Currently, the most typical biometric cryptosystem requires both identifier (e.g., ID) and biometrics. More specifically, the ID is utilized to retrieve the corresponding encrypted data of a specific subject. Then biometric is used to decode the secret embedded in the encrypted data retrieved from the previous step. Such a biometric cryptosystem is less user-friendly since an identifier has to present in such a verification process. The biometric cryptosystems with the identifier are accompanied by a high risk of identity loss, such as losing the id, token, etc.

Low accuracy of open-set FR: Closed-set face identification assumes all probe samples used in the query stage are registered in the gallery. On the contrary, open-set face identification includes identities that may not be enrolled in the gallery. The system should reject or ignore the corresponding unknown identities in the query [16]. Currently, it is still a big challenge to have an accurate open-set FR system.

1.2 Research Question

This thesis's main research question is: ***How can we develop an identifier-free biometric cryptosystem?*** The main question is subdivided into several sub-questions listed:

1. **How to design a face BTP technique to meet those desired requirements?** Designing a proper BTP scheme is still a big challenging task. Take one latest proposed BTP scheme Index-of-Max (IoM) hashing [17] as an example. IoM transforms the features into integers or a subspace, which can also be represented as a binary form for efficient matching speed. Despite theoretical advantages, the IoM hash codes must be long enough to achieve high accuracy performance and high reliability [17]. However, long hash code with higher accuracy will lead to more information leakage, thus cause privacy risks. Having a satisfied BTP technique that can preserve or improve the accuracy while ensuring security remains to be a big challenge.

2. **How to design an identifier-free cryptosystem?** Conventional biometric cryptosystem, which is designed for verification (1-to-1 matching) settings, is limited to certain usage cases. An identifier-free cryptosystem can alleviate such limitations and brings convenience to users. Besides, the security of traditional fuzzy vault is based on uniform mixing of user biometric data points and Chaff points, which will lead to high compromising risks [18–20]. An identifier-free cryptosystem with higher security is demanded.

3. **How to achieve high accuracy for the open-set system?** It is reported that the accuracy of FR usually drops in open-set settings [21]. How to improve the accuracy of the open-set protocol for a given BTP enabled FR system is still a big challenge.

1.3 Research Objectives

In this project, the below-listed objectives have been achieved to tackle the formulated research questions:

1. Design learning-based hashing scheme to improve accuracy performance while gaining strong privacy-preserving capability.

2. Build an identifier-free cryptosystem.

3. Study various techniques to improve the system accuracy.

1.4 Contribution Overview

This is a new research topic since no study is available to integrate both identification and verification systems in one body and incorporate biometric cryptosystem and template protection technologies.

In this thesis, a secure face cryptosystem without an identifier is proposed. The system integrates cancellable biometrics enabled 1-to-N search module that returns the top k id wherein one of them is associated with the claimant. They serve as an identifier to the next biometric cryptosystem (1-to-1) module. Therefore, the templates in the system are securely protected by the template protection technologies. The whole system is further fueled by machine learning technology to enable better accuracy, performance, and efficiency. The system has been evaluated in terms of verification performance, efficiency, security, and privacy.

The overall contribution can be summarized as:

1. To protect and compress deep learned face features, a learning-based Index-of-Max (LIoM) hashing is proposed. The LIoM transformed face features into a compact code vector. LIoM is a one-way locality-sensitive-hashing inherited from the random IoM [17], the major distinction between LIoM and IoM is the projection matrices, where the former are learned from training data based on a specifically designed loss function to ensure that samples from the same identity can produce similar hash codes. An AdaBoost-based sequential learning mechanism is also introduced to boot the performance.

2. To address the limitation of the conventional fuzzy vault, a chaff-less fuzzy vault (CFV) is proposed by coupling LIoM. CFV inherits classical fuzzy vault as a secret-biometric binding scheme via a finite field polynomial, and the security is based on computation hardness of polynomial reconstruction. LIoM hash codes are utilized in CFV as (ordered) point set instead of unordered point set in the conventional fuzzy vault. Besides, a Chaff set is not required for genuine set concealment in the proposed CFV.

3. To alleviate the limitation of the verification-based biometric cryptosystem, a novel identifier-free facial cryptosystem for identification (FCI) that only requires sole face as input for identifier retrieval is designed. The FCI is composed of a 1-to-N search subsystem and a CFV subsystem. Such an identifier-free face cryptosystem will benefit the face recognition systems and strengthen security and privacy protection.

4. It is worth highlighting that LIoM, CFV can work independently. For instance, CFV can be deployed by incorporating with other integer or binary-based features; the 1-to-N search subsystem can also be deployed independently if there are no secret binding demands in real settings. Besides, though the thesis adopts face as the demonstration biometric modality, the proposed LIoM, CFV, and FCI can also be applied to different biometric modalities such as fingerprint, iris and etc.

5. A systematic study is performed for a large-scale face identification problem in both open and closed-set evaluation protocols. The security of the proposed system is also evaluated and analyzed.

1.5 Practical Usage

Unlike other biometrics, face recognition application is versatile. It can be employed on not only identity management, but also other applications, such as entertainment, targeted advertisement, forensic, healthcare, etc. Therefore, face template protection is essential, and it is one of the essential parts in protecting security leakage and privacy intrusion of face systems.

The research outcome of this thesis can be applied to both secure face template protected verification and identification systems. On the other hand, it can also be used in key management system. The development of secure face verification and identification technology is expected to strongly impact and complement face recognition industry. This research will directly beneficial to the face recognition vendors who oblige to heighten the security and privacy protection of their products or systems. With this, the technology will further impact the downstream industry such as mobile devices, health care, consumer electronic, cloud service, and military etc. which require secure identity management.

1.6 Organization

The remaining chapters of this thesis are organized as follows:

Chapter 2 gives a literature review of deep model based face feature extractor, cancellable biometrics, biometric cryptosystem, and closed-set/open-set face identification. Since features extracted from deep models are utilized in this research, several state-of-the-arts face representation techniques based on deep learning are reviewed. The latest cancellable biometrics proposed on the face are also reviewed.

In chapter 3, a learning-based locality-sensitive hashing, namely learning-based Index-of-Max (LIoM), is proposed to protect the biometric template. A face identification (1-to-N search) system is built based on LIoM. Specifically, we utilized deep neural networks to extract discriminative face features; then, the deep features are compressed and secured by the proposed LIoM. Along with LIoM, we also explore several fusion strategies to improve performance.

In chapter 4, the LIoM and face identification system proposed in chapter 3 are tested on large-scale unconstrained face datasets with a closed-set/open-set face identification protocol. In addition, template protection criteria are also evaluated.

Chapter 5 is another key chapter. This chapter demonstrates a face cryptosystem for identification (FCI) where only sole input biometric is needed. The FCI composes a 1-to-N search subsystem from Chapter 3 and a 1-to-1 match chaff-less fuzzy vault (CFV) subsystem. The first subsystem stores N facial features protected by LIoM hashing and enhanced by a fusion module for searching accuracy. When a face image of the user is presented, the subsystem returns top k match scores, and thus, the corresponding vaults in the CFV subsystem will be activated. The 1-to-1 matching occurs among k vaults alongside query face, and an identifier associated with the user will be retrieved from the correct matched vault. The FCI system is evaluated under large-scale public unconstrained face datasets with respect to its accuracy and computation cost. Besides, it is also proved that the proposed system can prevent several high-risk attacks.

Chapter 6 discusses the future directions of this research and concludes the thesis.

Chapter 2

Literature Review and Background Study

Normally features used for face recognition can be divided into two main categories, i.e., handcraft and deep model based features. This chapter aims to review existing deep model based feature extractors due to its high performance achieved recently.The most recent works of face feature protection schemes such as cancellable biometrics and biometric cryptosystem are being discussed. Finally, closed-set and related works of open-set face identification are reviewed.

2.1 Deep Learning based Face Feature Extraction

With the rapid developments in GPU hardware, big data, and novel algorithms, deep FR techniques have fostered numerous startups with practical applications in the recent five years. An FR module consists of data preprocessing (augmentation and normalization), deep feature extraction, and similarity comparison. This thesis focuses on the latest deep feature extraction since we merely utilize deep face features in this thesis. We refer the readers interested in face alignment, detection, anti-spoofing, data preprocessing, database development, and feature classification methods of deep learning based face recognition to comprehensive survey papers such as [22] and [23].

Face Recognition is slightly different from other object classification tasks due to the particularity of faces: Numerous face images of large amounts of people make obtaining all classes for training impractical. Intra-personal variations could be more extensive than inter-personal differences due to different poses, illuminations, expressions, ages, and occlusions. Therefore, deep neural networks have not only been introduced but also adjusted for FR from two aspects. More extensive face databases are collected from the data aspect, and images are preprocessed to improve learning ability. The other is from the algorithm aspect, for which novel architectures and loss functions are designed to promote discrimination and generalization capability.

Face recognition algorithm is a very long-standing research in computer vision since 1964 [24]. Face recognition technology evolves from geometry [24], subspace representation (e.g. eigenface [25]), local descriptor (e.g, local binary pattern (LBP) [26]), shallow learning and more recently, deep learning approach [27]. The deep learning approach is especially exciting as it brings a remarkable quality leap to the overall face recognition technology [22].

In 2014, DeepFace [28] and DeepID [29] achieved state-of-the-art verification accuracy in LFW, for the first time surpassing humankind performance in the unconstrained scenario. Since then, deep-learning-based approaches became a hot research topic and achieved transcendental feature invariance progressively through nonlinear filters' stacking. Existing deep network architectures, such as convolutional neural networks (CNNs), deep belief networks (DBNs) [30], and stacked autoencoders (SAEs) [31], normally simulate the human brain perception process and can represent high-level abstractions by multiple layers of nonlinear transformations.

For deep feature extraction techniques, network architecture and loss function play a crucial role in determining its eventual performance. A lot of CNN architectures have been proposed and show excellence in the ImageNet challenge, such as AlexNet, VGGNet, GoogleNet, and ResNet [32–35]. In FR those architectures are widely used and have a deep influence on the current state-of-the-art deep FR systems, such as AlexNet based DeepFace, GoogleNet based FaceNet, and ResNet-based SphereFace.

Although a good CNN architecture can achieve a better performance, another factor, i.e., loss function design, is more important for deep feature embedding. The most common loss function in object recognition is softmax, which tries to encourage features' separability. However, softmax loss does not show good performance for the FR system since intra-variations could be larger than inter-differences of face images. Many works focus on creating novel loss functions to make features not only more separable but also discriminative. In general, they can be divided into three major categories (also see Table 2.1):

Table 2.1: Latest deep models of three loss types.

Loss Types	Description	Latest deep models
Softmax loss and its variations	Directly using softmax loss or modifying it to improve performance	NormFace [36]
Euclidean-distance-based loss	Compressing intra-variance and enlarging inter-variance based on Euclidean distance	Facenet [37],Deepid3 [38]
Angular/cosine-margin-based loss	Learning discriminative face features in terms of angular similarity	Arcface [39],Cosface [40]

- Euclidean-distance-based loss: compressing intra-variance and enlarging inter-variance based on Euclidean distance.

- Angular/cosine-margin-based loss: learning discriminative face features in terms of angular similarity, leading to potentially larger angular/cosine separability between learned features.

- Softmax loss and its variations: directly using softmax loss or modifying it to improve performance, e.g., L2 normalization on features or weights as well as noise injection.

In this project, we shall focus on deep features generated from Google FaceNet [37] and InsightFace (also known as ArcFace) [39], which will be discussed in the section 3.4.1.

2.2 Cancellable Biometrics

Cancellable biometrics protects the user's biometric data by employing features distortion transformation function (usually, it is a one-way transform). If the transformed biometric template is compromised, a new template can be generated from the same user by employing different transformation function parameters. In this section, several relevant schemes for cancellable face templates are reviewed. The summary of the cancellable biometrics reported on face is shown in Table 2.2.

Random projection (RP) is a process of projecting feature vectors from n dimensions to m $(n \gg m)$ dimensions $(n \gg m)$ in the Euclidean space by using random matrices [41]. RP is based on Johnson-Lindenstrauss lemma (J-L lemma) [42] which proves that points from a high-dimensional space can be embedded into low-dimensional space while preserving the distance approximately. The orthogonal projection matrix is one projection f proved and proposed in [43]. Briefly, Gram-Schmidt orthogonalization is performed on a $n \times m$ random matrix to generate a matrix $R \in R^{n \times m}$. Then, the feature vector $x \in R^n$ is projected onto $y \in R^m$ as $y = \sqrt{n/m} R^T x$. Specifically, the projection matrix R could be generated easily from Gaussian distributed sequences which are proven to have the characteristic of orthogonality [44].

BioHashing, an instance of RP, is a well-known scheme of salting-based generic cancellable biometrics scheme which is applied to face images [45]. Generally, BioHashing is a two-factor BTP technique based on user-specific token and biometric features and followed by a discretization procedure. The n bit BioHash code c of a biometric feature vector $x \in R^N$ is computed as $c = Sgn(\sum xb_i - \tau)$, where $Sgn(\cdot)$ is a signum function, and τ is an empirically determined threshold, $b_i \in R^N$ and $b_i \in R^N, i = 1, \dots, n(n \leq)$ is orthogonal pseudo-random vector. The Hamming distance is computed between two hash codes to indicate the similarity between two biometric vectors. A new template for the identical biometric feature vector can be re-issued by replacing it using newly generated pseudo-random numbers. However, BioHashing impractically assumes that the pseudo-random numbers would never be compromised, which is impractical, and hence there are high risks for BioHashing under key-stolen scenario [46].

A hybrid approach for face template protection was proposed in [47] based on random projection, class distribution preserving transform, and hash function. Specifically, face features are firstly transformed by a cancelable transformation such as random projection. Next, a discriminability enhancement transform is applied to improve the accuracy performance, and a binary template will be generated after the transform. Finally, a biometric cryptosystem such as a hash function is employed to protect the template.

Random permutation is another common approach to generate cancellable biometric template [48, 49].The feature vector is permutated with a randomly generated key. In [48], for example, permutation matrix is used as a parameterized transformation function in [48] to generate cancellable face templates. In [49], the principal component analysis (PCA) and independent component analysis (ICA) coefficients are extracted from face images and permuted by ID-specific parameters. And then, a feature level fusion is performed to generate the cancellable face templates. As for random permutation, the permutation key is assumed to be securely stored, which is impractical. If the key is stolen, the face template would be vulnerable for attackers. However, authentication accuracy is preserved since permutations are merely rearranging the feature vector.

Bloom filter is another generic transformation function been applied to handle face template [50]. In the bloom filter scheme, the biometric feature is mapped to a bit array b with several independent hashing functions, where b is a bit-array of length n. $b \in [0, 1]^n$. Specifically, $k(k \ll n)$independent hash functions denoted as by $h_1, h_2, \ldots, h_k$ are pre-defined first. Each element of a data set S is hashed by using those the hash functions and the resulting hash result is derived as k indices. Finally, set all k indices of the bit-array b to unity. At the verification stage, the bit-array of the query element y is matched with the stored template by hamming distance to indicate the similarity between two biometric data.

The work in [51] proposes a cancellable scheme for iris indexing based on Bloom-filter and search trees, and the scheme can also be applied to face images. In this scheme, the IrisCode is first randomly permutated, and Bloom-filter is applied to generate binary hash code as the PI. The enrolled templates are organized into tree-based search structures. Specifically, N enrolled templates are split evenly and assigned to T trees firstly; next, each node of a tree is created through a union of templates, i.e., the binary OR of all the subsequent children nodes. The root node of a tree is computed by ORing all the templates assigned to this tree. A small portion of all the constructed trees is pre-selected in the retrieval stage by comparing the probe and root nodes. The searching is done on the selected trees in a binary search manner by matching the probe with the tree's nodes and choosing the path with the best score. When a leaf is reached, a final comparison will be made to make the final decision.

Index-of-Max (IoM) is a new recently proposed ranking-based locality sensitive hashing technique for template protection [17], and then recording the indices of the maximum value is output as the hash code (A detailed explanation in section 3.3.3).

Deep Table-based Hashing (DTH) [52] is an instance of LSH based CB schemes for face templates. The DTH employs an end-to-end trained CNN to generate binary hash code from the raw face image directly. A new PI can be re-issued by re-shuffling the hash table associated with the network. While the DTH is shown to offer high accuracy, it is unrealistic and insecure as the entire network is trained directly from the enrollees' face images.

SecureFace [53] proposes a randomized CNN, which can generate a binary PI directly based on the face image input and user-specific key (SD) for verification. The network is trained independently from the enrollee's face images to avoid the security issue mentioned above. The highlight of the SecureFace is that it rectifies the SD management issue by securing the SD with the fuzzy commitment [54], where the latter is primarily designed for secret protection.

Table 2.2: Summary of selected cancellable biometrics on face.

Reference	Technique	Dataset	Result	Drawbacks
[45]	BioHashing	FERET[55]	EER=0.002	information leakage under key compromised
[49]	Random permutation	AR face[56]	-	security depends on key
[57]	Random projection (RP)	AR face	EER=10.9	performance degradation
[58]	Random projection	ORL[59], GT[60]	EER=0	performance degradation
[61]	BioHashing	Eigen-face[62]	EER=0 (stolen token)	information leakage under key compromised
[50]	Bloom filter	DDMB[63]	EER=5.50	fragile to brute force attack
[51]	Bloom filter indexing	DDMB	EER=5.50	fragile to brute force attack
[17]	Ranking based RP	FVC2002DB1[64]	EER=0.22	long hash codes required
[47]	Hybrid approach	FERET	EER=8.55	complicate implementation
[52]	Deep Table-based Hashing	YouTube Faces[65], FaceScrub[66]	EER=.0048(72-bit)	overlap identity training
[53]	Randomized CNN and Secure Sketch	CFP[67]	GAR (%) @ (FAR=0.1%)=96.87	deep based complex system

However, the matching performance could be affected due to the complicated interaction of cancellable transformation and fuzzy commitment.

2.3 Biometric Cryptosystem

Traditional cryptography guarantees to communicate in a manner that meets the following objectives - confidentiality, data integrity, authentication, and non-repudiation. Conventional cryptography technology utilizes keys, such as identification cards or licenses, instead of a person for authentication. The cryptography system can only be secure under a large size secret key situation. In addition, such a long secret key can easily vanish in the human mind. Hence a simple password is often employed to encrypt the public keys. This will cause a potential risk of attack on the password to retrieve the public keys. Concerning the drawbacks, biometric authentication is an ideal option to build a reliable and comfortable key management system. Both biometrics and cryptography are complementary for modern key management applications. Thus the integration of them lead to the new schemes: Biometric cryptosystems (BC) or Biometric Encryption. Generally, a BC:

- Encrypt the original templates to a helper data (Auxilary Data)
- Apply error-correcting coding methods to handle intra-class variance
- Require input in finite fields

According to the utilization or generation of the key, BC can be classified as key generation and key binding. BC either binds a key to a biometric or generates a key from the biometric and stores the help data in the database instead of the biometric data. Retrieving the key or the original biometric data should be computationally hard based on the help data, which means there is no or limited information leakage about the key and biometrics from the helper data. In BC, the key is released or reconstructed only if the right person's biometric data do the verification. Several popular BC schemes are proposed in the literature, such as fuzzy vault and fuzzy commitment under the key binding group, secure sketch, and fuzzy extractor under the key generation group.

Fuzzy Commitment

The fuzzy commitment scheme [54] is motivated by cryptographic bit commitment. In traditional cryptographic bit commitment, a sender usually commits an encrypted message to the receiver. Assume a stock market broker Alice and investors Bob as an example [68], Alice wants to convince Bob that her investment strategy can achieve a good profit rate:

Bob said: "Choose five stocks for me, please. If they are profitable, I will go with you."

Alice said: "If I choose five stocks for you, you can invest in them instead of paying me, that's unfair. Why don't I show you the stock list I chose last month?"

Bob: "How do I know that you didn't change last month's stock list? I can only trust the stock list you selected now because you can not change them. Before we sign the contract, I will not deploy your stock list. Believe me!"

Alice: "I would rather tell you the stock I chose last month. I won't change, believe me."

Alice wants to commit a prediction to Bob but does not reveal her prediction until some time later. On the other hand, Bob wants to make sure that she hasn't changed the prediction content after Alice sends the prediction to him. Usually, the cryptographic bit commitment can be implemented by using asymmetric passwords.

First, Bob generates a random bit string R and sends it to Alice. Then Alice generates a message consisting of bits b she wants to promise (b may actually be a few bits) and a random string of Bob (R). She encrypts it with a random key K and sends the result, denotes $enc(b+R,K)$, back to Bob. $enc(b+R,K)$ is the promise which Bob can't decrypt. When Alice decides to release her message (the stock list), Alice sends a key to Bob. Then Bob decrypts the message to reveal the bit. He can verify the validity of the message by matching his random string.

However, biometric data is stochastic, which is hard to integrate with the bit commitment. For example, the same individual's biometric data has a large intra-class variation and may vary in different acquisition sessions due to noise or other environmental conditions. Tradition cryptography can not be simply applied to encrypt the biometric data.

To eliminate the noise and intra-class variations, Error Correction Code (ECC), which is based on error-tolerant mechanisms, can be employed with biometric data. ECCs consists of a set of codewords C, where each codeword $c \in C$ is an n-bit sequence and wherein the k-bit messages $m \in M(n > k)$ are mapped information. The $n - k$ bits, namely parity bits, can be utilized to restore the corrupted codeword.

The idea of bit comment can be employed in biometric cryptosystem context with the help of ECCs. Juels and Wattenberg proposed a fuzzy commitment scheme in 1999 [54], which is an extension of bit commitment and based on ECCs and fuzzy biometric data. The fuzzy commitment scheme consists of two steps. i.e., commitment and de-commitment. In the commitment step, given a biometric vector ω, a set of n bit codewords C, select a codeword $c \in C$, where C is a set of n bit codewords generated from certain ECCs, and the length of c and ω are equal. The difference between the biometric and the codeword is defined as $\delta = \omega - c$, then the commitment will be: $\{hash(c), \delta\}$, where the $hash(\cdot)$ usually is a one-way hashing function. Since the $hash(\cdot)$ is a one-way hashing, the commitment $\{hash(c), \delta\}$ will not leak any information about the biometric data.In the de-commitment step, given a input biometric data denoted as ω', calculate a codeword c' from the commitment, i.e., $c' = \omega' - \delta = \omega' - \omega + c$. The codeword c' can be restored as the original c by the EEC system if the distance between ω' and ω is smaller than a certain threshold, i.e., $dist(\omega' - \omega) < \theta$, where θ is a threshold, $dist(\cdot)$ is a distance function, such as Hamming distance.

Different biometric modalities can be used in fuzzy commitment, such as iris [69], face [70–72], fingerprint [73] and etc. Commonly the biometric feature used in fuzzy commitment is represented in binary vector, and the δ will be calculated as $\delta = \omega \oplus c$ where $\oplus$ is XOR.

The theoretical soundness of fuzzy commitment stems from its error tolerance mechanism that is intended to address biometric variation. The security relies on the reconstruction complexity of codeword c. However, in reality, it is often challenging to find the perfect ECCs, and it is challenging to validate the security over the non-uniformity of binary biometrics. Fuzzy commitment is commonly associated with iris biometrics [69] as it is represented in binary string form, e.g., IrisCode [74] and it is well recognized as the most discriminative biometrics. For face biometrics, the challenge to apply fuzzy commitment (or known as helper data scheme, HDS by some authors [75]) is that face biometrics has significant intra-class variation than iris. This makes ECC design becomes more challenging. Furthermore, template conversion from real-value vector or matrix (vectorial biometrics) to binary form is a non-trivial problem [76], as it may cause severe information loss [77]. Several attempts on this challenge are [75, 78–80] where most of the works focus on face features design and quantization.

Fuzzy Vault

To overcome the drawbacks of fuzzy commitment, a new BC scheme, namely fuzzy vault, is proposed for unordered biometrics by Jules [81]. The fuzzy vault consists of two steps (see Figure 2.1): vault encryption and decryption. In the encryption step, Alice stores the secret K in a vault and locks it with the unordered set A. In the decryption step, Bob can unlock the same vault with an unordered set B to access the secret K if most elements in A and B have coincided together. Specifically, given a secret K, a secret sharing polynomial $P(x)$ is constructed over a finite field less than k degree by encoding the K as the coefficients. Then the projection $P(A)$ of the unordered biometric feature set A is computed on the polynomial P to obtain a finite point set $(A, P(A))$ and are collectively known as genuine set G where $\|G\| = t$. At last randomly generated points (chaff points), denoted as , $(a, b) \in C$ where $\|C\| = c \gg t$, are generated. The union set of G and C forms a vault V where $\|V\| = n$.

During decryption or secret retrieval, given an unordered biometric feature set B as the query, if enough elements of A and B are overlapped, the polynomial P can be reconstructed based on the feature set B by Lagrange interpolation and then obtain key K by using error correction code technology. However, if only a small proportion of elements have coincided, it will be difficult to reconstruct P, so K cannot be obtained. This algorithm's security is based on polynomial reconstruction because it works with unordered sets, so it is especially suitable

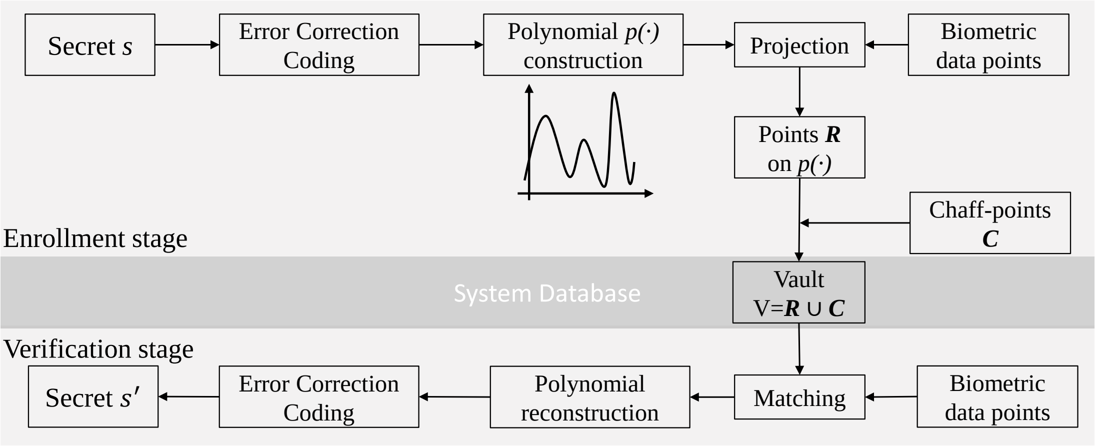


FIGURE 2.1: Diagram of fuzzy vault.

for unordered biometric data such as fingerprint minutiae and can tolerate errors between data sets.

For key retrieval performance, the success rate is often related to the intra-class variation of the biometric feature and the correction capacity of Error Detection (such as cyclic redundancy check CRC)/Correction Code (error correction code ECC). It will be difficult to find a suitable ECC which has enough error correction capacity due to the large intra-variation. Besides that [82–85] reported that secret retrieve rate is strong negatively correlated with the security.

The security of the fuzzy vault is based on several assumptions: 1) uniform mixing of G and C where $c \gg t$, which implies separating genuine set from chaff set in the vault would be difficult if not impossible; 2) small intra-class variations of A and A' to ensure they are sufficiently close for reconstruction and 3) computation hardness separating G from V provided $t \ll \sqrt{(p-1)n}$. In practice, it is usually hard to generate a Chaff set that meets the first assumption [18–20]. In addition, the intra-class variation of biometrics is usually large especially for face biometric. This impacts severely on secret retrieval performance, and it is also reported that poor retrieval rate affects security negatively [86, 87]. That is to say: poor biometrics can reduce the attack complexity of the fuzzy vault. The third assumption can also be violated simply by the brute attack. The brute-force attack on fuzzy vault can utilize list-decoding to enumerate valid solutions, thus permit secret recovery from the errors well beyond the barrier [88] . In [18, 19] the brute-force is launched and only 2^{32} to 2^{35} iterations are required to reconstruct the secret. Due to the low polynomial order constraints $p < 10$ attributed to the

TABLE 2.3: Summary of fuzzy commitment and fuzzy vault.

BCs	Description	Related References	Remark
Fuzzy commitment	Bit commitment + biometrics	[69–73]	Security relies on the reconstruction complexity of codeword c
Fuzzy vault	Polynomial construction and reconstruction	[18, 82–85]	Security relies on polynomial reconstruction.

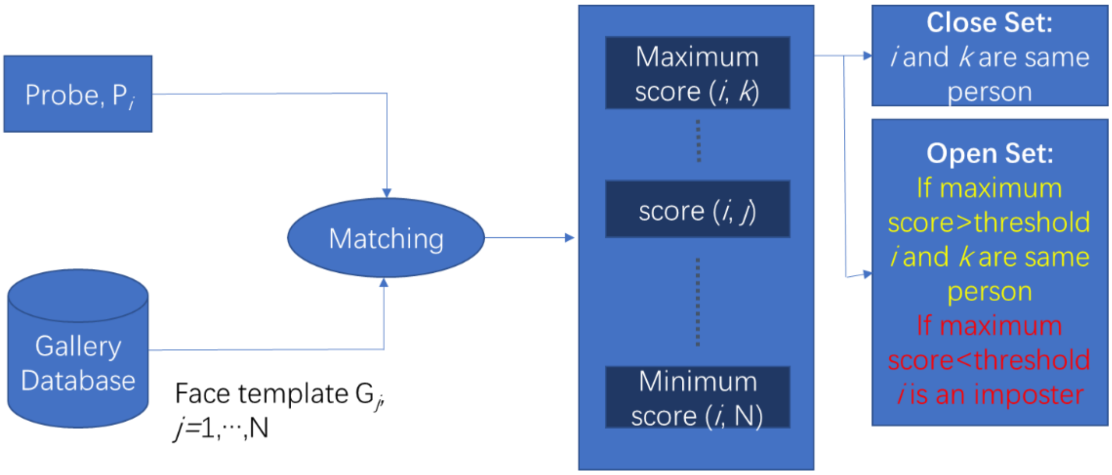

Probe, P_i
Gallery Database
Face template G_j, j=1,⋯,N
Matching
Maximum score (i, k)
score (i, j)
Minimum score (i, N)
Close Set:
i and k are same person
Open Set:
If maximum score>threshold i and k are same person
If maximum score<threshold i is an imposter

FIGURE 2.2: Close and Open Set Face Identification (1-to-N Match) System.

scarcity of minutiae and the limited number of chaff points, the fuzzy vault is fragile under brute-force attack [89]. A summary of BCs is shown in Table 2.3.

2.4 Closed-set and Open-set Face Identification

Unlike the face verification task, which is only restricted to one-to-one matching, face identification, or 1-to-N matching where N is the number of subjects in the gallery, it is a much more challenging problem when N turns huge. Face identification can be further divided into closed-set and open-set problems as depicted in Figure 2.2 where the latter is a much less explored subject.

Many works envisioned to solve closed-set face identification problem have been around for years [90–92]. The US National Institute of Standards and Technology (NIST) releases its face recognition internal benchmark every several years. Both research labs and commercial vendors participated in the NIST contest. A lot of FR algorithms are reported to be able to achieve excellent accuracy. According to the Multiple Biometric Evaluation (MBE) report in 2010 [93], three top-ranked Commercial Off the Shelf (COTS) can reach 82%-92% accuracy rate when matching the face probe against a gallery with 1.6 million identities.

However, the same COTS system failed to preserve the accuracy rate on a closed-set identification task with images from the Labeled Faces in the Wild (LFW) [94, 95]. Though the gallery only contains few thousand identities, the rank-1 accuracy of the COTS system deteriorated to about 56% [94]. Best-Rowden et al. [95] showed that simple thresholding of a COTS algorithm works perfectly for verification but does not provide satisfactory open-set identification performance. This suggests that a lot of algorithms may not reach good accuracy performance under the unconstrained situation in spite they can show good performance in constrained situations, the unconstrained task remains to be solved.

On the other hand, open-set face identification receives attention gradually. The open-set FR research has been recognized for over a decade so far [96]. The typical exploration is carried out on LFW, due to the bias of the LFW standard evaluation protocol, which doesn't utilize all information of the dataset and also include the open-set situation. To establish a better large-scale evaluation protocol, Liao et al. [21] proposes a protocol that utilize all face images in the LFW dataset under closed-set verification and open-set identification scenarios. However, in [21], several state-of-art FR algorithms were evaluated with open-set identification protocol, and it proves that the open-set FR is still a challenge for large galleries. Another new protocols for open-set unconstrained face identification has been also proposed by [95].

To tackle the open-set face identification problem, a number of studies [21, 38, 95, 97–102] explores to introduce a similarity score as the rejection threshold. However, the thresholding solution can only work well under verification, while a satisfactory performance on open-set face identification can not be achieved [95] .

Another line of research exploits a one-class classifier to filter out the imposter and only permit those enrolled for matching. This, in turn, changes the open-set to closed-set setup where the latter is known to be simpler and yield better performance. One-class support vector machine (SVM) is a common option [103–105] as it looks promising to train a classifier with only enrolled subject data. However, one-class SVM is not scalable since one-class SVM can not be trained on a large dataset, and the training time is costly. Santos et al. explored five different methods in [106] to filter out the imposter in a probe. Among those methods, one is focusing on discriminating faces between the known and background sets, while the remaining methods focus on identification responses. However, neither of them can attain good accuracy performance in a large gallery.

A formal definition of the open-set problem is outlined by Bendale et al. [107]. Apart from that, an algorithm that can reject a query as an outlier when it is too far from the training sample, namely Nearest Non-Outlier (NNO) was proposed in their work. NNO is an algorithm that can update the model continuously with additional unseen objects, and it is unnecessary to retrain the model. The query will be classified as unknown if all training classes reject the probe.

More recently,a lightweight classifier based on the extreme value machine (EVM) [108] [16] is proposed. EVM is established to obtain a probability of sample inclusion of each probe sample with respect to a gallery. Based on the statistical extreme value theory (EVT) calibrations over margin distributions, EVM can be applied to varying data bandwidths and also show superior performance on open-set protocol compared with thresholded similarity methods. Besides that, [16] also outlined a refined open-set face identification protocol on LFW that considered known probes, known unknown probes and unknown unknown probes.

Another task that pretty close to the face identification problem is face search (retrieval) in the social media [109, 110]. However, face search differs from face identification in the background set, which is included in the gallery without a label or identity. The inclusion of massive background images complicates the searching problem. The gallery that comprises the union of known identities and background set can go up to a multi-million scale. Early studies on face search problems primarily focused on faces captured under constrained conditions and of a small scale, e.g., the FERET dataset [55]. However, due to the growing need for strong face recognition capability in the social media context, ongoing research is focused on large-scale in-the-wild datasets.

In [111], Wu et al. propose a face inverted indexing system based on a component-based local face representation. The aligned face images are split into small blocks based on the facial landmarks. Then each block is quantized into a visual word by an identity-based quantization scheme. The candidate images corresponding to the face probe are retrieved from the inverted index of visual words. Furthermore, by leveraging human attributes, Chen et al. improved the search performance [112].

Table 2.4: The list of works on Face Search and Face Retrieval.

	Probe		Gallery		Dataset	Protocol
	#Images	#Subjects	#Images	#Subjects		
Wu et al. [111]	220	N/A	1M+	N/A	LFW + web facesa	Close
Chen et al. [112]	120	12	13,113	5749	LFW + Pubfig	Close
	4300	43	54,497	200	LFW + Pubfig	Close
Miller et al. [113]	4000	80	1M+	N/A	FaceScrub + Yahoo	Close
Yi et al. [114]	1195	N/A	201,196	N/A	FERET+ web faces	Close
Yan et al. [115]	16,028	N/A	116,028	N/A	FRGC + web faces	Close
Klare et al. [116]	840	840	840	840	LFW	Close
	25,000	25,000	25000	25000	PCSO	Close
Best et al. [95]	10,900	5153	3143	596	LFW	Close + Open
Liao et al. [21]	8,707	4249	1000	1000	LFW	Close + Open
Wang et al. [101]	7,370	5507	80M+	N/A	LFW + web faces	Close + Open
	14,868	4500	80M+	N/A	IJB-A + web faces	Close + Open

Recently, a multi-million scale face search system is developed by [101]. In their work, which is the largest face search experiment conducted to date, the gallery encompasses 80+ million images mixed with known subjects and background sets where the former is just a small fraction of the total gallery. The system adopts deep facial features extracted from a Convolutional Neural Network (CNN) and followed by product quantization for features compression. A simple re-ranking algorithm based on the score level fusion and COTS is incorporated to gain better performance.

A summary of the work on face search and retrieval is summarized in Table 2.4.

Chapter 3

Face Identification System with Learning-based Index-of-Max Hashing

In this chapter, to protect face features, the learning-based Index-of-Max (LIoM) hashing is proposed. To be specific, the existing random IoM hashing is advanced to a data-driven based hashing technique, where the hashed face code can be made compact and matching can be easily performed by the Hamming distance, which can offer highly efficient matching. Text, a face identification system with LIoM is developed. Besides, several fusion strategies are explored to address the open-set face identification problem.

3.1 Introduction

Over the past years, face recognition has become prevalent in the interest of surveillance and convenience. Face recognition systems can work under 1-to-1 match mode and 1-to-N match (identification) mode. The identification FR can further be divided into closed-set and open-set identification. All probes are enrolled in the gallery in closed-set settings, while some probes not enrolled in open-set settings [117]. Research on closed-set identification has been commonplace for a few decades, whereas open-set face identification has been gaining attention only recently, since it is a more practical scenario for face identification [117].

On another hand, there is a public concern about the privacy risks when the face template stored in the gallery is compromised [118]. For instance, if an attacker can intercept a person's face template, it might be used to impersonate that person (e.g. [119], [120]). This will raise severe security and privacy risks. Considering that biometrics are strongly associated with the owner, and thus cannot be changed when compromised, these concerns are aggravated.

Therefore, to design a practical open-set face identification (1-to-N matching) system, the system should fulfill the following essential properties:

1. The system must be able to differentiate between the identities enrolled in the gallery and unknown identities.

2. Once the system decides that the probe is not an imposter, the system is capable of determining the identity of a subject, where a top k ranked list retrieved from the gallery may suffice to identify the subject. The value of k is application-dependent. For instance, $k = 1$ is required for access control applications and perhaps $k = 200$ for forensic, surveillance and watch list applications. Conventionally, a system that uses $k = 1$ is called an identification system.

3. The face feature representation and matcher are to be as simple as possible for speedy matching yet still give a high accuracy performance for large-scale identification.

4. The face features should be protected without jeopardizing the original system performance. This demands non-invertible transformation on the face representation, yet the original face features information should be largely preserved to avoid performance deterioration.

Besides that, the facial images may be acquired from the unconstrained environment or "in the wild" [94], where the pose, lighting, expression, age, image resolution can be varied widely and occlusion may occur. The face identification system should be able to achieve satisfactorily accurate performance in the unconstrained environment. This induces more difficulties in the system design.

In this chapter, we address the above challenges by utilizing a deep neural network to extract discriminative face features, then the deep features are compressed and secured by the learning-based locality sensitive hashing method, namely learning-based Index-of-Max (LIoM), whose ancestor random IoM [17] was designed for biometric template protection. The hashed face code is compact enough and enables a simple matcher to be used for matching.

3.2 Motivations and Contributions

In this chapter, we aim to explore the large-scale open-set face identification problem in-the-wild. Our pipeline is composed of a deep neural network-based face feature extractor, a feature hashing module and a fusion module, which are meant to address the massive numbers of face images that are acquired under the unconstrained environment, privacy-protected facial features compression, and the open-set evaluation, respectively. For more realistic and comprehensive evaluation, we follow a new open-set evaluation protocol suggested by [117], wherein the probe samples can be categorized into three types: known, known unknown, and unknown unknown (detailed in section 4.1.2.1, chapter 4).

The open-set protocol employed corresponds to a lot of scenarios in real life. For instance, a surveillance camera that can capture people and compare their faces with the watch-list of criminals (known probes) in the database was deployed at a concert given by Jacky Cheung in China, and snagged many fugitives[1]. The singer Jacky Cheung and his staff were not on the watch-list but could be detected by the cameras regularly, and would not be recognized as criminals. Thus, this group of people can be regarded as known unknown probes when training the model, since they are not people of interest for the system. As for the normal audience not on the criminal list, they needed to be ignored by the system, and were taken as unknown unknown probes. Apart from that, we also evaluate the system based on the closed-set protocol [117] for benchmarking.

The main contributions of this scheme in this chapter are as follows:

1. A systematic study is performed for a large-scale face identification problem in both open and closed-set evaluation protocols. For the former, a more practical scenario that includes three types of probes is considered.
2. A supervised learning-based IoM (LIoM) hashing for deep facial feature protection and compression is presented. The LIoM is shown to be able to deliver a more accurate

[1] https://www.washingtonpost.com/news/innovations/wp/2018/05/22/facial-recognition-cameras-lead-to-arrest-of-a-man-wanted-for-allegedly-stealing-17000-worth-of-potatoes/

performance with less hashed code than for random IoM hashing, especially when LIoM is trained by the same group of identities.

3. Several feature-level and score-level fusion strategies for the face identification problem are explored.

3.3 Preliminary

In this subsection, three hashing concepts upon which LIoM is based, namely Locality-Sensitive-Hashing, Winner-Take-All hashing, and index-of-max hashing are presented.

3.3.1 Locality Sensitive Hashing (LSH)

LSH is different from conventional and cryptographic hashing. In traditional hashing, even one single bit difference will lead to completely different hash codes. In contrast, the LSH aims to map similar items into the same "buckets" with a maximized probability. LSH can be used to reduce the dimensionality of the high-dimensional data by projecting the original data into a much smaller number of "buckets" [121] (Figure 3.1).

The LSH family $\mathcal{H}$ is defined as follows:

Definition 3.1 (Locality Sensitive Hashing [122])**.** A LSH is a probability distribution on a family $\mathcal{H}$ of hash functions such that$\mathbb{P}_{h\in\mathcal{H}}\left[h\left(X\right)=h\left(Y\right)\right]=S(X,Y)$. With a similarity function S define on the collection of object X and Y.

The key ingredient of LSH is the hashing of object collection X and Y by means of multiple hash functions h_i. The use of h_i enables decent approximation of the pair-wise distance of X and Y in terms of collision probability. LSH ensures that X and Y with high similarity renders higher probability of collision in the hashed domain; on the contrary, the data points far apart each other result a lower probability of hash collision. Specifically, given the LSH family $\mathcal{H}=\{h_i:\mathbb{R}^d\rightarrow S\}$ which maps data points from $\mathbb{R}^d$ to a bucket $s\in S$, and the similarity function $d(\cdot)$, the LSH family satisfies the following conditions for any two given points $X,Y\in\mathbb{R}^d$:

$$\begin{aligned}\mathbb{P}_{h\in\mathcal{H}}(h(X)=h(Y))\leq\gamma, if\ d(X,Y)<\alpha\\ \mathbb{P}_{h\in\mathcal{H}}(h(X)=h(Y))\geq\delta, if\ d(X,Y)>\beta\end{aligned}\tag{3.1}$$

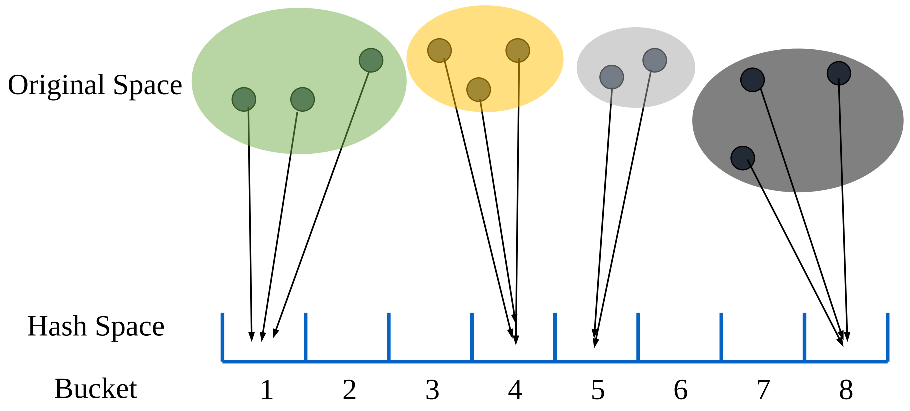


FIGURE 3.1: Locality sensitive hashing (From [3]).

where $\delta > \gamma$.

3.3.2 Winner-Take-All Hashing

Winner-Take-All Hashing (WTA) is one instance of LSH proposed in [123]. In WTA hashing, the partial order statistics-based embedding is computed from the input data as the final hash codes. As a non-linear transformation based on implicit order, WTA can tolerate certain numerical perturbations and preserve the matched items' similarity after the transformation. A diagram of the WTA hashing is shown in Figure 3.2. The overall WTA hashing procedure can be summarized into five steps:

1. Perform H random permutations on the input vector with dimension d, $x \in R^d$.
2. Select the first k items of the permuted x.
3. Choose the largest element within the k items.
4. Record the corresponding index values in bits.
5. Step 1 – step 4 is repeated m times, yielding in a hash code of length m, which can be compactly represented using $m \cdot \log_2 k$ bits.

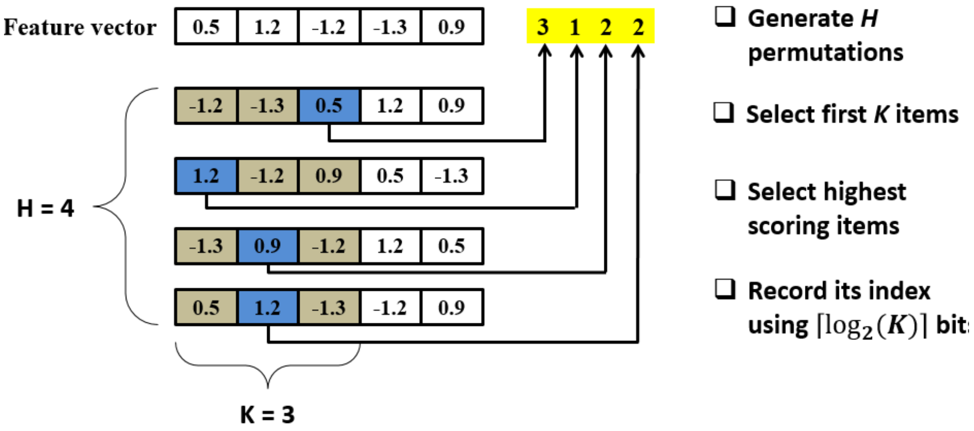


FIGURE 3.2: Winner-Take-All Hashing.
(Credit to Dr. Jin Zhe)

3.3.3 Index-of-Max (IoM) Hashing

IoM hashing, inspired by WTA, is a means of generic cancellable biometrics that can be perceived as a special instance of LSH portray in def. 3.1. In general, IoM embeds the biometric features onto the rank space. In rank space, the similarity functions S of LSH is characterized by the rank correlation of indices recorded from the maximum values (top-ranked) of two transformed biometric features. The rank representation nature of IoM hashed features empower strong concealment to biometric features. While as an LSH instance, the accuracy performance of biometric features can be largely preserved after hashing. This trait can be justified through below lemma and its connection with LSH.

Lemma 3.2 ([124]). *For* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ *be the unit vector* $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$ *at angle* θ. *Let* $\rho = \mathbf{u} \cdot \mathbf{v} = \cos\theta$ *and* $\mathbf{r}_1, \dots, \mathbf{r}_q$ *be a sequences of iid standard Gaussian random vectors, the probability for* $\mathbf{u}$ *and* $\mathbf{v}$ *be not separated by* $\mathbf{r}_1, \dots, \mathbf{r}_q$ *is designated as* $k_q(\mathbf{u}, \mathbf{v})$. *The Taylor series expansion*

$$k_q(\mathbf{u}, \mathbf{v}) = \sum_{i=0}^{\infty} a_i(q)\rho^i \tag{3.2}$$

of $k_q(\mathbf{u}, \mathbf{v})$ *around* $\rho = 0$ *converges for all* ρ *in the range of* $|\rho| \leq 1$. *The coefficients* $a_i(q)$ *are non-zero and their sum converges to 1. The first three coefficients can be expressed as* $a_0(q) = \frac{1}{q}, a_1(q) = \frac{h_1^2(q)}{q-1}$, *and* $a_2(q) = \frac{q h_2^2(q)}{(q-1)(q-2)}$, *where* $h_i(q)$ *is the expectation of* $\phi_i(x_{\max})$ *where* $\phi_i(\cdot)$ *be the normalized Hermite polynomials and* $x_{\max}$ *is the maximum entry of* q *iid standard Gaussian random variables.*

Remark 3.3. Let $H : \mathrm{S}^{d-1} \rightarrow \quad \{1, \ldots, q\}^m$, for $\mathbf{u}$ and $\mathbf{v} \in \mathrm{S}^{d-1}$ and $H(\mathbf{u}) = [\arg\max_{j=1\ldots q} \left\langle \mathbf{w}_j^1 \mathbf{u} \right\rangle, \ldots, \arg\max_{j=1\ldots q} \left\langle \mathbf{w}_j^m \mathbf{u} \right\rangle]$ where $\left\{ \mathbf{w}_j^i \in \mathbb{R}^d \mid i = 1 \ldots, m, j = 1, \ldots, q \right\} \sim \mathcal{N}(0, I_d)$. The expectation can be estimated as:

$$\mathbf{E}\left(\frac{1}{m}\sum_{i=1}^{m} 1_{h_i(\mathbf{u})=h_i(\mathbf{v})}\right) = \mathbb{P}\left\{\arg\max_{j=1\ldots q}\left\langle \mathbf{w}_j^l \mathbf{u}\right\rangle = \arg\max_{j=1\ldots q}\left\langle \mathbf{w}_j^l \mathbf{v}\right\rangle\right\} \tag{3.3}$$

where $h \in \mathcal{H}$. As m becomes large, $k_q(\mathbf{u}, \mathbf{v})$ can be approximated by:

$$\mathbb{P}\left\{\arg\max_{j=1\ldots q}\langle \mathbf{w}_j^i \mathbf{u}\rangle = \arg\max_{j=1\ldots q}\langle \mathbf{w}_j^i \mathbf{v}\rangle\right\} \approx k_q(\mathbf{u}, \mathbf{v}) \tag{3.4}$$

Hence, $h(\mathbf{u}) = \max_{j=1\ldots q} \left\langle \mathbf{w}_j^i \mathbf{u} \right\rangle$ resembles the LSH where the similarity function $S(\cdot, \cdot) = k_q(\mathbf{u}, \mathbf{v}) \in \mathbb{R}$ [125, 126].

The method to generate IoM hashed code can be condensed into a two-step procedure as follows and its pseudo code is given in Algorithm 1:

1. For a feature vector $x \in \mathbb{R}^d$, generate a set of random matrices $\mathbf{R}_i \in \mathbb{R}^{d \times q}, i = 1, \ldots, m$ where each entry in the matrix is drawn from standard Gaussian distribution $\mathcal{N}(0, 1)$.
2. Perform $y_i = \mathbf{R}_i \mathbf{x} \in \mathbb{R}^q, i = 1, \ldots, m$ which is equivalent to linear random projection [38].
3. For each $y_i \in \mathbb{R}^q$, determine the maximum value and record its index value, $h_i \in [1\ q]$. Thus, a set of IoM hashed code can be obtained by repeating the step 2 and 3 and yield $\mathbf{h} = \{h_i | i = i, \ldots, m\}$.

Given enrolled and query biometric vectors as $\mathbf{x}$ and $\mathbf{x}'$, respectively. The similarity function of IoM hashing is given as $S(\mathbf{x}, \mathbf{x}') = \frac{1}{q} + \sum_{i=1}^{\infty} a_i(q)(\cos\theta)$ where $\cos\theta = \frac{\mathbf{x} \cdot \mathbf{x}'}{\|\mathbf{x}\| \|\mathbf{x}'\|}$ and $a_i(q)$ is the coefficient that satisfies $\frac{1}{q} + \sum_{i=1}^{\infty} a_i(q) = 1$. Operationally, since the entries of IoM hashed vector are integers, the matching score is mere the total number of collisions (number of matched entries) of $\boldsymbol{T}$ and $\boldsymbol{T}'$ and normalized by m. Therefore, the similarity function $S(\mathbf{x}, \mathbf{x}')$ for IoM hashing is simply normalized Hamming distance over a finite field, a variant of Jaccard similarity coefficient.

Algorithm 1 IoM Hashing

```
Input:
    Feature vector $\mathbf{x}$ with $d$ dimensions, number of Gaussian random matrices $m$ and the number
    of Gaussian random projection vector $q$.
Output:
    Hashed code $\mathbf{T} = \{t_i \in \mathcal{U} | i = 1, \ \ldots, m\}$
Generate $m$ Gaussian random matrices $\mathbf{W}^i = (\mathbf{w}_1^i, \ldots, \mathbf{w}_q^i)$, $i = 1, 2 \ldots, m$.
Initialize $i^{th}$ hashed code $t_i$=0
Perform random projection and record the maximum index in the projected feature vector.

for $k$ = 1: $m$ do
    $\overline{\mathbf{x}}^k = \mathbf{W}^k \boldsymbol{x}$
    Find $\overline{\mathbf{x}}_j^k = \max\left(\overline{\mathbf{x}}^k\right), j = 1, \ldots, q$
    Then $t_i = j$ ($j$ refers the index of $\overline{\mathbf{x}}^k$)
end for
```

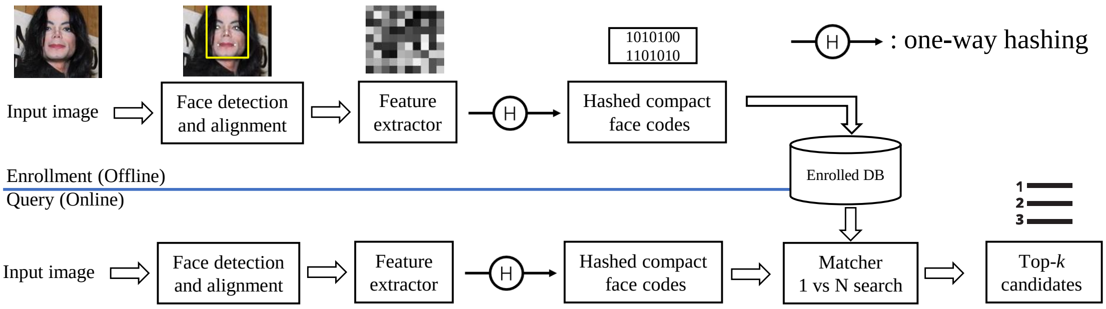


FIGURE 3.3: Face identification module.
(adopted from [1])

3.4 Face Identification Framework

This section outlines a face identification structure, which devised to accelerate the matching while achieving decent accuracy. Figure 3.3 illustrates the system consisting of three parts: i) face feature extraction with deep neural networks for the N gallery samples (offline) as well as for the probe samples (online); ii) face code hashing by means of learning based IoM hashing that transforms the deep face vector into a compact hash codes and iii) fusion and matching that compares the probe against the gallery samples simply with Hamming matcher retrieve the top-k most similar candidates ($k << N$).

3.4.1 Deep Face Feature Extraction

The face feature extractor is essential to extract a robust and discriminative feature representation. This is especially crucial for a large-scale face identification system in the unconstrained environment. In this work, we adopt two pre-trained convolution neural networks dedicated to face recognition, namely FaceNet [37] and InsightFace (also known as ArcFace) [39] (see Table 3.1.We refer deep features as the features that extracted by deep neural networks.

ArcFace loss, which is proposed in [39], is an improved softmax classification loss where the weight vector of j^{th} identity $\mathbf{w}_j$ is L_2 normalized. Specifically, target logit (activation of the output layer before applying softmax function) of the identification branch is re-defined as $\mathbf{w}_j^T \mathbf{z}_i = \|\mathbf{w}_j\| \, \|\mathbf{z}_i\| \cos\theta_j$ where $\mathbf{z}_i$ is the L_2 normalized embedding features of the i^{th} sample, belonging to the j^{th} identity. The normalization of the embedding features and weights makes the predictions only rely on the angle between $\mathbf{z}_i$ and $\mathbf{w}_j$, denoted as θ_j. The prediction of y_i th identity is now mere θ_{y_i} dependence. The ArcFace loss is defined as:

$$\mathcal{L}_{\text{ArcFace}} = -\frac{1}{B}\sum_{i=1}^{B} \log \frac{e^{s\left(\cos\left(\theta_{y_i}+\beta\right)\right)}}{e^{s\left(\cos\left(\theta_{y_i}+\beta\right)\right)} + \sum_{j=1, j\neq y_i}^{N} e^{s\cos\theta_j}} \tag{3.5}$$

where B is the batch size, β is an angular margin introduced to force the classification boundary closer to that specific weight vector, and s is a feature rescaling factor. In this manner, the learned embedding features are thus distributed on a hypersphere with a radius of s. With InsightFace pre-trained with MS-Celeb-1M, a face vector with size 512 can be obtained[2].

Triplet loss in FaceNet is proposed by Google Inc meant for face verification, identification and clustering. Triplet loss takes in a triplet of deep features, (a, p, m), where (a, p) have similar product labels and (a, n) have dissimilar product labels and tunes the network so that distance between anchor (a) and positive (p), $d(a, p)$ to be less than the distance between the anchor (a) and negative $(n), d(a, k)$, by at least distance margin m. To be specific, the triplet loss function is defined as follows:

$$\mathcal{L}_{\text{Triplet}}(a, p, n) = \max(0, m - d(a, p) - d(a, n)) \tag{3.6}$$

The key idea of FaceNet is mapping the facial images to Euclidean (L2) space through convolutional neural networks. Unlike other CNN, FaceNet does not follow the common approach in extracting face features from CNN, such as taking the layer before the output layer as features but perform end-to-end learning from input image space to feature space directly. The spatial coding is then used for recognition and verification. The features learned in this way are highly compact with size 128 or 256. . In this work, a face vector extracted by the

[2] Official implementation (https://github.com/deepinsight/insightface) is adopted in this study.

Table 3.1: Deep features used in this study.

Deep Features	Architecture	Loss Function	Input Images	Output Embedding Size	Training
FaceNet	Inception-Resnet-v1	SoftMax loss	160x160	256	MS-Celeb-1M
ArcFace/Insight Face	LResNet50E-IR	Additive Margin Softmax	112x112	512	MS-Celeb-1M

MS-Celeb-1M [127] pretrained FaceNet with embedding size 256 is utilized. In our research, we adopt David's open-source implementation of FaceNet and its pre-trained model[3].

In this thesis, all facial images are first aligned by MTCNN [128] and then cropped to 160x160 for FaceNet and 112x112 for InsightFace. Images that cannot be aligned will be discarded.

3.4.2 Learning based Index-of-Max hashing

As described in section 3.3.3, simply speaking, IoM hashing encodes a feature vector x as the index of the random subspace $R_i \in \mathbb{R}^{d \times q}$ that generates the largest projected value. The IoM hashing exploits the ranking order among random projected values instead of feature values. Due to the characteristic advantage, the generated IoM hash codes can be resistant to the noise of the biometric. In addition, it also show resilience to the scaling of the biometric vector. At last, IoM hashing is a non-linear projection which grants the security to the biometric system.

Despite theoretical soundness, the IoM hashing often require long enough hash code to achieve similar accuracy performance as its original counterpart. For instance, as reported in [17], a fingerprint vector with size 299 requires 600 hash codes or above to achieve similar accuracy performance with a suitable value of q. This is attributed to IoM hashing relying on random projection, which is not optimized according to the characteristic of the feature vectors. Therefore, it is judicious to replace the random subspace projection with a supervised learning mechanism from the feature vectors [129]. This makes IoM hashing be transformed from data-agnostic to data-driven algorithm, in which we coined learning-based IoM hashing or LIoM hashing. To be specific, a single IoM hashed code can be compactly formulated as a function projection matrix W as:

$$h(x; W) = arg \max_{1 \leq i \leq q} w_i^T x \in \mathbb{R} \tag{3.7}$$

[3] https://github.com/davidsandberg/facenet

where $w_i \in \mathbb{R}^d, i = 1, 2...q$,and $W = [w_1, w_2, \ldots, w_q]^T \in \mathbb{R}^{d\times q}$. w_i is one projection metric which used to perform subspace projection. In this sense, random IoM hashing is a special case of the LIoM where $R \in \mathbb{R}^{d\times q}$ is now replaced with W. Based on such a generalization, a optimized projection, or subspace can be tackle by learning strategy, while the final goal can be regarded as finding the hash functions characterized by the projections W as in (3.7).

Suppose $T = \{x_i \in \mathbb{R}^d\}_{i=1}^M$ be the set of m deep face vectors and $S = \{s_{ij} \in \{0, 1\}\}_{i,j}^M$be the set of pair-wise similarity matrix where $s_{ij} = 1$ signifies pair (x_i, x_j) is similar to each other, while $s_{ij} = 0$ signifies nonmatched pair. The similarity matrix S can be generated by labels or by measuring the distance between two vectors (e.g. Cosine distance). In our program, the label similarity matrix approach is adopted. Given s_{ij} of each training pair, an error function induced by a hash function $h(x; W)$ (or simply h for brevity), can be defined as:

$$\epsilon(h_i, h_j, s_{ij}) = \begin{cases} I(h_i \neq h_j), & s_{ij} = 1 \\ I(h_i = h_j), & s_{ij} = 0 \end{cases} \tag{3.8}$$

where $I(\cdot)$ is an indicator function. The objective of the learning is to find a W which can minimize the accumulation errors over T:

$$\epsilon(W) = \sum_{s_{ij}\in S} \epsilon(h_i, h_j, s_{ij}) \tag{3.9}$$

By introducing the constraint, (3.7) can be written as

$$\begin{aligned} & h(x; W) = arg \max_P p^T W x \in \mathbb{R} \\ & subject\ to\ p \in \{0, 1\}^q, 1^T p = 1, \end{aligned} \tag{3.10}$$

where the constraints enforce that only one entry of non-zero is existed in the hash code h while the remaining are entry of 0. Following (3.10 and 3.8) can be rewritten in a matrix form accordingly,

$$\begin{aligned} \epsilon(W) &= \sum_{s_{ij}=1} (1 - p_i^T p_j) + \sum_{s_{ij}=0} (p_i^T p_j) \\ &= \sum_{s_{ij}\in S} (1 - 2s_{ij}) p_i^T p_j + const \\ &= trace(PAP^T) + const \end{aligned} \tag{3.11}$$

where $P = [p_1, \ldots, p_M]$ is the $q \times M$ matrix, A is $M \times M$matrix and $a_{ij} = 1 - 2s_{ij} \in A$, $const$ is a constant. Unfortunately, since P is non-convex and highly discontinuous with respect to W, (3.11) is difficult to optimize. As a remedy, (3.10) can be conveniently approximated by integrating with softmax function ,

$$h(x; W) \approx SoftMax(Wx) \tag{3.12}$$

where $SoftMax(z)$ is a q-dimensional vector:

$$softmax(z)_j = \frac{e^{z_j}}{\sum_{i=1}^{q} e^{z_k}} \; for \; j = 1, \ldots, q \tag{3.13}$$

It is easy to attest that the softmax function approximations alters the constraints of p_i only while leaving the error function (3.11) unaffected. Note that the (3.11) can be interpreted probabilistically with the softmax approximation. In what follows, the probability of generating same hash code, i.e., same index of the maximum dimension from Wx_i and Wx_j (equivalently h_i and h_j) is

$$\pi_{ij} \equiv Pr(h_i = h_j | W, x_i, x_j) \sum_{k=1}^{q} p_{ik} p_{jk} = p_i^T p_j \tag{3.14}$$

where p_ik and p_jk are the kth element of p_i and p_j, respectively. By introducing the probability, the expected error for two hash codes from pair (x_i, x_j) is:

$$E[\epsilon_{ij}] = \begin{cases} 1 - \pi_{ij}, & s_{ij} = 1 \\ \pi_{ij}, & s_{ij} = 0 \end{cases} \tag{3.15}$$

where ϵ_{ij} is short for $\epsilon(h_i, h_j, s_{ij})$ which defined in (3.8). According to above discussion, the (3.11) can become the expected cumulative error over T. In a nutshell, the overall objective reduces to solving the following problem,

$$\min_{M} \sum_{i,j} a_{ij} \pi_{ij} = trace(PAP^T) \tag{3.16}$$

By introducing softmax approximation, we convert the objective function (3.11) to a continuous function of W, despite the problem remains non-convex. To search the local minima, one can adopt standard gradient descent algorithms. Since full T can be huge, computing the gradient over T could be prohibitive, hence mini-batch stochastic gradient decent algorithm is opted instead. Specifically, the gradient of π_{ij} can be computed as:

$$-\Delta_W \pi_{ij} \propto [(p_i^T p_j)p_i - p_i \odot p_j]x_i^T + [(p_j^T p_i)p_j - p_j \odot p_i]x_j^T \tag{3.17}$$

where $\odot$ is the point-wise product.The mini-batch update rules can be written as:

$$W \leftarrow W + \eta[Pdiag(P_s^T P) - P_s \odot P]X^T \tag{3.18}$$

where η is the learning rate and $diag(\cdot)$ returns diagonal matrix with the elements of the vector on the main diagonal. $X = [x_1, x_2, \ldots, x_M]$ is the $d \times M$ training data matrix,$P = [p_1, p_2, \ldots, p_M]$ is a q$\times M$ matrix containing the softmax vectors of each training vector and $P_s = PA$. After W is found, (3.7) is used to compute the hash code, which ranks from 0 to $q - 1$.

Since each hash code is learned independently with Algorithm 2, the entire m hash codes could be suboptimal. This is due to different random initial solutions leading to the same optimal point, resulting in redundant hash codes. In order to maximize the information contained among m hash codes, the hash functions could be learned sequentially so each hash function can offer complementary information to previous ones [130]. The crux of sequential learning is each hash code can be perceived as a weak classifier that assigns similarity labels to an input pair. The obtained ensemble classifier is related to the Hamming distance between hashing codes. Formally, each weak classifier corresponding to the lth code is

$$sim_l(x_i, x_j) = 1 - Hm(h(x_i; W_i), h(x_j; W_j)) \tag{3.19}$$

where $Hm(\cdot)$ is the bitwise Hamming distance. Then, the Hamming distance of two hash codes with size m can be seen as the vote of an ensemble of m weak classifiers on them. The AdaBoost-based sequential learning algorithm is shown in Algorithm 3. Specifically, a sampling weight $\omega_{ij}^{(l)}$ is assigned to each training pair and is updated before training each new hash function. The projection matrix W is updated in the similar fashion as in (3.18) but weighted by $\omega_{ij}^{(l)}$. When all the hash functions have been trained, the voting results of the related week classifiers are fused with a weighted combination

$$sim_l(x_i, x_j) = \sum_{l=1}^{m} \Phi_l(1 - Hm(h(x_i; W_i), h(x_j; W_j)) \tag{3.20}$$

where Φ_l are the weighted training error of the lth hash function.

Algorithm 2 Learning IoM (LIoM)

```
Input:
    Deep face vectors $X$, pair-wise similarity matrix $S$, subspace dimension $q$.
Output:
    LIoM projections matrix $W$.
Initialization: Random $W$ drawn from zero-mean unit-variance Gaussian distribution.
repeat
    Randomly select a training batch $X_b$ and obtain the batchwise label matrix $S_b$ accordingly

    Set $A = \lambda E - (\lambda + 1)S_b$ /* $E$ is a matrix of ones. */
    Compute $P$ by applying the softmax function to each column of $WX_b$
    Set $P_s = PA^T$
    Update $W$ according to (3.18)
until Convergence
```

Algorithm 3 Sequential learning IoM

```
Input:
    Deep face vectors $X$, pair-wise similarity matrix $S$,length hash code $m$, subspace dimension $q$.
Output:
    LIoM projections $N$
Initialization: Set the weight $\omega_{ij}$ of all pairs to one
for $l = 1$ to $m$ do
    Solve $W$ using Algorithm 2
    Compute hash codes for all samples by IoM hashing
    Calculate the weighted error $\epsilon_l$ where $\epsilon_l = \frac{\varepsilon(W)}{M}$ where $\varepsilon(W)$ is defined in (3.9)
    Evaluate $\phi_l = \ln(\frac{1-\epsilon_l}{\epsilon_l})$
    Update the weighting coefficients using $\omega_{ij}^l = \omega_{ij}^l \exp(\phi_l \in (h_i, h_j, s_{ij}))$
    Normalize $\omega_{ij}$
end for
```

Similar to IoM hashing, the similarity score between two hash codes $\mathbf{h}_\mathrm{x}$ and $\mathbf{h}_\mathrm{y}$ generated by the LIoM hashing can be computed as the ratio of the total number of collisions of two hash codes (the number of matched entries) divided by the total number of entries, which in turn can be represented as the Jaccard similarity $\mathbf{s}(\mathbf{x}, \mathbf{y}) = J(\mathbf{h}_x, \mathbf{h}_y) = \frac{|\mathbf{h}_\mathrm{x} \cap \mathbf{h}_y|}{|\mathbf{h}_x| + |\mathbf{h}_y|}$. Note that the hash code $\mathbf{h}_\mathrm{x}$ and $\mathbf{h}_\mathrm{y}$ can also be stored in binary string form and hence $\mathbf{s}(\mathbf{x}, \mathbf{y}) = 1 - Hm(\mathbf{h}_x, \mathbf{h}_y)$, where Hm$O$ is the bitwise Hamming distance. In this sense, the Hamming distance is regarded as a special case of the Jaccard distance in Hamming space.

In a nutshell, the projection matrices employed in the IoM hashing are generated randomly. In contrast, the projection matrices in LIoM are learned from the gradient descent algorithm, which can deliver a set of optimal ranking subspaces, and hence improve the accuracy over random IoM hashing. While the proposed LIoM is inspired by [129], it has several distinctions as follows:

1) The LIoM hashing transforms deep face features into binary (with only single modality) representation, while the LSRH hashing in [129] aims to transform multimodal (i.e. images and text) data sources into a common subspace (cross-modalities), typically in Hamming space, by exploiting a rank correlation measure.

2) Biometric data is private. LIoM hashing, inherited from random IoM hashing, can protect face templates and retains the original system performance. In contrast, the LSRH hashing in [129] deals with the image retrieval task with publicly available data generally.

3) The LIoM hashing is compact enough to satisfy the efficiency concern on the large-scale open-set face identification problem.

3.4.3 Fusion Strategy

For large-scale face identification, expecting good accuracy of performance with the sole use of hash code is not realistic. Fusion is one popular direction to boost the accuracy (e.g. [131, 132]). While a simple matcher is favored for speedy matching, we introduce and explore several fusion strategies for 1-to-N matching. In this thesis, three samples (LIoM hash codes or simply the hash vector for brevity) per subject are stored in the gallery to achieve better retrieval accuracy, and three fusion options are devised based on the score level and feature-level fusion techniques. Specifically, the first option is designed based on the score-level fusion, while option 2 is devised based on the fusion of hash codes (feature-level fusion) and option 3 is invented based on the fusion of deep face vectors (see Figure 3.4).

Suppose g, p, t to be a face image in the gallery $\mathbf{G}$, probe $\mathbf{P}$ and training-set $\mathbf{T}$ respectively. Two deep face vectors of that person, i.e. I_* and F_*, are derived based on InsightFace and FaceNet, respectively and $* \in \{g, p, t\}$. The face vectors are transformed by LIoM hashing, i.e. $h(I_*)$ and $h(F_*)$.

Option 1: Given a probe face sample p and its corresponding hash codes $h(I_p)$ and $h(F_p)$, two matching scores can be obtained via $S^I_{p,g}$=1- Hm$(h\left(I_p\right), h(I_g))$, $S^F_{p,g} = 1-$ Hm$(h\left(F_p\right), h(F_g))$, where Hm$(\cdot)$ is the Hamming distance. Finally, two sets of scores can be fused via one of the following score fusion rules:

1. $S_{p,g} = mean\left(S^I_{p,g}, S^F_{p,g}\right)$

2. $S_{p,g} = max\left(S^I_{p,g}, S^F_{p,g}\right)$

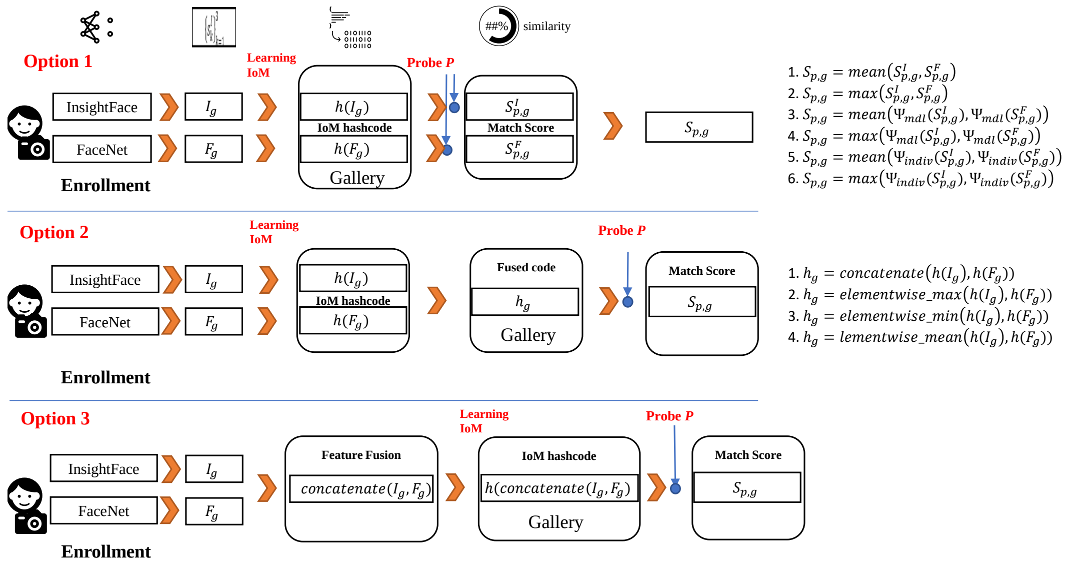

##%
similarity
Option 1
Learning IoM
Probe *P*
InsightFace
FaceNet
I_g
F_g
$h(I_g)$
IoM hashcode
$h(F_g)$
Gallery
$S^I_{p,g}$
Match Score
$S^F_{p,g}$
$S_{p,g}$
Enrollment
1. $S_{p,g} = mean(S^I_{p,g}, S^F_{p,g})$
2. $S_{p,g} = max(S^I_{p,g}, S^F_{p,g})$
3. $S_{p,g} = mean(\Psi_{mdl}(S^I_{p,g}), \Psi_{mdl}(S^F_{p,g}))$
4. $S_{p,g} = max(\Psi_{mdl}(S^I_{p,g}), \Psi_{mdl}(S^F_{p,g}))$
5. $S_{p,g} = mean(\Psi_{indiv}(S^I_{p,g}), \Psi_{indiv}(S^F_{p,g}))$
6. $S_{p,g} = max(\Psi_{indiv}(S^I_{p,g}), \Psi_{indiv}(S^F_{p,g}))$
Option 2
Learning IoM
InsightFace
FaceNet
I_g
F_g
$h(I_g)$
IoM hashcode
$h(F_g)$
Fused code
h_g
Gallery
Probe *P*
Match Score
$S_{p,g}$
Enrollment
1. $h_g = concatenate(h(I_g), h(F_g))$
2. $h_g = elementwise\_max(h(I_g), h(F_g))$
3. $h_g = elementwise\_min(h(I_g), h(F_g))$
4. $h_g = lementwise\_mean(h(I_g), h(F_g))$
Option 3
InsightFace
FaceNet
I_g
F_g
Feature Fusion
$concatenate(I_g, F_g)$
Learning IoM
IoM hashcode
$h(concatenate(I_g, F_g))$
Gallery
Probe *P*
Match Score
$S_{p,g}$
Enrollment

FIGURE 3.4: Different fusion strategies.
(adopted from [1])

where $mean(\cdot)$ and $max(\cdot)$ are the averaging and maximum operations, respectively.

Apart from the above simple score fusion rules, we also incorporate a more sophisticated outlier (an imposter in the open-set identification context) modeling approach – Extreme Value Machine (EVM) [108] to the fusion rule. EVM is essentially a non-linear kernel-free classifier that is designed based on the statistical extreme value theory (EVT) [133]. The EVM fits an EVT distribution per hash vector over several of the nearest fractional radial distances to hash vectors from other classes, and uses a statistical rejection model to model the probability of sample inclusion on the resulting cumulative distribution function. Taking a fixed number of hash vectors and distribution pairs per class that optimally summarize each class of interest yields a compact probabilistic representation of each class in terms of extreme vectors (EVs).

For the open-set face identification problem, EVM is tailored to a similarity function Ψ by letting each hash vector be associated with an identity. In this thesis, Ψ is modeled by two different strategies, i.e. individually and deep model-wise. Generally, the resultant probability of sample inclusion that probe p is associated with gallery g is given by:

$$\Psi(g, p; k, \lambda) = e^{-\left(\frac{s^*_{p,g}}{\lambda}\right)^k} \tag{3.21}$$

where k, λ are the shape and scale parameters of the Weibull distribution, $S^*_{p,g}$ is the similarity score between two face samples which can be $S_{p,g}^{F}$ or $S_{p,g}^{I}$. These parameters can be estimated individually or by the model. To train the Weibull distribution and estimate the parameters of each g in the gallery individually, the distance between g and $\boldsymbol{T}'$ is computed as

$$\operatorname{dist}_g=\left\{0.5 \times d\left(g, \boldsymbol{T}'\right) \mid t' \neq g ; t' \in \boldsymbol{T}'\right\} \tag{3.22}$$

for all training samples in $\boldsymbol{T}'$, where $d\left(g, \boldsymbol{T}'\right)=1-\operatorname{Hm}\left(h(g), h\left(t'\right)\right))$ for each $t' \in \boldsymbol{T}' \cdot \boldsymbol{T}'$ is a subset of $\mathbf{T}$ by excluding the same subject g. A Weibull distribution is fitted to the low tail of dist g for the subject g :

$$\operatorname{dist}_{g, \tau}=\left\{d \mid d \in \operatorname{dist}_g \wedge d<\theta^\tau\right\} \text { with } \theta^\tau=\max _\theta\left|\left\{d \mid d \in \operatorname{dist}_g \wedge d<\theta\right\}\right|=\tau \tag{3.23}$$

where the tail size τ represents a hyperparameter of EVM ($\tau=500$ in this thesis). By modeling the distribution individually, each subject in the gallery will be associated with a set of unique distribution parameters.

Apart from modeling the distribution individually, the Weibull distribution parameters can also be estimated easily model-wise, which only takes account of a different deep model. To train the model, the distances between each g in the gallery $\mathbf{G}$ and all the training subjects g are computed and collected at one time:

$$\text{dist }_{mdl}=\{0.5 \times d(g, t) \mid T \neq g ; T \in \mathbf{T} ; g \in \mathbf{G}\} \tag{3.24}$$

for all training samples, and all gallery samples, except that the distances between the same subjects are excluded. The Weibull distribution is fitted to the low tail of dist $_{mdl}$:

$$\text{dist }_{mdl, \tau}=\left\{d \mid d \in \text { dist }_{mdl} \wedge d<\theta^\tau\right\} \text { with } \theta^\tau=\max _\theta\left|\left\{d \mid d \in \operatorname{dist}_{mdl} \wedge d<\theta\right\}\right|=\tau \tag{3.25}$$

By modeling the distribution deep model-wise, only two sets of distribution parameters are generated, i.e. one for FaceNet and another for InsightFace.

Overall, we denote the final similarity between g and p as $\Psi_{\text{ind}}\left(S_{p,g}^*\right)$ based on the EVM model trained individually for each subject, and $\Psi_{mdl}\left(S_{p,g}^*\right)$ based op the EVM model trained by the corresponding deep model. The fusion strategies of option 1 incorporated with EVM can be devised as follows:

3. $S_{p,g} = mean\left(\Psi_{\text{mdl}}(S^I_{p,g}), \Psi_{\text{mdl}}(S^F_{p,g})\right)$
4. $S_{p,g} = max\left(\Psi_{\text{mdl}}(S^I_{p,g}), \Psi_{\text{mdl}}(S^F_{p,g})\right)$
5. $S_{p,g} = mean\left(\Psi_{\text{ind}}(S^I_{p,g}), \Psi_{\text{ind}}(S^F_{p,g})\right)$
6. $S_{p,g} = max\left(\Psi_{\text{ind}}(S^I_{p,g}), \Psi_{\text{ind}}(S^F_{p,g})\right)$

Option 2: The gallery hash codes derived from deep face vectors are fused at *feature level* with one of the four following rules:

1. $h_g = concatenate(h(I_g), h(F_g))$
2. $h_g = elementwise\_max(h(I_g), h(F_g))$
3. $h_g = elementwise\_min(h(I_g), h(F_g))$
4. $h_g = elementwise\_mod(h(I_g), h(F_g))$

The probe p is also submitted to the same process, i.e. LIoM transformed and feature-level fusion, denoted as h_p, matched with h_g via $S_{p,g} = 1 - \text{ Hm}(h_g, h_p)$.

Option 3: The deep face features I_g and F_g are normalized and concatenated directly and then transformed with LIoM, denoted as $h_g = h(concatenate\,(I_g, F_g))$. The probe pair P is also submitted to the same process to generate the hash code h_p and matched with h_g via $S_{p,g} = 1 - \text{ Hm}(h_g, h_p)$.

3.5 Chapter Conclusion

This chapter addressed several challenges of large-scale unconstrained face identification or the 1-to-N face matching problem. In particular, we considered open-set identification that consists of three kinds of probes, namely known, known unknown and unknown unknown probes. We utilized deep neural networks to extract face features and transform them to protected hash code via learning-based Index-of-Max (LIoM) hashing for privacy protection. The hash code is compact enough and matching can be simply carried out via the Hamming distance. To compensate the performance degradation due to LIoM hashing, several fusion strategies have been introduced to restore its original counterpart performance.

Chapter 4

Evaluation and Analysis of the Learning-based Index-of-Max Hashing

In this chapter, the large scale closed-set and open-set protocols are utilized to evaluate the LIoM's performance. Three large unconstrained face datasets of increasing complexity: LFW, VGG2 and the IJB-C dataset, are adopted in this study. Besides, other aspects such as storage, matching efficiency, and template protection criteria are also analysed.

4.1 Performance Evaluation

We present a comprehensive evaluation for the proposed LIoM hashing in this section. Our experiments are conducted on the LFW [94], VGGFace2 [134] and IARPA Janus Benchmark-C (IJB-C) [135] datasets (see Table 4.1).

LFW is the first dataset that was designed to study the large-scale unconstrained face recognition problem. The dataset contains 13,233 face images with 5,749 subjects collected from the web. Each face is labeled with the subject name. 1,680 subjects have two or more distinct images and the rest only have one image. The only constraint on these faces is that they were detected by the Viola–Jones face detector.

Table 4.1: Databases used in this project.

Datasets	Characteristics	We use
LFW (2007)	• First database for unconstrained face recognition. • 5k+subject, 10k images. • Collected from the web by the Viola-Jones face detector.	13,233 images, 5749 subjects
IJB-C (2018)	• For unconstrained face recognition and detection. • Emphasis on occlusion and diversity of subject occupation and geographic origin with the goal of improving representation of the global population. • IJB-C dataset is significantly more challenging.	131,967 images, 3530 subjects
VGG2 (2018)	• For unconstrained face recognition and detection. • 9131subjects (train: 8631, test: 500), 3.31 million images. • Downloaded from Google Image Search, large variations in pose, age, illumination, ethnicity, and profession.	train:8631*50 test:500*50

VGGFace2 is another large-scale dataset for unconstrained face recognition and was published in 2018. Images are downloaded from Google Image Search and have large variations in pose, age, illumination, ethnicity and profession. There are 9,131 subjects with 3.31 million images in this dataset.

The IJB-C face dataset is also a very recent dataset dedicated to unconstrained face recognition and detection. It improves upon the previous public domain IJB-B dataset, with a larger dataset size and an emphasis on occlusion and diversity of the subjects' occupation and geographic origin. Therefore, IJB-C is significantly more challenging compared to many existing public face datasets. IJB-C has 31,967 images with 3530 subjects.

In this section, we present the experiments in the following order. First, we evaluate the learning-based IoM, followed by closed-set and open-set identification experiments. We compare the identification performance without LIoM hashing, i.e. uncompressed and unprotected face feature representation, deep face features with the LIoM hashing but without fusion, and with both LIoM hashing and fusion. This is to distinguish how the LIoM hashing and fusion contribute to the face identification in terms of accuracy of performance and time cost. Finally, the accuracy of our proposed system is compared with the latest state-of-the-art techniques.

4.1.1 Learning-based IoM Hashing

To verify the accuracy of the LIoM hashing performance, several evaluations have been carried out with the LFW and VGG2 datasets. The experiments are conducted based on the verification protocol and 1-to-N matching protocol, where the equal error rate (ERR) and mean average precision (mAP) are respectively adopted as performance metrics. For comprehensive

Table 4.2: Dataset configuration for identity-dependent evaluation.

Dataset	Sample Number				
VGG2(1000 subjects)	1st - 10th	11th – 20th	21st - 30th	31st – 40th	41st – 50th
	Training	Probe	Gallery		

comparison, deep face vectors from FaceNet (baseline), random IoM and LIoM hash codes are compared.

Since LIoM hashing is a supervised model, we assess its performance under three scenarios:

(1) **Identity-dependent scenario**: the LIoM hashing is trained and tested by the same subject samples, which means the images in the training and testing sets come from the same subjects but are not overlapped.

(2) **Identity-independent scenario**: the images in the training and testing sets come from different subjects and there is no overlapping between the sets, although all the images come from the same dataset.

(3) **Dataset-independent scenario**: the training and testing sets come from different subjects from different datasets.

To create the training and testing datasets for the identity-dependent scenario, we select the first 1000 subjects with 50 images per subject from VGG2, then the first 10 images of each subject are used for LIoM hashing training and the remainder for testing. Since deep face features and random IoM hashing are learning-free, they are evaluated solely with the testing set (Table 4.2).

To generate the evaluation dataset for the identity-independent scenario, we select the first 4000 subjects and 50 images per subject from VGG2, then the first 2 images of each subject are chosen as the training set. Then, another 4000 subjects and 50 images per subject are selected, and the first 2 images of each subject are chosen as the testing set (Table 4.3). Testing for the dataset-independent scenario is done by LFW.

The EER is calculated based on the following protocol:

- In the testing set, each sample of a subject is matched against the remaining samples of the same subject to compute the genuine scores by the Hamming distance, and this process is repeated for all subjects. In practice, a random number of samples are selected to compute the score, since using all the samples will lead to a huge number of scores.

Table 4.3: Datasets configuration for identity-independent and dataset-independent evaluation.

Dataset	Remark
Training samples	First 4000 subjects of VGG2, 2 images per subject, yielding 8000 testing samples in total.
Identity-independent testing samples	Remaining 4000 subjects of VGG2, 50 images per subject, yielding 200,000 testing samples in total.
Dataset-independent testing samples	Pairs (probe and gallery) generated following LFW protocol.

Based on the collected genuine scores, the false acceptance rates (FARs) can be computed with respect to various predefined threshold values.

- The first sample of each subject is matched against the first sample of the remaining subjects to compute the imposter score by the Hamming distance, then the second sample is matched against the second sample of the remaining subjects and so on. In practice, a random number of subjects are selected to compute the score since using all the subjects will lead to a huge number of scores. Based on the collected imposter scores, the false rejection rates (FRRs) can be computed with respect to various predefined threshold values. The specific numbers of scores generated under different scenarios are shown in Table 4.4.
- EER can be estimated when FAR = FRR.

mAP is a widely used metric for search system evaluation [101]. Unlike EER which only consider the number of correctly matched items while the rank of retrieved items is ignored, mAP takes the ranks into account. Given a set of n face probes i.e. $Q = \{q_i\}_{i=1}^{n}$ and a galley set with N subjects, the average precision of q_i is defined as:

$$\text{avgP}\left(q_i\right) = \sum_j P\left(q_i, j\right) * \left[R\left(q_i, j\right) - R\left(q_i, j-1\right)\right] \tag{4.1}$$

where $P\left(q_i, j\right)$ is the precision at the j-th position for q_i and $R\left(q_i, j\right)$ is the recall rate at the j-th position for q_i. The mean Average Precision (mAP) of the entire probe set Q is:

$$\text{mAP}(Q) = \text{mean}\left(\text{avgP}\left(x_q^i\right)\right), \quad i = 1, 2 \ldots, n \tag{4.2}$$

From Figure 4.1, Figure 4.2 and Figure 4.3, we observe that LIoM hashing can improve the verification performance under different scenarios. In general,

Table 4.4: Number of matching scores under different scenarios.

Scenarios	EER	mAP
Identity-dependent	Genuine scores: $C_{40}^2 \times 1000 = 780k$ Imposter scores: $C_{200}^2 \times 40 = 796k$, 200 subjects taken randomly from 1k subjects	$30 \times 10/2 \times 1000 = 150k$
Identity-independent	Genuine scores: $C_{25}^2 \times 4000 = 1200k$, 25 samples taken randomly from 50 samples for each subject. Imposter scores:$C_{400}^2 \times 10 = 798k$, 400 subjects taken randomly from 4k subjects	4k samples taken randomly are matched with another randomly taken 4k samples to generate 4000k scores
Dataset-independent	Genuine scores: *3k*, Imposter scores: *3k*. The testing is performed according to LFW official pairs.	4k samples taken randomly are matched with all 12k samples to generate 48000k scores

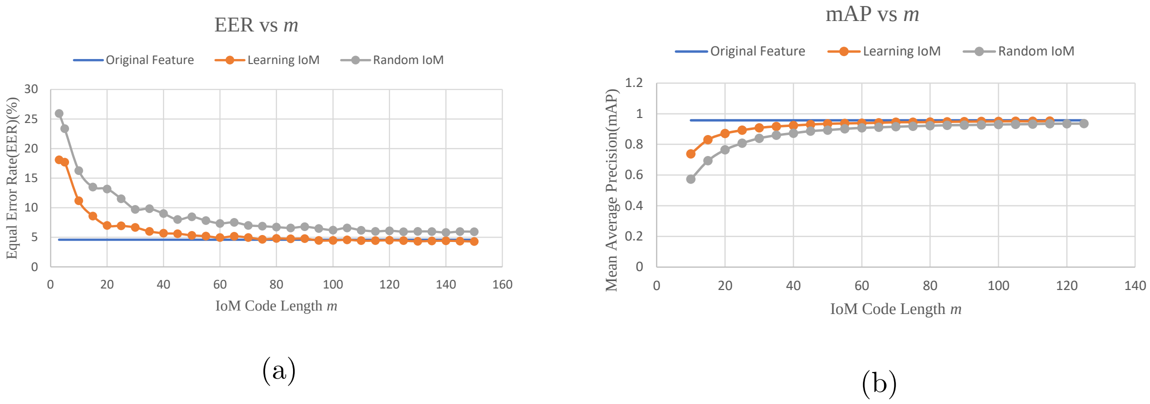
- LIoM hashing significantly outperforms random IoM hashing under the identity-dependent scenario. However, the advantage is not as significant under the identity-independent scenario and the dataset-independent scenario. This could be because LIoM hashing is less capable when generalizing an identity that is not seen in the training data.

- LIoM hashing demonstrates better accuracy of performance than random-IoM under the identity-independent and dataset-independent scenarios, although this is not as significant as under the identity-dependent scenario.

- LIoM hashing requires less hash code than random IoM to achieve the same accuracy of performance for all scenarios. For instance, the former manages to reach EER=5% at $m = 60$, while the latter needs $m = 140$ or more in the identity-dependent scenario.

- LIoM hashing can preserve deep face features better than random IoM hashing in all scenarios.

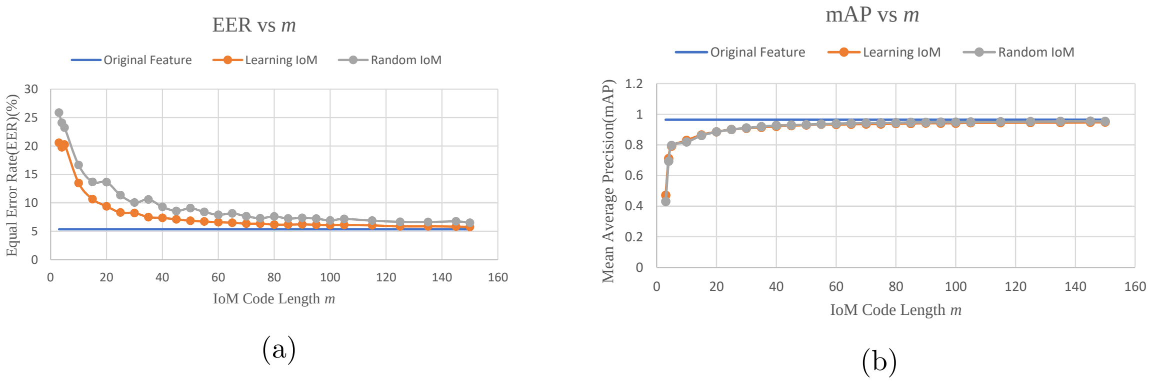
For the 1-to-N matching circumstance, LIoM hashing also outperforms random IoM hashing under the identity-dependent scenario, especially for small m. However, random IoM is comparable under the identity-independent and dataset-independent scenarios, as illustrated in Figure 4.2 and Figure 4.3. In the subsequent experiments, we choose m =100 and q =16 for experiments as they are sufficient to preserve the performance accuracy and for speedy matching.



(a)

(b)

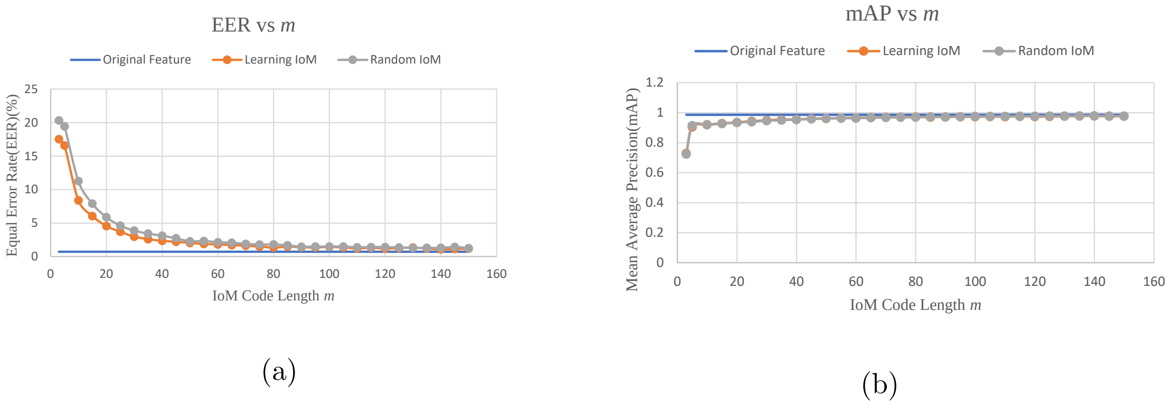
FIGURE 4.1: Performance comparisons under identity-dependent scenario. (adopted from [1])



(a)

(b)

FIGURE 4.2: Performance comparisons under identity-independent scenario. (adopted from [1])



(a)

(b)

FIGURE 4.3: Performance comparisons under database-independent scenario. (adopted from [1])

Table 4.5: Risk of different situations.

Probe Class / Model Result	Known	Known unknown	Unknown unknown
Positive	Identified no risk	Identified certain risk	Identified uncertain risk
Negative	Unidentified no risk	Unidentified risk	Unidentified uncertain risk

4.1.2 Open-set and Closed-set Identification

4.1.2.1 Database Configuration

The subjects in the dataset are classified into three classes according to the characteristic of the subject:

- **Known**-subjects who are enrolled in the gallery, e.g. the fugitive on the watch-list.
- **Known unknown**-subjects used for LIoM training but not included in the gallery, e.g. the staff at the concert.
- **Unknown unknown**-subjects in neither the training set nor the gallery, e.g. the normal audience.

For biometric identification system, it is crucial to determine a probe is not in the gallery.Take door access systems as an instance, the biometric system should identify the imposters who is not in the authorized list and issue a warning. The unknown unknown implies an uncertain risk for the system, while know unknown is a certain risk, different situations have different kind of risks that is summarized in Table 4.5 inspired from [136]). While how to manage those risks is still a non-trivial task for the biometric systems (e.g. detect the uncertain risk).

Accordingly, based on the LFW dataset, subjects who have more than three face samples are named **Known**; subjects who have two or three samples are called **Known unknown** and finally **Unknown unknown** is dedicated to the subjects who only have one sample (Table 4.6). Based on the above subject classes, we divide the considered datasets into training, gallery and four types of probe: Closed-set probe C, open-set probes O1, O2 and O3. The closed-set C is used for closed-set identification as well as verification evaluation, while the open-set probes O1, O2 and O3 are used for open-set identification evaluation.

Specifically, as for the LIoM training set, three samples of each **Known** subject and one sample of each **Known unknown** subject are adopted randomly. In what follows, three samples

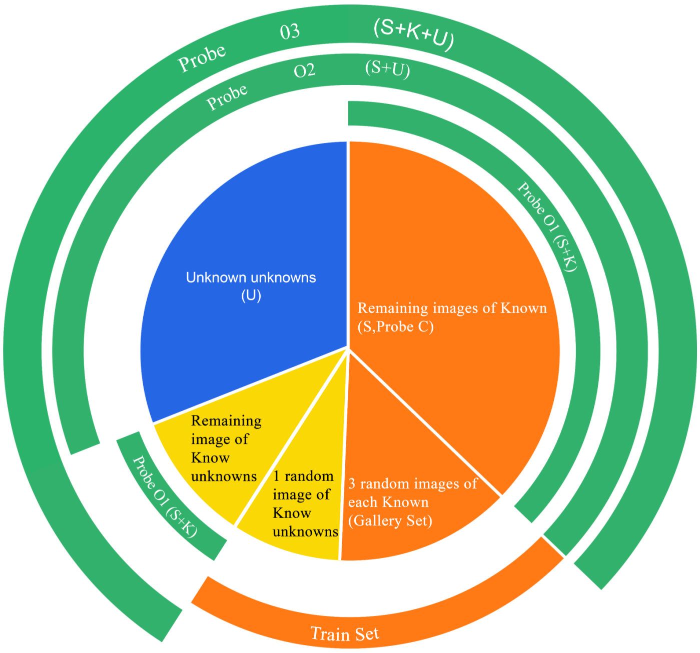

Probe 03 (S+K+U)
Probe O2 (S+U)
Probe O1 (S+K)
Unknown unknowns (U)
Remaining images of Known (S,Probe C)
Remaining image of Know unknowns
1 random image of Know unknowns
3 random images of each Known (Gallery Set)
Probe O1 (S+K)
Train Set

FIGURE 4.4: LFW configuration for face identification evaluation. (adopted from [1])

of each **Known** subject are enrolled randomly as the gallery set. It is worth noting that our protocol of LFW is originated from [117], while in this work randomness is introduced instead of taking a fixed order of samples to avoid bias.

The closed-set probe C consists of the remaining images S of the **Known** subjects after excluding the gallery images. The open-set O1 contains the same images as in probe set C (S) and images K from known unknowns, where K are images from known unknowns after excluding the training images. The open-set O2 consists of probe C (S) and all images from unknown unknowns U. Finally, open-set O3 consists of images from known, known unknown and unknown unknown subjects, denoted as O3 = $(S \cup K \cup U)$. The probe sets C, O1, O2 and O3 are illustrated in Table 4.7 and Figure 4.4.

For VGG2 and IJB-C, their respective dataset configurations can be found in Table 4.8, Table 4.9, Table 4.10, and Table 4.11.

Table 4.6: Division of subjects in LFW according to knowns, known unknowns and unknown unknowns sets.

Set	Selection Criteria	Subjects	Images
Knowns	4≤ subjects' images	610	6,733
Known unknowns	2≤ subjects' images ≤3	1,070	2,431
Unknown unknowns (U)	Subjects' images=1	4,069	4,069

Table 4.7: Division of subjects in LFW according to training, gallery and probes sets.

Description	Images	
Training	Taking 3 images of knowns + 1 image of known unknowns randomly	2,900
Gallery	Taking 3 images of knowns randomly	1,830
Probe C	C=S	4,903
Probe O1	O1 =S∪K	6,264
Probe O2	O2 =S∪U	8,972
Probe O3	O3 =S∪K∪U	10,333

Table 4.8: Division of subjects in VGG2 according to knowns, known unknowns and unknown unknowns.

Subject Set	Selection Criteria	Identities	Images
Knowns	1st – 2000th subject	2000	100,000
Known unknowns	2001st – 4000th subject	2000	100,000
Unknown unknowns (U)	4001st – 6000th subject	2000	100,000

Table 4.9: Division of subjects in VGG2 according to training, gallery and probes.

Description	Images	
Training set	2 random images of knowns + 1 random image of known unknowns	6000
Gallery set	3 random images of knowns	6000
Probe C	C=S, S is generated by collecting 3 images of remaining knowns randomly	6000
Probe O1	O1 =S∪K (K consists of 3 random images of known unknowns)	12k
Probe O2	O2 =S∪U (U consists of 1st images of unknown unknowns)	8k
Probe O3	O3 =S∪K∪U	14k

Table 4.10: Division of subjects in IJB-C according to knowns, known unknowns and unknown unknowns.

Subject Class	Selection Criteria	Subjects	Images
Knowns	30≤ subjects' images	1,797	106,737
Known unknowns	16≤ subjects' images ¡ 30	755	16,517
Unknown unknowns (U)	1≤ subjects' images ≤15	978	8,713

Table 4.11: Division of subjects in IJB-C according to training, gallery and probes.

Dataset Name	Description	Images
Training set	3 random images of knowns + 1 random image of known unknowns	4,349
Gallery set	3 random images of knowns	3,594
Probe C	Remaining images of knowns (represented as S)	5,391
Probe O1	O1 =S∪K (K is not part of the training set known unknowns)	7,656
Probe O2	O2 =S∪U	8,290
Probe O3	O3 =S∪K∪U	10,555

4.1.2.2 Performance Metrics

In this thesis, we adopt a cumulative matching characteristic (CMC) curve [137] and the detection and identification rate (DIR) [138, 139] to measure the performance accuracy for closed-set and open-set face identification evaluations, respectively.

The CMC plots the identification rate, a.k.a. the recognition rate, with respect to a given rank in the closed-set. It illustrates the relative number of probes that have reached at least rank k. Specifically, given a gallery with N samples, q testings are performed by matching the probe with all the gallery, and the targeted gallery samples are sorted according to the matching score in descending order, denoted as $r = (r_1, r_2 \ldots, r_N)$, then the identification rate at the rank k is computed as

$$\mathrm{CMC}(k) = \frac{1}{N} \sum_{i=1}^{q} \begin{cases} 1, & r_i \leq k \\ 0, & r_i > k \end{cases} \tag{4.3}$$

DIR curves plot the identification rates with respect to the false alarm rates in an open-set [139]. Given a probe which consists of known subjects, denoted as P^k, and unknown subjects denoted as P^{uk}, the matching score between a probe p and a gallery g is computed as $s(p, g)$. Then the DIR is given by

$$\mathrm{DIR}(\theta, k) = \frac{|\{s(p, g) \mid s(p, g) \geq \theta, \mathrm{rank}(p) \leq k\}|}{|P^k|} \tag{4.4}$$

where θ is a threshold and k is the identification rank. When the similarity of an unknown probe to any of the gallery subjects is higher than θ, a false alarm is issued

$$\mathrm{FAR}(\theta) = \frac{\left|\left\{s(p, g), \forall p \in P^{uk}, s(p, g) \geq \theta\right\}\right|}{|P^{uk}|} \tag{4.5}$$

4.1.2.3 Unprotected Identification System Performance

In this subsection, we present the baseline (without using a protection method, i.e. LIoM hashing) identification performance by using sole deep face vectors and their feature level fusion (by concatenation) extracted from FaceNet and InsightFace. The results on LFW, VGG2 and IJB-C are shown in Table 4.12 in terms of the identification rate (IR) at rank one (CMC with only rank one is taken) and the DIR at FAR=0.1% and FAR=1%. From all the closed-set

TABLE 4.12: Unprotected closed-set & open-set system performance in LFW, VGG2 and IJB-C.

DB		IR@ rank=1	DIR@ FAR=0.1% (O1)	DIR@ FAR=0.1% (O2)	DIR@ FAR=0.1% (O3)	DIR@FAR=1% (O1)	DIR@ FAR=1% (O2)	DIR@ FAR=1% (O3)
LFW	FaceNet	98.95	40.38	38.24	38.79	88.48	85.56	86.65
	InsightFace	99.80	48.42	88.10	88.10	99.48	99.51	99.51
	Fused	99.85	45.31	90.66	90.66	99.48	99.40	99.40
VGG2	FaceNet	89.38	47.71	46.81	47.89	71.14	69.51	70.78
	InsightFace	96.42	85.22	82.96	85.84	93.99	93.83	93.95
	Fused	99.29	97.11	96.57	97.05	98.14	98.05	98.10
IJB-C	FaceNet	75.73	16.76	11.44	11.40	35.91	26.25	27.64
	InsightFace	72.42	30.92	24.78	26.40	42.30	39.15	38.71
	Fused	79.81	43.47	41.72	40.47	58.61	54.05	53.61

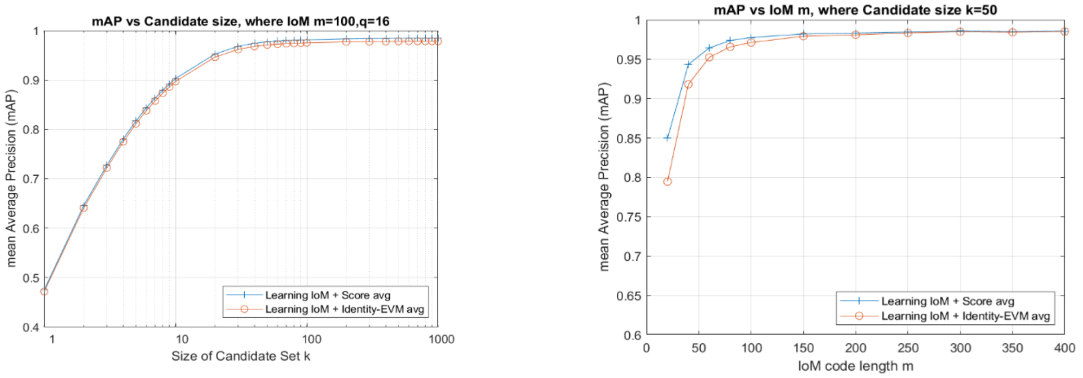
experiments, we find that InsightFace is generally more accurate than FaceNet, and their fusion outperforms individual InsightFace and FaceNet by a large margin, except for a few cases with comparable performance to InsightFace.

For the LFW open-set evaluation, fusion features do not always show better performance but are comparable with InsightFace in certain circumstances. Nevertheless, IJB-C and VGG2 favor fusion features and are shown to outperform InsightFace in all cases. In short, it is recommended to use fusion features rather than InsightFace and FaceNet individually.

4.1.2.4 Parameters Tuning

Before we proceed to present the protection systems' performance, the system's parameters are set first. The main parameters of the 1-to-N matching module are the candidate size top-k and LIoM parameters, i.e. m and q. From section 4.1.1, the parameters are set to $m = 100$ and $q = 16$. To find the best top-k, a top-k screening is run on the LFW closed-set protocol, and the mAP of the corresponding top-k candidate size is recorded in Figure 4.5(a).

As shown in Figure 4.5(a), when k becomes large, the mAP becomes stable, but large k indicates that more subsequent processing will be performed. Therefore, we opt for $k = 50$ since the performance is still sufficiently good while remaining efficient. From Figure 4.5(b), we find that the matching performance is reasonably good when m is set as 100.



(a) (b)

FIGURE 4.5: mAP vs k and mAP vs LIoM length m on fusion feature of LFW (closed-set). (adopted from [1])

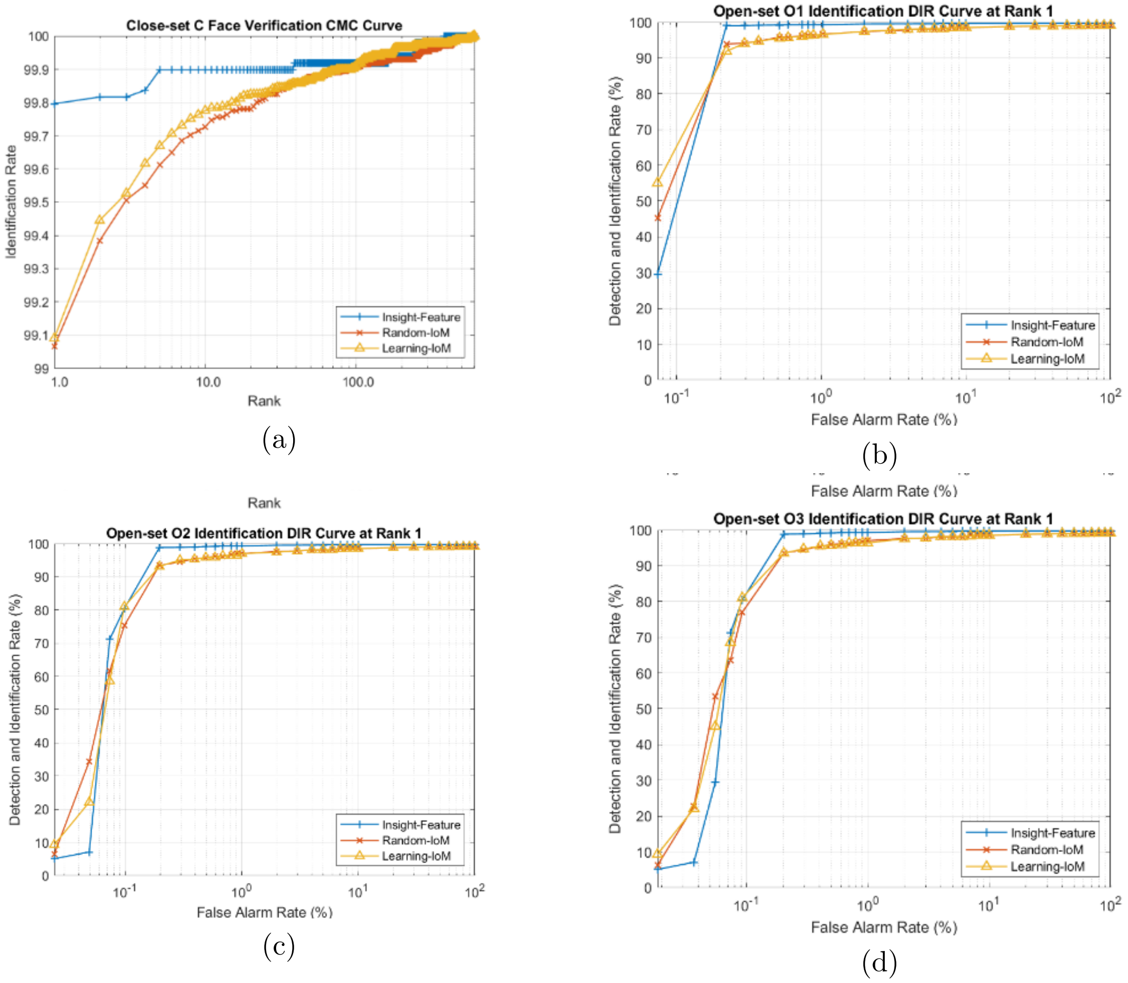
TABLE 4.13: Protected closed-set & open-set system performance without fusion strategy in LFW.

		IR@ rank=1	DIR@ FAR=0.1% (O1)	DIR@ FAR=0.1% (O2)	DIR@ FAR=0.1% (O3)	DIR@F AR=1% (O1)	DIR@ FAR=1% (O2)	DIR@ FAR=1% (O3)
FaceNet	Original	98.95	40.38	38.24	38.79	88.48	85.56	86.65
	Random IoM	97.53	35.86	35.11	36.51	81.20	79.56	79.56
	LIoM	97.44	39.78	34.43	37.14	77.01	78.10	78.10
InsightFace	Original	99.80	48.42	88.10	88.10	99.48	99.51	99.51
	Random IoM	99.14	59.53	89.03	89.03	97.63	97.63	97.63
	LIoM	99.18	54.19	85.63	82.97	97.57	97.75	97.75

4.1.2.5 Protected System Performance without Fusion Strategy

In this subsection, the performance of the protected system is explored on LFW. Firstly, the closed-set and open-set performance metrics are computed on the original deep features generated by different deep models, i.e. InsightFace and FaceNet, and then LIoM and random IoM's accuracy of performance are analyzed. The results suggest that LIoM can significantly improve the accuracy compared with random IoM in closed-set settings, as depicted in Figure 4.6 and Figure 4.7.

As shown in Table 4.13, when the deep features are hashed, the performance accuracy will be degraded as a trade-off between compact representation for fast matching with the Hamming distance. To be specific, the LIoM hashed code is of 400 ($m \log_2^q = 100 \log_2^{16}$) bits, while the FaceNet and InsightFace feature length in floating points are of 256 (256x32 bits) and 512 (512x32 bits), respectively. However, when LIoM with InsightFace is applied, the performance can be largely restored in the closed-set system.



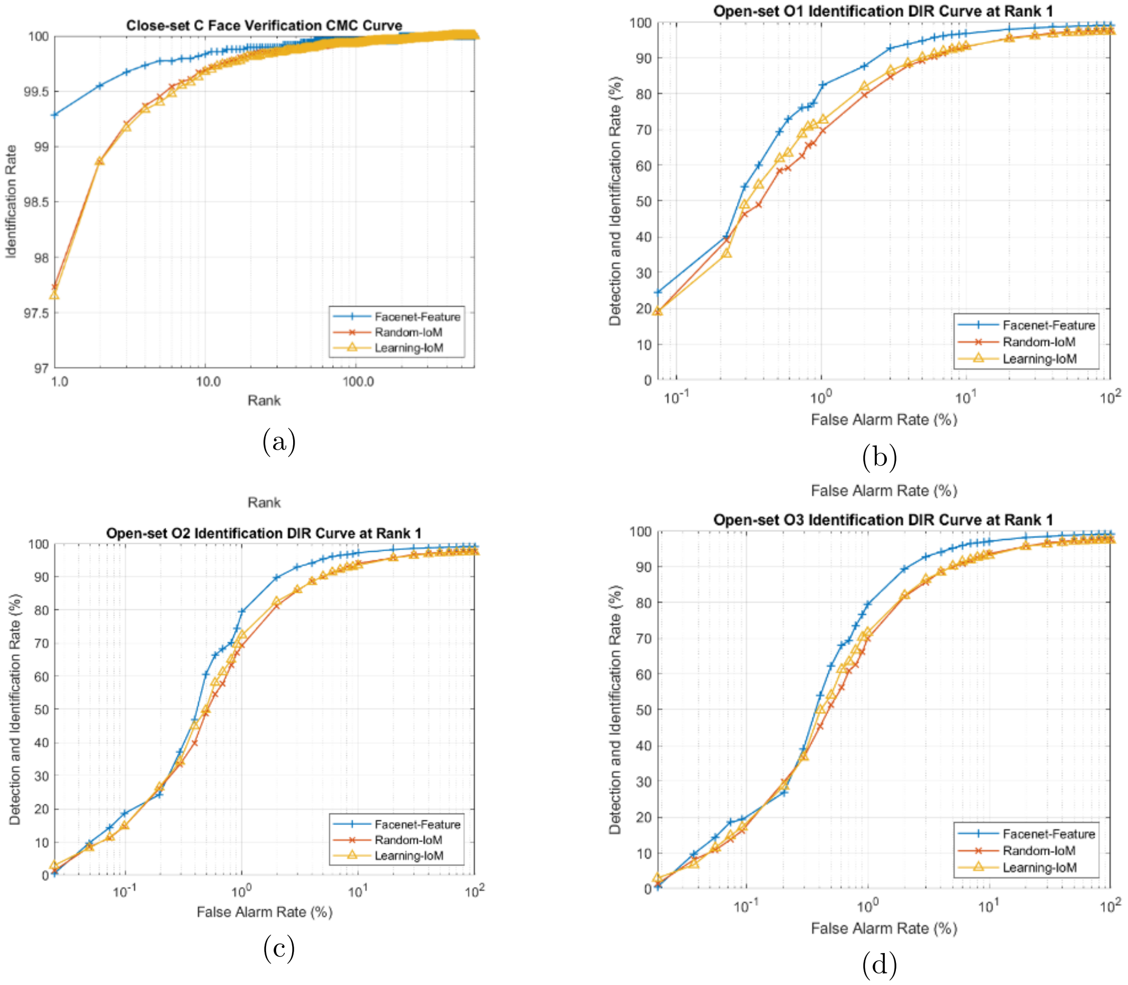
(a)

(b)

(c)

(d)

FIGURE 4.6: Protected identification system performance on LFW with InsightFace where (a) is for closed-set evaluation and (b), (c), (d) are for open-set evaluation.
(adopted from [1])



(a)

(b)

(c)

(d)

FIGURE 4.7: Protected identification system performance on LFW with FaceNet where (a) is for closed-set evaluation and (b), (c), (d) are for open-set evaluation. (adopted from [1])

Table 4.14: Protected closed-set & open-set system performance with fusion strategy in LFW.

Option		IR@ rank=1	DIR@ FAR=0.1% (O1)	DIR@ FAR=0.1% (O2)	DIR@ FAR=0.1% (O3)	DIR@FAR=1% (O1)	DIR@ FAR=1% (O2)	DIR@ FAR=1% (O3)
1	Avg	99.67	59.60	80.40	80.58	97.79	97.72	97.70
	Max	99.15	46.92	46.14	47.43	83.06	84.30	83.81
	Model EVM (Avg)	97.93	44.62	47.01	47.01	81.46	82.50	82.50
	Model EVM (Max)	97.92	44.62	47.01	47.01	81.46	82.50	82.50
	Identity EVM (Avg)	99.47	55.86	78.20	78.31	97.81	97.83	97.82
	Identity EVM (Max)	99.38	48.98	44.81	45.98	85.25	85.90	85.90
2	Concatenate	99.75	53.57	77.28	77.28	97.98	97.99	97.99
	Mean	98.88	61.05	72.18	72.84	94.31	94.22	94.55
	Max	98.93	61.46	82.16	82.16	94.54	94.79	95.04
	Min	98.88	61.05	72.18	72.84	94.31	94.22	94.55
3	-	99.47	49.64	72.78	76.19	96.85	97.49	97.34

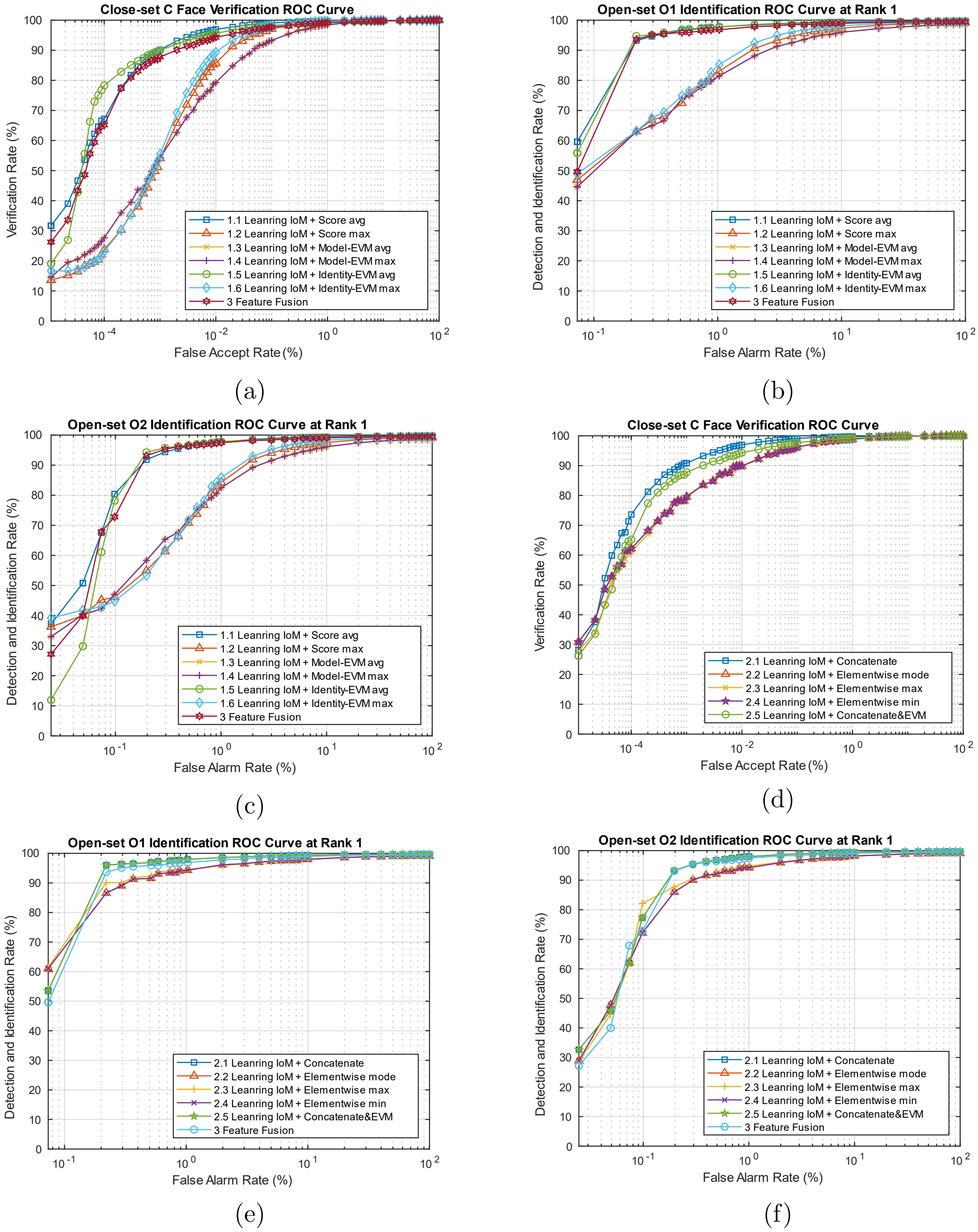
Table 4.15: Protected closed-set & open-set system performance with fusion strategy in VGG2.

Option		IR@ rank=1	DIR@ FAR=0.1% (O1)	DIR@ FAR=0.1% (O2)	DIR@ FAR=0.1% (O3)	DIR@FAR=1% (O1)	DIR@ FAR=1% (O2)	DIR@ FAR=1% (O3)
1	Avg	98.94	90.26	87.86	89.75	95.72	95.62	95.79
	Max	96.16	49.49	41.76	49.20	75.49	73.44	73.66
	Model EVM (Avg)	83.24	35.83	35.10	36.45	57.77	57.16	57.70
	Model EVM (Max)	83.15	35.83	35.10	36.45	57.77	57.16	57.70
	Identity EVM (Avg)	94.50	87.60	87.15	87.72	91.85	91.89	91.89
	Identity EVM (Max)	97.13	50.12	47.13	49.81	78.36	76.10	77.84
2	Concatenate	99.03	91.67	89.46	91.13	96.03	95.93	95.96
	Mean	95.10	77.25	74.89	77.25	86.13	86.85	86.85
	Max	94.92	75.97	76.16	75.22	86.94	85.72	86.18
	Min	95.10	77.25	74.89	77.25	86.13	86.85	86.85
3	-	96.83	89.69	87.48	89.69	93.87	93.23	93.51

4.1.2.6 Protected System Performance with Fusion Strategy

As presented in section 3.4.3, three fusion options are employed on LFW, VGG2 and IJB-C, depicted in Figure 4.8, Figure 4.9 and Figure 4.10, respectively. In these evaluations, we apply LIoM with $m = 100$ and $q = 16$. As observed from Table 4.14, Table 4.15 and Table 4.16, for LFW, the average score and identity-EVM of the average score of option 1 show the best overall performance, which is consistent with what [117] reported, and followed by the concatenation method in option 2. However, both model-EVM and identity-EVM on VGG2 and IJB-C do not excel as in LFW but score fusion on average in option 1 shows consistently good accuracy. Nevertheless, the concatenation in option 2 performs well in VGG2 and IJB-C after score fusion in option 1.

To better demonstrate the separability of the scores fused by average, the distributions of genuine and imposter similarity scores are shown in Figure 4.11. It can be seen that the fused score in Figure 4.11 (c) can achieve lower variance.



(a)

(b)

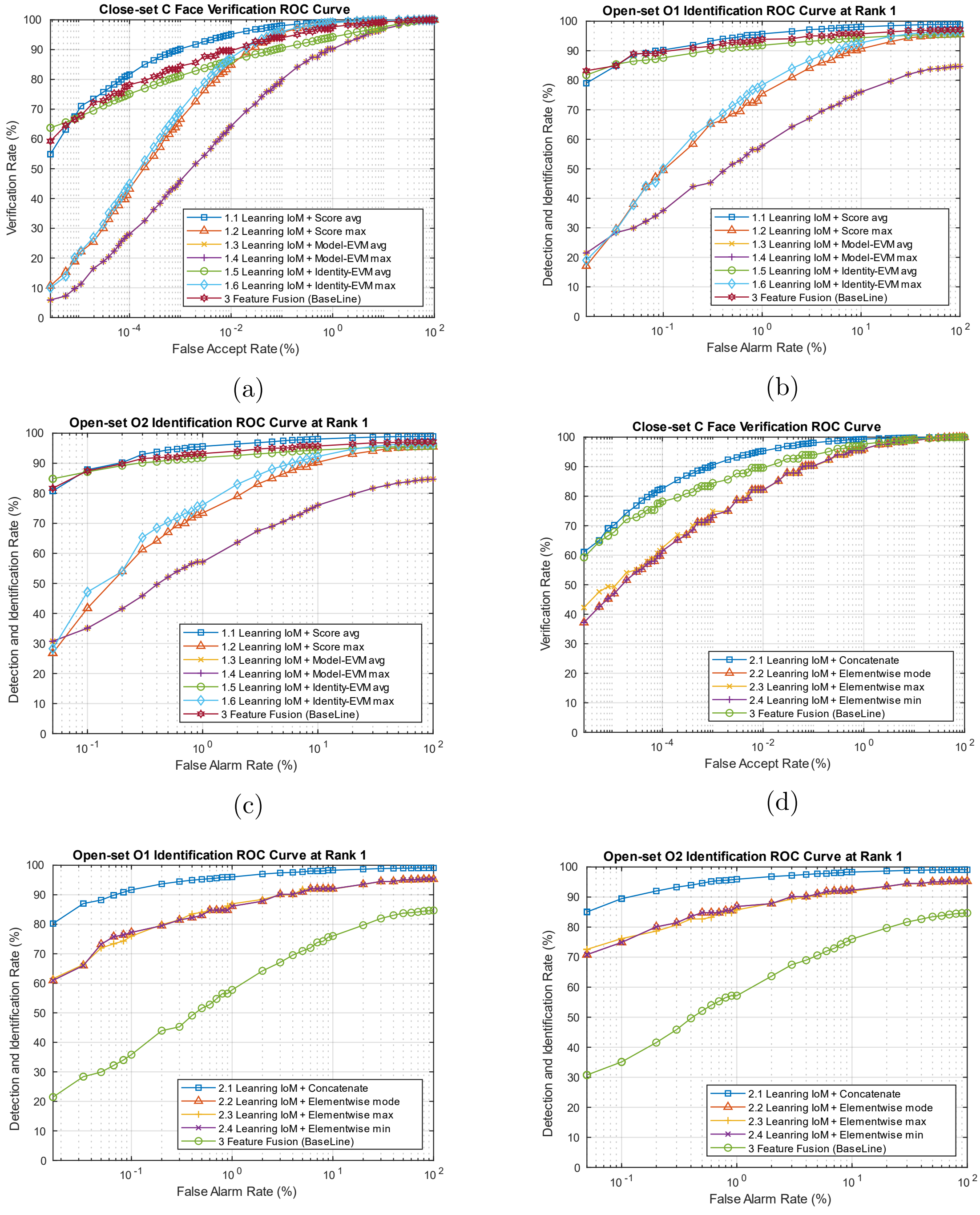
(c)

(d)

(e)

(f)

Figure 4.8: Accuracy of performance on LFW score with IoM hashing (protection+fusion). (adopted from [1])



(a)

(b)

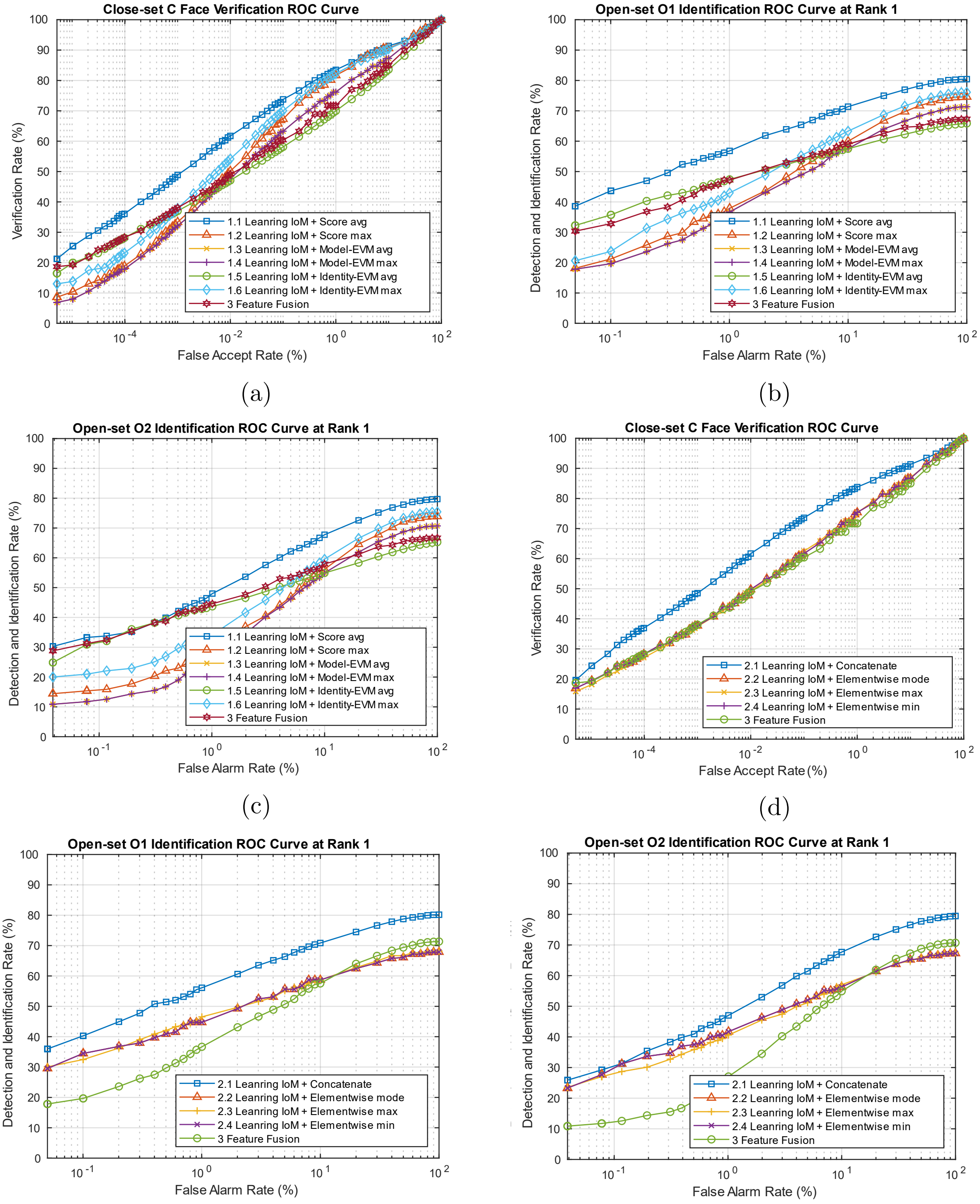
(c)

(d)

(e)

(f)

FIGURE 4.9: Accuracy of performance on VGG2 score with IoM hashing). (adopted from [1])



(a) (b)

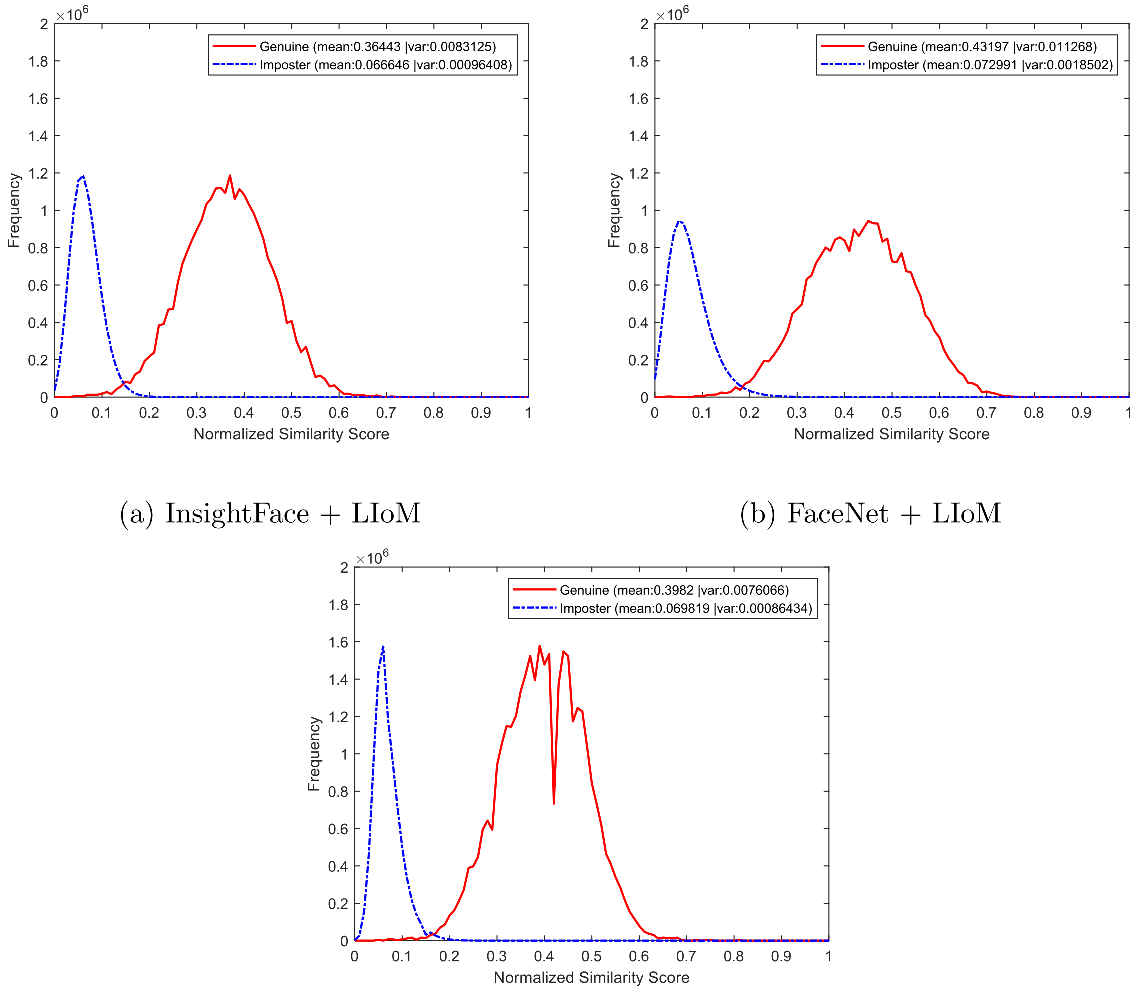
(c) (d)

(e) (f)

Figure 4.10: Accuracy of performance on IJB-C score with IoM hashing). (adopted from [1])

Table 4.16: Protected closed-set & open-set system performance with fusion strategy in IJB-C.

Option		IR@ rank=1	DIR@ FAR=0.1% (O1)	DIR@ FAR=0.1% (O2)	DIR@ FAR=0.1% (O3)	DIR@FAR=1% (O1)	DIR@ FAR=1% (O2)	DIR@ FAR=1% (O3)
1	Avg	80.55	43.65	33.80	33.31	56.80	48.01	49.09
	Max	75.10	21.23	15.89	16.09	37.90	28.09	31.14
	Model EVM (Avg)	73.34	19.67	12.58	13.71	36.67	27.06	29.52
	Model EVM (Max)	73.34	19.67	12.58	13.71	36.67	27.06	29.52
	Identity EVM (Avg)	54.38	35.74	32.05	32.31	47.43	43.70	42.76
	Identity EVM (Max)	66.17	23.70	22.10	21.36	42.99	34.20	36.01
2	Concatenate	80.57	40.29	31.30	32.63	56.11	47.02	47.75
	Mean	69.44	34.49	31.21	31.07	44.73	41.71	40.44
	Max	69.71	32.47	28.70	28.28	46.53	40.60	39.81
	Min	69.44	34.49	31.21	31.07	44.73	41.71	40.44
3	-	67.85	32.85	32.49	31.57	47.18	44.60	43.62



(a) InsightFace + LIoM

(b) FaceNet + LIoM

(c) Score fusion by average

Figure 4.11: Similarity score distribution on LFW.

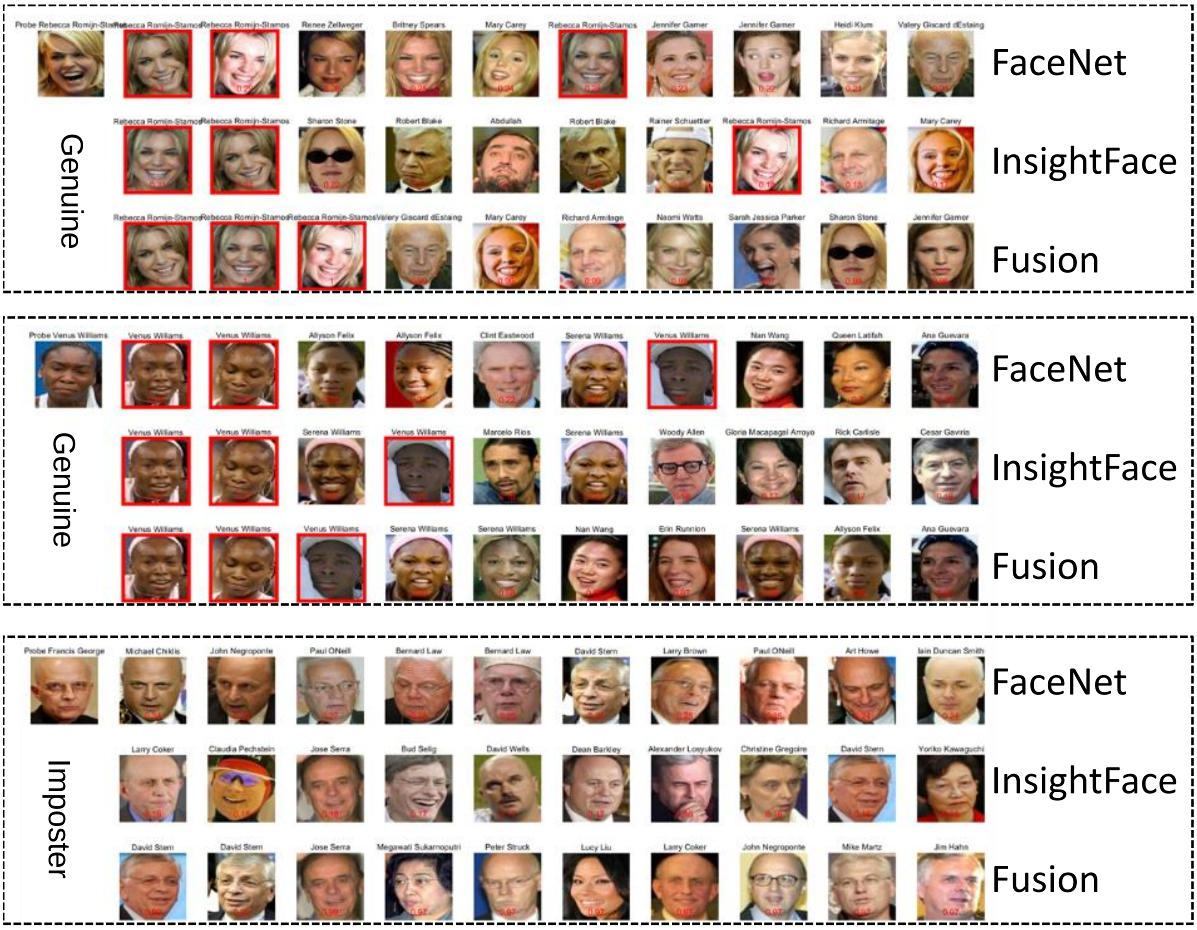


FIGURE 4.12: Illustration of face search in open-set 1 (LFW, O1).
The images of the first column are probes, the right 10 images are gallery search result. The first row of the gallery result is from FaceNet, while the second and third are from InsightFace and Feature Fusion Face respectively. The first two probes are genuine probe, while the last one is imposters. Best view in color.

4.1.2.7 Visual Evaluation of the system

To illustrate the effect of fusion to 1-to-N match performance, some return exemplars (id) are depicted in Figure 4.12. The images of the first column are probes, the right 10 images are gallery search result. The first row of the gallery result is from FaceNet, while the second and third are from InsightFace and Feature Fusion Face respectively. Due to high discriminative of FaceNet and InsightFace features, at most time the top three retrieved images would be the target probe. In these images we pick up some cases while fusion may help to improve the search result.

To evaluate how well the genuine and imposter scores are separated, d prime (d') is utilized to measure the separability. The larger the value of d', the more separable the two distributions, and the better the detection performance [140]. In our setting, d'= 5.11 when FAR=1%, while

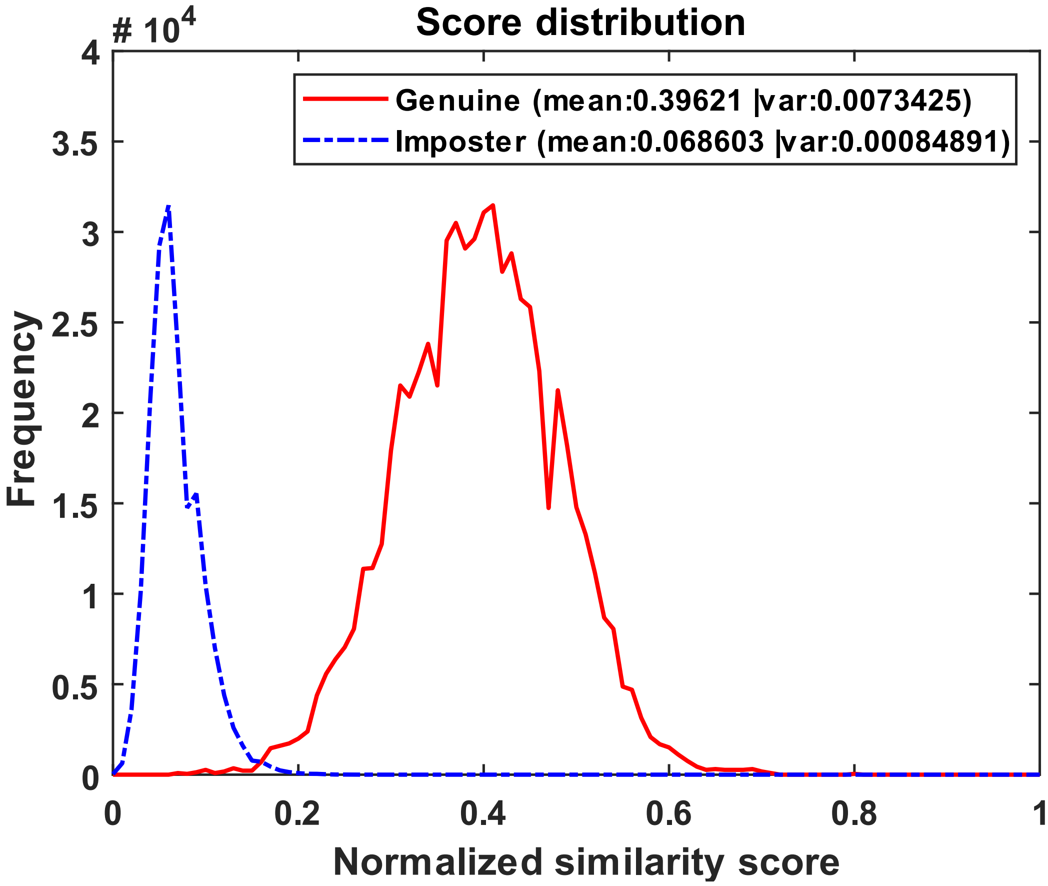


FIGURE 4.13: Genuine and imposter score distribution on LFW. (adopted from [2])

for the original deep feature, $d' = 5.32$ when FAR =1%. To better demonstrate the separability, the distributions of genuine and imposter similarity scores are shown in Figure 4.13.

4.1.2.8 Time Cost of the System

Time efficiency is another important concern associated with face identification systems. To investigate the time efficiency of the proposed system, the enroll time (deep feature extraction and hashing), 1-to-N matching time, including fusion operation as well as training time, are recorded for LIoM on FaceNet deep face vectors. The machine we use for simulation is equipped with MATLAB Ver. 2018b, Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz and 16GB RAM. In this simulation, the open-set protocol as discussed in section 4.1.2.1 is followed. 1830 gallery images, 2900 training set images and 4903 probe C images are used to evaluate the average processing time. Both random IoM and LIoM are fixed at 100 (400bit) of length.

Table 4.17: Average time of processing in training, enrollment and matching on FaceNet face vector.

	Deep Face Vector	Random IoM	Learning IoM
Training time	–	–	147.078 s
Enroll time	–	0.017 ms	0.021 ms
Matching time	0.176 ms	0.130 ms	0.148 ms
EER (identity-dependent)	1.25%	2.01%	0.59%
mAP (identity-dependent)	0.9833	0.9627	0.9944

The timing readings are tabulated in Table 4.17. The enroll time for one face image takes only around 0.02 ms. The training time of LIoM depends on the training set and hash code size, and it is observed to be quite lengthy. However, the training process is offline and would not affect the system's efficiency. Compared with the original unprotected deep face vector, the matching time of random IoM and LIoM is much less, reduced by around 25%. This is because the matching of the former is done by cosine distance, while the IoM and LIoM hash code matching is done by the Hamming distance, which is simpler and can be optimized in hardware.

4.1.2.9 Summary

To sum up, we conduct a cross-performance comparison for an unprotected system, a protected system without fusion and with fusion with their respective best performing settings. The identification rate at rank 1 for the closed-set protocol and DIR@DIR@1% for the open-set protocol are used as performance metrics.

From Table 4.18, Table 4.19 and Table 4.20, we notice the following:

1. The deep face vector with feature level fusion (corresponding to option 3 in section 3.4.3) performs pretty well in LFW and VGG2 in both closed- and open-set protocols. However, the performances on IJB-C are not satisfactory, especially in the open-set protocol.

2. When the deep face vector is transformed by IoM hashing, performance degradation is inevitable. The deterioration on LFW and VGG2 with the closed-set protocol is not significant, but on IJB-C it is apparent. Different levels of degradation are observed for the open-set protocol depending on the combinations of known, known unknown and unknown unknown probes. However, LIoM hashing is essential for template protection as well as speedy matching compared to the use of deep face vectors only.

Table 4.18: Comparison of unprotected, protected without and with fusion strategies closed-set identification systems in terms of IR(%)@1 performance.

DB	Unprotected with Deep Features Fusion	Protected with learning IoM	
		without fusion (InsightFace)	with fusion
LFW	99.85	99.18	99.67 (Opt 1 - Avg)
VGG2	99.29	96.42	99.03 (Opt 2 - Concatenate)
IJB-C	79.81	72.41	80.55 (Opt 1 - Avg)

Table 4.19: Comparison of unprotected, protected without and with fusion strategies open-set identification system in terms of DIR(%)@FAR=1%.

DB	Unprotected with Deep Features Fusion			Protected with learning IoM					
				without fusion (InsightFace)			with fusion		
	O1	O2	O3	O1	O2	O3	O1	O2	O3
LFW	99.48	99.40	99.40	97.57	97.75	97.75	97.81 (Identity EVM-Avg)	97.83 (Identity EVM-Avg)	97.82 (Model EVM-Avg)
VGG2	98.14	98.05	98.10	84.83	84.12	84.83	95.72 (Opt 1 - Avg)	95.62 (Opt 1 - Avg)	95.79 (Opt 1 - Avg)
IJB-C	58.61	54.05	53.61	33.24	30.78	30.73	56.80 (Opt 1 - Avg)	48.01 (Opt 1 - Avg)	49.09 (Opt 1 - Avg)

3. Nevertheless, the performance, as in the unprotected system, can be largely restored, especially for LFW and VGG2, when the fusion method is applied to the protected system, while still enjoying speedy matching. In general, we can observe that plain averaging in option 1 (see section 3.4.3) is the best fusion method, followed by the EVM approach and concatenation approach in option 2 in our experiments.

4. IJB-C remains a difficult dataset on which to achieve good performance, especially under the open-set identification protocol, relative to LFW and VGG2, due to the large variation in the nature of the face images.

Table 4.20: Comparison of unprotected, protected without and with fusion strategies open-set identification system in terms of DIR(%)@FAR=0.1%.

DB	Unprotected with Deep Features Fusion			Protected with learning IoM					
				without fusion (InsightFace)			with fusion		
	O1	O2	O3	O1	O2	O3	O1	O2	O3
LFW	99.85	45.31	90.66	99.18	54.19	85.63	59.60 (Opt 1 - Avg)	80.40 (Opt 1 - Avg)	80.58 (Opt 1 - Avg)
VGG2	97.11	96.57	97.05	75.29	63.25	73.11	90.26 (Opt 1 - Avg)	87.76 (Opt 1 - Avg)	89.75 (Opt 1 - Avg)
IJB-C	43.47	41.72	40.47	21.78	20.08	20.56	43.65 (Opt 1 - Avg)	33.80 (Opt 1 - Avg)	33.31 (Opt 1 - Avg)

4.1.3 State-of-the-Art Comparison

In this section, we provide a comparison of our proposed methods with the state of the art. Since VGG2 and IJB-C are very new, we only compare with those works that were conducted under the LFW dataset under LFW standard [94] and BLUFR verification and open-set identification protocols [21], as tabulated in Table 4.21. It is noted that the proposed method performs comparably to the state of the art in terms of accuracy of performance. However, it is significant that most of the works in Table 4.21 do not consider time efficiency on identification, face features storage complexity and protection issues.

In Table 4.22, a comprehensive comparison of the matching accuracy, storage, matching time and template protection based on FaceNet is provided. The comparison suggests that LIoM can achieve reasonably good accuracy of performance, with less storage space, and a higher matching speed compared with the deep facial features.

4.2 Unlinkability and Revocability Analysis of LIoM Hashing

Despite LIoM hashing is used for face features compression and protection, it plays the same role as random IoM for biometric template protection (BTP). The BTP should meet four essential criteria, i.e. performance, non-invertibility, unlikability and revocability [12]. The performance criterion requires accuracy performance preservation before and after the hashing, which have been evaluated in section 4.1.1. Since LIoM hashing is inherited from the random IoM where the major distinction is the projection matrices of the former are generated from training data while

Table 4.21: Accuracy comparison with state of the art.

Scheme	LFW standard protocol	BLUFR VR@ FAR=0.1%	BLUFR DIR@ Rank = 1, FAR = 1%
Baidu [141]	99.77%	99.41%	92.09%
Facebook [142]	98.37%	-	-
Light CNN [143]	99.33%	98.88%	92.29%
HighDimLBP [21]	-	41.66%	18.07%
CASIA [144]	96.13%	80.26%	28.90%
Centerloss [145]	99.28%	93.35% [143]	67.86% [143]
Range Loss [146]	98.63%	92.10%	63.69%
FaceSearchAtScale [147]	98.2%	89.8%	55.9%
NormFace [36]	-	95.83%	77.18%
TypicFace [148]	99.32%	97.82%	85.71%
Customized weighted constraint [149]	99.12%	94.79%	73.69%
FaceNet [37]	99.63%	98.14%	68.12%
InsightFace [39]	99.85%	99.70%	98.05%
Proposed method			
LIoM - InsightFace	99.72%	98.99%	78.88%
Random IoM - InsightFace	99.69%	98.57%	90.13%
LIoM - FaceNet	99.84%	98.88%	70.84%
Random IoM -FaceNet	99.83%	95.74%	61.45%
Random IoM - Feature Fusion	99.93%	99.25%	89.56%
LIoM - Feature Fusion	99.95%	99.69%	88.78%

Table 4.22: Summary of the proposed method performance on FaceNet.

	Deep Face Vector	Random IoM	Learning IoM
Matching accuracy	99.63%	99.83%	99.84%
Storage	256*32 bits	400bits	400bits
Matching time	0.176 ms	0.130 ms	0.148 ms
Template protection	×	✓	✓

the hashing process remains identical as in random IoM, we leave non-invertibility analysis for LIoM hashing and refer readers to [17] for further reading. This section focuses on unlinkability and revocability evaluations as they are associated with projection matrices used in the IoM hashing approach. All the experiments are carried out under the VGG2 dataset.

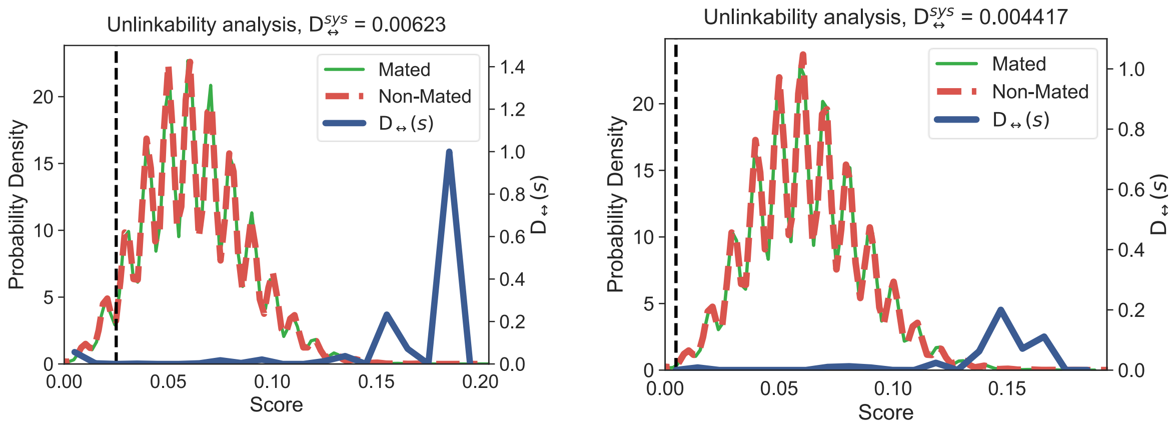
4.2.1 Unlinkability Evaluation

The unlinkability criterion demands that the LIoM hashed face code vector should not be differentiated whether they are generated from the same subject's deep face vector. This is to prevent matching across different applications (cross-matching). For LIoM hashing, this can be done by using *application-specific* seeds to initiate projection matrix **W** in algorithm 2 and 3, so the generated face code vectors are unique despite $\mathbf{W}$ are trained from the same training pool. Another vital reason for the face code vector to satisfy this property is that in the CFV, each subject has to generate two independent LIoM hashed face codes as genuine entries for polynomial projection and as keys for encryption (See section 5.4, chapter 5).

For evaluation, we follow a protocol that outlined by [15]. The experiments are executed with m=100 and q=16 for both LIoM hashing and random IoM hashing that transformed from FaceNet features. The cross-matching attack can be launched by comparing the face code vectors generated from the same person in different applications (different $\mathbf{W}$ in LIoM hashing). Under this attack, an adversary is assumed familiar with the LIoM hashing algorithm and holds the face code vectors of different applications. The adversary can exploit the matching score distributions of face code vectors to learn the face code vector is from the same person. Here we refer matching score of the same subject in different applications as ***mated score***, while the matching score from different subjects in different applications is referred as ***non-mated scores***.

To address the resistance to cross-matching attack, two different measures for the linkability are defined in [15], namely Local measure $D_{\leftrightarrow}(s) \in [0,1]$ and Global measure $D_{\leftrightarrow}^{sys}$. The $D_{\leftrightarrow}(s)$ evaluates the linkability for each specific linkage score s in a score-wise level. Given a score $s_k, D_{\leftrightarrow}\left(s_k\right) = 1$ means the adversary can decide which face code vector is from the same person with almost all certainty while $D_{\leftrightarrow}\left(s_k\right) = 0$ indicates it is hard to determine which two face code vectors are from the same person for this particular score s_k. $D_{\leftrightarrow}^{sys} \in [0,1]$ evaluates the unlinkability of the whole system and can be used as a benchmark for different systems independently of the score. $D_{\leftrightarrow}^{sys} = 1$ indicates the system is fully linkable for all scores of the mated subjects, while $D_{\leftrightarrow}^{sys} = 0$ suggests the system is fully unlinkable for all scores.

The unlinkability of LIoM hashing in our research is evaluated on identity-dependent settings. To simulate the different face code vectors in different applications, the first sample of each subject is selected firstly, denoted u_i where $i = 1, \ldots, 1000$. Next, each sample is transformed into face code vector with 10 different $\mathbf{W}$, which represent 10 applications, denoted as $h_i^j = h\left(u_i; \mathbf{W}^j\right)$, where $j = 1, \ldots, 10$. The mated scores and non-mated scores are computed among those templates and their distributions are illustrated in Figure 4.14. It is proved that both LIoM and random IoM hashing can achieve good unlinkability where the $D_{\leftrightarrow}^{sys}$ can achieve 0.006 and 0.004 respectively. Our result proves that LIoM can achieve unlinkability by using application-specific seeds to initiate algorithm 2 and 3 .



(a) Learning IoM

(b) Random IoM

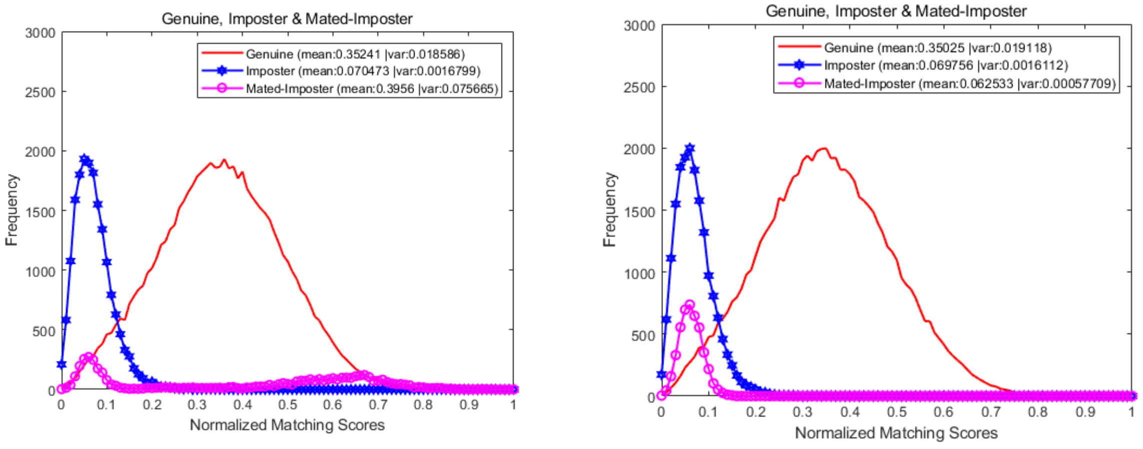
FIGURE 4.14: Unlinkability of LIoM and random IoM hashing under VGG2.

4.2.2 Revocability Evaluation

Unlike unlinkability, revocability indicates the capability of revoking a compromised face code vector. It should be computationally easy to generate numerous distinctive face code vectors whenever required. In our scheme, a large number of face code vectors can be generated by different random seeds.

To evaluate the revocability of LIoM hashing, a user-specific key scenario is considered. This means each subject should has his specific projection matrix **W** (seed), hence only one subject's face code vector needs to be revoked when the face code vector is compromised. The revocability property of LIoM hashing can be examined through the distributions of *Mated-imposter scores*, *Genuine scores,* and *Imposter scores.* In one system, genuine scores are generated by matching face code vectors from the same subject while imposter scores are computed by matching face code vectors from different subjects. The mated-imposter scores are generated by matching template from the same user but different specific projection matrices **W** to simulate face code vector replacement. The experiments are carried out with m=100 and q=16 for both LIoM hashing and random IoM hashing that transformed from FaceNet features.

As shown in Figure 4.15(a), the distribution of imposter and mated-imposter scores are overlapped largely, which suggest that there is no difference between the face code vectors generated from same individual face or different individual face by different **W**. Thus, the revocability property is justified. The distribution curves from random IoM shown in Figure 4.15(b) also vindicate the revocability of random IoM hashing due to the overlapping distribution of mated-imposter and imposter scores.



(a) Learning IoM

(b) Random IoM

FIGURE 4.15: Revocability of LIoM and random IoM Hashing under VGG2.

4.3 Chapter Conclusion

In this chapter, we evaluated our systems with three large public unconstrained face databases, namely LFW, VGG2 and IJB-C. Good results can be achieved on LFW and VGG2, the performance on IJB-C, which is the most challenging dataset by far today, was not satisfactory under the open-set setting.

We also compared the accuracy of our secured method against the latest state-of-the-art methods and other aspects such as storage, matching efficiency, and template protection. The result demonstrates the usefulness of using LIoM as a compact hashing algorithm. The benefits of the proposed approach were showcased as an open-set face identification system based on LIoM. It can achieve economic storage and an efficient matching speed, while adding an additional layer of template protection.

Chapter 5

Secure Chaff-less Fuzzy Vault for Face Identification System

In the this chapter, a brief introduction of the biometric cryptosystem is discussed first. Followed by the motivations of this chapter. Subsequently, a chaff-less fuzzy vault based on LIoM is described in details. Based on the chaff-less fuzzy vault, a large-scale face identification based on LIoM and the chaff-less fuzzy vault is designed. The system protects both the secret and the face template, and the user simply needs to present his or her face to retrieve the identifier. Finally a security analysis based on several attacks is discussed.

5.1 Introduction

Biometric cryptosystems (BCs) [150, 151] are one category of biometric template protection techniques. These systems can be divided into two categories: key binding and key generation. In the former, the biometric data is bound to a key (secret) to generate helper data (HD) during enrollment. The secret, for example, a random string generated during enrollment, can be treated as an identifier of the user. The HD should not leak information about the biometric or secret and is stored as public information. When a genuine probe biometric is present, the secret is retrieved, but this should be infeasible for an adversary. In key generation schemes, the HD is generated directly from biometric data. At the time of the query, a random key or identifier is generated on the fly, based on the probe biometric and the HD. Fuzzy commitment

[54] and fuzzy vaults [81] are two representative techniques used in key binding schemes, while fuzzy extraction [152] and FE-SViT [153] are typical schemes for key generation.

Although a BC works in verification/authentication mode, it differs from a conventional biometric verification/authentication system since the latter uses a system threshold to determine the similarity of the template and probe features, thus validating the identity of the claimant, while a BC delivers an exact identifier that is bound to or based directly on biometrics. Hence, in addition to being used as the normal means of identity verification, the identifier also can be used as a cryptographic key for encryption/decryption. BCs have been implemented by GenKey, a company based in the Netherlands, for elections and in the digital healthcare sector, mainly in emerging economies [154]. The most notable BC deployment of face biometrics thus far is its application for watch-list purposes in the self-exclusion program of most Ontario gaming sites [155].

To the best of our knowledge, BC was solely designed for use in verification (authentication) rather than identification. The primary difference between verification and identification is that the latter makes a claim to identity by performing a one-to-many search, whereas the former authenticates this identity claim using one-to-one matching. More specifically, a user is required to present identity credentials (e.g., ID, PIN) to support his or her claim to identity at the time of the query, and the corresponding HD is then retrieved based on these credentials. Next, the secret (identifier) is retrieved from the HD based on the corresponding face data. However, face identification is preferred over authentication in some circumstances, and the question of how to adapt BC for identification is, therefore, a non-trivial problem. For a BC that operates in identification mode, the user is only required to present his or her biometric data. The system is then expected to retrieve the correct identifier associated with the user and fail otherwise (for an adversary).

5.2 Motivations and Contributions

This chapter aims to propose a novel facial cryptosystem for identification where only a sole facial image is required for input. In such a setting, the system is expected to retrieve the correct identifier (secret bind to biometrics) that is associated with the user and fail otherwise for an adversary. To address this challenge, a chaff-less fuzzy vault scheme based on LIoM hashing is proposed.

To be precise, the system composes of a 1-to-N search subsystem and a 1-to-1 match CFV subsystem. The first subsystem stores N facial features that are protected by the LIoM hashing in which we coined face code vector. When a face code vector of the user is presented, the first subsystem returns top $k(<< N)$ match scores, and thus the corresponding k vaults in the CFV subsystem will be activated. The 1-to-1 matching occurs among k vaults alongside query face code vector, and an identifier associated with the user will be retrieved from the correct matched vault.

Apart from the above, we consider a challenging scenario where the facial images are taken from the unconstraint environment and the number of subjects is large in the gallery. We again adopt two deep convolutional neural networks, namely FaceNet and InsightFace, as a means for facial feature extractor. These powerful networks effectively alleviate face intra-class problems that directly impact searching and identifier retrieval in the first and second subsystems, respectively. In order to further enhance the performance, a biometric fusion module proposed in Chapter 3 is introduced to the first subsystem.

LIoM hashing is inherited from the random IoM hashing [17] where the latter is a biometric template protection method to protect the biometric template from being inverted while preserves the original template (before hash) accuracy performance. However, random IoM hashing is data-agnostic, and hence the size has to be sufficiently large to achieve decent accuracy performance. The LIoM hashing proposed in Chapter 3 is meant to resolve these issues where a data-driven supervised learning mechanism is utilized for better performance and feature compaction.

Chaff-less fuzzy vault (CFV) is proposed as a variant of fuzzy vault meant to address the uniform mixing genuine and chaff set problem, which has long plagued the fuzzy vault design. The CFV takes the best part of fuzzy vault and fuzzy commitment to eliminate the need for a chaff set for genuine data concealment. By coupling with the LIoM hashing, vectorial biometrics such as face biometrics can be applied to fuzzy vault seamlessly, and security can also be greatly enhanced. The security of CFV is based on computation hardness of polynomial reconstruction. Unlike conventional fuzzy vault, the CFV treats LIoM hash face code vectors as (ordered) point set where they are naturally in the finite field representation, while the fuzzy vault is meant for unordered point sets that required quantization. This eliminates the risk of information loss due to quantization. Another major distinction of the CFV from the fuzzy vault is that the chaff set is not required for genuine set concealment.

Besides, combining the 1-to-N searching module in Chapter 3, we propose a face cryptosystem for identification (FCI) system where only sole input biometric is needed. The FCI composes of a 1-to-N search subsystem and a 1-to-1 match CFV subsystem. The first subsystem stores N facial features protected by a learning-based Index-of-Max hashing and enhanced by a fusion module for searching accuracy. When a face image of the user is presented, the subsystem returns top k match scores, and thus the corresponding vaults in the CFV subsystem will be activated. The 1-to-1 matching occurs among k vaults alongside query face, and an identifier associated with the user will be retrieved from the correct matched vault.

In summary, the main contributions of this scheme are as follows:

1. A novel facial cryptosystem for identification (FCI) is outlined.
2. A chaff-less fuzzy vault for facial biometrics is proposed.
3. To improve the overall system accuracy, two deep learning-based facial feature extractor is adapted to generate two feature vectors for each face instance, and several fusion mechanisms are introduced.
4. Three large unconstrained face datasets of increasing complexity: LFW dataset, VGG2, and the IJB-C dataset are adopted in this study.

5.3 Overview of the Facial Cryptosystem for Identification

The FCI composes of 1-to-N search and chaff-less fuzzy vault (CFV) subsystems. In the first subsystem, the input facial images are first detected, aligned by the MTCNN proposed in [128] and cropped to a canonical size. Face features are extracted by pre-trained networks, i.e., FaceNet or/and InsightFace (see section 3.4.1). The face features are transformed into a face code vector by means of LIoM hashing. This enables the face features to be compressed and protected. During query time, the face code vector (probe) is matched with N hashed galleries via Hamming matcher. The N matching scores are then sorted, and the top k scores are taken. To boost the matching accuracy, a fusion module is introduced to the 1-to-N subsystem (see chapter 3, section 3.4.3). In the CFV subsystem, the corresponding k vaults are activated, and k rounds of 1-to-1 matching are carried out alongside the probe face code vector. Finally, an identifier of the probe user will be retrieved from the matched vault via polynomial interpolation.

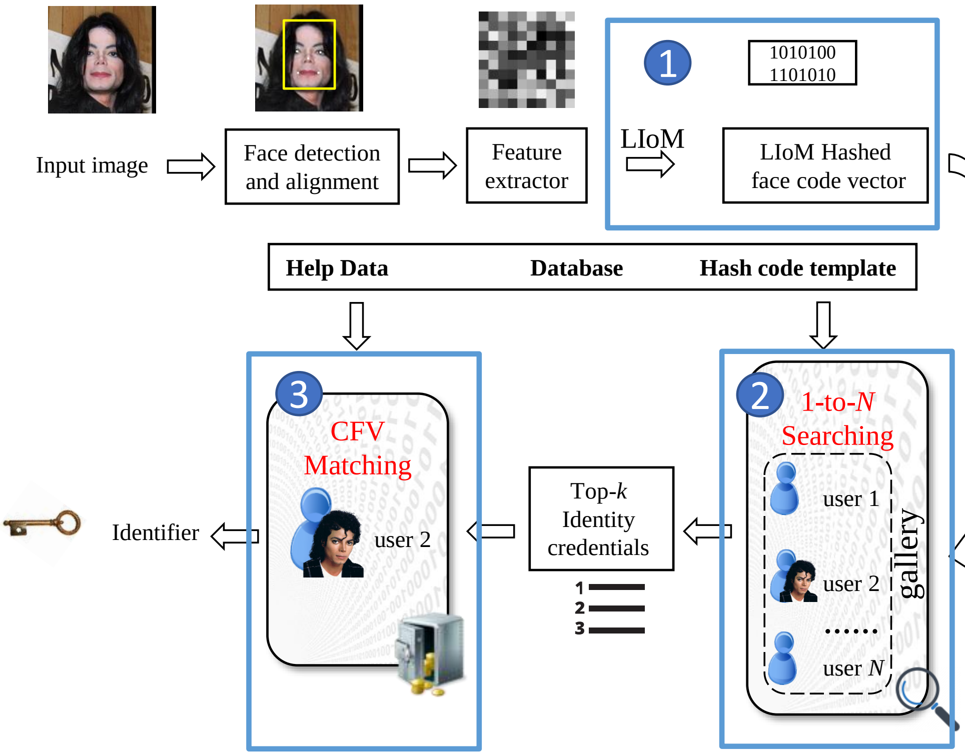


Figure 5.1: Overview of the face cryptosystem for identification system architecture. ① is the face code vector generation by means of IoM hashing; ② is the 1-to-N searching module which returns top k scores. A fusion module is introduced to the 1-to-N subsystem; ③ is the 1-to-1 matching module, where the corresponding k vaults are activated and k rounds of 1-to-1 matching are carried out alongside probe face code vector to release the identifier. (adopted from [2])

Note identifier and secret use interchangeably. The pipeline of the proposed FCI is shown in Figure 5.1.

5.4 Chaff-less Fuzzy Vault

Suppose $\Psi = \Lambda(\mathbf{x}; \mathbf{W}) \in [1\ q]^m, \varphi = \Lambda(\mathbf{x}; \mathbf{U}) \in [1\ q]^m$ be the enrolled face code vectors that generated from a user and the LIoM projection matrices $\mathbf{W} \neq \mathbf{U}$, where $\Lambda()$ is a LIoM hashing generator. This implies both Ψ and $\boldsymbol{\varphi}$ are independent and unlinkable and it can be done by initiated the algorithm 2 with distinctive seeds for $\mathbf{W}$ and $\mathbf{U}$. Given a finite field polynomial $P(\cdot) \in \mathcal{F}_q^p$ of $p-1$ order where the secret (identifier) K is encoded as coefficients of $P(\cdot)$ and q is the subspace dimension of LIoM, Ψ is projected onto $P(\Psi)$ and an ordered genuine set $G = [(\psi_i, P(\psi_i))\, |i = 1, \ldots, m]$ where $\psi_i \in \Psi$ can be acquired. To conceal G, XOR encryption is adopted to encrypt $P(\Psi)$ with φ yield a vault, $\mathcal{V} = [\varphi_i \oplus P(\psi_i)\, |i = 1, \ldots, m]$ where $\varphi_i \in \varphi$.

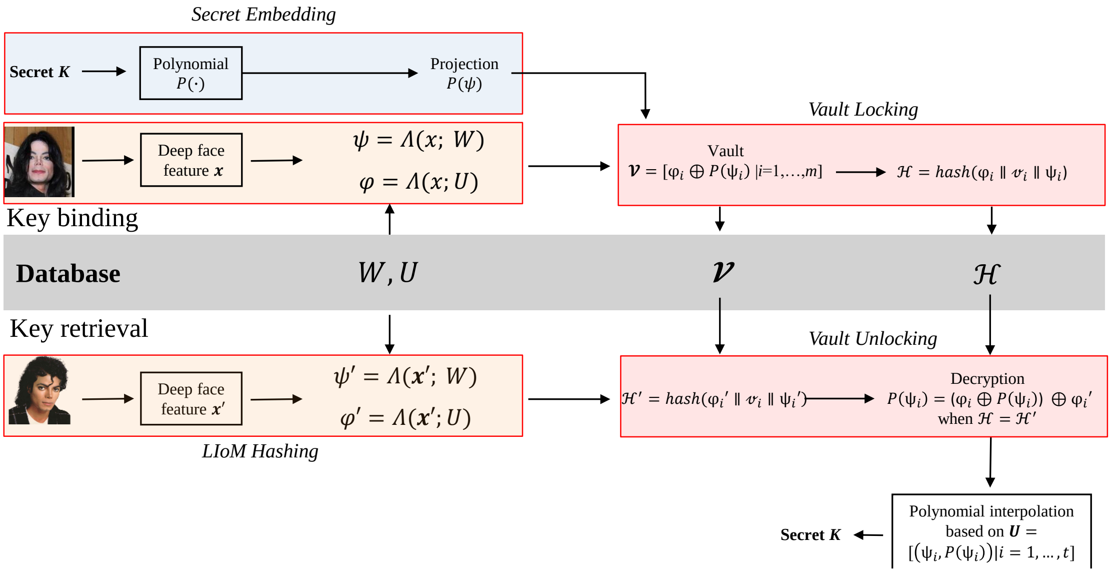

Secret Embedding
Secret K
Polynomial $P(\cdot)$
Projection $P(\psi)$
Vault Locking
Deep face feature x
$\psi = \Lambda(x;\ W)$
$\varphi = \Lambda(x; U)$
Vault
$\mathcal{V} = [\varphi_i \oplus P(\psi_i)\ |i=1,\ldots,m]$
$\mathcal{H} = hash(\varphi_i \parallel v_i \parallel \psi_i)$
Key binding
Database
W, U
$\mathcal{V}$
$\mathcal{H}$
Key retrieval
Vault Unlocking
Deep face feature x'
$\psi' = \Lambda(x';\ W)$
$\varphi' = \Lambda(x'; U)$
$\mathcal{H}' = hash(\varphi_i' \parallel v_i \parallel \psi_i')$
Decryption
$P(\psi_i) = (\varphi_i \oplus P(\psi_i)) \oplus \varphi_i'$
when $\mathcal{H} = \mathcal{H}'$
LIoM Hashing
Polynomial interpolation based on $\boldsymbol{U} = [(\psi_i, P(\psi_i))|i = 1, \ldots, t]$
Secret K

FIGURE 5.2: Chaff-less Facial Fuzzy Vault Illustration. (adopted from [2])

This mechanism resembles fuzzy commitment where φ and Ψ corresponds to binary biometrics and encoded secret (codeword), respectively. Similar to fuzzy commitment where codeword is hashed, the CFV owns $\mathcal{H} = \text{hash}\left(\varphi_i \,\|\boldsymbol{v}_i\|\ \ \psi_i\right)$ where $v_i \in \mathcal{V}$, $hash(\cdot)$ is a cryptographic hash function such as SHA-family function and $||$ denotes concatenation operator. However, in the CFV, the secret is embedded in the polynomial, biometric features is protected by LIoM hashing and the SHA-hashed entity contains $\{\boldsymbol{\varphi}, \boldsymbol{v}, \Psi\}$, which is in contrast to the fuzzy commitment where the secret is merely encoded with the ECC, biometric left unprotected and hashed codeword, respectively. In addition, the CFV merely applies polynomial interpolation to address the intra-class variation problem. This eliminates many issues caused by the ECC in fuzzy commitment [150].

During secret (identifier) retrieval, a query pair $(\boldsymbol{\varphi}', \boldsymbol{\Psi}')$ is generated with their respective $\mathbf{U}$ and $\mathbf{W}$ from a given biometric vector $\mathbf{x}'$. The vault $\boldsymbol{v}$ can be unlocked by comparing $\mathcal{H} = \text{hash}\left(\boldsymbol{\varphi}_i \,\|\boldsymbol{v}_i\|\, \boldsymbol{\psi}_i\right)$ and $\mathcal{H}' = \text{hash}\left(\boldsymbol{\varphi}_i' \,\|\boldsymbol{v}_i\|\ \ \psi_i'\right)$ via a filtering algorithm in the element-wise manner. If a sufficiently large number of elements say $t(\leq m)$ from $\mathcal{H}$ and $\mathcal{H}'$ are matched, decryption i.e. $\left(\varphi_i \oplus P\left(\psi_i\right)\right) \oplus \varphi_i'$ would succeed and $P\left(\psi_i\right)$ is output or fails otherwise. Upon success decryption, one construct an unlocking set $\boldsymbol{u} = \left[\left(\psi_i, P\left(\psi_i\right)\right)|i = 1, \ldots, t\right] \subseteq G$ where $t \geq k$, the identifier can be retrieved via polynomial interpolation. A high-level overview of the CFV is illustrated in Figure 5.2.

The detail steps of enrollment and identifier retrieval is given as follows:

Identifier Binding (Enrollment):

Given deep face vector $\mathbf{x} \in \mathbb{R}^d$, identifier K, LIoM projection matrices $\mathbf{U}$ and $\mathbf{W}$, face code vector with size m, subspace dimension q, a finite field polynomial $P(\cdot) \in \mathcal{F}_q^p$ with $p-1$ order and a one-way hash function: $\{0,1\}^* \rightarrow \{0,1\}^\ell$:

1. Encode K as the coefficients of $P(\cdot)$
2. Generate $\Psi = \Lambda(\mathbf{x}; \mathbf{W})$ and $\boldsymbol{\varphi} = \Lambda(\mathbf{x}; \mathbf{U})$
3. Perform polynomial projection $P\left(\psi_i\right), i = 1, \ldots, m$ where $\psi_i \in \Psi$
4. Encrypt $P\left(\psi_i\right)$ with $\varphi_i \in \boldsymbol{\varphi}$ to generate a vault $\boldsymbol{V} = \left[\varphi_i \oplus P\left(\psi_i\right) | i = 1, \ldots, m\right]$
5. Generate $\mathcal{H} = \text{hash}\left(\varphi_i \left\|v_i\right\| \psi_i\right)$ where $v_i \in \mathcal{V}$
6. Store $\mathbf{U}, \mathbf{W}, \boldsymbol{V}$ and $\mathcal{H}$ as public helper data.

Identifier Retrieval:

Given query deep face vector $\mathbf{x}' \in \mathbb{R}^d$, helper data, face code vector size m and hash function hash.

1. Generate $\Psi' = \Lambda\left(\mathbf{x}', \mathbf{W}\right)$ and $\boldsymbol{\varphi}' = \Lambda\left(\mathbf{x}', \mathbf{U}\right)$
2. Compute $\mathcal{H}' = \text{hash}\left(\varphi_i' \left\|v_i\right\| \psi_i'\right), i = 1, \ldots, m$ where $\psi_i \in \Psi$ and $\varphi_i \in \varphi$
3. Run Algorithm 4 with $\mathcal{H}$ and $\mathcal{H}', \Psi, \boldsymbol{\varphi}'$ and $\mathcal{V}$ to acquire $\boldsymbol{U} = \left[\left(\psi_i, P\left(\psi_i\right)\right) | i = 1, \ldots, t\right]$ where t is the number of match entries in $\mathcal{V}$. Note that Unique($\boldsymbol{U}$) in Algorithm 3 is a function that warrants only unique genuine pairs present in $\boldsymbol{U}$. A failure signal will issue if $t \leq p$
4. Execute polynomial reconstruction with $\boldsymbol{u}$.

In step 4, the secret can be retrieved via Lagrange interpolation if $p \leq t \leq m$. since there will be at most $q(< t)$ unique elements only in $\boldsymbol{u}$ since $\left(\psi_i, P\left(\psi_i\right)\right) \in [1, q] \times [1, q]$, Unique $(\boldsymbol{u})$ is required to prevent unnecessary computation overhead on polynomial interpolation. On the other hand, $\mathcal{H}$ is crucial for the CFV to perform a conditional check to ensure only relevant $v_i \in \mathcal{V}$ can be decrypted. This can be perceived as a kind of error-checking mechanism in the CFV. Besides that, $\mathcal{H}$ warrant data integrity of vault $\mathcal{V}$. This is because if any alteration occurs

Algorithm 4 Genuine Pairs Filtering

```
Input:
    $\mathcal{H}, \psi', \varphi', \mathcal{V}$
Output:
    unlocking set $\mathcal{U}$
Initialization: $u \leftarrow \emptyset, \mathcal{H}' = \text{hash}\left(\varphi_i' \| v_i \| \psi_i'\right)$
for $i = 1$ to $m$ do
    /* $k_i \in H$, $k_i' \in H'$ */
    if $k_i = k_i'$ then
        $P(\psi_i) = (\varphi_i \oplus P(\psi_i)) \oplus {\varphi_i}'$ /* Decryption */
        $\mathcal{U} \leftarrow \mathcal{U} \cup (\psi_i,\ P(\psi_i))$ /* Collect unlocking set */
    end if
end for
$u \leftarrow$ Unique $(\mathcal{U})$
```

in $\mathcal{V}$, the decryption will fail. Lastly, $\mathcal{H}$ facilitates polynomial interpolation by ensuring that only genuine pairs are found in $\boldsymbol{U}$, and hence the identifier can be retrieved simply with a single step of polynomial reconstruction rather than iterative decoding that practiced by ordinary fuzzy vault scheme.

It is worth highlighting that the proposed CFV can work with either random version and learning-based version of IoM hashing since both LIoM and random IoM can produce integer/binary templates. Adopting random IoM will avoid training and system complications. Besides, other hashing which can produce integer/binary templates can also be adopted in CFV; hence the deployment of CFV shall depend on different biometric modalities and hashing techniques.

5.5 Performance Evaluation

Since there are three samples per subject stored in the gallery for a given probe, the secret would be retrieved successfully when matched with the corresponding genuine samples. If more than one sample in the gallery of the same subject is matched successfully, the system outputs an accepting signal and releases the mode of the retrieved secret. An example for this circumstance is illustrated in Table 5.1. In this example, G_k^i refers to kth subject's ith $(i = 1, 2, 3)$ template in the gallery while P is a probe. The result will be mode(G_1^1,G_1^2,G_1^3) .

In this chapter, identical datasets and protocols in chapter 4 are adopted. From subsection 4.1.2.6 in chapter 4, we notice either score fusion by averaging or feature fusion by concatenation offers the overall best performance. However, two LIoM face code vectors, where each code vector is the resulting of fused features, are required for the CFV. Hence, we opt for feature

Table 5.1: An example of secret retrieval matching with gallery.

Probe	G_1			...	...
Gallery	G_1^1	G_1^2	G_1^3	...	G_k^i
Subject P	✓	✓	×	...	×
Retrieval Result	mode(G_1^1,G_1^2,G_1^3) → user $G1$				

fusion in implementation as it incurs less computation overhead than that of score fusion for the FCI.

5.5.1 Performance Metrics

In the CFV subsystem, for a given probe P, the k vaults $\mathcal{V}$ that associated to the gallery in the search subsystem will be activated and followed by decryption for each $\mathcal{V}$. A unique, genuine entry will be unlocked upon successful decryption and then followed by polynomial reconstruction. Therefore, we enumerate four possible cases as follows:

Case 1. Identifier is correctly retrieved when P and his/her associated vault is presented.

Case 2. Identifier is incorrectly retrieved when P and his/her associated vault are presented.

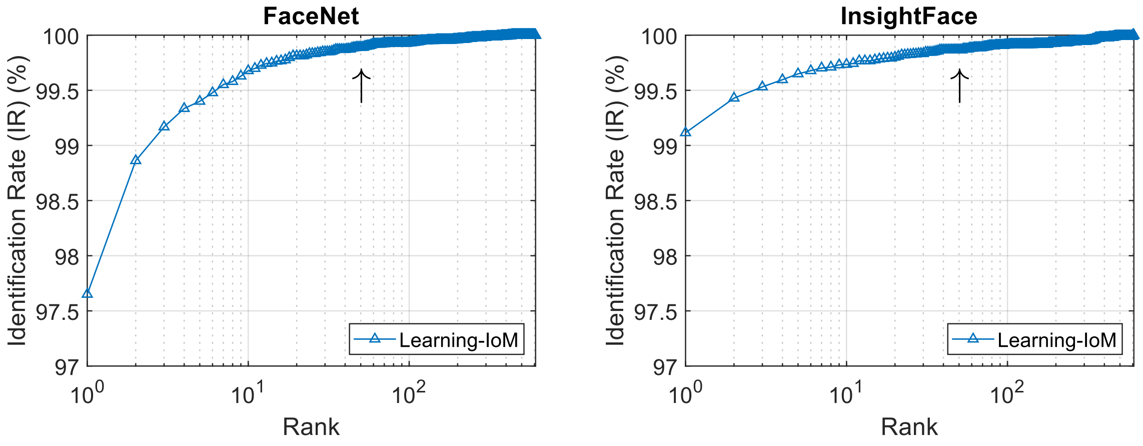
Case 3. Identifier is correctly retrieved despite P, and vault are not associated.

Case 4. Failure of identifier retrieval due to unsuccessful decryption. Specifically, the unsuccessful decision is given when the number of match entries in $\mathcal{V}$ is not more than p, i.e., $t \leq p$ as stated in step 3 in **Identifier Retrieval** process (section 5.4).

Note case 2 and case 3 are analogous to false rejection and false acceptance in the ordinary biometric systems, respectively. Case 4 is unique in our context since it may happen to genuine subjects and imposter.

Suppose there are Q probes, and based on the four cases above, three performance metrics with respect to p can be defined to evaluate the identifier retrieval rate:

1. True Identify Rate, TIR(p) $= \frac{\#Correct\ Retrieval}{Q}$ for case 1.
2. Misidentify Rate, MIR(p) $= \frac{\#Incorrect\ Retrieval}{Q}$ for case 2 and 3.
3. Failure Identify Rate, FIR(p) $= =\frac{\#Fail\ Retrieval}{Q}$ for case 4.



(a) FaceNet

(b) InsightFace

FIGURE 5.3: IR (%) vs k of LIoM hashing on the LFW: (a) FaceNet; (b) InsightFace. (adopted from [2])

Apart from the three metrics above, we further define an indicator, namely TIR@MIR. In the FCI, the TIR is expected to be high for correct identifier retrieval and otherwise for MIR. This is because an adversary can exploit MIR to launch zero effort false accept attack (will be detailed in section 5.6).

5.5.2 Parameter Tuning for 1-to-N Searching

The key parameters for the CFV subsystem include polynomial order $p-1$, m, and q from LIoM hashing. From section 4.1.2.4 in chapter 4, we find that the performance of searching is reasonably good for LIoM hashing with $m = 100$ and $q = 16$. As indicated in chapter 4, the LIoM hashing performance could be critical for identifier retrieval, we choose a larger m and q where $m = 200$ and $q = 32$ and p is tuned from 5 to 32.

Another parameter for 1-to-N searching subsystem is top-k candidate size. The k value can be inferred from the CMC curve by referring to the IR at different rank k. From the CMC that based on the LFW shown in Figure 5.3, the IR improves and level off when k increases. We opt for k=50 for subsequent experiments as overly large k indicates higher computation overhead in the CFV.

Table 5.2: TIR@0.1%MIR of FCI key retrieval (%).

	Close-set	Open-set O1	Open-set O2	Open-set O3
LFW	98.57	74.53	52.04	44.20
VGG2	98.92	48.93	73.68	42.10
IJB-C	43.02	30.30	25.11	19.72

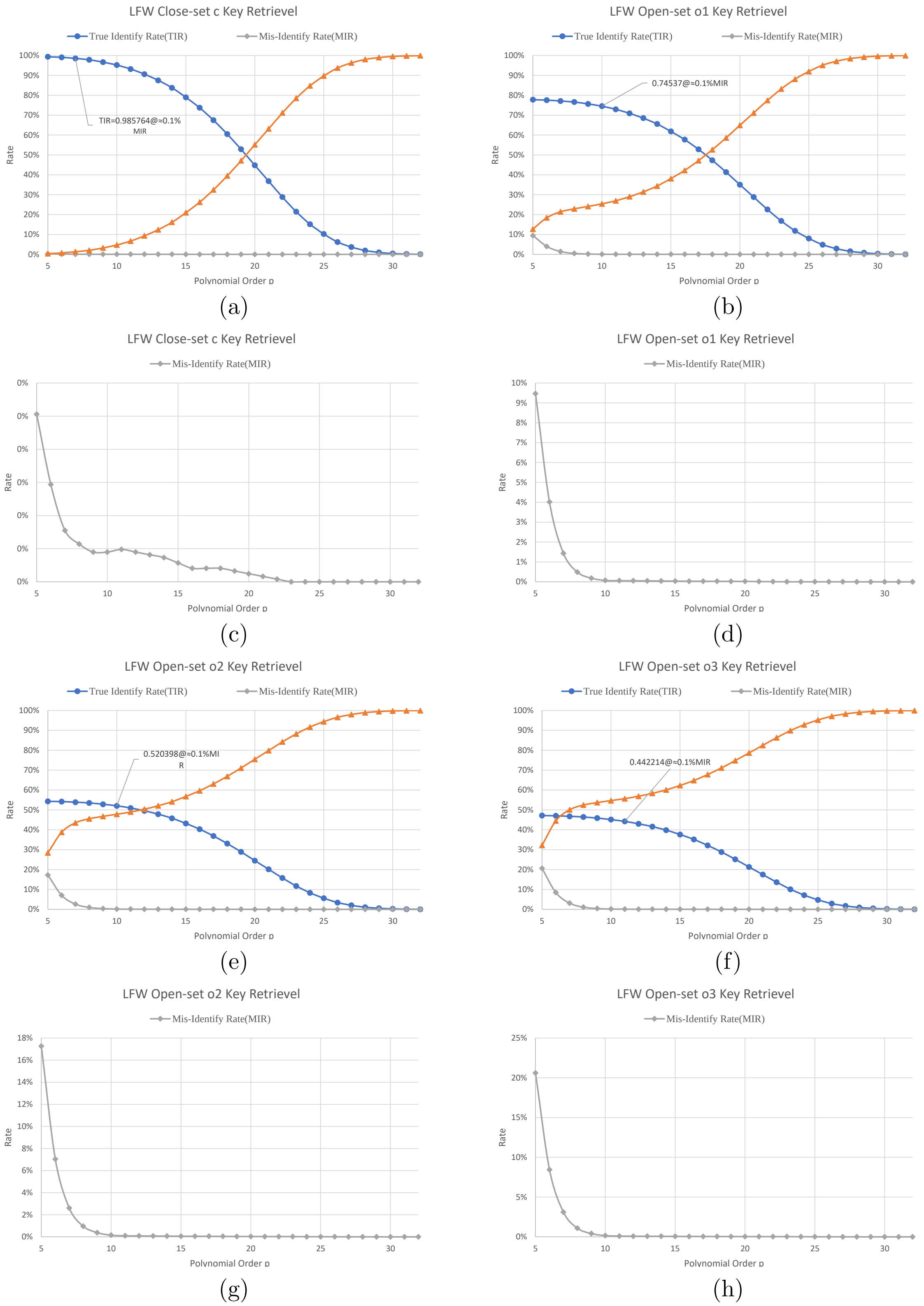
5.5.3 Secret Retrieval Performance

To have an evaluation of the FCI, the secret retrieval experiment is carried out on LFW, VGG2 and IJB-C to record the accuracy performance metrics (Figure 5.4, Figure 5.5, and Figure 5.6). As we can expect, when p is small, more probes will be retrieved successfully, and some of them may be retrieved with a wrong key, thus lead to higher MIR. When p becomes larger, fewer probes will pass the key retrieval process, and the MIR will decrease gradually. The TIR@0.1%MIR is recorded in Table 5.2.

We can find that close-set is easy to achieve a good accuracy performance, except IJB-C dataset, which is more challenging due to its characteristic. However, it is still hard to achieve a good performance in the open-set protocol, which remains a big challenge.

Since MIR is a critical factor for the system, we can choose a p that makes MIR $\approx 0.1\%$ and TIR is the largest among all p. As anticipated, for small p, it is easier to satisfy vault unlocking subject to $t \leq p$, where t is the number of unlocked genuine entries from vault $\mathcal{V}$. Thus, high TIR and low FIR can be anticipated. Unfortunately, this may lead to higher MIR in return as the probability for an imposter to succeed increases due to the lower bar of collecting unique genuine entries for polynomial reconstruction. On the contrary, vault unlocking becomes difficult for large p and thus suppresses MIR as well as TIR, but heightens FIR in return. For instance, from Figure 5.4 (a) and Figure 5.4 (c), we observe that TIR decreases to 10% from $p = 5$ to $p = 25$, while MIR drops from 0.27% to 0%. It is also noted that the identifier retrieval performance is directly related to the discrimination of face code vectors (difficulty of face dataset) as indicated by the TIR@0.1MIR in Table 5.2.

In summary, the entire FCI system's identifier retrieval performance is regulated by four system parameters, namely top-k candidate size that returned by search subsystem, polynomial order $p - 1$ as well as m and q from LIoM hashing. Note that m and q can be set differently for two subsystems as they just depend on the number and size of projection matrices of LIoM hashing. In practice, we recommend choosing small m with a suitable q for the first subsystem due to searching time concern at the expense of loss of discrimination. However, k value can be



(a) (b)

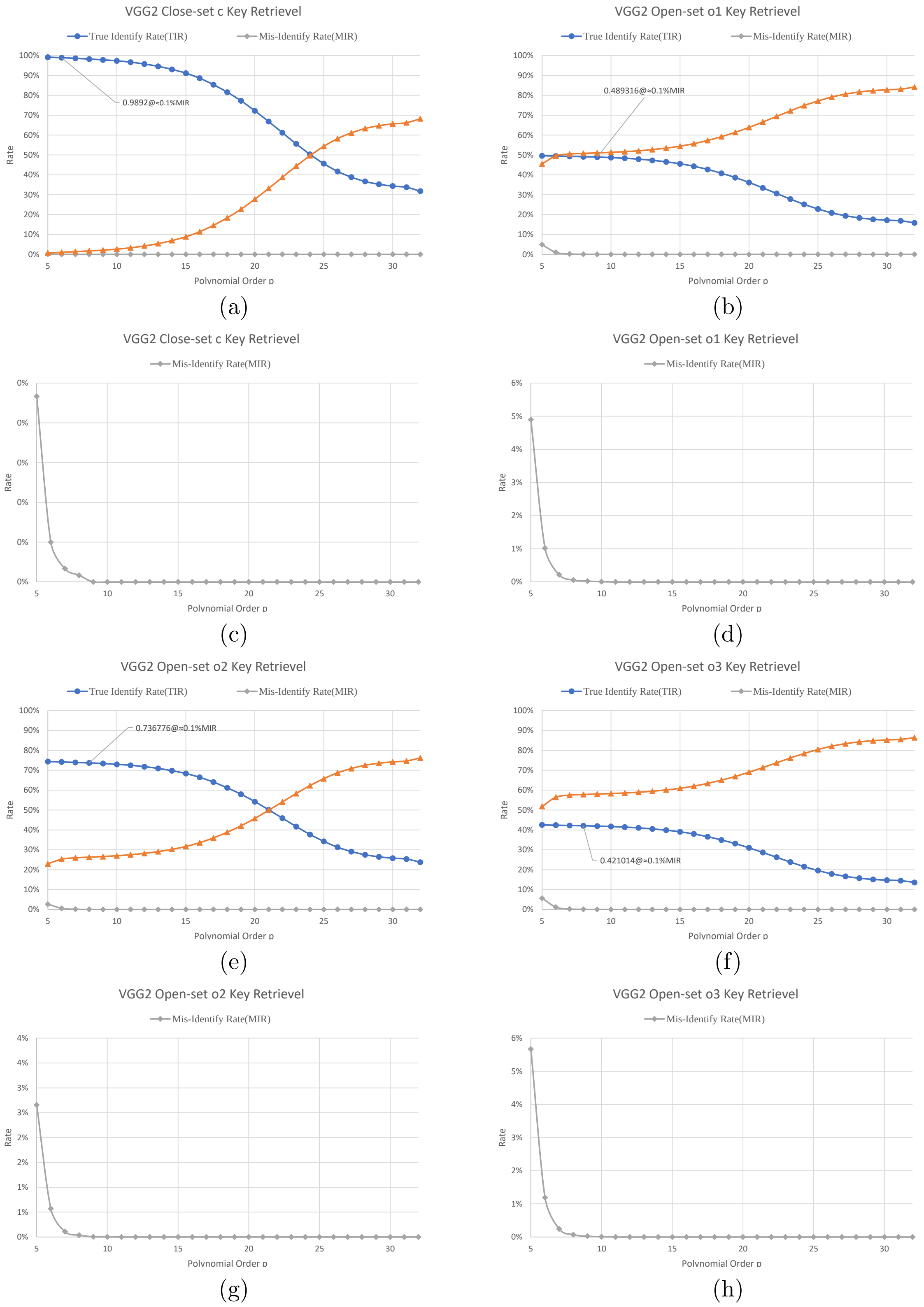
(c) (d)

(e) (f)

(g) (h)

FIGURE 5.4: Key retrieval performance of LFW by fusion.

tuned to compensate for the effect of m and q. On the other hand, m can be set larger alongside q for the CFV to increase the success probability of decryption. Finally, p is the most critical



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 5.5: Key retrieval performance of VGG2 by fusion.

to determining the eventual performance of the FCI.

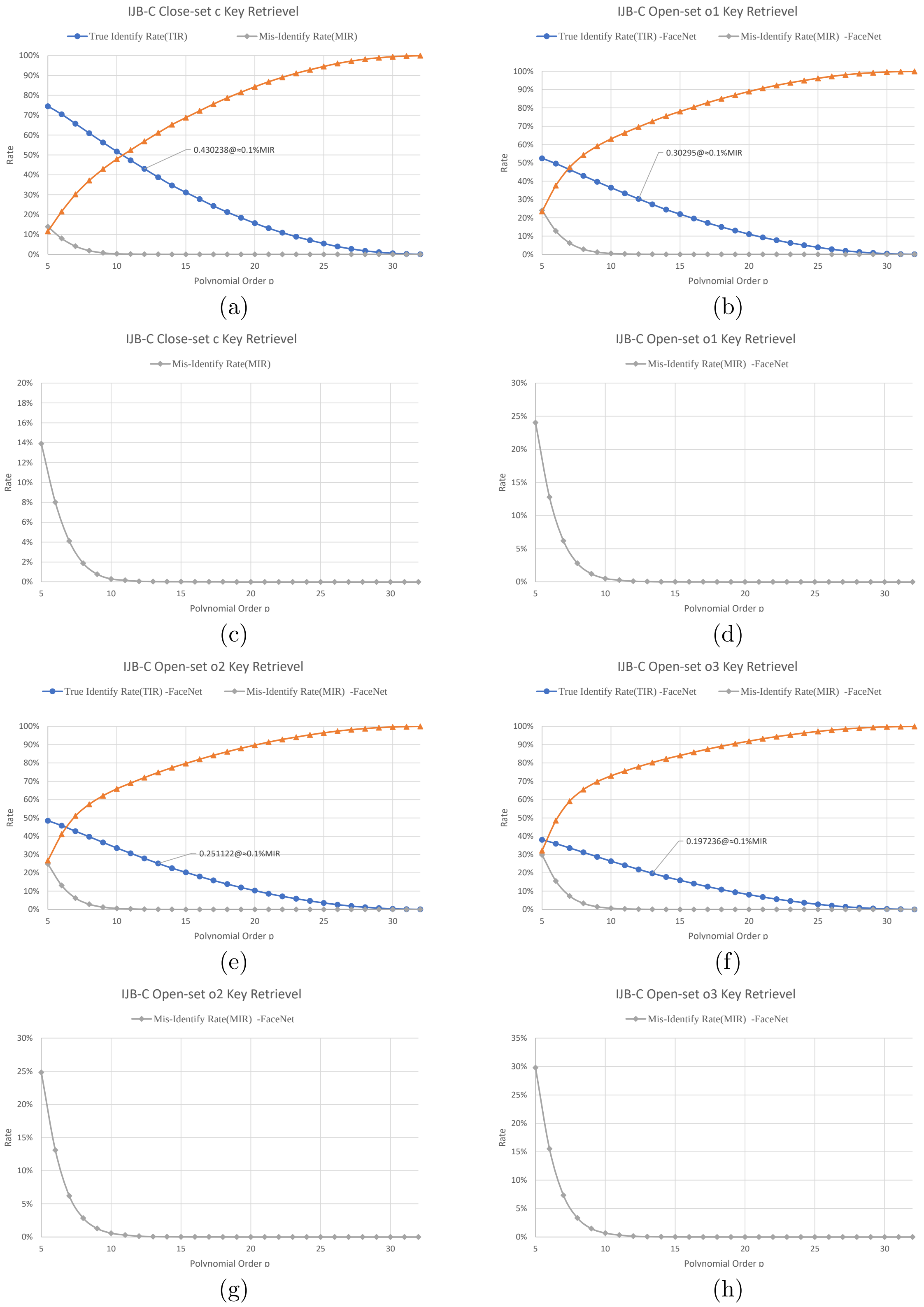


Figure 5.6: Key retrieval performance of IJB-C by fusion.

Table 5.3: Summary of Identifier Retrieval Time Cost of one probe with $p = 10$ and $p = 15$.

	Gallery images	Probes images	Average time (s) per probe (p=10)	Average time (s) per probe (p=15)
LFW	1830	4903	0.12	0.11
VGG2	6000	6000	0.19	0.19
IJB-C	3594	5391	0.05	0.04

5.5.4 Computation Efficiency

Computation efficiency is another key performance factor for the FCI system aside from accuracy. To record the time cost, the total computation time in second, which includes both search and the CFV subsystems, for all probes in LFW are first recorded. Machine and Software used in the evaluation are as follows:

- Matlab version R2017b, 64-bit.
- Win10 Enterprise 1709.
- Intel i7-6700 CPU @ 3.40GHz, RAM 16GB.

The average time for one probe can be acquired by dividing the total time with total probes number 4903 in LFW. From Figure 5.7, we note the time cost decreases with respect to the increment of p. This is due to the successful probability decryption becoming lower for large p, polynomial reconstruction, which is the most time-consuming, would not occur. Specifically, failure alarm will be issued when $t \leq p$, as stated in step 3 of Identifier Retrieval in section 5.4. To reconstruct the polynomial, at least p matched entries should be found. When p becomes larger, it will be hard for the system to collect sufficient match entries for polynomial reconstruction. However, the system will only process the queries when $t > p$, otherwise retrieval will be halted to save processing time.

On average, we can see that only around 0.1 seconds is needed to retrieve the identifier among 1830 samples in the gallery. A summary of the time cost can be found in Table 5.3. In summary, it is crucial to find a balance point for polynomial order p, which can lead to decent performance with a reasonable waiting time.

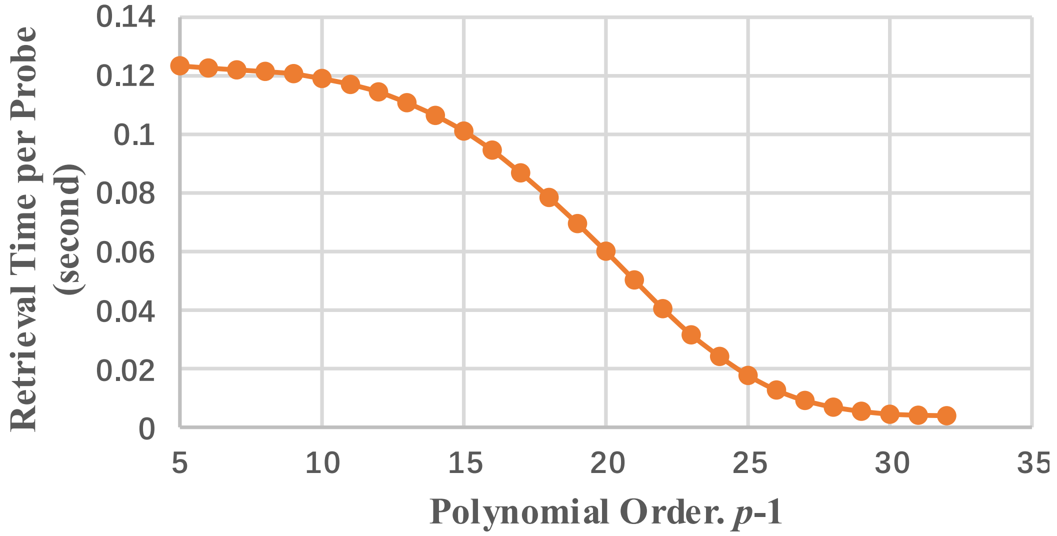


Figure 5.7: Identifier Retrieval Time Cost for One Probe against p on LFW. (adopted from [2])

5.6 Security Analysis for the CFV

Despite the FCI system composed of two subsystems, the search subsystem's security relies on LIoM hashing, which has been analyzed and presented in [17] and section 4.2. Therefore, we shall focus on the security of the CFV in this section. Specifically, the CFV security will be analyzed based on the four major attacks targeted to fuzzy vault and its variants. To do so, we first formalize the security model of the CFV.

5.6.1 Security Model

Ideally, the security of the CFV can be simply reduced to the computational hardness in seeking q out of m matches via a filtering procedure (algorithm 4) in the vault. This analysis is based on the random oracle model over a uniformly random enrolled $\mathbf{x}$ and query face features $\mathbf{x}'$. The security can be assured by choosing a large $\{q,\, m\}$ that satisfies performance-security-efficient trade-off.

However, the above assumption is impractical as real biometric features can never be uniformly random. To resolve this issue, the security is characterized by the similarity of $\mathbf{x}$ and $\mathbf{x}'$ or $S\left(\mathbf{x},\mathbf{x}'\right)$ more precisely, which follows certain non-uniform distribution. Formally, we adopt a *random error model* that outlined in [152], whereby the probability of matches is captured by

α^2 and $1-\alpha^2$ for mismatches probability where $\alpha^2 = S\left(\mathbf{x},\mathbf{x}'\right)^2$ that developed in section 5.4. This induces another issue in which the estimation of α can only be meaningful based on the massive biometric samples. To alleviate this problem, we adopt the *maximum match probability* notion, denoted as $P^{\max}_{M_I}$. Formally, $P^{\max}_{M_I} = \max\left\{P_i | i = 1, \ldots, M_I\right\}$ where P_i is the imposter matching score and M_I is the total number of imposter scores, which can be obtained when EER is computed (section 4.1.2.7). This enables us to demonstrate security over a moderate sample size that considers the non-uniform nature of biometric features.

5.6.2 Sneak Key Inversion (SKI) Attack

The SKI Attack [156] is a scenario where the adversary merely guesses the secret that bond to the biometric cryptosystems unlock the vault and eventually recover biometric data via the broken vault. An effective security measure to resist the SKI attack bound within the exhausting guesses of the secret. In our context, the secret (identifier to be exact) is typically encoded as the coefficients of the finite field polynomial with order p - 1 with a total of $\log\left(\beta\right)^p$ bits, where β is a prime number. In our experiment, the p varies from 5 to 32 and β =13, hence the guessing complexity range is of 13^5 bits to 13^{32} bits, which is sufficient to withstand the SKI attack on many occasions. Moreover, the deep face features in the FCI is protected by LIoM hashing, hence serve another layer of protection against SKI attack.

5.6.3 Surreptitious Key Inversion (SuKI) Attack

Unlike the SKI attack that depends on the guessing of secret to break the biometric cryptosystem, SuKI attack [156] exploits the compromised secret directly for the same purpose. In our context, the revelation of the identifier may lead to the disclosure of genuine set, **G** from the CFV vault $\mathcal{V}$. However, the genuine entries are encrypted in the vault thus disclosure is highly unlikely.

Similar to the SKI, the face features are also protected by the LIoM hashing. Here, we assume the adversary manages to retrieve the face code vectors and he/she knows well the hashing algorithm as well as the corresponding parameters e.g. q and m and projection matrices **W** and **U**. We note that the LIoM hashing converts the deep face feature into the integer indices, which has no clue to guess the actual face vector information directly from the stolen LIoM hashed face code vector.

5.6.4 Brute-force Attack

The brute-force attack on the CFV involves an adversary attempting to guess the genuine set $\mathbf{G} = \left[\left(\psi_i,\ P\left(\psi_i \right) \right) | i = 1, \ldots, m \right]$ directly, so that he or she can retrieve the identifier indirectly or guess the secret directly.

A secret $K \in \mathcal{F}_r^p$ is encoded as the coefficients of $P(\cdot)$ with polynomial order $p - 1$. The complexity of an attack that involves guessing the secret depends on the secret space, the polynomial order p, and r. We use $r = 8$, $p = 16$ as an example, which means that the secret is encoded in 8 bits with polynomial order 16, and the complexity of guessing the secret (the polynomial coefficients) is $\left(2^8\right)^p = 2^{128}$. A larger value of p increases the guessing complexity and guarantees resistance to this attack.

To guess the genuine set $\boldsymbol{G}$, the adversary must enumerate a minimum q of ψ_i randomly. Given $\mathcal{H}_i = \text{hash}\left(\varphi_i \left\| v_i \right\| \psi_i \right)$, where $\varphi_i \in \mathcal{F}_q$ and $\hat{\varphi}_i \in \mathcal{F}_q$, the adversary needs to find a collision on a hash subject to $\mathcal{H}_i = \mathcal{H}_i'$ after q^2 trials. To retrieve the secret, at least p collisions are needed, and hence the attack complexity is $q^{2*p} = 16^{2*p} = 2^{8*p}$. When $p = 16$, the guessing complexity is 2^{128}, which provides resistance to attacks involving guessing $\boldsymbol{G}$.

5.6.5 MisIdentify Attack

MisIdentify attack (MIA) refers to a scenario where an adversary can repeatedly attempt to verify $\mathcal{H}$ with the biometric instances from the compromised or public face databases. In this circumstance, the adversary is expected to retrieve the identifier with the probability equal to nonzero MIR. The MIA is equivalent to the False Accept Attack (also known as dictionary attack) in the normal fuzzy vault schemes and biometric systems. Besides that, the adversary can utilize an artificial template generator alone to launch MIA. Yet, it is also possible for the adversary who possesses sufficient computation power, allow him or her to launch MIA by sampling the face features randomly according to certain distribution. From this perspective, the MIA complexity can be formulated in terms of $\alpha^2 = \left(P_{M_I}^{\max}\right)^2$ that presented in the robustness analysis of the CFV (Section 5.4). Hence, MIA complexity in bit over M_I is defined as $\mathbf{MI}\left(m, \delta, P_{M_I}^{\max}\right) = -\log\left(\mathbb{P}\left(t_{\left(P_{M_I}^{\max}\right)} \geq \lceil \delta m \rceil\right)\right)$ where $t_{\left(P_M^{\max}\right)} \sim \text{Bin}\left(m, \left(P_{M_I}^{\max}\right)^2\right)$

With same setting used in section 5.6.4, the maximum MIA complexity is estimated as 37 bits with $p = 28$. This is considerably low but the complexity can be improved by decreasing

m or simply set MIR=0% by adjusting p (please refer section 5.5.4) at the expense of scarifying identifier retrieval rate. It is also noted $P_{M_I}^{\max}$, which is directly linked to the face feature discrimination plays a vital role in MIA. The higher discriminative of face features, the lower $P_{M_I}^{\max}$ can be acquired and thus contribute to higher MIA complexity. Unlike the brute-force attack, the MIA is independent from q. Increasing q would not contribute to better MIA security, and we deem the MIA is a much stronger attack that could serve as the lower bound to brute-force attack via a considerably large q.

5.7 Chapter Conclusion

In this chapter, a facial cryptosystem for identification (FCI) that only requires sole face as input for identifier retrieval is proposed. The FCI is composed of 1-to-N search subsystem and a chaff-less fuzzy vault (CFV) subsystem. To warrant sufficient discrimination on face features for decent accuracy performance, face feature extractors by means of deep neural networks and biometric fusion modules have been adopted. In response to searching efficiency and security of the FCI, the first subsystem employs LIoM hashing for deep learned face features compression and protection. The LIoM transformed face features is a compact code vector, which satisfies irreversibility, unlinkability, and revocability criteria for biometric template protection. We couple LIoM hashing with a novel fuzzy vault variant, i.e., CFV, which can achieve reasonably good identifier retrieval accuracy on LFW, VGG2, and IJB-C large-scale unconstrained face benchmark databases. It is worth noting that there is a strong correlation between accuracy performance and security of the FCI, or the CFV to be precise, and face feature discrimination. Therefore, it is a future work to improve face feature discrimination.

Chapter 6

Conclusions and Futures

In this chapter, a summary of this thesis is given, and some future directions are also discussed based on the current work.

6.1 Conclusion

Face recognition technology is growing rapidly in the interest of convenience and surveillance, and it might transform everything from policing to the way people interact every day with banks, stores, and transportation services. Face recognition is also now becoming one of the standard phone unlock machinery in mobile phones[1]. As FR becomes prevalent in daily life, more and more concerns about the privacy and security risks are rising. Biometric template attack is one of the risks associated with the biometric system that needs to be addressed. Though some template protection techniques are invented to protect the biometric template, there still has some limitations, such as inconvenience caused by the extra use of identifiers in the biometric cryptosystem. Besides, the accuracy under open-set settings will usually drop in most FR systems. This thesis has systematic research to address the challenges regarding the FR system's privacy and security and proposed several algorithms and systems. In summary, the main work and contribution of this thesis are list as below:

BTP is a critical compartment for the secure FR system to protect the biometric features. Existing BTP schemes usually suffer from performance degradation and other issues. In the latest proposed BTP scheme IoM hashing, the IoM hash codes must be long enough to achieve high

[1] Use Face ID on your iPhone or iPad Pro, https://support.apple.com/en-us/HT208109

accuracy performance and high reliability [17]. To address the biometric template protection issue, i.e., the performance degradation and compact integer template demanding, a learning based hashing, namely LIoM, is proposed. As discussed in chapter 3, both LIoM and random IoM exploit the ranking order among random projected values instead of feature values and are resistant to biometric noise. While the projection matrices employed in the IoM hashing are generated randomly, in LIoM, it is learned from the gradient descent algorithm. Specifically, to avoid the non-convex discontinuous optimization problem, the hashing function is relaxed (or approximated) by a designed SoftMax function, then a loss function, which ensures paired samples can generate similar hash codes, is utilized to learn the projection matrix based on the mini-batch update rules. The AdaBoost-based sequential learning algorithm is adopted to boost the performance further. The proposed LIoM can deliver a set of optimal ranking subspaces and improve the accuracy over random IoM hashing. LIoM hashing also achieves a more compact hash code which requires less storage and computation power.

Another challenge of FR is the performance degradation in the large-scale open-set identification settings. Based on the LIoM hashing, a BTP based 1-to-N searching FR system is implemented and evaluated in chapter 3 and chapter 4. We follow a new open-set evaluation protocol, and wherein the probes are categorized into known, known unknown, and unknown unknown. The conventional closed-set protocol is also adopted for benchmarking the proposed system. To improve the performance, several feature-level and score-level fusion strategies for the face identification problem are explored. The proposed system is evaluated based on three large unconstrained face datasets: LFW, VGG2, and the IJB-C dataset. The results validate that performance degradation is inevitable when the deep face vector is transformed by IoM hashing, while LIoM hashing can achieve better performance compared with random IoM. The LIoM can also contribute to template protection as well as speedy matching compared to the use of deep face vectors only. The deep face vector with feature level fusion and score fusion by average can improve the accuracy significantly.

Apart from the 1-to-N face identification application, biometrics are usually utilized in key management systems, which can bind a secret to the biometric data and generate the final protected template (Help Data, or PI in general). Such a system normally is designed for verification (1-to-1 matching) settings, and input of both biometric data and an associate ID are required. To address such limitation, an identifier-free face cryptosystem for identification (FCI) with higher security is proposed in chapter 5. The proposed identifier-free cryptosystem only requires sole face as input to retrieve the protected identifier. Specifically, to achieve the identifier-free face cryptosystem, by mere giving a sole face image, we integrated an LIoM hashing empowered open-set 1-to-N match module, which releases top k matched id, to be used as an identifier upon correct match and a CFV enabled 1-to-1 match module for secret retrieval. We utilized deep face features generated from FaceNet and InsightFace networks to cope with the system's accuracy and matching speed demands. Learning-based IoM, which can be trained based on the labeled data set, protects the face features. Random IoM can be considered as a special case of the learning based IoM. The learning-based IoM converted the long deep face feature vector into a compact binary code with mere 400 bits for speedy matching with a simple hamming matcher without significant performance drop when coupled with fusion methods at 1-to-N matching module.

6.2 Future work

However, there still exist some problems that need to be addressed. Some possible future research directions are discussed below.

6.2.1 Discriminative Features

Applying biometric template protection can lead to performance degradation unavoidably. How to preserve the accuracy while ensuring the privacy and security of users and systems still remains to be a big challenge. One possible solution is to generate high discriminative features from biometric input.

In this study, we found that a robust and discriminative feature representation is quite essential for the system performance, especially in a large-scale unconstrained environment setting. More stable LIoM hash codes can be generated with a robust and discriminative feature representation, and better accuracy can be achieved subsequently. Hence works on features discrimination shall be one of the direction which can promote the development of BTP.

6.2.2 Integrate LIoM to Deep Neural Network

The proposed LIoM can be regarded as an instance of projection operation, and it is natural to be portrayed as a one-hidden-layer network. The next promising work is to design an end-to-end LIoM deep hashing network, which takes face images as the input and outputs hash codes directly. This will reduce the system complexity and improve performance in the same time.

6.2.3 Apply FCI to Other Biometric Features

It is worth to highlight that other deep face features can also be adopted in the proposed scheme, as far as the face features are in fixed-length vector format. On the other hand, the proposed system is only implemented on the face, while other biometric modalities, such as iris and fingerprint, can also be adopted. In practise, as far as the features extracted from the corresponding biometric modalities are in fixed-length vector format, such biometric modalities then could be considered applicable to FCI.

However, it may be hard to generate fixed-length features for some modalities. For example, the most popular fingerprint descriptor, i.e., Minutia cylinder-code (MCC) [157], is size-variant. To utilize FCI on such size-variant features, extra processing such as [158] can be introduced to generate fixed-length features.

6.3 A Cat-and-Mouse Game

The security of the biometric system is just a game between "mouse" and "cat". With the development of technology, adversaries may gain more opportunities to attack the biometric system. There is no absolute secure system, while such biometric systems' security research shall persist and not be stopped. The research from the perspective of adversaries and attacks may be regarded as one of the preventive countermeasures.

In summary, it is endless for research work on security and privacy-preserving biometrics. In the future, the above-discussed directions shall be explored, and more studies will be conducted.

Bibliography

[1] Xingbo Dong, Soohyung Kim, Zhe Jin, Jung Yeon Hwang, Sangrae Cho, and Andrew Beng Jin Teoh. Open-set face identification with index-of-max hashing by learning. *Pattern Recognition*, 103:107277, 2020. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2020.107277. URL https://www.sciencedirect.com/science/article/pii/S0031320320300820.

[2] Xingbo Dong, Soohyung Kim, Zhe Jin, Jung Yeon Hwang, Sangrae Cho, and Andrew Beng Jin Teoh. A secure chaff-less fuzzy vault for face identification system. *ACM Transactions on Multimedia Computing Communications and Applications*, 2021.

[3] Xingbo Dong, Zhe Jin, and Andrew Teoh Beng Jin. A genetic algorithm enabled similarity-based attack on cancellable biometrics. In *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8. IEEE, 2019.

[4] Jeff Yan, Alan Blackwell, Ross Anderson, and Alasdair Grant. Password memorability and security: Empirical results. *IEEE Security & privacy*, 2(5):25–31, 2004.

[5] Szybalski Michal. *WHITE PAPER: BIOMETRICS - MODERN METHODS OF AUTHENTICATION*. Comarch ICT, 2016. URL https://www.comarch.co.uk/files-uk/file_14/comarch-ict-biometrics-white-paper.pdf.

[6] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, page 1, 2017.

[7] V. M. Patel, N. K. Ratha, and R. Chellappa. Cancelable Biometrics: A review. *IEEE Signal Processing Magazine*, 32(5):54–65, September 2015. ISSN 1053-5888. doi: 10.1109/MSP.2015.2434151.

[8] Mulagala Sandhya and Munaga V. N. K. Prasad. Biometric Template Protection: A Systematic Literature Review of Approaches and Modalities. In Richard Jiang, Somaya Al-maadeed, Ahmed Bouridane, Prof. Danny Crookes, and Azeddine Beghdadi, editors, *Biometric Security and Privacy: Opportunities & Challenges in The Big Data Era*, Signal Processing for Security Technologies, pages 323–370. Springer International Publishing, Cham, 2017. ISBN 978-3-319-47301-7. doi: 10.1007/978-3-319-47301-7_14. URL https://doi.org/10.1007/978-3-319-47301-7_14.

[9] E. Chandra and K. Kanagalakshmi. Cancelable biometric template generation and protection schemes: A review. In *2011 3rd International Conference on Electronics Computer Technology*, volume 5, pages 15–20, April 2011. doi: 10.1109/ICECTECH.2011.5941948.

[10] K. Nandakumar and A. K. Jain. Biometric Template Protection: Bridging the performance gap between theory and practice. *IEEE Signal Processing Magazine*, 32(5):88–100, September 2015. ISSN 1053-5888. doi: 10.1109/MSP.2015.2427849.

[11] 2018 reform of eu data protection rules. URL https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf.

[12] ISO/IEC 24745:2011 - Information technology – Security techniques – Biometric information protection. URL https://www.iso.org/standard/52946.html.

[13] Guangcan Mai, Kai Cao, C YUEN Pong, and Anil K Jain. On the reconstruction of face images from deep face templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[14] Emile JC Kelkboom, Jeroen Breebaart, Tom AM Kevenaar, Ileana Buhan, and Raymond NJ Veldhuis. Preventing the decodability attack based cross-matching in a fuzzy commitment scheme. *IEEE Transactions on Information Forensics and Security*, 6(1): 107–121, 2010.

[15] Marta Gomez-Barrero, Javier Galbally, Christian Rathgeb, and Christoph Busch. General framework to evaluate unlinkability in biometric template protection systems. *IEEE Transactions on Information Forensics and Security*, 13(6):1406–1420, 2017.

[16] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(7):1757–1772, 2013.

[17] Z. Jin, J. Y. Hwang, Y. Lai, S. Kim, and A. B. J. Teoh. Ranking-Based Locality Sensitive Hashing-Enabled Cancelable Biometrics: Index-of-Max Hashing. *IEEE Transactions on Information Forensics and Security*, 13(2):393–407, February 2018. ISSN 1556-6013. doi: 10.1109/TIFS.2017.2753172.

[18] Karthik Nandakumar, Anil K Jain, and Sharath Pankanti. Fingerprint-based fuzzy vault: Implementation and performance. *IEEE transactions on information forensics and security*, 2(4):744–757, 2007.

[19] Peng Li, Xin Yang, Kai Cao, Xunqiang Tao, Ruifang Wang, and Jie Tian. An alignment-free fingerprint cryptosystem based on fuzzy vault scheme. *Journal of network and computer applications*, 33(3):207–220, 2010.

[20] Johannes Merkle, Heinrich Ihmor, Ulrike Korte, Matthias Niesing, and Michael Schwaiger. Performance of the fuzzy vault for multiple fingerprints (extended version). *arXiv preprint arXiv:1008.0807*, 2010.

[21] Shengcai Liao, Zhen Lei, Dong Yi, and Stan Z Li. A benchmark study of large-scale unconstrained face recognition. In *Biometrics (IJCB), 2014 IEEE International Joint Conference on*, pages 1–8. IEEE, 2014.

[22] Mei Wang and Weihong Deng. Deep Face Recognition: A Survey. *v:1804.06655 [cs]*, April 2018. URL http://arxiv.org/abs/1804.06655. arXiv: 1804.06655.

[23] Rajeev Ranjan, Swami Sankaranarayanan, Ankan Bansal, Navaneeth Bodla, Jun-Cheng Chen, Vishal M Patel, Carlos D Castillo, and Rama Chellappa. Deep learning for understanding faces: Machines may be just as good, or better, than humans. *IEEE Signal Processing Magazine*, 35(1):66–83, 2018.

[24] WW Bledsoe. The model method in facial recognition: a technical report pri 15. panoramic research. *Inc., Palo Alto, California*, 1964.

[25] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

[26] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (12):2037–2041, 2006.

[27] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708, 2014.

[28] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708, 2014.

[29] Yi Sun, Yuheng Chen, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation by joint identification-verification. In *Advances in neural information processing systems*, pages 1988–1996, 2014.

[30] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[31] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(Dec):3371–3408, 2010.

[32] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012. URL http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[33] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]*, September 2014. URL http://arxiv.org/abs/1409.1556. arXiv: 1409.1556.

[34] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[35] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016. doi: 10.1109/CVPR.2016.90.

[36] Feng Wang, Xiang Xiang, Jian Cheng, and Alan Loddon Yuille. Normface: L2 hypersphere embedding for face verification. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 1041–1049, 2017.

[37] 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*, pages 815–823, 2015.

[38] Yi Sun, Ding Liang, Xiaogang Wang, and Xiaoou Tang. Deepid3: Face recognition with very deep neural networks. *arXiv preprint arXiv:1502.00873*, 2015.

[39] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.

[40] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5265–5274, 2018.

[41] Jaishanker K Pillai, Vishal M Patel, Rama Chellappa, and Nalini K Ratha. Secure and robust iris recognition using random projections and sparse representations. *IEEE transactions on pattern analysis and machine intelligence*, 33(9):1877–1893, 2011.

[42] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In Richard Beals, Anatole Beck, Alexandra Bellow, and Arshag Hajian, editors, *Contemporary Mathematics*, volume 26, pages 189–206. American Mathematical Society, Providence, Rhode Island, 1984. ISBN 978-0-8218-5030-5 978-0-8218-7611-4. doi: 10.1090/conm/026/737400. URL http://www.ams.org/conm/026/.

[43] P Frankl and H Maehara. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *Journal of Combinatorial Theory, Series B*, 44(3):355–362, June 1988. ISSN 0095-8956. doi: 10.1016/0095-8956(88)90043-3. URL http://www.sciencedirect.com/science/article/pii/0095895688900433.

[44] Rosa I. Arriaga and Santosh Vempala. An algorithmic theory of learning: Robust concepts and random projection. *Machine Learning*, 63(2):161–182, May 2006. ISSN 0885-6125, 1573-0565. doi: 10.1007/s10994-006-6265-7. URL https://link.springer.com/article/10.1007/s10994-006-6265-7.

[45] A. B. J. Teoh, A. Goh, and D. C. L. Ngo. Random Multispace Quantization as an Analytic Mechanism for BioHashing of Biometric and Random Identity Inputs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):1892–1901, December 2006. ISSN 0162-8828. doi: 10.1109/TPAMI.2006.250.

[46] Adams Kong, King-Hong Cheung, David Zhang, Mohamed Kamel, and Jane You. An analysis of BioHashing and its variants. *Pattern Recognition*, 39(7):1359–1368, July 2006. ISSN 0031-3203. doi: 10.1016/j.patcog.2005.10.025. URL http://www.sciencedirect.com/science/article/pii/S0031320305004280.

[47] Yi C Feng, Pong C Yuen, and Anil K Jain. A hybrid approach for generating secure and discriminating face template. *IEEE transactions on information forensics and security*, 5 (1):103–117, 2009.

[48] Jeonil Kang, DaeHun Nyang, and KyungHee Lee. Two-factor face authentication using matrix permutation transformation and a user password. *Information Sciences*, 269:1–20, June 2014. ISSN 00200255. doi: 10.1016/j.ins.2014.02.011. URL https://linkinghub.elsevier.com/retrieve/pii/S002002551400108X.

[49] MinYi Jeong, Chulhan Lee, Jongsun Kim, Jeung-Yoon Choi, Kar-Ann Toh, and Jaihie Kim. Changeable Biometrics for Appearance Based Face Recognition. In *2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, pages 1–5, Baltimore, MD, USA, September 2006. IEEE. ISBN 978-1-4244-0486-5 978-1-4244-0487-2. doi: 10.1109/BCC.2006.4341629. URL http://ieeexplore.ieee.org/document/4341629/.

[50] Marta Gomez-Barrero, Christian Rathgeb, Javier Galbally Herrero, Julian Fierrez, and Christoph H Busch. Protected facial biometric templates based on local gabor patterns and adaptive bloom filters. In *International Conference on Pattern Recognition*. IEEE, 2014.

[51] Pawel Drozdowski, Surabhi Garg, Christian Rathgeb, M Gomez-Barrcro, Donghoon Chang, and Christoph Busch. Privacy-preserving indexing of iris-codes with cancelable bloom filter-based search structures. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 2360–2364. IEEE, 2018.

[52] Young Kyun Jang and Nam Ik Cho. Deep face image retrieval for cancelable biometric authentication. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–8. IEEE, 2019.

[53] Guangcan Mai, Kai Cao, Xiangyuan Lan, and Pong C Yuen. Secureface: Face template protection. *IEEE Transactions on Information Forensics and Security*, 16:262–277, 2020.

[54] Ari Juels and Martin Wattenberg. A fuzzy commitment scheme. In *Proceedings of the 6th ACM Conference on Computer and Communications Security*, CCS '99, pages 28–36, New York, NY, USA, 1999. ACM. ISBN 1-58113-148-8. doi: 10.1145/319709.319714. URL http://doi.acm.org/10.1145/319709.319714.

[55] P Jonathon Phillips, Hyeonjoon Moon, Syed A Rizvi, and Patrick J Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on pattern analysis and machine intelligence*, 22(10):1090–1104, 2000.

[56] Aleix M Martinez. The ar face database. *CVC Technical Report24*, 1998.

[57] Youngsung Kim and Kar-Ann Toh. A method to enhance face biometric security. In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pages 1–6. IEEE, 2007.

[58] Yongjin Wang and KN Plataniotis. Face based biometric authentication with changeable and privacy preservable templates. In *Biometrics Symposium, 2007*, pages 1–6. IEEE, 2007.

[59] Att laboratories cambridge, orl face databse. URL www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

[60] Georgia tech face database. URL www.anefian.com/facereco.htm.

[61] Andrew BJ Teoh, Yip Wai Kuan, and Sangyoun Lee. Cancellable biometrics and annotations on biohash. *Pattern recognition*, 41(6):2034–2044, 2008.

[62] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

[63] Javier Ortega-Garcia, Julian Fierrez, Fernando Alonso-Fernandez, Javier Galbally, Manuel R Freire, Joaquin Gonzalez-Rodriguez, Carmen Garcia-Mateo, Jose-Luis Alba-Castro, Elisardo Gonzalez-Agulla, Enrique Otero-Muras, et al. The multiscenario multienvironment biosecure multimodal database (bmdb). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1097–1111, 2009.

[64] Davide Maltoni, Dario Maio, Anil K Jain, and Salil Prabhakar. *Handbook of fingerprint recognition, 2nd edition.* Springer Science & Business Media, 2009.

[65] Lior Wolf, Tal Hassner, and Itay Maoz. Face recognition in unconstrained videos with matched background similarity. In *CVPR 2011*, pages 529–534. IEEE, 2011.

[66] Hong-Wei Ng and Stefan Winkler. A data-driven approach to cleaning large face datasets. In *2014 IEEE international conference on image processing (ICIP)*, pages 343–347. IEEE, 2014.

[67] C.D. Castillo V.M. Patel R. Chellappa D.W. Jacobs S. Sengupta, J.C. Cheng. Frontal to profile face verification in the wild. In *IEEE Conference on Applications of Computer Vision*, February 2016.

[68] Bruce Schneier. Intermediate protocols. *Applied Cryptography, Second Edition: Protocols, Algorthms, and Source Code in C*, pages 75–100, 2015.

[69] Feng Hao, Ross Anderson, and John Daugman. Combining crypto with biometrics effectively. *IEEE transactions on computers*, 55(9):1081–1088, 2006.

[70] Michiel Van Der Veen, Tom Kevenaar, Geert-Jan Schrijen, Ton H Akkermans, and Fei Zuo. Face biometrics with renewable templates. In *Security, Steganography, and Watermarking of Multimedia Contents VIII*, volume 6072, page 60720J. International Society for Optics and Photonics, 2006.

[71] B Chen and Vinod Chandran. Biometric based cryptographic key generation from faces. In *Digital Image Computing Techniques and Applications, 9th Biennial Conference of the Australian Pattern Recognition Society on*, pages 394–401. IEEE, 2007.

[72] Emile JC Kelkboom, Berk Gökberk, Tom AM Kevenaar, Anton HM Akkermans, and Michiel van der Veen. "3d face": biometric template protection for 3d face recognition. In *International Conference on Biometrics*, pages 566–573. Springer, 2007.

[73] Yadigar Imamverdiyev, Andrew Beng Jin Teoh, and Jaihie Kim. Biometric cryptosystem based on discretized fingerprint texture descriptors. *Expert Systems with Applications*, 40 (5):1888–1901, 2013.

[74] John Daugman. Probing the uniqueness and randomness of iriscodes: Results from 200 billion iris pair comparisons. *Proceedings of the IEEE*, 94(11):1927–1935, 2006.

[75] Tom AM Kevenaar, Geert Jan Schrijen, Michiel van der Veen, Anton HM Akkermans, and Fei Zuo. Face recognition with renewable and privacy preserving binary templates. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, pages 21–26. IEEE, 2005.

[76] Margarita Osadchy and Orr Dunkelman. It is all in the system's parameters: Privacy and security issues in transforming biometric raw data into binary strings. *IEEE Transactions on Dependable and Secure Computing*, 16(5):796–804, 2018.

[77] MengHui Lim, Andrew Beng Jin Teoh, and Jaihie Kim. Biometric feature-type transformation: Making templates compatible for secret protection. *IEEE Signal Processing Magazine*, 32(5):77–87, 2015.

[78] Haiping Lu, Karl Martin, Francis Bui, Konstantinos N Plataniotis, and Dimitris Hatzinakos. Face recognition with biometric encryption for privacy-enhancing self-exclusion. In *2009 16th International Conference on Digital Signal Processing*, pages 1–8. IEEE, 2009.

[79] Orane Cole and Khalil El-Khatib. A privacy enhanced facial recognition access control system using biometric encryption. In *2017 13th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 199–206. IEEE, 2017.

[80] Hyunggu Lee, Andrew Beng Jin Teoh, Ho Gi Jung, and Jaihie Kim. A secure biometric discretization scheme for face template protection. *Future Generation Computer Systems*, 28(1):218–231, 2012.

[81] Ari Juels and Madhu Sudan. A fuzzy vault scheme. *Designs, Codes and Cryptography*, 38 (2):237–257, 2006.

[82] Wencheng Yang, Jiankun Hu, Song Wang, and Milos Stojmenovic. An alignment-free fingerprint bio-cryptosystem based on modified voronoi neighbor structures. *Pattern Recognition*, 47(3):1309–1320, 2014.

[83] Cai Li and Jiankun Hu. A security-enhanced alignment-free fuzzy vault-based fingerprint cryptosystem using pair-polar minutiae structures. *IEEE Transactions on Information Forensics and Security*, 11(3):543–555, 2016.

[84] Wencheng Yang, Jiankun Hu, and Song Wang. A delaunay quadrangle-based fingerprint authentication system with template protection using topology code for local registration and security enhancement. *IEEE transactions on Information Forensics and Security*, 9 (7):1179–1192, 2014.

[85] Wencheng Yang, Jiankun Hu, and Song Wang. A delaunay triangle group based fuzzy vault with cancellability. In *Image and Signal Processing (CISP), 2013 6th International Congress on*, volume 3, pages 1676–1681. IEEE, 2013.

[86] Cai Li and Jiankun Hu. A security-enhanced alignment-free fuzzy vault-based fingerprint cryptosystem using pair-polar minutiae structures. *IEEE Transactions on Information Forensics and Security*, 11(3):543–555, 2015.

[87] Wencheng Yang, Jiankun Hu, and Song Wang. A delaunay quadrangle-based fingerprint authentication system with template protection using topology code for local registration and security enhancement. *IEEE transactions on Information Forensics and Security*, 9 (7):1179–1192, 2014.

[88] Venkatesan Guruswami. *PhD Thesis: List decoding of error-correcting codes*, volume 3282. Massachusetts Institute of Technology, 2001.

[89] Preda Mihailescu. The fuzzy vault for fingerprints is vulnerable to brute force attack. *arXiv preprint arXiv:0708.2974*, 2007.

[90] Xiangyu Zhu, Junjie Yan, Dong Yi, Zhen Lei, and Stan Z Li. Discriminative 3d morphable model fitting. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, volume 1, pages 1–8. Citeseer, 2015.

[91] Xiaoyang Tan and Bill Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE transactions on image processing*, 19(6):1635–1650, 2010.

[92] Fen Miao, Shu-Di Bao, Ye Li, et al. Physiological signal based biometrics for securing body sensor network. In *New trends and developments in biometrics*. InTech, 2012.

[93] Patrick J Grother, George W Quinn, and P Jonathon Phillips. Report on the evaluation of 2d still-image face recognition algorithms. *NIST interagency report*, 7709:106, 2010.

[94] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[95] Lacey Best-Rowden, Hu Han, Charles Otto, Brendan F Klare, and Anil K Jain. Unconstrained face recognition: Identifying a person of interest from a media collection. *IEEE Transactions on Information Forensics and Security*, 9(12):2144–2157, 2014.

[96] Patrick Grother, Ross J Micheals, and P Jonathon Phillips. Face recognition vendor test 2002 performance metrics. In *International Conference on Audio-and Video-based Biometric Person Authentication*, pages 937–945. Springer, 2003.

[97] Hazım Kemal Ekenel, Lorant Szasz-Toth, and Rainer Stiefelhagen. Open-set face recognition-based visitor interface system. In *International Conference on Computer Vision Systems*, pages 43–52. Springer, 2009.

[98] Johannes Stallkamp, Hazim K Ekenel, and Rainer Stiefelhagen. Video-based face recognition on real-world data. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[99] Rafael Vareto, Samira Silva, Filipe Costa, and William Robson Schwartz. Towards open-set face recognition using hashing functions. In *Biometrics (IJCB), 2017 IEEE International Joint Conference on*, pages 634–641. IEEE, 2017.

[100] Claudio Ferrari, Stefano Berretti, and Alberrto Del Bimbo. Extended youtube faces: a dataset for heterogeneous open-set face identification. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 3408–3413. IEEE, 2018.

[101] Dayong Wang, Charles Otto, and Anil K Jain. Face search at scale. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(6):1122–1136, 2017.

[102] Fayin Li and Harry Wechsler. Open set face recognition using transduction. *IEEE transactions on pattern analysis and machine intelligence*, 27(11):1686–1697, 2005.

[103] Filipe de O Costa, Ewerton Silva, Michael Eckmann, Walter J Scheirer, and Anderson Rocha. Open set source camera attribution and device linking. *Pattern Recognition Letters*, 39:92–101, 2014.

[104] Hakan Cevikalp and Bill Triggs. Efficient object detection using cascades of nearest convex model classifiers. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3138–3145. IEEE, 2012.

[105] Behrooz Kamgar-Parsi, Wallace Lawson, and Behzad Kamgar-Parsi. Toward development of a face recognition system for watchlist surveillance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):1925–1937, 2011.

[106] Cassio Elias dos Santos Junior and William Robson Schwartz. Extending face identification to open-set face recognition. In *Graphics, Patterns and Images (SIBGRAPI), 2014 27th SIBGRAPI Conference on*, pages 188–195. IEEE, 2014.

[107] Abhijit Bendale and Terrance Boult. Towards open world recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1893–1902, 2015.

[108] Ethan M Rudd, Lalit P Jain, Walter J Scheirer, and Terrance E Boult. The extreme value machine. *IEEE transactions on pattern analysis and machine intelligence*, 40(3):762–768, 2018.

[109] Leizhong Zhang, Qiong Yang, Ta Bao, Dave Vronay, and Xiaoou Tang. Imlooking: image-based face retrieval in online dating profile search. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, pages 1577–1582. ACM, 2006.

[110] Prasetyawidi Indrawan, Slamet Budiyatno, Nur Muhammad Ridho, and Riri Fitri Sari. Face recognition for social media with mobile cloud computing. *International Journal on Cloud Computing: Services and Architecture*, 3(1):23–35, 2013.

[111] Zhong Wu, Qifa Ke, Jian Sun, and Heung-Yeung Shum. Scalable face image retrieval with identity-based quantization and multireference reranking. *IEEE transactions on pattern analysis and machine intelligence*, 33(10):1991–2001, 2011.

[112] Bor-Chun Chen, Yan-Ying Chen, Yin-Hsi Kuo, and Winston H Hsu. Scalable face image retrieval using attribute-enhanced sparse codewords. *IEEE Trans. Multimedia*, 15(5): 1163–1173, 2013.

[113] Ira Kemelmacher-Shlizerman, Steven M Seitz, Daniel Miller, and Evan Brossard. The megaface benchmark: 1 million faces for recognition at scale. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4873–4882, 2016.

[114] Dong Yi, Zhen Lei, Yang Hu, and Stan Z Li. Fast matching by 2 lines of code for large scale face recognition systems. *arXiv preprint arXiv:1302.7180*, 2013.

[115] Junjie Yan, Zhen Lei, Dong Yi, and Stan Z Li. Towards incremental and large scale face recognition. 2011.

[116] Brendan F Klare, Austin Blanton, and Benjamin Klein. Efficient face retrieval using synecdoches. In *Biometrics (IJCB), 2014 IEEE International Joint Conference on*, pages 1–7. IEEE, 2014.

[117] Manuel Gunther, Steve Cruz, Ethan M Rudd, and Terrance E Boult. Toward open-set face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 71–80, 2017.

[118] Ruud M Bolle, Jonathan H Connell, and Nalini K Ratha. Biometric perils and patches. *Pattern recognition*, 35(12):2727–2738, 2002.

[119] Javier Galbally, Chris McCool, Julian Fierrez, Sebastien Marcel, and Javier Ortega-Garcia. On the vulnerability of face verification systems to hill-climbing attacks. *Pattern Recognition*, 43(3):1027–1038, 2010.

[120] Marta Gomez-Barrero, Javier Galbally, and Julian Fierrez. Efficient software attack to multimodal biometric systems and its application to face and iris fusion. *Pattern Recognition Letters*, 36:243–253, 2014.

[121] Moses S Charikar. Similarity estimation techniques from rounding algorithms. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 380–388. ACM, 2002.

[122] Aristides Gionis, Piotr Indyk, Rajeev Motwani, et al. Similarity search in high dimensions via hashing. In *Vldb*, volume 99, pages 518–529, 1999.

[123] Jay Yagnik, Dennis Strelow, David A Ross, and Ruei-sung Lin. The power of comparative reasoning. In *2011 International Conference on Computer Vision*, pages 2431–2438. IEEE, 2011.

[124] Alan Frieze and Mark Jerrum. Improved approximation algorithms for max k-cut and max bisection. *Algorithmica*, 18(1):67–81, 1997.

[125] Anand Rajaraman and Jeffrey David Ullman. *Mining of massive datasets.* Cambridge University Press, 2011.

[126] Flavio Chierichetti and Ravi Kumar. Lsh-preserving functions and their applications. *Journal of the ACM (JACM)*, 62(5):1–25, 2015.

[127] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European Conference on Computer Vision*, pages 87–102. Springer, 2016.

[128] Jia Xiang and Gengming Zhu. Joint face detection and facial expression recognition with mtcnn. In *Information Science and Control Engineering (ICISCE), 2017 4th International Conference on*, pages 424–427. IEEE, 2017.

[129] Kai Li, Guo-Jun Qi, Jun Ye, and Kien A Hua. Linear subspace ranking hashing for cross-modal retrieval. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (9):1825–1838, 2017.

[130] Kai Li, Guo-Jun Qi, Jun Ye, Tuoerhongjiang Yusuph, and Kien A Hua. Supervised ranking hash for semantic similarity search. In *Multimedia (ISM), 2016 IEEE International Symposium on*, pages 551–558. IEEE, 2016.

[131] Kar-Ann Toh, Youngsung Kim, Sangyoun Lee, and Jaihie Kim. Fusion of visual and infra-red face scores by weighted power series. *Pattern recognition letters*, 29(5):603–615, 2008.

[132] Isaac Martín de Diego, Angel Serrano, Cristina Conde, and Enrique Cabello. Face verification with a kernel fusion method. *Pattern Recognition Letters*, 31(9):837–844, 2010.

[133] Ronald Aylmer Fisher and Leonard Henry Caleb Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society*, volume 24, pages 180–190. Cambridge University Press, 1928.

[134] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 67–74. IEEE, 2018.

[135] Brianna Maze, Jocelyn Adams, James A Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K Jain, W Tyler Niggel, Janet Anderson, Jordan Cheney, et al. Iarpa janus

benchmark-c: Face dataset and protocol. In *2018 International Conference on Biometrics (ICB)*, pages 158–165. IEEE, 2018.

[136] Characterizing unknown unknowns. URL https://www.pmi.org/learning/library/characterizing-unknown-unknowns-6077.

[137] Ruud M Bolle, Jonathan H Connell, Sharath Pankanti, Nalini K Ratha, and Andrew W Senior. The relation between the roc curve and the cmc. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, pages 15–20. IEEE, 2005.

[138] P Jonathon Phillips, Patrick Grother, and Ross Micheals. Evaluation methods in face recognition. In *Handbook of face recognition*, pages 551–574. Springer, 2011.

[139] Sébastien Marcel. Beat–biometrics evaluation and testing. *Biometric technology today*, 2013(1):5–7, 2013.

[140] Neil A Macmillan and C Douglas Creelman. *Detection theory: A user's guide*. Psychology press, 2004.

[141] Jingtuo Liu, Yafeng Deng, Tao Bai, Zhengping Wei, and Chang Huang. Targeting ultimate accuracy: Face recognition via deep embedding. *arXiv preprint arXiv:1506.07310*, 2015.

[142] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Web-scale training for face identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2746–2754, 2015.

[143] Xiang Wu, Ran He, Zhenan Sun, and Tieniu Tan. A light cnn for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security*, 13(11): 2884–2896, 2018.

[144] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014.

[145] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *European conference on computer vision*, pages 499–515. Springer, 2016.

[146] Xiao Zhang, Zhiyuan Fang, Yandong Wen, Zhifeng Li, and Yu Qiao. Range loss for deep face recognition with long-tailed training data. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5409–5418, 2017.

[147] Dayong Wang, Charles Otto, and Anil K Jain. Face search at scale. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1122–1136, 2016.

[148] Lei Li, Heng Luo, Lei Zhang, Qing Xu, and Hao Ning. Typicface: Dynamic margin cosine loss for deep face recognition. In *Pacific Rim International Conference on Artificial Intelligence*, pages 710–718. Springer, 2018.

[149] Monica MY Zhang, Kun Shang, and Huaming Wu. Learning deep discriminative face features by customized weighted constraint. *Neurocomputing*, 332:71–79, 2019.

[150] Christian Rathgeb and Andreas Uhl. A survey on biometric cryptosystems and cancelable biometrics. *EURASIP Journal on Information Security*, 2011(1):3, 2011.

[151] Kai Xi and Jiankun Hu. Bio-cryptography. In *Handbook of Information and Communication Security*, pages 129–157. Springer, 2010.

[152] Y Dodis, L Reyzin, and A Smith Fuzzy Extractors. How to generate strong keys from biometrics and other noisy, data april 13. EUROCRYPT, 2004.

[153] Kai Xi, Jiankun Hu, and BVK Vijaya Kumar. Fe-svit: A svit-based fuzzy extractor framework. *ACM Transactions on Embedded Computing Systems (TECS)*, 15(4):1–24, 2016.

[154] David Bissessar, Carlisle Adams, and Alex Stoianov. Privacy, security and convenience: biometric encryption for smartphone-based electronic travel documents. In *Recent advances in computational intelligence in defense and security*, pages 339–366. Springer, 2016.

[155] Ann Cavoukian, Michelle Chibba, and Alex Stoianov. Advances in biometric encryption: Taking privacy by design from academic research to deployment. *Review of Policy Research*, 29(1):37–61, 2012.

[156] Walter J Scheirer and Terrance E Boult. Cracking fuzzy vaults and biometric encryption. In *2007 Biometrics Symposium*, pages 1–6. IEEE, 2007.

[157] Raffaele Cappelli, Matteo Ferrara, and Davide Maltoni. Minutia cylinder-code: A new representation and matching technique for fingerprint recognition. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2128–2141, 2010.

[158] Zhe Jin, Meng-Hui Lim, Andrew Beng Jin Teoh, Bok-Min Goi, and Yong Haur Tay. Generating fixed-length representation from minutiae using kernel methods for fingerprint authentication. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 46(10): 1415–1428, 2016.