

2016-06-09

Biometrics for Cybersecurity and Unconstrained Environments

Mohammad Haghighat
University of Miami, haghighat@ieee.org

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation

Haghighat, Mohammad, "Biometrics for Cybersecurity and Unconstrained Environments" (2016). *Open Access Dissertations*. 1675.
https://scholarlyrepository.miami.edu/oa_dissertations/1675

This Open access is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.

UNIVERSITY OF MIAMI

BIOMETRICS FOR CYBERSECURITY AND UNCONSTRAINED
ENVIRONMENTS

By

Mohammad Haghghat

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

August 2016

©2016
Mohammad Haghghat
All Rights Reserved

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

BIOMETRICS FOR CYBERSECURITY AND UNCONSTRAINED
ENVIRONMENTS

Mohammad Haghghat

Approved:

Mohamed Abdel-Mottaleb, Ph.D.
Professor of Electrical and
Computer Engineering

Saman Zonouz, Ph.D.
Assistant Professor of Electrical and
Computer Engineering
Rutgers University

Shahriar Negahdaripour, Ph.D.
Professor of Electrical and
Computer Engineering

Jie Xu, Ph.D.
Assistant Professor of Electrical and
Computer Engineering

Anil K. Jain, Ph.D.
Distinguished Professor of
Computer Science and Engineering
Michigan State University

Guillermo Prado, Ph.D.
Dean of the Graduate School

HAGHIGHAT, MOHAMMAD (Ph.D., Electrical and Computer Engineering)

Biometrics for Cybersecurity and Unconstrained Environments (August 2016)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Mohamed Abdel-Mottaleb.

No. of pages in text. (155)

Biometric identification has been a challenging topic in computer vision for the past few decades. In this thesis, we study four main challenges in biometrics: 1) secure and privacy-preserving biometric identification in untrusted public cloud servers, 2) single sample face recognition in unconstrained environments, 3) multimodal biometrics using feature-level fusion, and 4) low-resolution face recognition.

In biometric identification systems, the biometric database is typically stored on a trusted server, which is also responsible for performing the identification process. However, if this server is a public cloud, maintenance of the confidentiality and integrity of sensitive data requires trustworthy solutions for its storage and processing. In the first part of our study, we present CloudID, a privacy-preserving cloud-based biometric identification solution. It links the confidential information of the users to their biometrics and stores it in an encrypted fashion. Making use of a searchable encryption technique, biometric identification is performed in the encrypted domain to make sure that the cloud provider or potential attackers do not gain access to any sensitive data or even the contents of the individual queries. The proposed approach is the first cloud-based biometric identification system with a proven zero data disclosure possibility. It allows different enterprises to perform biometric identification on a single database without revealing any sensitive information.

In the second part of this study, we present a fully automatic face recognition technique robust to face pose variations in unconstrained environments. The proposed method normalizes the face images for both in-plane and out-of-plane pose variations using an enhanced technique based on active appearance models (AAMs). We improve the performance of AAM fitting, not only by training it with in-the-wild images and using a powerful optimization technique but also by initializing the AAM with estimates of the locations of the facial landmarks obtained by a method based on flexible mixture of parts. The proposed initialization technique results in significant improvement in AAM fitting to non-frontal poses and makes the normalization process robust, fast and reliable. Owing to the proper alignment of the face images, made possible by this approach, we can use local feature descriptors, such as Histograms of Oriented Gradients (HOG), which makes the system robust against illumination variations. In order to improve the discriminating information content of the feature vectors, we also extract Gabor features from the normalized face images and fuse them with HOG features using Canonical Correlation Analysis (CCA). The proposed face recognition system is capable of recognizing faces from non-frontal views and under different illumination conditions using only a single gallery sample for each subject. This is important because of its potential applications in real life applications such as video surveillance.

In the third part of this study, we propose a real-time feature level fusion technique for multimodal biometric systems. The goal of feature fusion for recognition is to combine relevant information from two or more feature vectors into a single one with more discriminative power than any of the input feature vectors. In pattern recognition problems, we are also interested in separating the classes. In this study,

we present Discriminant Correlation Analysis (DCA), a feature level fusion technique that incorporates the class associations into the correlation analysis of the feature sets. DCA performs an effective feature fusion by maximizing the pair-wise correlations across the two feature sets, and at the same time, eliminating the between-class correlations and restricting the correlations to be within the classes. The proposed method can be used in pattern recognition applications for fusing features extracted from multiple modalities or combining different feature vectors extracted from a single modality. DCA has a very low computational complexity and it can be employed in real-time applications. Multiple sets of experiments performed on various biometric databases, and using different feature extraction techniques, show the effectiveness of the proposed method, which outperforms other state-of-the-art approaches.

In the fourth and last part of this thesis, we propose a novel real-time approach for matching Low Resolution (LR) probe face images with High Resolution (HR) gallery face images with an application to surveillance systems. The proposed method is based on DCA. It projects the LR and HR feature vectors in a common domain in which not only the LR and HR feature vectors have the same length, but also the correlation between corresponding features in LR and HR domain is maximized. In addition, the process of calculating the projection matrices considers the class structure of the data and it aims to separate the classes in the new domain, which is very beneficial from the recognition perspective. It is worth mentioning that the proposed method has a very low computational complexity and it can be employed for processing several faces that appear in a crowded image in real-time. Experiments performed on low-resolution surveillance images verify the effectiveness of our proposed method in comparison with the state-of-the-art LR face recognition techniques.

*Dedicated to Claire,
Badoum, April, Teatop, and Tasha*

Acknowledgements

I would like to thank Dr. Mohamed Abdel-Mottaleb and Dr. Saman Zonouz who supported me in the past few years through the research and completion of my degree. I believe their personality and technical capability was an indispensable factor for me to finish this endeavor.

I would also like to express my gratitude to the committee members, Professors Anil Jain, Shahriar Negahdaripour, and Jie Xu, for generously offering their time, support, guidance and good will throughout the preparation and review of this document.

MOHAMMAD HAGHIGHAT

University of Miami

August 2016

Table of Contents

LIST OF FIGURES	ix
LIST OF TABLES	xiv
1 INTRODUCTION	1
2 PRIVACY-PRESERVING BIOMETRIC IDENTIFICATION: SECURITY CHALLENGES IN CLOUD COMPUTING ENVIRONMENTS	11
2.1 Related Work in Privacy-Preserving Biometric Identification	14
2.2 System Overview and Access Control	17
2.3 Offline System Setup	21
2.3.1 Biometric Data Acquisition	22
2.3.2 Feature Extraction	22
2.3.3 Scalability Considerations and Feature Quantization	23
2.3.4 Cryptographic Key Generation	29
2.3.5 Biometric Record Encryption	30

2.4	Online System Behavior	31
2.4.1	Face Features Variations	31
2.4.2	Cryptographic Token Generation	32
2.4.3	Biometric Database Query	33
2.5	Evaluation	34
2.5.1	Experimental Setup	34
2.5.2	Accuracy	35
2.5.3	Performance	37
2.5.4	Scalability of CloudID	39
2.5.5	Case study	42
2.6	Conclusions and Future Work	46

3 FULLY AUTOMATED FACE NORMALIZATION AND SINGLE SAMPLE FACE RECOGNITION IN UNCONSTRAINED ENVIRONMENTS

		48
3.1	Related Work in Pose-Invariant Face Recognition	50
3.2	Preprocessing for Face Normalization	57
3.2.1	Active Appearance Models and Piece-wise Affine Warping	58
3.2.2	Proposed AAM Initialization	60
3.3	Feature Extraction and Fusion	64
3.3.1	Feature Extraction	65
3.3.2	Feature Fusion Using Canonical Correlation Analysis	67

3.4	Experimental Setup and Results	70
3.4.1	Experimental Setup: AAM Training	70
3.4.2	Normalization Performance	70
3.4.3	Experiments on FERET Database	72
3.4.4	Experiments on CMU-PIE Database	75
3.4.5	Experiments on LFW Database	76
3.5	Conclusions and Future Work	81
4	DISCRIMINANT CORRELATION ANALYSIS: REAL-TIME FEAT-	
	TURE LEVEL FUSION FOR MULTIMODAL BIOMETRICS	83
4.1	Introductory Remarks and Overview	83
4.2	Incorporating Class Structure in Multivariate Correlation Analysis . .	87
4.2.1	Feature-Level Fusion Using Discriminant Correlation Analysis	88
4.2.2	Multiset Discriminant Correlation Analysis	92
4.3	Experiments and Analysis	94
4.3.1	Unimodal Multi-Feature Fusion	95
4.3.2	WVU Database	99
4.3.2.1	Comparison of Methods	102
4.3.3	Multimodal Biometric Dataset, BIOMDATA	105
4.3.4	Multimodal Biometrics: AR Face Database and USTB Ear Database	108
4.3.5	AR Face Database	109

4.3.6	Scalability of DCA	111
4.3.7	Sketch to Mugshot Matching	113
4.4	Conclusions and Future works	115
5	LOW RESOLUTION FACE RECOGNITION USING DISCRIMI-	
	NANT CORRELATION ANALYSIS	117
5.1	Motivation	117
5.2	Proposed Method	119
5.2.1	Canonical Correlation Analysis (CCA)	120
5.2.2	Discriminant Correlation Analysis (DCA)	123
5.3	Experiments and Analysis	127
5.4	Conclusions	131
6	CONCLUSIONS	133
	APPENDIX	137
	BIBLIOGRAPHY	139

List of Figures

2.1	Framework of the CloudID in query processing.	18
2.2	Face detection using Viola and Jones method.	22
2.3	Gabor wavelets in five scales and eight orientations.	24
2.4	Creating the quantized feature vector using 13 k-d trees of 4 significant features.	28
2.5	Face images from FERET database after face detection and cropping. First row: frontal series (ba). Second row: frontal series with an alternative expression (bj). Third row: frontal series with illumination change (bk).	35
2.6	Comparison of the eigenface and Gabor+GDA face recognition algorithms	36
2.7	Average encryption time and query time using different number of features.	39
2.8	Average encryption time and query time using multi-threading.	40
2.9	Sample face images of a subject from WVU database after face detection and cropping.	41

2.10 Scalability of the CloudID in dealing with new and unseen subjects. The training is already done using 200 subjects and 202 new subjects are gradually added to the database.	42
2.11 Face detection and feature extraction, results of applying filters shown in Fig. 2.3 on a face image.	43
3.1 Warping the face image into the base (frontal) mesh. (a) Rotated face image. (b) Fitting mesh corresponding to the rotated face image. (c) Triangulated base (frontal) mesh, s_0 . (d) Face image warped into the base mesh.	58
3.2 Initialization problem in AAM fitting. (a) Initial shape used in POIC and SIC algorithms $\mathbf{p} = 0$. (b) Initialization of the base mesh on the target face image. (c) Fitting result of the Fast-SIC method after 100 iterations. (d) Result of the piecewise affine warping into the base mesh.	61
3.3 Top view perspective of a human head in frontal and rotated poses. .	63
3.4 Our proposed initialization method for AAM fitting. (a) Estimated landmarks using the flexible mixture of trees. (b) Triangularization of the initial mesh created by the estimated landmarks. (c) Fitting result of the Fast-SIC method after only 5 iterations. (d) Result of the piecewise affine warping into the base mesh.	65
3.5 Histogram of Oriented Gradients (HOG) features in 4×4 cells. . . .	66
3.6 Gabor features in five scales and eight orientations.	67
3.7 (a) Self-occluded face image with 60° rotation. (b) Normalized face image with a stretched half face.	71

3.8	Face images of a sample subject from FERET b-series database (upper row), and their normalized faces (lower row).	73
3.9	Symmetry issue in FERET database. The upper row includes the sample images at $+60^\circ$ (<i>bb</i>) and the lower row shows the corresponding images at -60° (<i>bi</i>).	74
3.10	Face images of a sample subject from CMU-PIE database (upper row), and their normalized faces (lower row).	75
3.11	Sample images of three subjects from LFW database.	77
3.12	Normalized face images corresponding to the ones shown in Fig. 3.11.	77
3.13	Performances of different approaches using BLUFR protocol. (a) Verification ROC curves. (b) Open-set identification ROC curves at rank 1.	81
4.1	Visualization of covariance matrices (black color represents zero values and the elements with higher values are illustrated brighter). (a) Covariance between features ($\tilde{X}^* \tilde{X}^{*T}$). (b) Covariance between samples ($\tilde{X}^{*T} \tilde{X}^*$).	92
4.2	Multiset discriminant correlation analysis techniques for five sample sets with $rank(X_1) > rank(X_2) > rank(X_3) > rank(X_4) = rank(X_5)$	93
4.3	Sample face images of a subject in AR database (Top row: first session; Bottom row: second session).	95
4.4	Accuracy of the unimodal biometric systems using Gabor and HOG features on AR face database.	96
4.5	Accuracy of the unimodal biometric systems using Gabor and SURF features on AR face database.	97

4.6	Accuracy of the unimodal biometric systems using Gabor, SURF and HOG features on AR face database.	98
4.7	Different frames of the subject #1 from profile to frontal equally distanced by 5°.	100
4.8	Profile/near-profile face images detected in different frames.	100
4.9	Frontal/near-frontal face images detected in different frames.	101
4.10	Ear normalization for in-plane rotations. (a) Detected Landmarks. (b) Normalized Ear.	101
4.11	Ear images detected in different frames.	102
4.12	Examples of challenging samples in BIOMDATA database. The images are corrupted with blur, occlusion, shadows, and sensor noise.	106
4.13	Preprocessing for iris images. (a) Original iris image from BIOMDATA database. (b) Segmented iris area. (c) 25 × 240 binary iris template.	107
4.14	Preprocessing for fingerprint images. (a) Original fingerprint image from BIOMDATA database. (b) Enhanced image using the method in [1]. (c) Core point of the fingerprint and the region of interest around it.	107
4.15	Sample ear images of a subject cropped from USTB database III.	109
4.16	Face mask used to crop out different modalities.	110
4.17	Scalability of the proposed DCA algorithm using different number of training subjects and testing on unseen populations.	112
4.18	Photo synthesis result: (a) sketch drawn by the artist; (b) real photo; and (c) pseudo-photo synthesized from the sketch.	115

5.1	Standard approaches to matching a LR probe to a HR gallery. (a) Downsampling the gallery and then matching. (b) Upsampling the probe (interpolation or super-resolution) and then matching. In the proposed approach, we project LR and HR images to a common linear space for matching.	119
5.2	Visualization of covariance matrices (black color represents zero values and the elements with higher values are illustrated brighter). (a) Covariance between features (H^*H^{*T}). (b) Covariance between samples ($H^{*T}H^*$).	127
5.3	Sample images from SCface database. Top row: high-resolution gallery images. Bottom row: low-resolution probe images.	128
5.4	Setting of the surveillance cameras in SCface database. (a) Surveillance cameras capturing distances. (b) Set-up of the five cameras. The images are from [2].	129
5.5	Sample images from SCface database. (a) HR gallery image (face size: 128×128). (b) LR probe image in distance 3 (64×64). (c) LR probe image in distance 2 (32×32). (d) LR probe image in distance 1 (16×16).130	

List of Tables

2.1	Face recognition accuracy of CloudID	37
2.2	System accuracy using different accepted range sizes	38
3.1	Face recognition rates of different approaches in confrontation with different face distortions on the FERET database. The frontal face images with neutral expression, labeled <i>ba</i> , are used for training.	73
3.2	Face recognition rates of different approaches in confrontation with different pose changes on the CMU-PIE database. The frontal faces captured by camera <i>c27</i> is used for training.	76
3.3	Average accuracy of different approaches following the first experiment on LFW database following the image-restricted with label-free outside data protocol. Note that the list is shown in chronological order.	78
3.4	Face recognition rates of different approaches following the second experiment on LFW database.	79
3.5	Overview of the experimental setting for the BLUFR protocol on the LFW database. Numbers are averaged over the 10 trials.	80

3.6	Benchmark performance of different state-of-the-art face recognition approaches for verification and open-set identification at rank 1. The reported numbers are the mean verification rates (%) subtracted by the corresponding standard deviations over 10 trials. The list is shown in chronological order.	82
4.1	Maximum rank-1 recognition rates using individual and fused feature vectors (H: HOG, S: SURF, G: Gabor).	95
4.2	Rank-1 recognition rates obtained by a KNN classifier (K=1) using individual modalities in WVU database.	103
4.3	Rank-1 recognition rates for multimodal fusion of face, ear and profile face biometrics in WVU database.	104
4.4	Average run-time values of different feature level fusion techniques for recognition of one multimodal face-ear pair in WVU database.	104
4.5	BIOMDATA multimodal biometric dataset.	105
4.6	Rank-1 recognition rates obtained by a minimum distance classifier using individual modalities on BIOMDATA database.	106
4.7	Rank-1 recognition rates for multimodal fusion of iris and fingerprint biometrics in BIOMDATA database.	108
4.8	Rank-1 recognition rates on the virtual multimodal database of face and ear images.	108
4.9	Rank-1 recognition rates obtained by a KNN classifier using individual modalities in AR database. Modalities include 1. left periocular, 2. right periocular, 3. nose, 4. mouth, and 5. face.	110

4.10	Rank-1 recognition rates for multimodal fusion of different modalities in AR database. Modalities include 1. left periocular, 2. right periocular, 3. nose, 4. mouth, and 5. face.	110
4.11	Rank-1 recognition rates for multimodal fusion of all modalities in AR database.	111
4.12	Maximum rank-1 recognition rates over the number of features in Fig. 4.17.	113
4.13	Rank-1 recognition rate for photo-sketch matching in CUHK database.	115
5.1	Rank-1 recognition rates for different camera/distance in SCface database.	131
5.2	Rank-1 recognition rates for different LR image resolutions in SCface database.	131

CHAPTER 1

Introduction

Substitution of biometrics for passwords in authentication and identification systems received attention in security systems [3]. Biometric identifiers are distinctive and measurable characteristics used to recognize individuals [4]. Some of the well-known biometric used for human identification are fingerprints, face, iris, voice and DNA. Some of the advantages of biometrics over passwords include their higher level of security, mobility, difficulty to forge, and user friendliness. According to a new study published by Javelin Research [5], smartphone users prefer using biometrics as an alternative for passwords, which brings more security to new technologies such as *Apple Pay*. In spite of all these advantages, there are some challenges that biometric systems face. In this study, three main challenges in biometrics are discussed: 1) secure and privacy-preserving biometric identification in untrusted public cloud servers, 2) single sample face recognition in unconstrained environments, and 3) feature-level fusion in multimodal biometric systems.

In biometric identification systems, the biometric database is typically stored in a trusted server, which is also responsible for performing the identification process. However, a standalone server may not be able to provide enough storage and processing power for big databases. Nowadays, cloud computing and storage solutions have

provided users and enterprises with various capabilities to store and process their data in third-party data centers. However, maintenance of the confidentiality and integrity of sensitive data requires trustworthy solutions for storage and processing of data with proven zero information leakage.

In case of having a cloud-based biometric identification system, the untrusted cloud needs to have access to the biometric database. This will raise the risk of identity theft, because biometric data of the enrolled subjects can be stolen and misused against their will. The biometric data is unique and irrevocable, and unlike passwords users cannot change their biometric characteristics. Consequently, the system must guarantee the preservation of the users' privacy, and therefore, the biometric database has to be encrypted.

In the literature, there are two sets of studies considering the privacy-preserving biometric identification. The first set of studies [6–12] only achieve privacy-preserving at the time of executing the query, protecting the confidentiality of both server and client. In these approaches, the server, in which the biometric database is stored, is considered to be trusted, and the biometric database is stored unencrypted. However, the server cannot learn neither the client's biometric information nor the query result. On the other hand, the client does not get any information about the biometric database stored in the trusted server. However, these approaches cannot be used in case of untrusted servers like public clouds because they allow the server to have access to the contents of the biometric database.

In the second set of studies, the server is not trusted and the biometric database is encrypted and the identification process is applied in the encrypted domain without decrypting the data [13–17]. However, these algorithms do not guarantee zero

data disclosure and suffer from information leakage. The mathematical framework to analyze the information leakage is provided in [18]. Pagnin *et al.* [18] employ a *center search attack* and prove that when the matching process is performed using a specific family of distances (such as Hamming and Euclidean distances), then information about the reference template is leaked, and the attacker can recover the biometric template even if it is stored encrypted. The details about all these methods and their limitations are provided in Section 2.1 of Chapter 2.

In the **first part** of this study, presented in **Chapter 2**, we propose a privacy-preserving cloud-based and cross-enterprise biometric identification solution. The proposed method, called CloudID, links the confidential information of the users to their biometrics and stores it in an encrypted fashion. Making use of a searchable encryption technique, biometric identification is performed in encrypted domain to make sure that neither the cloud provider nor potential attackers can gain access to any sensitive data or even the contents of the individual queries.

CloudID applies a conjunctive range query over encrypted gallery samples, which returns *true* response only if all the features of the gallery sample fall in certain ranges defined by predicates created using the query sample. This makes CloudID secure against the center search attacks. We proposed a k-d tree structure to quantize the biometric feature vectors and define the range predicates. This structure also allows the system to handle variations in the biometric data and improves the overall performance of the system.

The proposed approach allows different enterprises to perform biometric identification on a single database without revealing any sensitive information. It provides a solution to the concerns about the security and confidentiality of personal informa-

tion stored in the cloud through the use of biometrics, while guarding against identity theft. We implemented a working prototype of the CloudID and evaluated it using a face biometric database. Experimental results show that CloudID performs the identification of clients with high accuracy and minimal overhead and proven zero data disclosure.

The **second part** of this study, presented in **Chapter 3**, deals with the challenge of single sample face recognition in unconstrained environments. One of the major challenges of biometric systems is the variability in the characteristics of the biometrics for each individual. Human face, as an example biometric trait, is a complex object with features that change over time. Facial features change due to changes in illumination, head pose, facial expressions, cosmetics, aging, and occlusions because of beard or glasses. However, humans have an ability to recognize faces and identify persons at a glance. This natural ability does not exist in machines; therefore, we design intelligent and expert systems that can simulate the recognition artificially.

In real applications, on the other hand, single sample face recognition is very important because of the limitations on the availability of gallery images for subjects to be identified. For example, in many application scenarios in law enforcement such as passport or driver licence identification, there is only a single facial image per subject available. The variations between the single gallery face image and the probe face images make the single sample face recognition even more difficult.

Pose variations are considered to be one of the most challenging issues in face recognition. Due to the complex non-planar geometry of the face, the 2D visual appearance significantly changes with variations in the viewing angle. These changes are often more significant than the variations of innate characteristics, which distinguish

individuals [19]. Consequently, the accuracy of most of the current face recognition systems drops significantly in the presence of pose variations [20, 21].

In Chapter 3, we present a fully automatic face recognition technique robust to most common face variations, *i.e.*, pose and illumination, in unconstrained environments. The proposed method normalizes the face images for both in-plane and out-of-plane pose variations using an enhanced technique based on Active Appearance Models (AAMs). We improve the performance of AAM fitting, not only by training it with in-the-wild images and using a powerful optimization technique, but also by initializing the AAM with estimates of the locations of the facial landmarks obtained by a method based on flexible mixture of parts. The proposed initialization technique results in significant improvement in AAM fitting to non-frontal poses and makes the normalization process robust, fast and reliable.

Owing to the proper alignment of the face images, made possible by the proposed normalization approach, we can use local feature descriptors, such as Histograms of Oriented Gradients (HOG), for matching. The use of HOG features makes the system robust against illumination variations. In order to improve the discriminating information content of the feature vectors, we also extract Gabor features from the normalized face images and fuse them with HOG features using Canonical Correlation Analysis (CCA). The proposed face recognition system is capable of recognizing faces from non-frontal views and under different illumination conditions using only a single gallery sample for each subject.

There are lots of studies in the literature that consider the problem of pose invariant face recognition in-the-wild [22–41]. Most of these methods, however, have main limitations. The main limitations include: requiring manual intervention of the

user during face normalization or recognition [22–25, 30, 36, 37], reliance on more than one training images for either reconstruction of a 3D model or including the non-frontal poses in training [25–29, 34, 35, 39] and being restricted to some predetermined pose angles [25–27, 32]. The details about all these methods and their limitations are provided in Section 3.1 of Chapter 3.

Unlike most previous methods, the proposed system is fully automatic, and handles a wide and continuous range of poses, *i.e.*, it is not restricted to any predetermined pose angles. Furthermore, it is capable of recognizing subjects from non-frontal view images and from images with different illumination using only a single image as the gallery. Experimental results performed on the FERET [42], CMU-PIE [43] and Labeled Faces in the Wild (LFW) [44] databases verify the effectiveness of the proposed method in normalization and recognition of face images obtained in unconstrained environments, which outperforms the state-of-the-art algorithms.

In the **third part** of this study, presented in **Chapter 4**, we present a real-time feature level fusion technique for multimodal biometric systems. Most of the real-world biometric systems, so-called *unimodal*, rely on the evidence of a single source of biometric information. *Multimodal* biometric systems, on the other hand, fuse multiple sources of biometric information to make a more reliable recognition. Information fusion is a key step in multimodal biometric systems. Fusion of the biometric information can occur at different stages of a recognition system. In case of *feature level fusion*, the data itself or the features extracted from multiple biometrics are fused. *Matching-score level fusion* consolidates the scores generated by multiple classifiers pertaining to different modalities. Finally, in case of *decision level fusion*

the final results of multiple classifiers are combined via techniques such as majority voting.

Feature level fusion is believed to be more effective than the fusion at the other levels because the feature set contains richer information about the input biometric data than the matching score or the output decision of a classifier. Therefore, fusion at the feature level is expected to provide better recognition results [45,46]. However, matching-score level fusion and decision level fusion are more popular in the literature and there is not as much research on feature level fusion. The reason is the difficulty of feature level fusion in cases where the features are not compatible, *e.g.*, eigen-coefficients of faces and minutiae set of fingerprints, or when commercial biometric systems do not provide access to the feature sets (nor the raw data), which they use in their products [45].

Two well-known and typical feature fusion methods are: serial feature fusion [47] and parallel feature fusion [48, 49]. Serial feature fusion works by simply concatenating two sets of feature vectors into a single feature vector. Obviously, if the first source feature vector, x , is p -dimensional and the second source feature vector, y , is q -dimensional, the fused feature vector, z , will be $(p+q)$ -dimensional. Parallel feature fusion, on the other hand, combines the two source feature vectors into a complex vector $z=x+iy$ (i being an imaginary unit). Note that if the dimensions of the two input vectors are not equal, the one with the lower dimension is padded with zeros. Recently, feature fusion based on Canonical Correlation Analysis (CCA) [50–54] has attracted the attention in the area of multimodal recognition. CCA-based feature fusion uses the correlation between two sets of features to find two sets of transfor-

mations such that the transformed features have maximum correlation across the two data sets, while being uncorrelated within each data set.

The goal of the feature fusion for recognition is to combine relevant information from two or more feature vectors into a single one, which is expected to be more informative than any of the input feature vectors. In pattern recognition problems, we are also interested in separating the classes. In Chapter 4, we propose a feature fusion method that considers the class associations in data sets.

We present *Discriminant Correlation Analysis (DCA)*, a feature level fusion technique that incorporates the class associations into the correlation analysis of the feature sets. DCA performs an effective feature fusion by maximizing the pair-wise correlations across the two feature sets, and at the same time, eliminating the between-class correlations and restricting the correlations to be within the classes. The proposed method can be used in pattern recognition applications for fusing features extracted from multiple modalities or combining different feature vectors extracted from a single modality. It is worth mentioning that our method does not have the small sample size (SSS) problem faced by the CCA-based algorithms. Moreover, it has a very low computational complexity and it can be employed in real-time applications. We also propose a multiset method to generalize DCA to be applicable to more than two sets of variables. *Multiset Discriminant Correlation Analysis (MDCA)* follows a cascade approach and applies DCA on two sets of variables at a time. Multiple sets of experiments performed on various biometric databases, and using different feature extraction techniques, show the effectiveness of our proposed method, which outperforms other state-of-the-art approaches.

The **fourth part** of this study, presented in **Chapter 5**, presents a low resolution face recognition technique with an application to surveillance systems. With the growing installation of surveillance cameras in many areas, there is an increasing demand for face recognition technology for surveillance applications, ranging from small-scale stand-alone camera applications in banks and supermarkets to large-scale multiple networked closed-circuit televisions in law enforcement applications in public streets. Since the surveillance cameras are usually far from the subjects, the captured face images are often in very Low Resolution (LR). The discriminant properties of the face are usually degraded in the LR images, which significantly decreases the accuracy of traditional face recognition algorithms developed for High Resolution (HR) images [55]. Empirical studies have shown that minimum face image resolution between 32×32 and 64×64 is required for existing face recognition algorithms [56]. However, in surveillance videos, when the person is not close to the camera, the face region are usually smaller [57]. Another problem with the LR face images is the difference in resolution of the probe and gallery images, which makes the recognition process more challenging [58].

There are two standard approaches to address the LR face recognition problem. In the first approach, the gallery images are downsampled to the resolution of the probe images and then recognition is applied. However, in this approach, the discriminating information available in the HR gallery images is wasted. In the second approach, on the other hand, the probe images are upsampled to the resolution of the gallery images using super-resolution techniques [59–65]. In this method, the discriminating information of the HR images is preserved. However, the main goal of most super-resolution techniques is to obtain a visually appealing reconstruction, not to enhance

recognition performance [66]. Therefore, not only there is no discriminating information added to the LR images, but also the redundant information may actually reduce the recognition accuracy. Recently, there have been attempts to apply super-resolution and recognition simultaneously [55, 57]. But in these approaches, given a probe image, an optimization needs to be repeated for each gallery image in the database, which results in significant computational cost, especially in big databases.

In Chapter 5, we propose a novel real-time approach for matching LR probe face images with HR gallery face images. We use DCA to find a pair of transformations, which project the LR and HR feature vectors in a common subspace. In the new subspace not only the problem of feature vector length mismatch is solved, but also the correlation between corresponding features in LR and HR domain is maximized. In addition, the process of calculating the transformation matrices considers the class structure of the data and it aims to separate the classes in the new subspace, which is very beneficial from the recognition perspective. The face recognition is performed in the new subspace, which preserves the discriminant information content of the feature vectors and provides a better recognition accuracy. It is worth mentioning that the proposed method has a very low computational complexity and it can be employed for processing several faces that appear in a crowded image in real-time. Extensive experiments performed on low-resolution surveillance images from the Surveillance Cameras Face (SCface) database [2] verify the effectiveness of our proposed method, which outperforms the state-of-the-art LR face recognition techniques.

Finally, in **Chapter 6**, we conclude the approaches presented in all previous chapters.

CHAPTER 2

Privacy-Preserving Biometric Identification: Security Challenges in Cloud Computing Environments

The work presented in this chapter is a key step towards a cloud-based unified storage system for personal records. The idea is to create an encrypted database of personal records of individuals, *e.g.*, name, date of birth, educational information, banking or credit history, medical records, criminal records, insurance, etc., as a unified and privacy-preserving cloud-based database. Biometric information of individuals are used as a key attribute for this database. This database needs to be accessible by all enterprises that make use of biometric identification. In a large biometric system with lots of subjects, the system requires high storage capacity and processing power. The need to be accessed by multiple enterprises and to have a high processing power motivate the use of a cloud-based system to store and process the data.

The growing popularity of cloud-based systems has increased the importance of addressing the serious issue of the security of the data stored in the cloud [67–71]. In case of using biometrics to have access to records stored on the cloud, there is the risk of identity theft, because biometric data of the enrolled subjects can be stolen

and misused against their will. The biometric data is unique and irrevocable, and unlike passwords users cannot change their biometric characteristics. Consequently, the system must guarantee the preservation of the users' privacy, and therefore, the biometric database has to be encrypted.

Since the encrypted biometric database is stored in a public cloud, the identification process should be done with minimum amount of information leakage. That is, the comparisons need to be performed without the decryption of the data to prevent eavesdroppers from any access. However, the variations in the biometrics of each subject bring about a serious problem in the encrypted domain. Small changes in the data (plaintext) result in big differences in the ciphertext (encrypted plaintext). This difference can mislead the recognition process. Consequently, it is not feasible to just simply add an encryption scheme to a biometric identification system in order to secure the data and expect to obtain the same results that are obtained without the encryption.

In this chapter, we present a privacy-preserving cloud-based identification system (CloudID), which allows users to securely store their confidential information in untrusted public clouds¹. It gives them the opportunity of having effective and secure storage along with the computational power of the cloud infrastructures as well as controlled and flexible data access by multiple agents. Our proposed approach is the first cloud-based biometric identification system with a proven zero data disclosure possibility. CloudID performs the identification process in the encrypted domain without decrypting the data. This prevents the cloud provider or potential attack-

¹This work appeared in a conference and a journal paper [72] and [73].

ers from gaining access to any sensitive data or even the contents of the individual queries.

Unlike other privacy-preserving biometric identification methods, our approach does not apply a distance-based matching. However, using the query sample, it creates an encrypted conjunctive range query, which is applied on the encrypted gallery samples stored in the cloud. In this scenario, the only revealed piece of information is the binary matching result, *i.e.*, *match* or *not match*. This makes CloudID secure against *center search attack* [18] in which the attacker can recover the biometric template even if it is stored encrypted. In order to improve the performance of the biometric-based identification in the encrypted domain, we propose a k-d tree structure to create encrypted search queries. Applying this structure in the core of a searchable encryption technique helps the system not only to quantize the biometric features but also to handle the variations in the biometric data. Moreover, our algorithm is not limited to any specific type of biometric data and it can work with any biometric trait and any feature extraction method.

A working prototype of the CloudID framework is implemented and evaluated using a public biometric database. Our experimental results show the feasibility of CloudID for accurate biometric identification with no confidential data disclosure possibility, which enables building trustworthy storage systems for sensitive records.

The rest of the chapter is organized as follows. Section 2.1 presents related work from the literature. Section 2.2 provides a brief overview of the system whose complete design is described in Sections 2.3 and 2.4. The implementation details and overall performance of the system are presented in Section 2.5. Finally, Section 2.6 concludes the chapter.

2.1 Related Work in Privacy-Preserving Biometric Identification

In the literature, there are two sets of studies considering the privacy-preserving biometric identification. The first set of studies [6–12] only achieve privacy-preserving at the time of executing the query, protecting the confidentiality of both server and client. In these approaches, the server, in which the biometric database is stored, is considered to be trusted, and the biometric database is stored unencrypted. This allows the server to have access to the contents of the biometric database. However, these approaches cannot be used in case of untrusted servers such as clouds. In the second set of studies, the server is not trusted and the biometric database is encrypted [13–17]. However, these algorithms have some limitations and suffer from information leakage. In this section, we briefly describe these methods and compare them with our proposed approach.

To the best of our knowledge, Erkin *et al.* [6] considered the problem of privacy preserving biometric identification for the first time. They proposed a privacy-preserving face recognition system based on the well-known eigenface approach introduced by Turk *et al.* [74, 75]. They employed Pailliers cryptosystem [76], as an additive homomorphic encryption and calculated the Euclidean distance between face image feature vector from client and server’s face image database. The matching algorithm is performed between the client and the server without revealing the client’s biometric information or the result of the query to the server. At the same time, the client cannot learn from the database stored in the server. Later, Barni *et al.* [8] proposed a similar algorithm for a fingerprint recognition system, *FingerCodes* [77]. Both of

these protocols [6, 8] rely on homomorphic encryption and do not try to find the specific match but the group of the nearest matches within some threshold.

Sadeghi *et al.* [7] improved the efficiency of Barni’s algorithm by applying a hybrid approach where *garbled circuits* were used in conjunction with homomorphic encryption to find the minimum distance. Huang *et al.* [9] combined the algorithms proposed in [7] and [6] to further improve the computational and bandwidth efficiency of the system.

The main idea in [6, 7, 9] is to find the nearest match for a query in the biometric database based upon the Euclidean distance. In these references, each query is encrypted using the public key published by the client and sent to the server. The server also encrypts each biometric data in the database using an additive homomorphic encryption using the same public key. Then, the Euclidean distances between the query and each gallery in the database are calculated in the encrypted domain, d_1, d_2, \dots, d_n . In [7, 9], this information is fed into a garbled circuit, which finds the closest match by calculating $i^* = \operatorname{argmin}_i(d_1, d_2, \dots, d_n)$.

Blanton and Gasti [10] also proposed a secure protocol for iris codes based on additive homomorphic encryption and garbled circuits. The protocol uses Hamming distance to measure the similarity between iris codes. They also applied their technique on *FingerCode* calculating the Euclidean distances for fingerprint recognition.

Osadchy *et al.* [11, 12] designed a privacy-preserving face identification system called SCiFI. The implementation of this method is based on additive homomorphic encryption and oblivious transfer. SCiFI represents the facial images by binary feature vectors and uses Hamming distances to measure the image similarity.

In the above-mentioned scenarios [6–12], the server cannot learn neither the client’s biometric information nor the query result. On the other hand, the client does not get any information about the biometric database, which is saved unencrypted in the trusted server. However, these approaches cannot be used in case of untrusted servers like public clouds. In CloudID, our assumptions are quite different from these protocols. In our approach, the biometric database is encrypted and outsourced to an untrusted server (cloud), which does not have the key to decrypt the data. Therefore, not only the query process but also the biometric database stored in the server is secure. This makes our approach applicable for secure storage and processing of biometric data in public clouds.

Another group of researchers also considered the problem of privacy-preserving biometric identification [13–17]. These studies are much similar to our approach since the biometric database is encrypted and the identification process is applied in the encrypted domain without decrypting the data. Bringer *et al.* [13, 14] introduced an error-tolerant searchable encryption system to cope with the intra-class variations of the biometric data. This method is designed for a special type of biometric data, *IrisCode*, in which biometric templates are represented by binary vectors [78]. With the use of binary representation, comparisons are performed by Hamming distance. A locality sensitive hashing function is applied to narrow down the database search to a few candidates. The locality sensitive hashing function produces identical or very similar hash results for representations that are close to each other. This actually categorizes data in neighborhoods that have a few members. Given a query, it is assumed that the answer is among a set that has the same hash as the query. From this set, the nearest neighbor to the query is considered as the definite answer. Being

computationally expensive, they improved the performance of this algorithm using Oblivious RAM [15], and Oblivious Transfer [16]. In [17], they generalized the method proposed in [16] to compute several distance metrics, such as Hamming distance, Euclidean distance, and Mahalanobis distance.

The algorithms proposed in [13–17] do not guarantee zero data disclosure and suffer from information leakage. The mathematical framework to analyze the information leakage is provided in [18]. Pagnin *et al.* [18] employ a *center search attack* and prove that when the matching process is performed using a specific family of distances (such as Hamming and Euclidean distances), then information about the reference template is leaked, and the attacker can recover the biometric template even if it is stored encrypted. On the other hand, using the locality sensitive hashing function, the (potentially malicious) cloud provider is able to cluster the data into groups based on the pattern for searching/processing the data for every received query. Therefore, through observations of the data search procedure, the malicious provider may eventually be able to sort the data records, which could lead to potential data confidentiality breach. To address these vulnerabilities, our proposed method (CloudID) does not calculate any distances to match the biometric information. CloudID applies a conjunctive query over encrypted biometrics, and the only information that is revealed is the binary matching result, *i.e.*, *match* or *not match*.

2.2 System Overview and Access Control

CloudID is the first practical system that provides privacy-preserving capability to a biometric-based system for cloud-side query processing. Fig. 2.1 illustrates the different components of CloudID and their logical interconnections. The database of

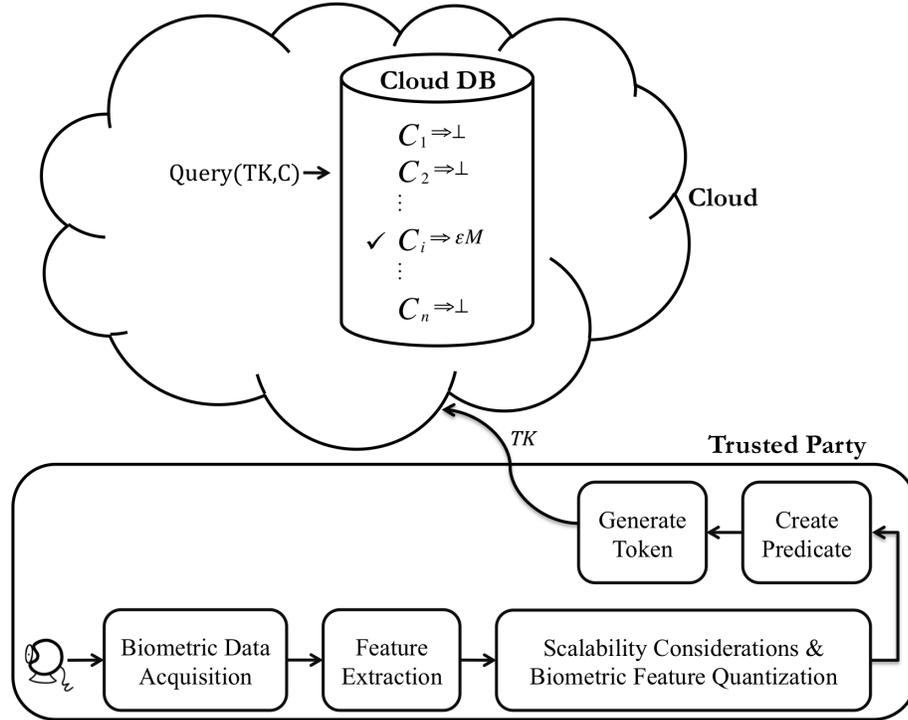


Figure 2.1: Framework of the CloudID in query processing.

personal records including their biometric data is stored encrypted in the cloud to prevent attackers from gaining unauthorized access. Biometric identification is applied in the encrypted domain to ensure that the cloud provider or potential attackers do not gain access to any sensitive data. On the other hand, the system must ensure that the cloud provider cannot learn the contents of individual queries either. When a query is performed, the provider should learn nothing but the value of a conjunctive predicate. We modify a search-aware encryption technique [79, 80] to make it applicable for biometric data.

The proposed system makes use of a *multiple-encryption* technique, *i.e.*, encrypting an already encrypted message. The two stages of the multiple-encryption are called *inner-level encryption* and *outer-level superencryption*, in order of occurrence. During the inner-level encryption, our system encrypts all personal information of

subject i , *e.g.*, his/her medical records (M_i), using the public key (PK_e) provided by each enterprise, *e.g.*, health care organization. εM_i denotes the encrypted data obtained by the inner-level encryption.

$$\varepsilon M_i \xleftarrow{\text{inner-level-enc.}} Enc(PK_e, M_i). \quad (2.1)$$

Then, during the outer-level superencryption, CloudID links the data encrypted in the inner-level (εM_i) to the subject's biometric record (B_i). That is, CloudID generates a public key based on the biometric record of the subject (PK_{B_i}) and encrypts all his/her information before storing it in the cloud. C_i denotes the ciphertext, corresponding to subject i , stored in the cloud.

$$C_i \xleftarrow{\text{outer-level-superenc.}} SuperEnc(PK_{B_i}, \varepsilon M_i). \quad (2.2)$$

The link to the biometric data enables the system to identify the subject and retrieve the corresponding data from the database.

Although CloudID can be extended to various biometric identification systems, we use facial data as an example in this study. As shown in Fig. 2.1, at the time of a query, a camera captures the image of the client and sends it to the biometric data acquisition system to localize the region of interest in the image. In case of facial data, a face detector localizes the face, then the face region is fed into the feature extraction stage. The number of the extracted features is usually high, which reduces the performance of the system. Therefore, a dimensionality reduction method is used to reduce the length of the feature vector. The feature vectors consist of real numbers; however, encryption algorithms are applied to discrete values. We propose a k-d tree structure to quantize the features and create encrypted search queries at the same time.

Since the biometric features of a subject change over time, the system derives range boundaries from the feature vector of the query to be compared with the feature vectors stored in the database. A conjunctive predicate, $P()$, is created for the comparison query. Eq. (2.3) shows the conjunctive proposition for a biometric feature vector $B = [b_1, b_2, \dots, B_n]^T$. The query is answered by records whose feature vectors fall in between these boundaries.

$$P(B) = \begin{bmatrix} l_1 < b_1 < u_1 \\ l_2 < b_2 < u_2 \\ \vdots \\ l_n < b_n < u_n \end{bmatrix} \quad (2.3)$$

In order to achieve maximum protection and minimum amount of information leakage, rather than verifying the individual comparison propositions separately, we verify the whole conjunctive proposition. For example, in the above predicate, where the length of the feature vector is n , $2n$ propositions must be verified conjunctively ($l_1 < b_1 \wedge b_1 < u_1 \wedge l_2 < b_2 \wedge b_2 < u_2 \wedge \dots \wedge l_n < b_n \wedge b_n < u_n$). From the cloud's point of view, such a query only reveals the Boolean value of the conjunctive predicate. That is, if all the propositions are true, then the query result is εM ; however, if any of the propositions is false, then the query will return NULL output (\perp), and the cloud provider will not know which comparison(s) is/are not true. For example, if a conjunctive query, *e.g.*, $P_1 \wedge P_2$, is false, the cloud provider should not be able to tell whether P_1 or P_2 or both were false for the conjunction to be false. All the steps illustrated in Fig. 2.1 are described in detail in Sections 2.3 and 2.4.

We assume that all enterprises have access to the public cloud on which all the users' encrypted confidential information is stored. For example, once Alice grants

access to a specific enterprise, the cloud gives the enterprise permission to make queries over her encrypted data. In other words, each enterprise has the right to make queries over the data of only a subset of the users who granted access to the enterprise. Alice can also revoke the access by asking the cloud to remove her data from the list of records accessible by the enterprise. The access control for the enterprises in the cloud is out of the scope of this study. Instead, we focus on the privacy of the encrypted information and provide a search over encrypted data scheme that can securely find the information of the users using their biometric data.

2.3 Offline System Setup

In this section, we describe the offline setup of the privacy-preserving cloud-based identification solution. As mentioned above, we use face images as the biometric identifiers.

The biometric identification in CloudID consists of two phases: training and query. During the training phase, the system stores the feature vectors of the individuals in a biometric database, while in the query phase, it identifies the closest match to the query from the database. The number and type of features used have a direct effect on the performance and accuracy of the identification.

After detecting the face, features are extracted from the face region, then the dimension of the feature space is reduced and the features are quantized in preparation for encryption. The components of the CloudID are described in the following subsections.



Figure 2.2: Face detection using Viola and Jones method.

2.3.1 Biometric Data Acquisition

The first stage of gathering face information for a subject is to capture a photo of the subject and localize the face area using a face detection algorithm [81]. CloudID employs the Viola-Jones method [81] for face detection. Fig. 2.2 shows a few examples of face detection results, obtained by this algorithm, on images from the FERET database [82].

2.3.2 Feature Extraction

CloudID uses Gabor filters to extract features from the detected face region. The most important advantage of Gabor filters is their invariance to rotation, scale, and translation. Furthermore, they are robust against photometric disturbances, such as illumination changes and image noise [83–88].

The Gabor filter-based features are directly extracted from the gray-level images. In the spatial domain, a two-dimensional Gabor filter is a Gaussian kernel function

modulated by a complex sinusoidal plane wave, defined as:

$$G(x, y) = \frac{f^2}{\pi\gamma\eta} \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp(j2\pi f x' + \phi)$$

$$x' = x \cos\theta + y \sin\theta$$

$$y' = -x \sin\theta + y \cos\theta$$
(2.4)

where f is the frequency of the sinusoid, θ represents the orientation of the normal to the parallel stripes of a Gabor function, ϕ is the phase offset, σ is the standard deviation of the Gaussian envelope and γ is the spatial aspect ratio which specifies the ellipticity of the support of the Gabor function.

CloudID employs forty Gabor filters in five scales and eight orientations as shown in Fig. 2.3. The size of the face images used in our experiments is 120×120 pixels. Using forty Gabor filters, the dimension of the feature vector is $120 \times 120 \times 40 = 576000$. Since the adjacent pixels in an image are usually highly correlated, we can reduce this information redundancy by downsampling the feature images resulting from Gabor filters [83, 84]. The feature images are downsampled by a factor of four, which means that the feature vector will have a size of $576000 / (4 \times 4) = 36000$. These vectors are then normalized to zero mean and unit variance. In addition to downsampling, we need to use dimensionality reduction methods to further reduce the size of the feature vectors.

2.3.3 Scalability Considerations and Feature Quantization

In biometrics, the number of extracted features is usually high, which increases the computational complexity and decreases the performance of the system because of the curse of dimensionality. In order to address these issues, dimensionality reduction

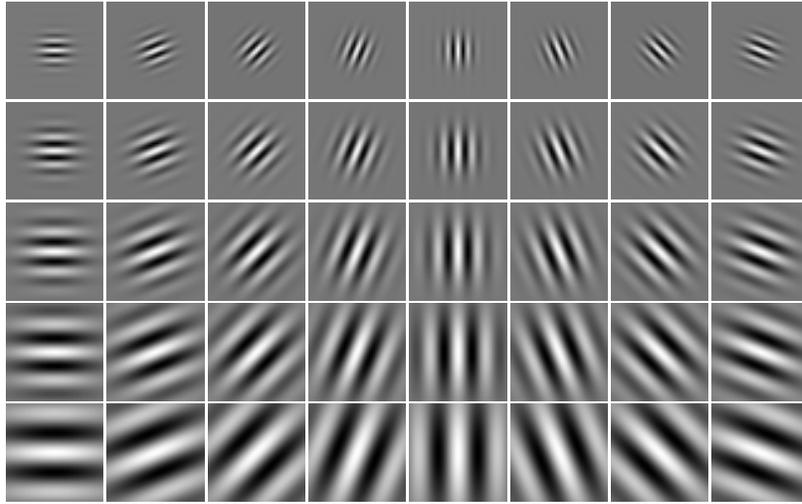


Figure 2.3: Gabor wavelets in five scales and eight orientations.

methods are applied to reduce the length of the feature vectors and their redundancies [89].

Classical dimensionality reduction techniques, such as principal component analysis (PCA) and linear discriminant analysis (LDA), have been widely utilized in face recognition techniques since Turks pioneering work [74]. However, each of these well known methods has its own shortcomings.

The Eigenface method, which uses PCA for dimensionality reduction, results in projection directions that maximize the total scatter across all classes, *i.e.*, across all images of all faces. During this projection which maximizes the total scatter, PCA preserves unwanted variations due to lighting and facial expressions [90]. As stated in [20], “the variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity”.

The Fisherface method, on the other hand, uses LDA to find a projection matrix W which is optimized to separate different classes, *i.e.*, maximizes the ratio of

between-class scatter (S_b) to within-class scatter (S_w):

$$W = \arg \max \frac{|W^T S_b W|}{|W^T S_w W|}. \quad (2.5)$$

W can be computed from the eigenvectors of $S_w^{-1} S_b$. However, due to the curse of dimensionality, especially in case of face recognition with low number of training data, S_w is usually singular, *i.e.*, the inverse of S_w does not exist.

To overcome these shortcomings, CloudID uses generalized discriminant analysis (GDA) [91], which is a non-linear feature reduction technique. Similar to LDA, the objective of GDA is to find a projection for the features into a lower dimensional space by maximizing the ratio of between-class scatter to within-class scatter. Suppose that the space X is mapped into the space F through a non-linear mapping function ϕ : $X \rightarrow F, x \rightarrow \phi(x)$. Considering C to be the number of classes and N_z to be the number of samples in class z , the S_w and S_b of the training set can be computed as follows:

$$S_w = \frac{1}{C} \sum_{z=1}^C \frac{1}{N_z} \sum_{k=1}^{N_z} \phi(x_{zk}) \phi^t(x_{zk}) \quad (2.6)$$

$$S_b = \frac{1}{C} \sum_{z=1}^C (\mu_z - \mu)(\mu_z - \mu)^t \quad (2.7)$$

where μ_z is the mean of the samples that belong to class z . GDA finds the eigenvalues λ and eigenvectors ν that satisfy:

$$\lambda S_w \nu = S_b \nu. \quad (2.8)$$

Since the eigenvectors lie in the span of $\phi(x_{11}), \dots, \phi(x_{zk}), \dots, \phi(x_{CN_z})$, there exists α_{zk} such that

$$\nu = \sum_{z=1}^C \sum_{k=1}^{N_z} \alpha_{zk} \phi(x_{zk}). \quad (2.9)$$

To generalize LDA to the nonlinear case, GDA considers an expression of dot product of a sample i from class p and another sample j from class q by the following kernel function

$$(k_{ij})_{pq} = \phi^t(x_{pi}) \cdot \phi(x_{qj}) = k(x_{pi}, x_{qj}) = e^{-|x_{pi} - x_{qj}|^2/r}. \quad (2.10)$$

Let \mathbf{K} be an $M \times M$ matrix defined on the class members by $(\mathbf{K}_{pq})_{p=1, \dots, C, q=1, \dots, C}$, where \mathbf{K}_{pq} is a matrix composed of dot products between class p and q in the feature space F :

$$\mathbf{K}_{pq} = (k_{ij})_{i=1, \dots, N_p, j=1, \dots, N_q}. \quad (2.11)$$

We also introduce an $M \times M$ block diagonal matrix:

$$\mathbf{U} = (\mathbf{U}_z)_{z=1, \dots, C} \quad (2.12)$$

where \mathbf{U}_z is an $N_z \times N_z$ matrix with all its elements equal to $\frac{1}{N_z}$.

By substituting Eqs. (2.6), (2.7), and (2.9) into (2.8) and taking inner-product with vector $\phi(x_{ij})$ on both sides, the solution of (2.8) can be obtained by solving:

$$\lambda \mathbf{K} \mathbf{K} \mathbf{a} = \mathbf{K} \mathbf{U} \mathbf{K} \mathbf{a}. \quad (2.13)$$

where \mathbf{a} denotes a column vector with entries α_{zk} , $z = 1, \dots, C$, $k = 1, \dots, N_z$. The solution of \mathbf{a} is computed by finding the eigenvectors of the matrix $(\mathbf{K} \mathbf{K})^{-1} \mathbf{K} \mathbf{U} \mathbf{K}$. If matrix \mathbf{K} is not reversible, the eigenvector is found by first diagonalising matrix \mathbf{K} [91]. After finding the L most significant eigenvectors, a projection matrix is constructed as:

$$W = [\mathbf{a}_1 \mathbf{a}_2 \dots \mathbf{a}_L]. \quad (2.14)$$

The projection of x in the L -dimensional GDA space is calculated by:

$$y = k_x W \quad (2.15)$$

where $k_x = [k(x, x_{11}) \dots k(x, x_{zk}) \dots k(x, x_{CN_c})]$.

Note that the number of features in LDA-based methods can be at most $C - 1$. In our experiments with the database of 200 subjects, the maximum size of the projected vectors is 199 which is a significant reduction in comparison to 36000.

CloudID combines the feature quantization, feature selection, and classification using parallel k-d trees. The feature vectors of the training set are organized in a k-d tree by partitioning the feature space along each feature using a hyperplane. In order to quantize L features, we need a k-d tree of the depth $L + 1$, which stores up to $2^L - 1$ samples. The maximum number of quantized features is calculated using Eq. (2.16).

$$L_{max} = \lceil \log_2(C + 1) \rceil - 1 \quad (2.16)$$

In our experiments, since the number of classes, C , is 200, the maximum possible depth of the k-d tree is 8, which can quantize 7 features. In order to make use of more features, CloudID employs several k-d trees in parallel for different sets of features. Fig. 2.4 illustrates the process of creating a quantized feature vector using several k-d trees.

Since GDA sorts the features according to their discriminative power, CloudID selects the most discriminative features of the GDA to construct the k-d trees. Moreover, it builds the k-d tree such that the less discriminative feature is used for the root and the most discriminative feature for the leaves. The feature used in the root of a tree only defines one hyperplane and divides the space into two partitions. Consequently, the second and third levels of the tree divide the feature space into just three and five partitions. Since the features near the root of the tree only define a few quantization levels, CloudID assigns the least discriminative features for those

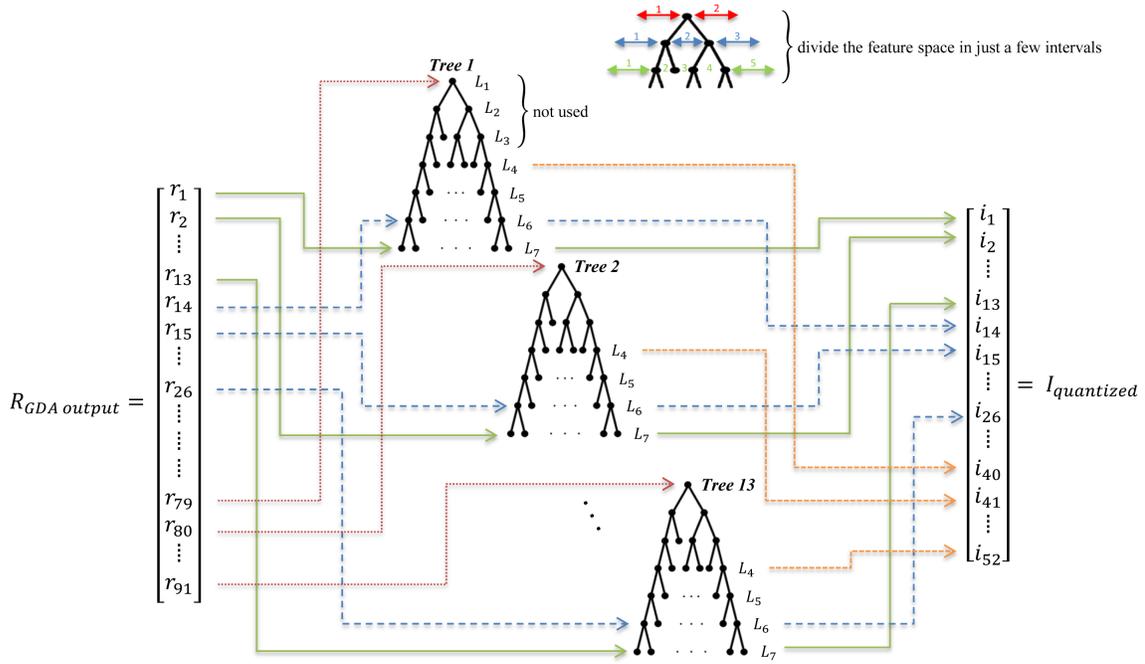


Figure 2.4: Creating the quantized feature vector using 13 k-d trees of 4 significant features.

levels and do not use them in the quantization process. On the other hand, the most discriminative features are used in deeper levels of the tree where the number of hyperplanes are high and a finer quantization is performed.

As shown in fig. 2.4, using the training set, CloudID makes use of 91 features to construct 13 parallel k-d trees. Each of the first 13 features of the GDA output, *i.e.*, the most discriminative ones, is used for the leaves of a tree. On the other hand, the last 13 features (79 to 91) are used for the roots. The first three levels of each tree are then disregarded and not included in the conjunctive predicate. In other words, each tree is used to quantize 4 features and therefore by using 13 trees we are quantizing the first 52 features of the GDA output. Note that, in real systems, the number of subjects are higher, which, based on Eq. (2.16), increases the number of features used in each tree.

2.3.4 Cryptographic Key Generation

Generally, owners of the biometric records can not trust cloud providers, even the well known providers, for storing their data in plain format, due both to internal security policies, and federal or state laws and requirements [92]. Therefore, data must be encrypted before storing it on a remote server.

While the usage of regular cryptographic encryption solutions prevents unauthorized access and helps ensure the confidentiality of biometric data, it prevents services such as searching over encrypted data that is required for secure identification in the CloudID framework. As a result, the whole biometric data set must be downloaded to a local system for every query. This method is not efficient in terms of bandwidth or computational requirements on the client side because the client has to store and decrypt all received data to be able to extract those that are actually needed.

To provide trustworthy data storage and data search capabilities, CloudID makes use of search-aware encryption [79,80] to prevent attackers from gaining unauthorized access to the confidentially-sensitive biometric data records. Let us assume that the trusted party intends to encrypt and send a user's biometric data record, denoted by B , to the cloud for secure storage and remote biometric matching that requires query processing over encrypted data. This approach consists of four major phases, namely *Setup*, *Encryption*, *Token Generation*, and *Query*. The first two phases are parts of the offline system setup and the latter two phases are parts of the online setup and will be described in the next section.

CloudID uses a public/private key scheme to encrypt and decrypt data. In the *Setup* phase, all the cryptographic parameters, *i.e.*, public key, PK , and private key,

SK , are generated:

$$PK \leftarrow (PK_1, PK_2, \dots, PK_t), \quad (2.17)$$

$$SK \leftarrow (SK_1, SK_2, \dots, SK_t), \quad (2.18)$$

where t denotes the number of all possible predicates for which the user can query. t directly depends on the number of features and the number of quantization levels for each feature.

2.3.5 Biometric Record Encryption

The *Encryption* phase performs the last offline setup step by encrypting each user's biometric data record. It is important to note that this step will be completed locally by the trusted party so that the (untrusted) cloud provider will only have access to ciphertexts, and not the encryption keys themselves. In particular, the *Encryption* step is carried by using the following assignments for $1 < j < t$

$$C_j \leftarrow \begin{cases} \text{Encrypt}(PK_j, \varepsilon M) & \text{if } P_j(B) = 1 \\ \text{Encrypt}(PK_j, \perp) & \text{o/w,} \end{cases} \quad (2.19)$$

where M is the relevant data of the subject that is encrypted using the public key, PK_j , corresponding to the predicate, $P_j()$, which is satisfied by user's biometric, B (refer to Eq. (2.3)). The other public keys corresponding to the *False* propositions will encrypt a NULL value (\perp). Once completed, this step will give us a ciphertext vector

$$C \leftarrow (C_1, C_2, \dots, C_t), \quad (2.20)$$

whose length is the number of possible predicates t . Note that, only one of the ciphertexts includes the data of the subject.

2.4 Online System Behavior

After the offline system setup and storing the biometric samples of the individuals in the database, the system is ready to respond to queries and identify a user via his/her biometric information. Here, we assume that all the subjects are already registered in the database by their face information. All the relevant data, M , belonging to a subject is linked to his/her biometric, B , and both the data and the biometric are encrypted. During the query time, a subject's biometric information is captured by a camera and CloudID compares it to the encrypted biometrics in the database to identify the person. The biometric matching is performed in encrypted domain so that neither an intruder nor the cloud provider can access the user information.

2.4.1 Face Features Variations

As mentioned before, the biometrics of an individual change over time. For example, facial features vary due to the changes in illumination, head pose, facial expressions, cosmetics, aging, and occlusions. Therefore, the feature vector of the query will not be exactly the same as the one saved in the database for the same person.

In the feature quantization section, we describe how the CloudID uses k-d trees to find the quantization thresholds. Some variations in the facial features may result in significant changes in the quantized data, *i.e.*, a particular feature value in the query

may not fall in the same quantization interval of the corresponding feature for the same subject enrolled in the database. Therefore, the system needs to accept not only the exact quantized values, but reasonably-sized ranges in order to include accepted variations around the quantized values. These ranges are defined by lower and upper boundaries, where the features of the query fall in between these two boundaries. Note that the size of these ranges has a direct influence on the accuracy of the system. A tight range makes the system very strict not accepting large variations; and a wide range makes the system more lenient accepting large variations. Therefore, in the first case, the recognition rate is low, however in the latter case, the recognition rate increases at the cost of a higher false positive rate.

As shown in Fig. 2.4, k-d trees divide the feature space into several intervals. The above-mentioned ranges are made of few adjacent intervals around the query’s feature values. The number of the intervals in each level, L , of the tree is $2^{L-1} + 1$; *i.e.*, the features near the root of the tree divide the space into less intervals than the ones near the leaves. Therefore, the intervals near the root of the tree are wide and coarse while at deeper levels they are fine. Taking this into account, CloudID increases the number of the accepted adjacent intervals in deeper levels of the tree to make sure that we accept almost the same size ranges in all levels. In our experiments, we have empirically chosen the size of the range to be $2^{L-2} + 2$ adjacent intervals.

2.4.2 Cryptographic Token Generation

To preserve the users’ privacy, CloudID employs a searchable encryption for comparison queries [80]. The *Token Generation* phase prepares the system for processing a given query. As discussed above, for each query, CloudID creates numerical ranges

defined by lower and upper boundaries. CloudID processes the comparison queries given the range boundaries applied in a conjunctive manner to create a predicate, $P()$. CloudID implements *GenToken* that uses each query predicate and the secret key to generate the corresponding *token TK* and sends it to the cloud.

$$TK \leftarrow GenToken(SK, P()). \quad (2.21)$$

This step is accomplished locally by the trusted party, and hence the cloud provider will only see the encrypted token. For instance, given predicate $P_i()$, the trusted party chooses the corresponding private key $SK_i \in SK^2$ to generate the token. This *token* will be used in the *Query* phase.

2.4.3 Biometric Database Query

The final step is processing the received query by the cloud provider that has access to the encrypted database records, and the single token, TK sent by the trusted party. The cloud provider uses TK to decrypt the corresponding ciphertext C_i for each individual's records. Consequently, the cloud provider will only retrieve the records if the biometric of the individual satisfies the predicate used for the token generation, and will get \perp as a result for all other decryptions (Eq. (2.19)):

$$\begin{cases} Query(TK, C) = \varepsilon M & \text{if } P(B) = 1 \\ Query(TK, C) = \perp & \text{if } P(B) = 0 \end{cases} \quad (2.22)$$

²Remember SK includes a private key for each possible predicate (query).

There might be more than one biometric record in the database that satisfy the query predicate. All the retrieved records are then sent back to the trusted party and filtered after decryption³.

In summary, using the above-mentioned four steps, CloudID makes sure that the cloud provider can perform and accomplish the search over the encrypted data, and in the meantime, cannot access any confidential data. See Appendix for the proof of user privacy preservation in CloudID.

2.5 Evaluation

2.5.1 Experimental Setup

We evaluated the accuracy and performance of CloudID’s various components through an extensive set of experiments. To that end, we used *Pegasus* high performance computing cluster at the center for computational science (CCS) at University of Miami [93]. Each node in Pegasus is equipped with eight Intel[®] Xeon[®] Sandy Bridge 2.6 GHz cores and 32 GB of memory.

In order to evaluate the system, the Facial Recognition Technology (FERET) database is used in our approach [82]. Six hundred frontal face images from 200 subjects are selected, *i.e.*, three images per subject: a frontal image with neutral expression, a frontal image with an alternative expression, and a frontal image with different illumination. In FERET database, these images are letter coded as ba , bj ,

³It is noteworthy that for absolute data disclosure prevention, using a multiple-encryption scheme, the trusted party encrypts the records using a symmetric key such that M is already encrypted (ϵM) and the cloud provider is not allowed to see the actual records.



Figure 2.5: Face images from FERET database after face detection and cropping. First row: frontal series (ba). Second row: frontal series with an alternative expression (bj). Third row: frontal series with illumination change (bk).

and *bk*. We applied Viola-Jones [81] face detection method on these images to extract the faces. Fig. 2.5 shows some samples from the database after face detection.

2.5.2 Accuracy

The accuracy of a biometric system depends on different factors, *e.g.*, the discriminative ability of the features, number of features, and classifier performance. The previous methods discussed in Section 2.1 mainly depend on a special type of biometric data or recognition technique [6–16]. However, CloudID is not limited to any biometric trait, feature extraction technique or dimensionality reduction method, and it just deals with feature vectors. For example, the face recognition algorithm used in our experiments utilizes Gabor features and GDA dimensionality reduction. This combination with a simple k-nearest neighbor (KNN) classifier has better accuracy than the eigenface method [74] used by the previous privacy preserving solutions [6, 7]. Fig. 2.6 clearly shows that this approach outperforms the eigenface method⁴.

⁴Note that this comparison is between the face recognition approaches before applying any security scheme on them.

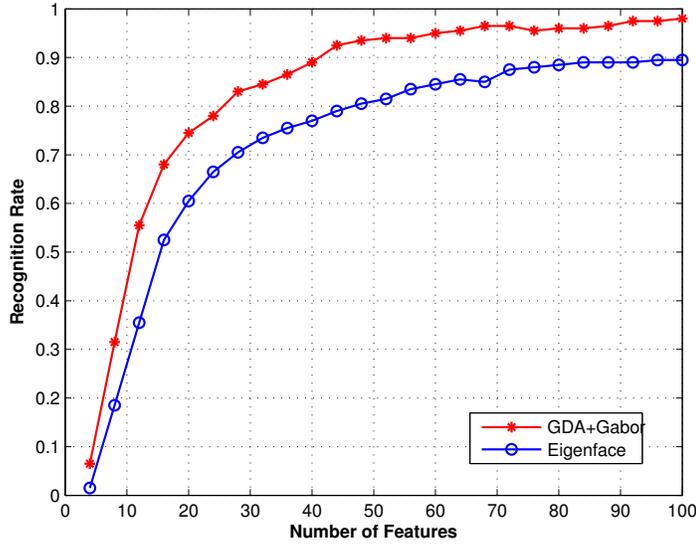


Figure 2.6: Comparison of the eigenface and Gabor+GDA face recognition algorithms

Table 2.1 shows the recognition and false positive rates of CloudID using different number of features. We chose the number of accepted adjacent intervals to be $2^{L-2}+2$. Each of the parallel trees may give different results for the identification problem. The matched subjects are the ones that majority of the trees have identified as the true matches. It is worth noticing that there is information loss in the process of quantizing the feature vectors. Therefore, the recognition accuracy also depends on the efficacy of the quantization method. The usage of quantization and range queries makes secure biometric systems feasible, but forces the designer to accept a reduction in accuracy in comparison to that which would have been achieved in an unconstrained scheme [94].

As shown in Table 2.1, increasing the number of features improves the recognition rate significantly. Moreover, since each proposition (matching feature) can be either *True* or *False*, if we hypothetically assume that each proposition has the same probability of being *True*, e.g. p , using n features in the conjunctive query, the probability of the conjunctive proposition being *True* is p^n . Hence, using more features decreases

Table 2.1: Face recognition accuracy of CloudID

No. of Trees	No. of Features	Recognition Rate (%)	False Positive Rate (%)
1	4	77.50	24.62
3	12	82.50	15.58
5	20	83.00	10.93
7	28	87.50	8.38
9	36	91.50	5.38
11	44	93.50	3.32
13	52	95.00	1.71

the probability of having a positive response exponentially which on the other hand leads to a reduction in false positive rate (See Table 2.1).

As mentioned before, we apply a range query over the encrypted database where the recognition rate is a function of the width of the accepted range. However, increasing the width of the range also increases the false positive rate. Table 2.2 clearly shows this effect for a fixed number of features ($n = 52$). Note that, the number of quantization levels increases exponentially relative to the level of the tree ($2^{L-1} + 1$); therefore, the number of accepted adjacent intervals should also increase with the increase of the tree level to make sure that we cover almost the same size range covered in previous levels. Therefore, in addition to the number of features, the width of the accepted range is another degree of freedom for the designer.

2.5.3 Performance

Quantizing the feature vectors and using range queries to have privacy-preserving properties forces a compromise between the discriminative power and the security

Table 2.2: System accuracy using different accepted range sizes

No. of Accepted Adjacent Intervals	Recognition Rate (%)	False Positive Rate (%)
$2^{L-2} - 3$	28.50	0.00
$2^{L-2} - 2$	44.00	0.01
$2^{L-2} - 1$	59.50	0.07
2^{L-2}	70.00	0.26
$2^{L-2} + 1$	84.00	0.95
$2^{L-2} + 2$	95.00	1.71
$2^{L-2} + 3$	98.00	6.39

level of the feature values. This, in turn, affects the accuracy of the system, highlighting the fact that privacy comes at the price of performance [94]. For real-world deployment, CloudID needs to respond to queries efficiently. Fig. 2.7 shows the average encryption time needed for each subject along with the average time needed to perform a query using different number of levels in each tree. The experimental results clearly verify the exponential increase in complexity with the increase in the number of features in each tree, which is related to the number of predicates. On the other hand, the numbers of trees brings only a linear increase in complexity.

For a fixed number of predicates, the average encryption time for each subject is constant. In our experiments, using the last 4 levels of each tree, the average encryption time of each subject is about 18 seconds. However, as the number of subjects in the database increases, the number of comparisons for query processing grows as well, which increases the query time. Note that, in authentication problems, we would not have this complexity and the query could be made in real-time. But to avoid any information leakage, we do not label data stored in the cloud; therefore, we resort to the identification problem.

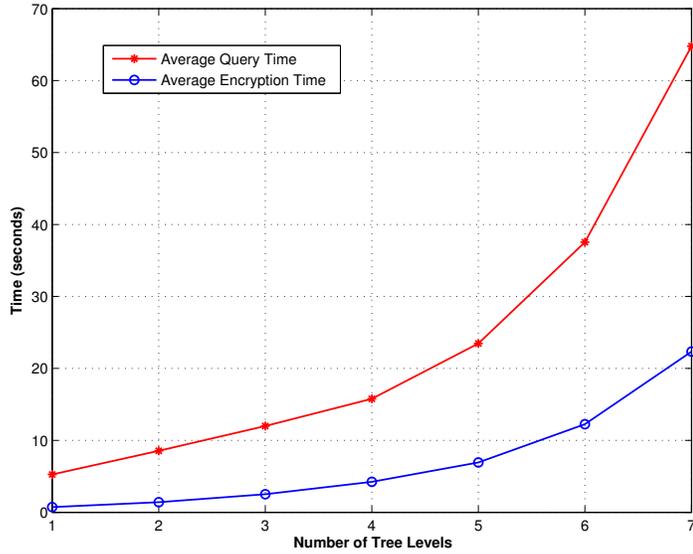


Figure 2.7: Average encryption time and query time using different number of features.

The main weakness of the proposed method is its computational complexity. To further improve CloudID’s performance, we implemented the system in parallel threads on eight cores. Due to the independence of the comparisons in CloudID, parallel computation can be effectively applied. Fig. 2.8 shows the encryption and query times of the system using different numbers of threads. Since not all the processes are threaded, using all of the cores does not necessarily reduce the time by one eighth. It is obvious that in real cloud systems, more resources can be employed, which can further improve the performance of the system.

2.5.4 Scalability of CloudID

In this section, we evaluate the scalability of the proposed method in dealing with new subjects that were not used for training. The goal is to examine if CloudID is trained on a separate a population of subjects whether the GDA transformation matrix and the quantization levels will still perform well on new subjects. We use a

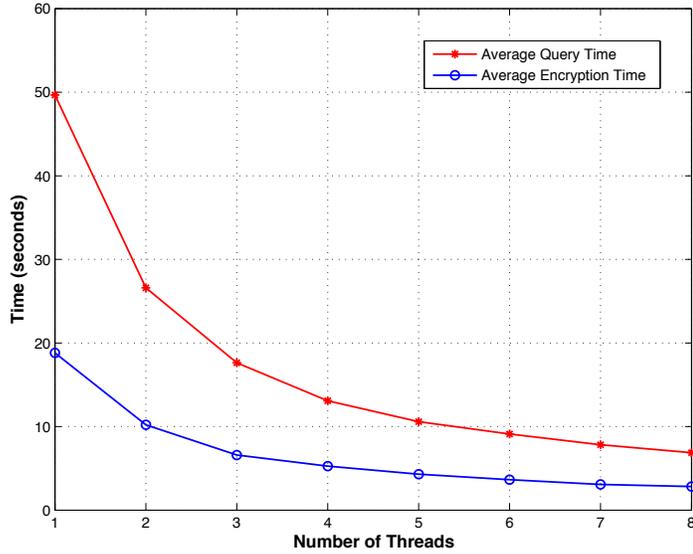


Figure 2.8: Average encryption time and query time using multi-threading.

population of subjects to train CloudID and obtain the GDA transformation matrix and the quantization levels. Another population of subjects, which is not used for training, is used for evaluating the recognition performance.

For this purpose, we use the WVU database [95] with 402 subjects. This database consists of almost 110 seconds long video clips captured with a camera that rotates around the face. We extract three face images per subject in frontal pose and $\pm 10^\circ$ of rotation from frontal. Fig. 2.9 shows the detected and cropped face images of a sample subject from WVU database. In this experiment, we use a 10-fold random subsampling validation process. Two randomly chosen images of 200 randomly selected subject are used for training and obtaining the GDA transformation matrix and the quantization levels. Based on Eq. 2.16, we can only quantize seven features in each tree. Disregarding the three root levels and using 13 parallel trees, we get the quantization levels for 52 features. The remaining images of the training subjects are used for testing on which we achieve 97.44% accuracy.

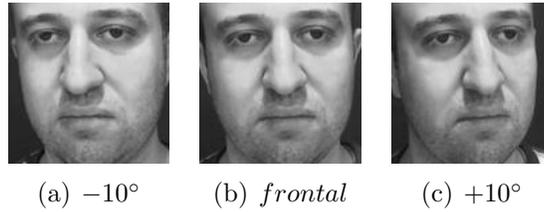


Figure 2.9: Sample face images of a subject from WVU database after face detection and cropping.

Later, we gradually add new subjects to the database but we use the same GDA transformation matrix and the quantization levels. Note that two randomly selected images of the new subjects are used for the enrollment and the remaining one is added to the query set. As shown in Fig. 2.10, the average recognition rate slowly decreases by adding new subjects to the system such that after doubling the number of subjects it decreases by 1.37%.

An interesting point is that when we reach 255 subjects, we can retrain the system to have one level deeper k-d trees that can quantize eight features (See Eq. 2.16). We performed another experiment retraining the system at this point. As shown in Fig. 2.10, retraining boosts the average recognition rate of the system such that with 255 subjects we achieve even better accuracy than that of with 200 subjects. The main reason is obviously using more features, *i.e.*, 65 features in this case, and also a finer quantization that is achieved in deeper levels of the trees.

It can be seen that the slope of the new graph is lower than the previous one and increasing the subjects to 402 reduces the accuracy by only 0.53%. Therefore, we can conclude that the retraining not only increases the accuracy of the system but also makes the system more stable in dealing with new unseen subjects.

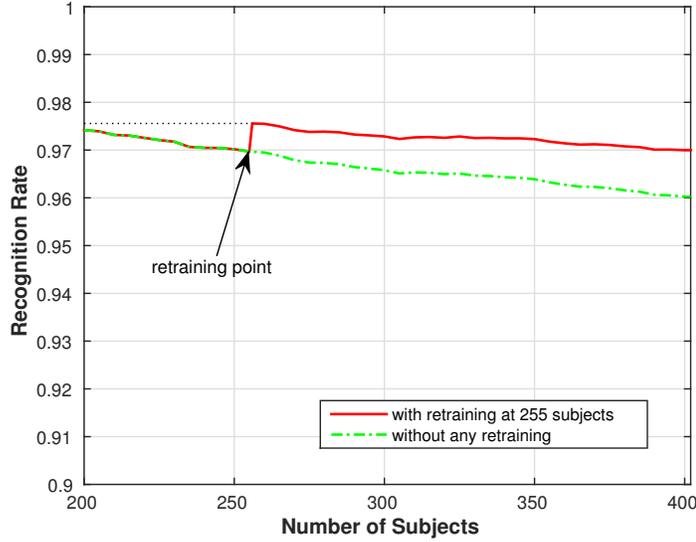


Figure 2.10: Scalability of the CloudID in dealing with new and unseen subjects. The training is already done using 200 subjects and 202 new subjects are gradually added to the database.

2.5.5 Case study

In this section an overall case study of the system is presented using the first subject in Fig. 2.5. Viola-Jones face detection [81] is applied to crop the face area of the image. Then, features are extracted using the Gabor wavelets shown in Fig. 2.3. Fig. 2.11 illustrates the face detection step along with the real parts of the results of applying Gabor filters to the face image. As can be seen in Fig. 2.11, Gabor filters extract the variations in different frequencies and orientations in the face. Here, the size of the resulting feature vector is the size of the image (120×120) multiplied by the number of scales and orientations (5×8) divided by the row and column downsampling factors (4×4) which is $120 \times 120 \times 5 \times 8 / (4 \times 4) = 36000$ in total.

After applying GDA dimensionality reduction, the number of features is reduced to $C - 1$, where C is 200 in our experiments. As mentioned in Section 2.3, the quantization is done using k-d trees. Each level of the k-d tree has 2^{L-1} boundary points which divide the space into $2^{L-1} + 1$ intervals. Therefore, each level will

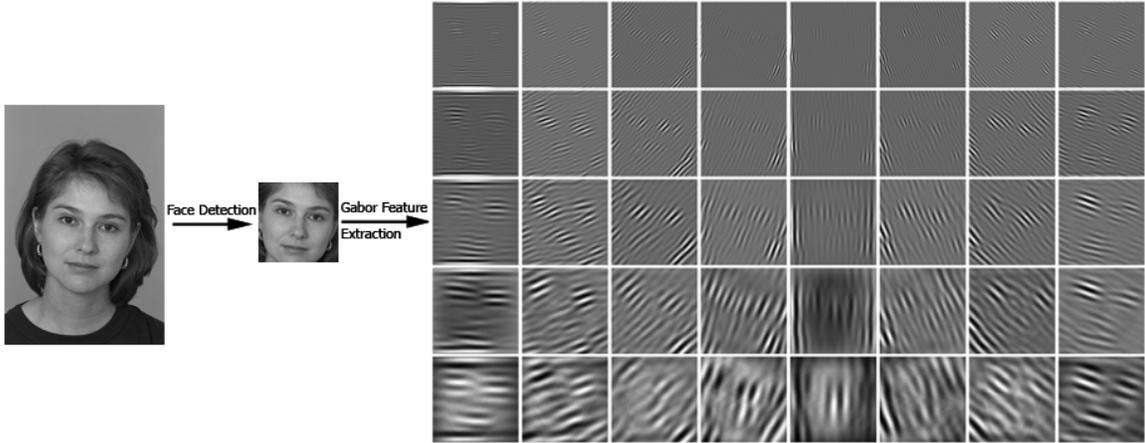


Figure 2.11: Face detection and feature extraction, results of applying filters shown in Fig. 2.3 on a face image.

have $2^{L-1} + 1$ quantized values which are denoted by the integers 1 to $2^{L-1} + 1$. Assuming that we only use one tree with seven levels, below is the feature vector and its quantized form for the first image of the first subject, which is used for training and enrollment.

$$\mathbf{B}_{\text{enroll}} = \begin{bmatrix} -4.9309 \\ -3.7765 \\ 5.1355 \\ 7.9747 \\ 5.7201 \\ 7.7592 \\ 1.719 \end{bmatrix} \xrightarrow{\text{Quantization}} \begin{bmatrix} 1 \\ 1 \\ 5 \\ 9 \\ 17 \\ 33 \\ 52 \end{bmatrix}$$

CloudID links the above vector to the records of the subject, encrypts them and stores them in the public cloud. During the time of query, let us assume that the image shown in the second row of the Fig. 2.5 is the image of the client's face. In this case, the feature vector for this sample is:

$$\mathbf{B}_{\text{query}} = \begin{bmatrix} -4.1877 \\ -2.7772 \\ 3.6925 \\ 1.5925 \\ 1.8369 \\ 3.564 \\ 5.228 \end{bmatrix} \xrightarrow{\text{Quantization}} \begin{bmatrix} 1 \\ 1 \\ 5 \\ 9 \\ 15 \\ 30 \\ 59 \end{bmatrix}$$

The system feeds the query feature vector, B_{query} , into the k-d tree and extracts lower and upper boundaries to create the predicate. Here in our experiments, we chose the number of accepted adjacent intervals to be $2^{L-2} + 2$ to make sure that they cover almost the same size range for each feature. Note that in the quantization, we have considered the extremes. *i.e.*, minus infinity and plus infinity, to be 0 and $2^{L-1} + 2$. Therefore, the boundaries will not exceed these extreme values. For example, for level 1, plus infinity is mapped to 3, while for levels 2 and 3, it is mapped to 4 and 6, respectively. As you can see in (2.23), the first three levels cover almost the whole quantized space from minus infinity (0) to plus infinity ($2^{L-1} + 2$), *i.e.*, very coarse quantization, which is not useful for identification. Therefore, the system uses only the last 4 levels of the tree to create the predicate:

$$\begin{bmatrix} 0 < 1 < 3 \\ 0 < 1 < 4 \\ 1 < 5 < 6 \\ 3 < 9 < 10 \\ 5 < 15 < 18 \\ 12 < 30 < 34 \\ 25 < 59 < 66 \end{bmatrix} \Rightarrow P_o() = \begin{bmatrix} 3 < b_4 < 10 \\ 5 < b_5 < 18 \\ 12 < b_6 < 34 \\ 25 < b_7 < 66 \end{bmatrix} \quad (2.23)$$

The above predicate is used to create the token, which is used to query the encrypted database in the cloud. Any ciphertext whose corresponding feature vector falls in the intervals of the predicate will return *True* result for the conjunctive comparison. However, even if one of the propositions is *False*, the whole conjunctive comparison will be *False*.

Eq. (2.24) shows the proposition of the enrolled feature vector, B_{enroll} , in the predicate created by the query feature vector, $P_o()$. Here the result of the conjunctive comparison is *True* for the query. However, there might be feature vectors of other subjects, which also satisfy the above predicate. The number of these incorrectly identified subjects affects the false positive rate of the system.

$$P_o(B_{enroll}) = \begin{bmatrix} 3 < p_4 = 9 < 10 \\ 5 < p_5 = 17 < 18 \\ 12 < p_6 = 33 < 34 \\ 25 < p_7 = 52 < 66 \end{bmatrix} = True \quad (2.24)$$

Similarly, the above steps are applied to the other trees using other features. The overall response for the conjunctive query will be *True* only if majority of the trees have *True* response.

2.6 Conclusions and Future Work

In this chapter, we presented a privacy-preserving cloud-based and cross-enterprise biometric identification solution. Our proposed system is the first cloud-based biometric identification system with a proven zero data disclosure possibility. In this approach, all biometric information are encrypted and the identification process is performed in the encrypted domain without decrypting the biometric data.

Unlike other privacy-preserving biometric identification methods, our approach is not limited to any special type of biometric data and it can work with any biometric trait and feature extraction techniques. Moreover, it does not use distance-based matching, which is proven to have information leakage. However, it applies a conjunctive range query over encrypted gallery samples, which returns *true* response only if all the features of the gallery sample fall in certain ranges defined by predicates created using the query sample. This makes CloudID secure against *center search attack* in which the attacker can recover the biometric template even if it is stored encrypted. We proposed a k-d tree structure to quantize the biometric feature vectors and define the range predicates. This structure also allows the system to handle variations in the biometric data.

The proposed system enables clients to securely store their confidential information in the cloud and facilitates remote biometric-based identification by enterprises

that are granted access to these confidential records. It provides a solution to the concerns about the security and confidentiality of personal information stored in the cloud through the use of biometrics, while guarding against identity theft. We implemented a working prototype of the CloudID and evaluated it using a face biometric database. Our experimental results show that CloudID can be used in practice for biometric identification with a proven zero data disclosure.

The main weakness of our proposed method is its complexity and the size of the ciphertext. In order to perform real-time identification in case of large databases, more computing resources need to be allocated by the cloud provider, which might be costly. Another issue is that the identification accuracy is dependent on the efficiency of the quantization approach. In the future, we plan to solve the complexity problem of the algorithm using more efficient searchable encryption techniques. Moreover, we plan to design more intelligent and possibly adaptive quantization methods to reduce the information loss in this step and consequently increase the recognition rate. We also plan to design an expert system that can integrate multiple sources of biometric information, *e.g.*, face, ear, and fingerprint, to make a more reliable recognition.

CHAPTER 3

Fully Automated Face Normalization and Single Sample Face Recognition in Unconstrained Environments

Single sample face recognition have become an important problem because of the limitations on the availability of gallery images. In many real-world applications such as passport or driver license identification, there is only a single facial image per subject available. The variations between the single gallery face image and the probe face images, captured in unconstrained environments, make the single sample face recognition even more difficult.

Although face recognition has been a challenging topic in computer vision for the past few decades, most of the attention was focused on recognition based on face images captured in controlled environments. Capturing a face image naturally without controlling the environment, so-called *in the wild* [44, 96], may result in images with different illumination, head pose, facial expressions, and occlusions. The accuracy of most of the current face recognition systems drops significantly in the presence of these variations, specially in case of pose and illumination variations [20, 21].

Building deterministic or stochastic face models is a challenging task due to the face variations. However, normalization can be used in a preprocessing step to reduce the effect of these variations and pave the way for building face models. Pose variations are considered to be one of the most challenging issues in face recognition. Due to the complex non-planar geometry of the face, the 2D visual appearance significantly changes with variations in the viewing angle. These changes are often more significant than the variations of innate characteristics, which distinguish individuals [19].

In this chapter, we propose a fully automated face identification system suitable for images captured in unconstrained environments¹. The system is robust to pose and illumination variations, which usually affect images captured in unconstrained environments. The system includes a face normalization method based on an enhanced active appearance model approach. We propose a novel initialization technique for AAM, which results in significant improvements in its fitting to non-frontal poses and makes the normalization process robust and fast. Our AAM is trained using face images in-the-wild, which cover a vast range of illumination, pose and expression variations.

In contrast with majority of the algorithms encountered in the literature, the proposed normalization algorithm is fully automatic and handles a continuous range of poses, *i.e.*, it is not restricted to any predetermined pose angles. Relying on the competence of our algorithm in normalizing the face images, we can assume that the face images are properly aligned. This alignment allows us to use corresponding local feature descriptors such as Histogram of Oriented Gradients (HOG) [98] for feature extraction, which makes the system robust against illumination variations. In

¹This work appeared in a journal paper [97].

addition, we fuse the HOG features with Gabor features using Canonical Correlation Analysis (CCA) to have a more discriminative feature set.

It is worth mentioning that our system is capable of recognizing a face from a non-frontal view and under different illumination conditions using only a single gallery image for each subject. This is important because of its potential applications in many realistic scenarios like passport identification and video surveillance. Experimental results performed on the FERET [42], CMU-PIE [43] and Labeled Faces in the Wild (LFW) [44] databases verify the effectiveness of our proposed method, which outperforms the above-mentioned state-of-the-art algorithms.

This chapter is organized as follows: Section 3.2 describes our face normalization technique. Section 3.1 presents related work from the literature. Section 3.3 describes the feature extraction and fusion approaches used in the proposed system. The implementation details and experimental results are presented in Section 3.4. Finally, Section 3.5 concludes the chapter.

3.1 Related Work in Pose-Invariant Face Recognition

The Active Appearance Models (AAMs) proposed by Cootes *et al.* [99, 100], have been used in face modeling for recognition. After fitting the model to a face image, either the model parameters, the location of the landmarks, or the local features extracted at the landmarks are used for face recognition [101–104] or facial expression analysis [105–109]. For face recognition, Guillemaut *et al.* [22] and Heo and Savvides [23] proposed using the normalized face images created by warping the face images

into the frontal pose. Gao *et al.* [24] improved the performance of this technique using a modified piecewise affine warping. None of these methods, however, is fully automatic and they require a manual labeling or manual initialization.

Chai *et al.* [25] assumed that there is a linear mapping between a non-frontal face image and the corresponding frontal face image of the same subject under the same illumination. They create a virtual frontal view by first partitioning the face image into many overlapped local patches. Then, a local linear regression (LLR) technique is applied to each patch to predict its corresponding virtual frontal view patch. Finally, the virtual frontal view is generated by integrating the virtual frontal patches. Li *et al.* [26] proposed a similar patch-based algorithm; however, they measured the similarities of the local patches by correlations in a subspace constructed by Canonical Correlation Analysis. Du and Ward [27] proposed a similar method based on the facial components. Unlike [25] and [26], where the face image is partitioned into uniform blocks, the method in [27] divides it into the facial components, *i.e.*, two eyes, mouth and nose. The virtual frontal view of each component is estimated separately, and finally the virtual frontal image is generated by integrating the virtual frontal components. The common drawback of these three patch-based approaches, [25–27], is that the head pose of the input face image needs to be known. Moreover, these methods require a set of prototype non-frontal face patches, which are in the same pose as the input non-frontal faces; hence, they cannot handle a continuous range of poses and are restricted to a discrete set of predetermined pose angles.

Blanz and Vetter [28] proposed a face recognition technique that can handle variations in pose and illumination. In their method, they derive a morphable face model by transforming the shape and texture of example prototypes into a vector space rep-

resentation. New faces at any pose and illumination are modeled by forming linear combinations of the prototypes. The morphable model represents shapes and textures of faces as vectors in a high-dimensional space. The knowledge of face shapes and textures is learned from a set of textured 3D head scans. This method requires a set of manually annotated landmarks for initialization and the optimization process often converges to local minima due to a large number of parameters, which need to be tuned. Breuer *et al.* [29] presented an automatic method for fitting the 3D morphable model; however, their method seems to have a high failure rate [33].

Castillo and Jacobs [30] used the cost of stereo matching as a measure of similarity between two face images in different poses. This method does not construct a 3D face or a virtual frontal view; however, using stereo matching, it finds the correspondences between pixels in the probe and gallery images. This method requires manual specification of feature points and in case of automatic feature matching, it is fallible in scenarios where an in-plane rotation is present between the image pair.

The method proposed by Sarfraz and Hellwich [31] handles the pose variations for face recognition by learning a linear mapping from the feature vector of a non-frontal face to the feature vector of the corresponding frontal face. However, their assumption of the mapping being linear seems to be overly restrictive [33].

Asthana *et al.* [33] used several AAMs each of which covering a small range of pose variations. All these AAMs are fitted on the query face image and the best fit is selected. The frontal view is then synthesized using the pose-dependent correspondences between 2D landmark points and 3D model vertices. Mostafa *et al.* [34, 35] constructed 3D face shapes from stereo pair images. These 3D shapes are used to synthesize virtual 2D views in different poses, *e.g.*, frontal view. A 2D probe image

is matched with the closest synthesized images using the local binary pattern (LBP) features [110]. The drawback of this method is the need for stereo images. In order to solve this problem, the authors developed another method where the 3D shapes are constructed using only a frontal view and a generic 3D shape created by averaging several 3D face shapes.

Sharma *et al.* [36] proposed the Discriminant Multiple Coupled Latent Subspace method for pose-invariant face recognition. They propose to obtain pose-specific representation schemes so that the projection of face vectors onto the appropriate representation scheme will lead to correspondence in the common projected space, which facilitates direct comparison. They find the sets of projection directions for different poses such that the projected images of the same subject in different poses are maximally correlated in the latent space. They claim that the discriminant analysis with artificially simulated pose errors in the latent space makes it robust to small pose errors due to subjects incorrect pose estimation.

De Marsico *et al.* [37] proposed a face recognition approach, called “FACE”, in which an unknown face is identified based on the correlation of local regions from the query face and multiple gallery instances, that are normalized with respect to pose and illumination, for each subject. For pose normalization, the facial landmarks are first located by an extension of the active shape model [111] and then the in-plane face rotation is normalized using the locations of the eye centers. The rows in the best exposed half of the face are then stretched to a constant length. Then, the other side of the face image is reconstructed by mirroring the first half. The illumination normalization is performed using the Self-Quotient Image (SQI) algorithm [112], in

which the intensity of each pixel is divided by the average intensity of its $k \times k$ square neighborhood.

Ho and Chellappa [38] proposed a patch-based method for synthesizing the frontal view from a given nonfrontal face image. In this method, the face image is divided into several overlapping patches, and a set of possible warps for each patch is obtained by aligning it with frontal faces in the training set. The alignments are performed using an extension of the Lucas-Kanade image registration algorithm [113,114] in the Fourier domain. The best warp is chosen by formulating the optimization problem as a discrete labeling algorithm using a discrete Markov random field and a variant of the belief propagation algorithm [115]. Each patch is then transformed to the frontal view using its best warp. Finally, all the transformed patches are combined together to create a frontal face image. A shortcoming of this method is that they divide both frontal and non-frontal images into the same regular set of local patches. This division strategy results in the loss of semantic correspondence for some patches when the pose difference is large; therefore, the learnt patch-wise affine warps may lose practical significance.

Yi *et al.* [39] proposed an approach for unconstrained face recognition that is robust against pose variations. A 3D deformable model is generated and a fast 3D model fitting algorithm is proposed to estimate the pose of the face image. Then, a set of Gabor filters is transformed according to the pose and shape of the face image for feature extraction. Finally, Principal Component Analysis (PCA) is applied on the Gabor features to eliminate the redundancies, then, the dot product is used to compute the similarity between the feature vectors.

Most recently, Guo *et al.* [40] extended the Linear Discriminant Analysis (LDA) approach to multi-view scenarios. Multi-view Linear Discriminant Analysis (MiLDA) is a subspace learning framework for multi-view data analysis based on graph embedding [116]. The authors introduced a new measure of distance between projected vertex sets of intrinsic graphs to mitigate the effect of the differences between views and preserve the intrinsic graphs. This distance is defined as the weighted sum of squared Euclidean distances between every cross-view data pair in two graph embedding models. Having sets of multi-view data, MiLDA aims to find a common subspace of higher discriminability between classes. The transformed feature vectors in the common subspace are classified using a nearest neighbor classifier.

Most recently, deep learning has attracted the interest of many researchers, such that deep networks has become the model of choice for unconstrained face recognition. Various deep learning architectures such as deep neural networks, convolutional deep neural networks, deep belief networks and recurrent neural networks have been applied to fields like computer vision, automatic speech recognition, natural language processing, audio recognition and bioinformatics. A deep neural network is an artificial neural network with multiple hidden layers between the input and output layers. Deep learning is especially suitable for dealing with large training sets [117].

One of the well-known deep learning based algorithms, *DeepFace* [117], presented by Facebook Research, focuses on solving the unconstrained face recognition problem by learning a set of features in the image domain. They propose a multi-stage approach that aligns faces to a general 3D shape model. A multi-class network is trained to perform the face recognition task on over four thousand identities. The authors also experimented with a so-called Siamese network where they directly optimize the

L1-distance between two face features. Their best performance on LFW (97.35%) comes from combination of three deep neural network structures with more than 120 million parameters, which are trained using a large private training database of 4.4 million labeled faces from 4,030 subjects.

Sun *et al.* [118,119] propose a compact and therefore relatively cheap to compute network. They use an ensemble of 200 of these networks, each operating on a different face patch. In [119], they present DeepID2+, which uses both PCA and a Joint Bayesian model [120] that effectively correspond to a linear transform in the embedding space. Their method does not require explicit 2D/3D alignment and they reach an accuracy of 99.47% on LFW database. Note that they use a private database of 202,595 images from 10,117 subjects.

Another deep learning based face recognition algorithm is recently proposed by Google [121]. This method called FaceNet presents a unified system for face verification, identification and clustering. It is based on learning a Euclidean embedding per image using a deep convolutional network. The network is trained such that the squared L2 distances in the embedding space directly correspond to face similarity. FaceNet is trained using an enormous private database of 260 million images of 8 million subjects. The images are tight crops of the face area, so no 2D or 3D alignment, other than scale and translation is performed. The training time for this network is reported to be between 1,000 to 2,000 hours. FaceNet reports the state-of-the-art accuracy of 99.63% on LFW.

All the above-mentioned deep learning based methods use large number of images per subject. In contrast, in a recent publication, Gao *et al.* [41] presented a face recognition approach based on deep learning using a single training sample per per-

son. In [41], the authors propose a supervised auto-encoder to build the deep neural network by training a nonlinear feature extractor at each layer. After the layer-wise training of each building block and building a deep architecture, the output of the network is used for face recognition. One of the shortcomings of this method is the manual cropping and alignment of the face images. It is also tested only on near frontal face images.

3.2 Preprocessing for Face Normalization

As stated in [20], “the variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity”. Pose variations cause major problems in real-world face recognition systems. In an unconstrained environment, there are usually in-plane and out-of-plane face rotations. In order to achieve better recognition results, we preprocess the facial images to handle these variations.

In this section, we present a pose normalization technique based on piece-wise affine warping, which can normalize both in-plane and out-of-plane pose changes. The warping is applied on triangular pieces determined by enhanced active appearance models described below. The overall process is illustrated in Fig. 3.1. In the following subsections, we describe the fitting and warping process of the active appearance models and present a novel initialization technique for AAMs, which results in significant improvement in the fitting accuracy.

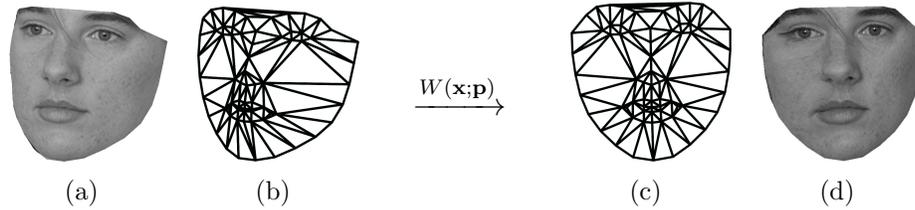


Figure 3.1: Warping the face image into the base (frontal) mesh. (a) Rotated face image. (b) Fitting mesh corresponding to the rotated face image. (c) Triangulated base (frontal) mesh, s_0 . (d) Face image warped into the base mesh.

3.2.1 Active Appearance Models and Piece-wise Affine Warping

Active appearance models have been widely used in pattern recognition research [99]. Face modeling has been the most ubiquitous application of AAMs. Given the model parameters, AAMs reconstruct a specific face via statistical models of shape and appearance. The model parameters are obtained by maximizing the match between the model instance and the face by fitting the AAM to the input face image.

The shape, S , of an AAM, is defined by the coordinates of a set of landmarks on the face. Learning the shape model requires annotating these landmarks on a training set of face images, then, applying principal component analysis (PCA) to these shapes. The shape model of a specific face is expressed as a base shape, s_0 , plus a linear combination of the n shape eigenvectors, s_i , $i = 1, \dots, n$, that correspond to the n largest eigenvalues:

$$S = s_0 + \sum_{i=1}^n p_i s_i, \quad (3.1)$$

where p_i s are the shape parameters.

The appearance of an AAM is defined within the base shape, s_0 , which means that learning the appearance model requires removing the shape variations. The appearance of an AAM is an image $A(\mathbf{x})$, where \mathbf{x} is the set of pixels inside the

base mesh s_0 ($\mathbf{x} \in s_0$). In order to obtain the appearance model, PCA is applied on these shape-free images. The appearance model of a specific face is expressed as a base appearance, a_0 , plus a linear combination of m appearance eigenvectors, a_i , $i = 1, \dots, m$ corresponding to the m largest eigenvalues:

$$A(\mathbf{x}) = a_0(\mathbf{x}) + \sum_{i=1}^m q_i a_i(\mathbf{x}), \quad (3.2)$$

where q_i s are the appearance parameters.

The shape and appearance parameters for a given face image are obtained in the process of AAM fitting. Project-Out Inverse Compositional (POIC) algorithm [122] and Simultaneous Inverse Compositional (SIC) algorithm [123] are two well-known algorithms for AAM fitting. SIC performs significantly better than POIC on images of subjects that are not included in the training. However, the computational cost of SIC is very high [124]. Recently, Tzimiropoulos and Pantic [125] proposed Fast-SIC, which reduces the computational complexity of SIC. In our experiments, we use the Fast-SIC optimization technique for fitting the AAM.

Let $\mathbf{p} = \{p_1, p_2, \dots, p_n\}$ be the set of shape parameters obtained from AAM fitting. As shown in Fig. 3.1, a piecewise affine warp, $W(\mathbf{x}; \mathbf{p})$, transfers a face instance into the base shape. After fitting the AAM, each triangle in the AAM mesh has a corresponding triangle in the base (frontal) mesh. Using the coordinates of the vertices in the AAM mesh, the coordinates of the corresponding triangle in the base mesh are computed from the current shape parameters \mathbf{p} using Eq. 3.1. Using the coordinates of the vertices in corresponding triangles, we compute an affine transformation for each triangle, such that the vertices of the first triangle map to the vertices of the second triangle [122]. For every pixel inside the target triangle in the frontal mesh, the corresponding location in the AAM mesh is calculated. Then,

the value of this pixel is obtained based on a nearest neighbor interpolation in the calculated location. This process is applied to all the triangles and the synthesized frontal face is created in the base mesh s_0 . In our approach, we use the warped face within the base shape as the normalized face image. This step results in a shape-free facial appearance ($\mathbf{p} = 0$), which allows face identification to be performed in the coordinates of the base shape.

3.2.2 Proposed AAM Initialization

Despite the popularity of the AAMs, there is no guarantee for obtaining correct fitting, specially when the images are not in near-frontal pose. As mentioned before, both POIC and SIC algorithms use the base mesh s_0 , when $\mathbf{p} = 0$, as the initial shape model. The base mesh represents the mean shape of all the training samples, which is usually in frontal pose as shown in Fig. 3.2(a). Typical fitting methods use a face detection algorithm to find the face and then scale the base mesh to the size of the detected face and use it as the initial shape model. However, in semi-profile poses, this initialization sometimes falls out of face region and if the algorithm starts with this mesh, it may not converge to the actual shape. Fig. 3.2(b) shows the initialization of the base mesh on a sample face image. The result of the AAM fitting using Fast-SIC method after 100 iterations is shown in Fig. 3.2(c). Fig. 3.2(d) shows the result of the piecewise affine warping into the base mesh, which is supposed to represent the normalized face image.

For better initialization, in this work, we use the flexible mixture of parts proposed in [126] to automatically initialize the locations of the landmarks. Every facial landmark with its predefined neighborhood patch is defined as a *part*. The landmarks

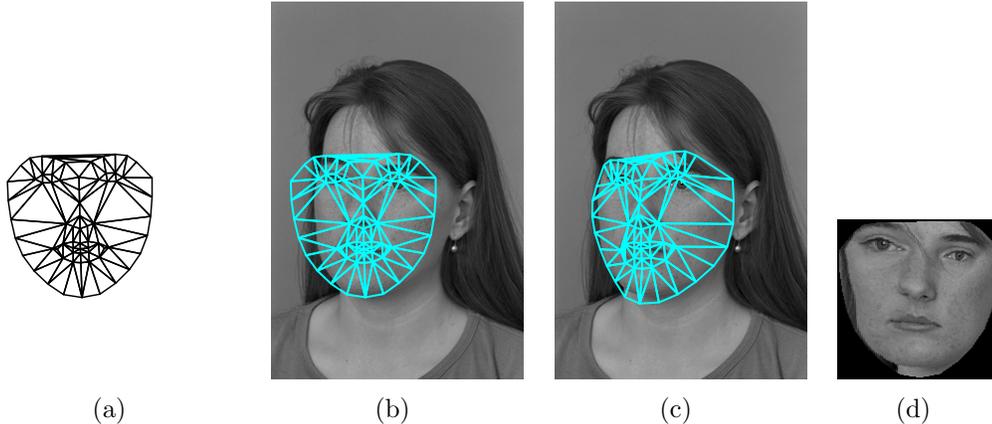


Figure 3.2: Initialization problem in AAM fitting. (a) Initial shape used in POIC and SIC algorithms $\mathbf{p} = 0$. (b) Initialization of the base mesh on the target face image. (c) Fitting result of the Fast-SIC method after 100 iterations. (d) Result of the piecewise affine warping into the base mesh.

on a face define a mixture of these parts, which are used to build a tree graph to represent the spatial structure of the landmarks. Due to the topological changes caused by pose variations, Zhu and Ramanan [127] proposed a model based on mixture of trees with a shared pool of parts for face detection, pose estimation, and landmark localization. We modified this approach to initialize the landmark locations for our AAM.

Let I denote the facial image, in which $l_i = (x_i, y_i)$ is a landmark location in part i . For each viewpoint t , we define a tree graph $G_t = (V_t, E_t)$, where $V_t \subseteq V$, and V is the shared pool of parts. A configuration of parts $L = \{l_i : i \in V\}$ is scored as:

$$S(I, L, t) = \sum_{i \in V_t} \omega_i^{t_i} \cdot \phi(I, l_i) + \sum_{i, j \in E_t} \lambda_{i, j}^{t_i, t_j} \cdot \psi(l_i, l_j) + \alpha^t. \quad (3.3)$$

The first term in Equation (3.3) is an appearance evaluation function, indicating how likely a landmark is in an aligned position. $\phi(I, l_i)$ is a feature vector extracted from a neighborhood centered at l_i , where in our experiments, we use HOG features [98]; and $\omega_i^{t_i}$ is a template for part i tuned for the mixture for viewpoint t_i . The second term is the shape deformation cost, *i.e.*, computes the cost associated with

the relative positions of neighboring landmarks. $\lambda_{i,j}^{t_i,t_j}$ is used to encode parameters of rest location and rigidity, controlling the shape displacement of part i relative to part j defined as $\psi(l_i, l_j) = [dx \ dx^2 \ dy \ dy^2]^T$, where $dx = x_i - x_j$ and $dy = y_i - y_j$. Finally, the last term α^t is a scalar bias associated with the mixture for viewpoint t .

We seek to maximize $S(I, L, t)$ over the landmark locations, L , and viewpoint, t , and find the best configuration of parts. Since each mixture is a tree-structured graph, maximization can be efficiently done with dynamic programming [128] to find the global optimum solution.

Learning: To learn the model, a fully supervised scenario using labeled positive and negative samples is used. Assume that $\{I_n, L_n, t_n\}$ and $\{I_n\}$ denote the n^{th} positive and negative samples, respectively. The scoring function, Equation (3.3), is linear in its parameters. Concatenating the parameters, we can write $S(I, k) = \mu \cdot \Phi(I, k)$, where $\mu = (\omega, \alpha)$ and $k_n = (l_n, t_n)$. Now, learning the model can be formulated as:

$$\begin{aligned} \arg \min_{\mu, \xi_n \geq 0} \quad & \frac{1}{2} \|\mu\| + C \sum_n \xi_n & (3.4) \\ \text{s.t. } \forall n \in \text{pos} \quad & \mu \cdot \Phi(I_n, k_n) \geq 1 - \xi_n \\ \forall n \in \text{neg}, \forall k \quad & \mu \cdot \Phi(I_n, k) \leq -1 + \xi_n. \end{aligned}$$

Zhu and Ramanan [127] trained their model in 13 viewpoints spanning 180° with sampling every 15° . They used images from CMU Multi-PIE face database [129] with 68 facial landmarks in poses between -45° and $+45^\circ$, and 39 facial landmarks in poses $\pm 60^\circ$, $\pm 75^\circ$ and $\pm 90^\circ$. In order to cover the whole range of pose variations, we used the model in [127], which uses 900 positive samples from Multi-PIE, and

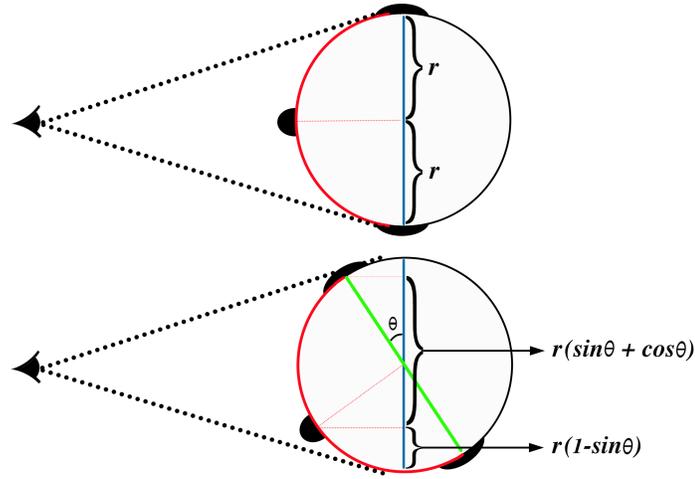


Figure 3.3: Top view perspective of a human head in frontal and rotated poses.

1218 negative samples from INRIA Person database [98], including outdoor scenes with no people in them.

AAM Initialization: In the testing stage, since we use the landmarks for the initialization of our AAM, in cases of detecting a mixture with 39 vertices (landmarks), we estimate the location of the remaining 29 landmarks based on the information obtained from the topology of the facial landmarks in the viewpoint corresponding to the detected mixture. Without loss of generality, if we assume that the top view of a human head is a circle with radius r , Fig. 3.3 shows the visible area of the left and right sides of the face in frontal and rotated poses. As illustrated, the ratio between the visible areas in two sides of the face is

$$\gamma = \frac{1 - \sin(\theta)}{\sin(\theta) + \cos(\theta)}, \quad (3.5)$$

θ being the pose angle.

In cases where the landmark localization stage selects a mixture of 39 vertices, these landmarks are fitted on the best exposed half of the face. The selected mixture provides an estimation of the pose angle, θ . γ , obtained from the Equation (3.5),

is used as a scaling factor to roughly calculate the location of the landmarks on the other half of the face by relatively mirroring the current landmarks across the face mid-line.

The landmark localization algorithm based on the flexible mixture of parts works very well in finding the contour of the face but it is not accurate enough in the more detailed regions such as the eyes or the mouth. Fig. 3.4(a) shows the result of this method on a sample face image.

In this chapter, instead of using the base mesh, s_0 , we create the initial shape model for AAM using the estimated landmarks obtained from the flexible mixture of parts model. Fig. 3.4(b) shows the triangularized initial mesh using these landmarks. The result of the AAM fitting using Fast-SIC method after only five iterations is shown in Fig. 3.4(c). It is clear from Fig. 3.4(c) that, using this initialization, the fitting is much more accurate. Fig. 3.4(d) shows the result of the piecewise affine warping into the base mesh, which in comparison with Fig. 3.2(d), provides a better representation of the face. In the rest of this chapter, we use these warped images as the normalized face images.

3.3 Feature Extraction and Fusion

The face images of an individual subject are similar to each other and different from the face images of other subjects. However, face images of an individual are not exactly the same either. The question is how these changes are different from the changes between different subjects. The proper alignment of the face images made possible by the proposed normalization technique reduces the variations between fea-

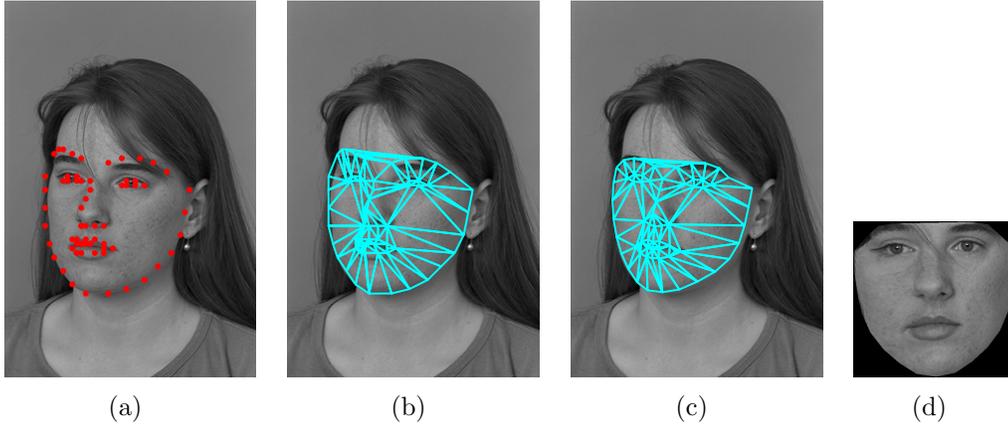


Figure 3.4: Our proposed initialization method for AAM fitting. (a) Estimated landmarks using the flexible mixture of trees. (b) Triangularization of the initial mesh created by the estimated landmarks. (c) Fitting result of the Fast-SIC method after only 5 iterations. (d) Result of the piecewise affine warping into the base mesh.

ture vectors of the samples of the same subject, which facilitates building a more accurate face model. In this section we describe the feature extraction techniques as well as the feature fusion method employed in our approach.

3.3.1 Feature Extraction

In our experiments, the normalized face images are resized to 120×120 pixels. We use two different techniques to extract features from the normalized images. These techniques include Gabor wavelet features [72, 130] and Histogram of Oriented Gradients (HOG) [98].

Since the face images are aligned, we can make use of local descriptors such as the histograms of oriented gradients (HOG) [98] for feature extraction. Here, we extract the HOG features in 4×4 cells for nine orientations. We use the UOCTTI variant for the HOG presented in [131]. UOCTTI variant computes both directed and undirected gradients as well as a four dimensional texture-energy feature, but

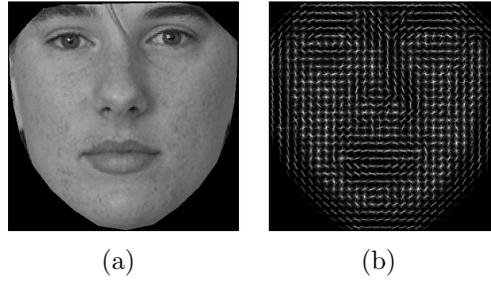


Figure 3.5: Histogram of Oriented Gradients (HOG) features in 4×4 cells.

projects the result down to 31 dimensions, 27 dimensions corresponding to different orientation channels and 4 dimensions capturing the overall gradient energy in square blocks of four adjacent cells. Fig. 3.5(b) shows the HOG features extracted from a sample face image in Fig. 3.5(a)².

On the other hand, we employ forty Gabor filters in five scales and eight orientations. The most important advantage of Gabor filters is their invariance to rotation, scale, and translation. Furthermore, they are robust against photometric disturbances, such as illumination change and image noise [72, 133]. Since the adjacent pixels in an image are usually correlated, the information redundancy can be reduced by downsampling the feature images that result from Gabor filters [130]. In our experiments, the feature images are downsampled by a factor of five. Fig. 3.6 shows the Gabor features for the normalized face image in Fig. 3.5(a). The dimensionality of both Gabor and HOG feature vectors are reduced using principal component analysis (PCA) [75].

²VLFeat open source library is used to extract and visualize the HOG features [132].

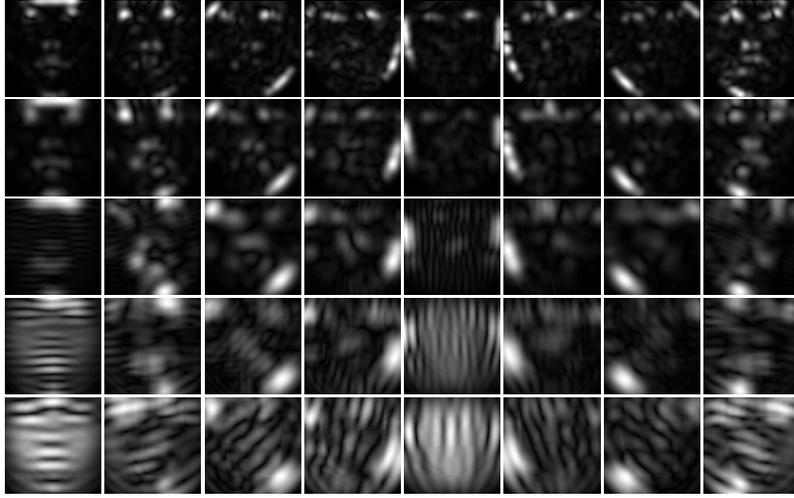


Figure 3.6: Gabor features in five scales and eight orientations.

3.3.2 Feature Fusion Using Canonical Correlation Analysis

We combine the two feature vectors to obtain a single feature vector, which is more discriminative than any of the input feature vectors. This is achieved by using a feature fusion technique based on Canonical Correlation Analysis (CCA) [50].

Canonical correlation analysis has been widely used to analyze associations between two sets of variables. Suppose that $X \in \mathbb{R}^{p \times n}$ and $Y \in \mathbb{R}^{q \times n}$ are two matrices, each contains n training feature vectors from two different modalities. In other words, there are n samples for each of which $(p + q)$ features have been extracted. Let $S_{xx} \in \mathbb{R}^{p \times p}$ and $S_{yy} \in \mathbb{R}^{q \times q}$ denote the within-sets covariance matrices of X and Y and $S_{xy} \in \mathbb{R}^{p \times q}$ denote the between-set covariance matrix (note that $S_{yx} = S_{xy}^T$). The overall $(p + q) \times (p + q)$ covariance matrix, S , contains all the information on associations between pairs of features:

$$S = \begin{pmatrix} cov(x) & cov(x, y) \\ cov(y, x) & cov(y) \end{pmatrix} = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix}. \quad (3.6)$$

However, the correlation between these two sets of feature vectors may not follow a consistent pattern, and thus, understanding the relationships between these two sets

of feature vectors from this matrix is difficult [134]. CCA aims to find the linear combinations, $X^* = W_x^T X$ and $Y^* = W_y^T Y$, that maximize the pair-wise correlations across the two data sets:

$$\text{corr}(X^*, Y^*) = \frac{\text{cov}(X^*, Y^*)}{\text{var}(X^*) \cdot \text{var}(Y^*)}, \quad (3.7)$$

where $\text{cov}(X^*, Y^*) = W_x^T S_{xy} W_y$, $\text{var}(X^*) = W_x^T S_{xx} W_x$ and $\text{var}(Y^*) = W_y^T S_{yy} W_y$. Maximization is performed using Lagrange multipliers by maximizing the covariance between X^* and Y^* subject to the constraints $\text{var}(X^*) = \text{var}(Y^*) = 1$. The transformation matrices, W_x and W_y , are then found by solving the eigenvalue equations [134]:

$$\begin{cases} S_{xx}^{-1} S_{xy} S_{yy}^{-1} S_{yx} \hat{W}_x = \Lambda^2 \hat{W}_x \\ S_{yy}^{-1} S_{yx} S_{xx}^{-1} S_{xy} \hat{W}_y = \Lambda^2 \hat{W}_y \end{cases}, \quad (3.8)$$

where \hat{W}_x and \hat{W}_y are the eigenvectors and Λ^2 is the diagonal matrix of eigenvalues or squares of the *canonical correlations*. The number of non-zero eigenvalues in each equation is $d = \text{rank}(S_{xy}) \leq \min(n, p, q)$, which will be sorted in decreasing order, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$. The transformation matrices, W_x and W_y , consist of the sorted eigenvectors corresponding to the non-zero eigenvalues. $X^*, Y^* \in \mathbb{R}^{d \times n}$ are known as canonical variates. For the transformed data, the sample covariance matrix defined in Eq. (3.6) will be of the form:

$$S^* = \left(\begin{array}{cccc|cccc} 1 & 0 & \dots & 0 & \lambda_1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \lambda_2 & \dots & 0 \\ \vdots & & \ddots & & \vdots & & \ddots & \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & \lambda_d \\ \hline \lambda_1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & & \ddots & & \vdots & & \ddots & \\ 0 & 0 & \dots & \lambda_d & 0 & 0 & \dots & 1 \end{array} \right).$$

The above matrix shows that the canonical variates have nonzero correlation only on their corresponding indices. The identity matrices in the upper left and lower right corners show that the canonical variates are uncorrelated within each data set.

As defined in [50], feature-level fusion is performed either by concatenation or summation of the transformed feature vectors:

$$Z_1 = \begin{pmatrix} X^* \\ Y^* \end{pmatrix} = \begin{pmatrix} W_x^T X \\ W_y^T Y \end{pmatrix} = \begin{pmatrix} W_x & 0 \\ 0 & W_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix}, \quad (3.9)$$

or

$$Z_2 = X^* + Y^* = W_x^T X + W_y^T Y = \begin{pmatrix} W_x \\ W_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix}, \quad (3.10)$$

where Z_1 and Z_2 are called the Canonical Correlation Discriminant Features (CCDFs). In this chapter, we use the concatenation method defined in Eq. 3.9. The fused feature vectors (Z) are used to build face models following the face modeling approach presented in [135]. The query sample is then classified as the nearest neighbor based on the Euclidean distance between the query's model and the models in the gallery.

3.4 Experimental Setup and Results

3.4.1 Experimental Setup: AAM Training

In our experiments, we trained the AAMs using in-the-wild databases. For this purpose, we use three of the training sets provided for *300 Faces in-the-Wild Challenge* [136]. These images contain large variations in pose, expression, illumination and occlusion. These databases are Labeled Face Parts in-the-Wild (LFPW) [137], Helen [138], and a database collected by Intelligent Behavior Understanding Group (IBUG) [136]. LFPW database consists of 1,035 annotated images collected from *Yahoo*, *Google*, and *Flickr*. HELEN database contains 2,330 annotated faces downloaded from *Flickr*. Most of the expressions in these two databases are neutral and smile. Therefore, IBUG database, which contains 135 highly expressive face images, is added to include a larger variety of facial expressions. In total, 3,500 in-the-wild face images are used to train the AAM. Note that these databases are only used for training the AAM and since they are not labeled, they are not employed in evaluating the recognition accuracy of our system.

3.4.2 Normalization Performance

Here we discuss the self-occlusion problem in case of large pose variations. Fig. 3.7 shows a semi-profile face image with a large pose angle, where only a small fraction of the right side of the face is visible. According to Equation (3.5), for instance in case of a 60° pose angle, the visible area of the *occluded* side of the face shrinks by a factor of $1 - \sin(60^\circ) = 0.13$, while for the other side of the face, the visible area



Figure 3.7: (a) Self-occluded face image with 60° rotation. (b) Normalized face image with a stretched half face.

stretches by a factor of $\sin(60^\circ) + \cos(60^\circ) = 1.36$. The ratio between these two areas is less than 10%.

In the proposed normalization technique, after fitting the AAMs, the face image is warped into the base frontal mesh. Since the areas of the left and right halves of the base mesh have the same size, the occluded side of the face will be over-sampled (stretched) in the process of piecewise warping. In this case, a small misalignment in the AAM fitting may cause a large error in the warped face image, which will result in a distorted half-face. Even if a semi-profile face is perfectly fitted, the warped frontal view will still be distorted due to the stretching [24]. This phenomenon is clearly seen in Fig. 3.7, which has a 60° of face rotation. In the normalization process, the right half of the face, *i.e.*, the occluded half, is stretched, which results in a distorted half face. This distortion will have negative effect on the recognition accuracy. Therefore, in these cases, we only use half of the face that corresponds to the visible side and ignore the distorted half.

In order to automatically distinguish between the well-normalized and the distorted half-faces in semi-profile images, we trained a two class minimum distance

classifier using Discrete Cosine Transform (DCT) features. This classifier is trained using 400 well-normalized half-faces generated from frontal faces in the *ba* set of FERET database [42], and 400 distorted half-faces randomly chosen from the *hl* and *hr* sets of FERET database, which include poses at -67.5° and $+67.5^\circ$ rotations. After face normalization, this classifier uses the DCT features extracted from each half of the face to determine whether it is well-normalized or distorted. Based on the outcome, we either use only the well-normalized side or the whole face for identification. The complexity of this step is negligible not only because DCT features are very simple to calculate [139, 140], but also because the decision is made based on the Euclidean distances from the centroids of only two classes.

In this following, we present several sets of experiments to demonstrate the performance of our proposed face normalization and recognition system. We conduct three sets of experiments, on three databases: Facial Recognition Technology (FERET) [42], CMU-PIE [43] and Labeled Faces in the Wild (LFW) [44].

3.4.3 Experiments on FERET Database

The first set of experiments was performed on the FERET b-series database [42]. It contains 2,200 face images for 200 subjects, *i.e.*, eleven images per subject. Three of the images include frontal faces with different facial expressions and illuminations. These images are letter coded as *ba*, *bj*, and *bk*. The other eight images are faces in different poses with $+60^\circ$, $+40^\circ$, $+25^\circ$, $+15^\circ$, -15° , -25° , -40° , and -60° degrees of rotation. These images are letter coded as *bb*, *bc*, *bd*, *be*, *bf*, *bg*, *bh*, and *bi*, respectively. Fig. 3.8 shows these images for a sample subject along with the results of the proposed



Figure 3.8: Face images of a sample subject from FERET b-series database (upper row), and their normalized faces (lower row).

Table 3.1: Face recognition rates of different approaches in confrontation with different face distortions on the FERET database. The frontal face images with neutral expression, labeled *ba*, are used for training.

Method	Face Alignment	Trained on FERET	bb +60°	bc +45°	bd +25°	be +15°	bf -15°	bg -25°	bh -45°	bi -60°	bj expr.	bk illum.
LGBP [141]	Automatic	No	-	51.0	84.0	96.0	98.0	91.0	62.0	-	-	-
PAN [24]	Manual	Yes	44.0	81.5	93.0	97.0	98.5	91.5	78.5	52.5	-	-
Asthana [32]	Manual	Yes	32.5	74.0	95.5	98.5	98.0	93.0	87.0	48.0	-	-
Sarfraz [31]	Automatic	Yes	78.0	89.0	97.0	98.6	100	89.7	92.4	84.0	-	-
3DPN [33]	Automatic	No	-	91.9	97.0	97.5	98.5	98.0	90.5	-	-	-
CLS [36]	Manual	Yes	70.0	82.0	90.0	95.0	96.0	94.0	85.0	79.0	-	-
FRAD [34]	Automatic	No	-	82.35	98.47	98.97	100	97.98	87.5	-	-	-
PIMRF [38]	Automatic	No	-	91.5	96.5	98.5	98.0	97.3	91.0	-	-	-
PAF [39]	Automatic	No	93.75	98.0	98.5	99.25	99.25	98.5	98.0	93.75	-	-
FAR [142]	Automatic	No	-	96.0	100	100	100	99.0	96.5	-	-	-
Proposed Method	Automatic	No	91.5	96.0	100	100	100	100	99.0	93.0	99	100

normalization approach. Note that, our proposed normalization approach is fully automatic and no manual adjustments were needed in any of the 2,200 samples.

In our experiments, only a single image, *i.e.*, the frontal face image with neutral expression labeled *ba*, is used for enrollment and the remaining ten images with different poses, expressions, and illumination conditions are used for testing. Table 3.1 shows the accuracy of our proposed method for each set in comparison with previous methods in the literature. Note that, the proposed method is evaluated with all the pose angles presented in FERET database. However, only five of the previous



Figure 3.9: Symmetry issue in FERET database. The upper row includes the sample images at $+60^\circ$ (bb) and the lower row shows the corresponding images at -60° (bi).

methods used the images from all the pose angles [24, 31, 32, 36, 39], and the other studies [33, 34, 38, 141, 142] only used a subset of the pose angles.

The recognition rates for $+60^\circ$ and $+45^\circ$ poses (bb & bc) are less than those for -60° and -45° poses (bi & bh). The reason goes back to the setup of the FERET database in which the positive rotations are slightly more than the negative ones. Fig. 3.9 shows examples of this difference. The upper row shows the sample images at $+60^\circ$ (bb) and the lower row shows the corresponding images at -60° (bi) for the same subjects.

As seen in Table 3.1, our proposed algorithm outperforms the previous algorithms [24, 31–34, 36, 38, 39, 141, 142] in most of the pose angles. In case of high rotations ($\pm 60^\circ$), the recognition rates are comparable with the best method PAF [39]. It is worth mentioning that some of the methods in Table 3.1 are not fully automatic and they require manual intervention, some of these methods also use the same database (FERET) in training their normalization approach. However, our approach is fully automatic and does not use FERET database in training the normalization technique.



Figure 3.10: Face images of a sample subject from CMU-PIE database (upper row), and their normalized faces (lower row).

Note that in [38] and [33], if the face and both eyes are not detected using the cascade classifiers, a Failure to Acquire (FTA) is reported and the image is not included in the test set. However, we tested the recognition rate on all the 200 images of each set and no images were excluded in the evaluation process (no FTA is considered).

3.4.4 Experiments on CMU-PIE Database

The second set of experiments were performed on CMU-PIE database [43]. This database consists of face images taken from sixty eight subjects under thirteen different poses. Similar to the previous methods [25, 30, 31, 33, 34, 38, 141], seven poses are used in our experiments. The frontal pose, labeled *c27*, is used as the gallery image. The probe set consists of six non-frontal poses labeled as *c37* and *c11* (the yaw angle about $\pm 45^\circ$), *c05* and *c29* (the yaw angle about $\pm 22.5^\circ$), and *c07* and *c09* (the pitch angle about $\pm 22.5^\circ$). Fig. 3.10 shows these images for a sample subject along with the results of applying the proposed normalization method to them.

The performance of the proposed system is compared with the state-of-the-art approaches in [25, 30, 31, 33, 34, 36, 38–41, 141]. Table 3.2 shows the outstanding accuracy

Table 3.2: Face recognition rates of different approaches in confrontation with different pose changes on the CMU-PIE database. The frontal faces captured by camera *c27* is used for training.

Method	Face Alignment	Trained on PIE	Gallery Size	c11 −45°	c29 −22.5°	c07 22.5° <i>up</i>	c09 22.5° <i>down</i>	c05 +22.5°	c37 +45°
LGBP [141]	Automatic	No	67	71.6	87.9	78.8	93.9	86.4	75.8
LLR [25]	Manual	No	34	89.7	100	98.5	98.5	98.5	82.4
3ptSMD [30]	Manual	No	34	97.0	100	100	100	100	100
Sarfraz [31]	Automatic	No	68	83.8	86.8	-	-	94.1	89.7
3DPN [33]	Automatic	No	67	98.5	100	98.5	100	100	97.0
CLS [36]	Manual	Yes	34	100	100	100	100	100	100
FRAD [34]	Automatic	No	68	95.6	100	100	100	100	100
PIMRF [38]	Automatic	No	67	97.0	100	98.5	100	100	97.0
PAF [39]	Automatic	No	68	100	100	100	100	100	100
MiLDA [40]	Automatic	No	68	90.30	99.58	-	-	98.73	92.55
SSAE [41]	Manual	Yes	48	-	68.06	71.45	71.96	67.52	-
Proposed Method	Automatic	No	68	100	100	100	100	100	100

of our proposed method for each pose in comparison with these methods. We obtain 100% accuracy in all sets. In our experiment, all the 68 subjects were employed for the evaluations; however, in some of the previous methods, *e.g.*, [33, 38, 141], the probe size is 67, because when their algorithm fails to normalize an image, they do not consider it as a recognition error and exclude that image from the test set. Some methods, in Table 3.2, only used 34 subjects out of the 68, *e.g.*, [25, 30, 36]. Gao *et al.* [41] used 20 subjects for training their proposed deep neural network and the remaining 48 subjects were for evaluation. It is important to note that the deep learning based face recognition algorithm presented in [41] is not robust to pose variations and it is only tested in near frontal poses.

3.4.5 Experiments on LFW Database

Our last experiment is on the Labeled Faces in the Wild (LFW) [44] database. LFW is one of the most challenging databases for evaluating the performance of face verification systems in unconstrained environments. This database contains 13,233



Figure 3.11: Sample images of three subjects from LFW database.



Figure 3.12: Normalized face images corresponding to the ones shown in Fig. 3.11.

face images of 5,749 subjects labeled by their identities. 1,680 of these subjects have more than one face images. The images are collected from *Yahoo! News* in 2002-2003, and have a wide variety of variations in pose, illumination, expression, scale, background, color saturation, focus, etc. Fig. 3.11 shows some sample images from this database and Fig. 3.12 shows the results of the proposed normalization method on these images. It is obvious from the figure that even with the changes in pose, expression, illumination and occlusion, the normalization results are impressive as the faces are precisely detected and aligned.

Table 3.3: Average accuracy of different approaches following the first experiment on LFW database following the image-restricted with label-free outside data protocol. Note that the list is shown in chronological order.

Method	Accuracy
BIF [143]	88.13 ± 0.58
I-LQP [144]	86.20 ± 0.46
PAF [39]	87.77 ± 0.51
Sub-SML [146]	89.90 ± 0.38
VMRS [147]	91.10 ± 0.59
DDML [148]	90.68 ± 1.41
LM3L [149]	89.57 ± 0.02
Proposed Method	91.46 ± 0.51
HPEN [150]	92.80 ± 0.47
TSML [151]	89.80 ± 0.47
SLBFLE [152]	92.97 ± 1.20
MSBSIF-SIEDA [153]	94.63 ± 0.95

In order to compare with a wide range of methods, we evaluated our proposed algorithm in three different experiments. The first experiment follows the directions used in [39, 143, 144] following the image-restricted with label-free outside data protocol. As in [143], the LFW dataset is organized into two disjoint sets: ‘View 1’ is used training whereas ‘View 2’ is used for testing. Although [39, 143, 144] use the aligned version of the faces provided by [145], we use the original version of the LFW database and all face images are aligned using our normalization technique described in Section 3.2. The average accuracy and the standard error of the proposed method and the methods following the same protocol are shown in Table 3.3. Note that we have also included the results of the most recent methods that are presented after our work. The methods are shown in chronological order.

Although LFW is basically designed for metric learning for face verification, De Marsico *et al.* [37] evaluated some of the most popular face recognition algorithms as well as their own method on a subset of this database. This subset is made from the first fifty subjects who have at least eight images. Five of the images are used

Table 3.4: Face recognition rates of different approaches following the second experiment on LFW database.

PCA [75]	ICA [154]	ILDA [155]	SVM [156]	HMLBP [157]	FACE [37]	Proposed Method
37	41	48	45	49	61	87.3

as gallery images and three as probes. We used the same setting to evaluate the performance of our proposed method. Table 3.4 shows the performance of our proposed system in comparison with the Eigenface approach [75], which is based on PCA, Independent Component Analysis (ICA) method proposed in [154], Incremental Linear Discriminant Analysis (ILDA) approach [155], a method using Support Vector Machines (SVM) [156], a recent approach based on Hierarchical Multiscale LBP (HMLBP) [157], and the method called “FACE” proposed in [37], which is the most recent method evaluated using this protocol.

Table 3.4 shows that our proposed system outperforms all the above-mentioned methods including the recent method proposed in [37] with an impressive margin of 26% in the recognition rate. Note that the experiments are performed using the original, not the aligned, version of the LFW database.

Finally, we evaluated our proposed method following the evaluation protocol of [158], which contains both verification and open-set identification scenarios, with a focus at low false accept rates (FARs). Benchmark of Large-scale Unconstrained Face Recognition (BLUFR) [158] is a new protocol to fully exploit all the 13,233 LFW face images for large-scale unconstrained face recognition evaluation. The experimental setting of BLUFR is summarized in Table 3.5. It adopts 10 random trials for training and testing. For each of the 10 trials, the whole LFW database is randomly divided into a training set and a testing set. The training set of each trial includes 1,500 subject, 3,524 images on average, while the testing set contains the remaining

Table 3.5: Overview of the experimental setting for the BLUFR protocol on the LFW database. Numbers are averaged over the 10 trials.

Image Set		# Classes	# Images	# Genuine matches	# Imposter matches
Train		1,500	3,524	85,341	6,122,185
Test	All	4,249	9,708	156,915	46,960,863
	Gallery	1,000	1,000	-	-
	Genuine probe	1,000	4,357	-	-
	Imposter probe	3,249	4,357	-	-

4,249 subjects, 9,708 images on average. Each trial contains about 156,915 genuine matching scores and 46,960,863 impostor matching scores on average for performance evaluation.

The representations of faces in LFW are extracted in the same way as the previous experiment. We use the Joint Bayesian [120] supervised subspace learning method to extract discriminative information for face recognition and further reduce the dimension. Among the leading learning methods, Joint Bayesian has shown to achieve the best accuracies [158, 159]. Then the results are calculated using the standard benchmark toolkit [160]. Fig. 3.13 shows the resulting Receiver Operating Characteristic (ROC) curves for verification and open-set identification at rank 1 for our proposed method and several state-of-the-art methods. Table 3.6 shows the Verification Rate (VR) with FAR=0.1% and rank-1 open set Detection and Identification Rate (DIR) with FAR=1%. The numbers in the table are measured in $(\mu - \sigma)$ of 10 trials, where μ is the mean accuracy and σ is the standard deviation.

The performance of some methods under BLUFR protocol are reported in [160]. The best reported method is the High Dimensional LBP + Joint Bayes, which achieves VR = 41.66% (at FAR = 0.1%). Unfortunately, our proposed method is not comparable with the best result currently reported.

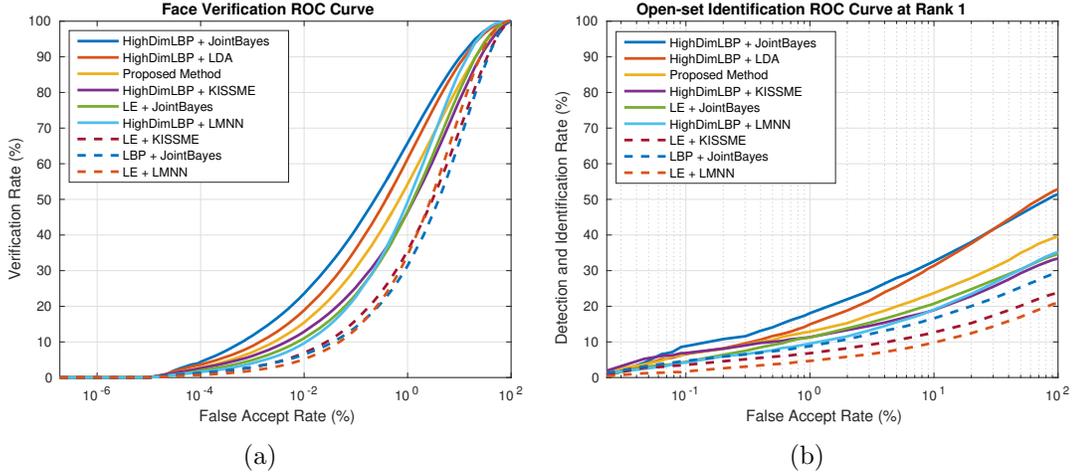


Figure 3.13: Performances of different approaches using BLUFR protocol. (a) Verification ROC curves. (b) Open-set identification ROC curves at rank 1.

3.5 Conclusions and Future Work

In this chapter, we presented a fully automatic system for pose and illumination invariant face recognition based on a single gallery image. The system starts by synthesizing the frontal views using a piece-wise affine warping. The warping is applied to triangular pieces determined by an enhanced AAM. The accuracy of the model fitting has direct influence on the quality of the synthesized frontal view, especially in case of semi-profile view images. In order to enhance the fitting accuracy, we initialize the AAM using estimates of the facial landmark locations obtained by a method based on flexible mixture of parts. The fitting accuracy is further improved by training the AAM with in-the-wild images and using powerful optimization technique, Fast-SIC.

Experimental results demonstrated the efficacy of our proposed fitting approach. Histograms of oriented gradients and Gabor wavelet features are extracted from these aligned frontal views. We use canonical correlation analysis to fuse these two features of a face image into a single but more discriminative feature vector. Our system handles a wide and continuous range of poses, *i.e.*, it is not restricted to any predeter-

Table 3.6: Benchmark performance of different state-of-the-art face recognition approaches for verification and open-set identification at rank 1. The reported numbers are the mean verification rates (%) subtracted by the corresponding standard deviations over 10 trials. The list is shown in chronological order.

Method	VR (%) @FAR=0.1%	DIR (%) @FAR=1%, Rank=1
LE + LMNN [161]	13.57	4.66
LBP + JointBayes [120]	14.18	8.82
LE + KISSME [162]	16.12	6.83
LE + JointBayes [120]	23.31	11.26
HighDimLBP + LMNN [159, 161]	22.68	9.53
HighDimLBP + KISSME [159, 162]	25.35	11.34
Proposed Method + JointBayes	30.76	12.86
HighDimLBP + LDA [158]	36.12	14.94
HighDimLBP + JointBayes [158]	41.66	18.07
DFD-SID + JointBayes [163]	40.70	15.28
LDA-SID + JointBayes [163]	48.99	19.13

mined pose angles. Furthermore, it is capable of recognizing subjects from non-frontal view images and from images with different illumination using only a single image as the gallery. Experimental results performed on FERET, CMU-PIE and LFW databases demonstrated the effectiveness of our proposed method in comparison with the state-of-the-art algorithms. In the future, we plan to investigate the possibility of synthesizing frontal faces with neutral expression to make the system invariant to facial expressions. Moreover, we plan to increase the biometric recognition accuracy by using multiple sources of biometrics information, *e.g.*, face, ear, and fingerprint, in a multimodal fashion.

CHAPTER 4

Discriminant Correlation Analysis: Real-Time Feature Level Fusion for Multimodal Biometrics

4.1 Introductory Remarks and Overview

Most of the real-world biometric systems, so-called *unimodal*, rely on the evidence of a single source of biometric information. *Multimodal* biometric systems, on the other hand, fuse multiple sources of biometric information to make a more reliable recognition. Fusion of the biometric information can occur at different stages of a recognition system. In case of *feature level fusion*, the data itself or the features extracted from multiple biometrics are fused. *Matching-score level fusion* consolidates the scores generated by multiple classifiers pertaining to different modalities. Finally, in case of *decision level fusion* the final results of multiple classifiers are combined via techniques such as majority voting.

Feature level fusion is believed to be more effective than the other levels of fusion because the feature set contains richer information about the input biometric data than the matching score or the output decision of a classifier. Therefore, fusion at the feature level is expected to provide better recognition results [45, 46]. However,

matching-score level fusion and decision level fusion are more popular in the literature and there is not much research on feature level fusion. The reason is the difficulty of feature level fusion in cases where the features are not compatible, *e.g.*, eigen-coefficients of faces and minutiae set of fingerprints, or when commercial biometric systems do not provide access to the feature sets (nor the raw data), which they use in their products [45]. The goal of the feature fusion for recognition is to combine relevant information from two or more feature vectors into a single one, which is expected to be more discriminative than any of the input feature vectors.

Two well-known and typical feature fusion methods are: serial feature fusion [47] and parallel feature fusion [48,49]. Serial feature fusion works by simply concatenating two sets of feature vectors into a single feature vector. Obviously, if the first source feature vector, x , is p -dimensional and the second source feature vector, y , is q -dimensional, the fused feature vector, z , will be $(p + q)$ -dimensional. Parallel feature fusion, on the other hand, combines the two source feature vectors into a complex vector $z=x+iy$ (i being an imaginary unit). Note that if the dimensions of the two input vectors are not equal, the one with the lower dimension is padded with zeros.

Recently, feature fusion based on Canonical Correlation Analysis (CCA) [50] has attracted the attention in the area of multimodal recognition. CCA-based feature fusion uses the correlation between two sets of features to find two sets of transformations such that the transformed features have maximum correlation across the two data sets, while being uncorrelated within each data set. This method was described in details in Section 3.3.2 of Chapter 3. Recently, CCA-based methods have become popular and other related and improved methods have also been proposed [51–54,97].

Kettenring [164] proposed a generalized extension of CCA for several sets of variables. Nielsen [165] improved Kettenrings method to present a multiset canonical correlation analysis (MCCA), which can be used to analyze relationships between more than two sets of variables. Although Kettenrings and Niensens methods [164, 165] are able to analyze multi-group variables, they do not demonstrate the integral relation among the multi-set variables, and the constraints do not guarantee that the transformed variables are statistically uncorrelated [53]. Recently, Yuan *et al.* [166] proposed a multi-set integrated canonical correlation analysis (MICCA) framework for the multi-set problems. MICCA can distinctly express the integral correlation among multi-set features. However, it follows an iterative approach, which reduces its efficiency.

Most recently, sparse representation has attracted the interest of many researchers, both for reconstructive and discriminative tasks [64, 167, 168]. The assumption is that a query sample belonging to a specific class can be represented with a linear combination of the training samples from that class. Therefore, it aims to find a sparse vector having non-zero elements only in the indices corresponding to that class. As indicated in the definition of feature level fusion, “the feature sets originating from multiple biometric algorithms are consolidated into a single feature set” [169]. Although not following this definition in building a single feature set that can be used by any classifier, Joint Sparse Representation Classification (JSRC) [170] is considered as a feature level fusion technique. JSRC builds multiple corresponding dictionaries each using training samples of a modality. Having a query consisting of multiple modalities, it aims to find joint sparse vectors that share the same sparsity pattern and have non-zero values only in the indices corresponding to a mutual class

in multiple modalities. That is, training samples of the same class from the different modalities are used to reconstruct the query data. Bahrampour *et al.* [171] improved the performance of this method by using a multimodal task-driven dictionary learning algorithm.

In this chapter, we propose a feature fusion method that considers the class associations in data sets¹. Our method, called *Discriminant Correlation Analysis (DCA)*, eliminates the between-class correlations and restricts the correlations to be within classes. DCA has the characteristics of the CCA-based methods in maximizing the correlation of corresponding features across the two data sets and in addition decorrelates features that belong to different classes within each data set. To the best of our knowledge, no other feature fusion method in the literature considered the class structure, and our method is the first to incorporate the class structure into the feature level fusion. It is worth mentioning that our method does not have the small sample size (SSS) problem faced by the CCA-based algorithms. Moreover, we propose a multiset method to generalize DCA to be applicable to more than two sets of variables. *Multiset Discriminant Correlation Analysis (MDCA)* follows a cascade approach and applies DCA on two sets of variables at a time. Extensive experiments performed on several multimodal biometric databases verify the effectiveness of our proposed method, which outperforms the state-of-the-art feature level fusion techniques².

This chapter is organized as follows: Section 4.2 presents our proposed discriminant correlation analysis method. The implementation details and experimental results on several databases are presented in Section 4.3. Finally, Section 4.4 concludes the chapter.

¹This work appeared in a conference and a journal paper [172] and [173].

²The MATLAB® source code for the proposed DCA method is provided at <http://www.mathworks.com/matlabcentral/fileexchange/55405>.

4.2 Incorporating Class Structure in Multivariate Correlation Analysis

The feature fusion method described in the previous section has two disputable issues. The first issue is encountered in case of a small sample size problem. In many real world applications, the number of samples is usually less than the number of features ($n < p$ or $n < q$). This makes the covariance matrices singular and non-invertible. Therefore, we will face a major problem in inverting the S_{xx} and S_{yy} matrices used in Eq. (3.8). A solution to overcome this issue is to reduce the dimensionality of the feature vectors before applying CCA. Therefore, a two stage PCA + CCA approach can be considered [51].

The second issue in CCA-based approaches is their negligence of the class structure among samples. CCA decorrelates the features, but in pattern recognition problems, we are also interested in separating the classes. The dimensionality reduction approaches based on Linear Discriminant Analysis (LDA) [174] consider this matter by finding projections that best separate the classes. However, a *two stage* LDA + CCA will not be an effective solution due to the fact that the transformation applied by the second stage, *i.e.*, CCA, will not preserve the properties achieved by the first stage, *i.e.*, LDA. Therefore, we need transformations that not only maximize the pair-wise correlations across the two feature sets, but also *simultaneously* separate the classes within each set of features. In this section, we present a solution to achieve this goal.

Correlation analysis and discriminant analysis have been previously used in a combined way in [175] and [176]. However, the problem definition and the presented methods are totally different from our problem setting and proposed technique. These

methods do not consider the problem of multimodal recognition or feature level fusion, which is the problem discussed in this chapter. In [175] and [176], the correlation analysis is used for the cross-domain matching problem in unimodal recognition systems. For example, [175] proposes a cross-view face recognition system, where the query face image is in a different view angle than the one given for enrollment. In the cross-domain matching problem, the correlation analysis aims to extract the correlated features from feature vectors of the different domains.

In our method, we incorporate the class structure, *i.e.*, memberships of the samples in classes, into the correlation analysis, which helps in highlighting the differences between classes and at the same time maximizing the pair-wise correlations between features across the two feature sets. This helps fusing the relevant information captured by different modalities in multimodal recognition systems. Our proposed approach, called Discriminant Correlation Analysis (DCA), is described below.

4.2.1 Feature-Level Fusion Using Discriminant Correlation Analysis

Let's assume that the samples in the data matrix are collected from c separate classes. Accordingly, the n columns of the data matrix are divided into c separate groups, where n_i columns belong to the i^{th} class ($n = \sum_{i=1}^c n_i$). Let $x_{ij} \in X$ denote the feature vector corresponding to the j^{th} sample in the i^{th} class. \bar{x}_i and \bar{x} denote the means of the x_{ij} vectors in the i^{th} class and the whole feature set, respectively. That is, $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$ and $\bar{x} = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n_i} x_{ij} = \frac{1}{n} \sum_{i=1}^c n_i \bar{x}_i$. The between-class

scatter matrix is defined as

$$S_{bx_{(p \times p)}} = \sum_{i=1}^c n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T = \Phi_{bx} \Phi_{bx}^T, \quad (4.1)$$

where

$$\Phi_{bx_{(p \times c)}} = [\sqrt{n_1}(\bar{x}_1 - \bar{x}), \sqrt{n_2}(\bar{x}_2 - \bar{x}), \dots, \sqrt{n_c}(\bar{x}_c - \bar{x})]. \quad (4.2)$$

If the number of features is higher than the number of classes ($p \gg c$), it is computationally easier to calculate the covariance matrix as $(\Phi_{bx}^T \Phi_{bx})_{c \times c}$ rather than $(\Phi_{bx} \Phi_{bx}^T)_{p \times p}$. As presented in [74], the most significant eigenvectors of $\Phi_{bx} \Phi_{bx}^T$ can be efficiently obtained by mapping the eigenvectors of $\Phi_{bx}^T \Phi_{bx}$. Therefore, we only need to find the eigenvectors of the $c \times c$ covariance matrix $\Phi_{bx}^T \Phi_{bx}$.

If the classes were well-separated, $\Phi_{bx}^T \Phi_{bx}$ would be a diagonal matrix. Since $\Phi_{bx}^T \Phi_{bx}$ is symmetric positive semi-definite, we can find transformations that diagonalize it:

$$P^T (\Phi_{bx}^T \Phi_{bx}) P = \hat{\Lambda}, \quad (4.3)$$

where P is the matrix of orthogonal eigenvectors and $\hat{\Lambda}$ is the diagonal matrix of real and non-negative eigenvalues sorted in decreasing order.

Let $Q_{(c \times r)}$ consist of the first r eigenvectors, which correspond to the r largest non-zero eigenvalues, from matrix P . We have:

$$Q^T (\Phi_{bx}^T \Phi_{bx}) Q = \Lambda_{(r \times r)}. \quad (4.4)$$

The r most significant eigenvectors of S_{bx} can be obtained with the mapping: $Q \rightarrow \Phi_{bx} Q$ [74]:

$$(\Phi_{bx} Q)^T S_{bx} (\Phi_{bx} Q) = \Lambda_{(r \times r)}. \quad (4.5)$$

$W_{bx} = \Phi_{bx} Q \Lambda^{-1/2}$ is the transformation that unitizes S_{bx} and reduces the dimensionality of the data matrix, X , from p to r . That is:

$$W_{bx}^T S_{bx} W_{bx} = I, \quad (4.6)$$

$$X'_{(r \times n)} = W_{bx(r \times p)}^T X_{(p \times n)}. \quad (4.7)$$

X' is the projection of X in a space, where the between-class scatter matrix is I and the classes are separated. Note that there are at most $c - 1$ nonzero generalized eigenvalues; therefore, an upper bound on r is $c - 1$ [177]. Other upper bounds for r are the ranks of the data matrices, *i.e.*, $r \leq \min(c - 1, \text{rank}(X), \text{rank}(Y))$.

Similar to the above approach we solve for the second feature set, Y , and find a transformation matrix W_{by} , which unitizes the between-class scatter matrix for the second modality, S_{by} and reduces the dimensionality of Y from q to r :

$$W_{by}^T S_{by} W_{by} = I, \quad (4.8)$$

$$Y'_{(r \times n)} = W_{by(r \times q)}^T Y_{(q \times n)}. \quad (4.9)$$

The updated Φ'_{bx} and Φ'_{by} are non-square $r \times c$ orthonormal matrices. Although $S'_{bx} = S'_{by} = I$, the matrices $\Phi'_{bx}{}^T \Phi'_{bx}$ and $\Phi'_{by}{}^T \Phi'_{by}$ are strict diagonally dominant matrices ($\forall i : |a_{ii}| > \sum_{j \neq i} |a_{ij}|$), where the diagonal elements are close to one and the non-diagonal elements are close to zero. This makes the centroids of the classes have minimal correlation with each other, and thus, the classes are separated.

Now that we have transformed X and Y to X' and Y' , where the between-class scatter matrices are unitized, we need to make the features in one set have nonzero correlation only with their corresponding features in the other set. To achieve this, we need to diagonalize the between-set covariance matrix of the transformed feature

sets, $S'_{xy} = X'Y'^T$. We use singular value decomposition (SVD) to diagonalize S'_{xy} :

$$S'_{xy(r \times r)} = U \Sigma V^T \quad \Rightarrow \quad U^T S'_{xy} V = \Sigma. \quad (4.10)$$

Note that X' and Y' are of rank r and $S'_{xy(r \times r)}$ is nondegenerate. Therefore, Σ is a diagonal matrix whose main diagonal elements are non-zero. Let $W_{cx} = U \Sigma^{-1/2}$ and $W_{cy} = V \Sigma^{-1/2}$, we have:

$$(U \Sigma^{-1/2})^T S'_{xy} (V \Sigma^{-1/2}) = I, \quad (4.11)$$

which unitizes the between-set covariance matrix, S'_{xy} . Now, we transform the feature sets as follows:

$$\tilde{X}^* = W_{cx}^T X' = \underbrace{W_{cx}^T W_{bx}^T}_{I} X = W_x X, \quad (4.12)$$

$$\tilde{Y}^* = W_{cy}^T Y' = \underbrace{W_{cy}^T W_{by}^T}_{I} Y = W_y Y. \quad (4.13)$$

where $W_x = W_{cx}^T W_{bx}^T$ and $W_y = W_{cy}^T W_{by}^T$ are the final transformation matrices for X and Y , respectively.

It can be easily shown that the between-class scatter matrices of the transformed feature sets are still diagonal; hence, the classes are separated. The between-class scatter matrix for \tilde{X}^* is calculated as:

$$\tilde{S}_{bx}^* = W_{cx}^T \underbrace{W_{bx}^T S_{bx} W_{bx}}_{I} W_{cx}. \quad (4.14)$$

From Eq. (4.6), $W_{bx}^T S_{bx} W_{bx} = I$ and since U is an orthogonal matrix, we have:

$$\tilde{S}_{bx}^* = (U \Sigma^{-\frac{1}{2}})^T (U \Sigma^{-\frac{1}{2}}) = \Sigma^{-1}. \quad (4.15)$$

Similarly, we can show that $\tilde{S}_{by}^* = \Sigma^{-1}$, which is diagonal.

Fig. 4.1(a) shows the covariance between features in a transformed feature set ($\tilde{X}^* \tilde{X}^{*T}$), which is a strict diagonally dominant matrix. Black color represents zero

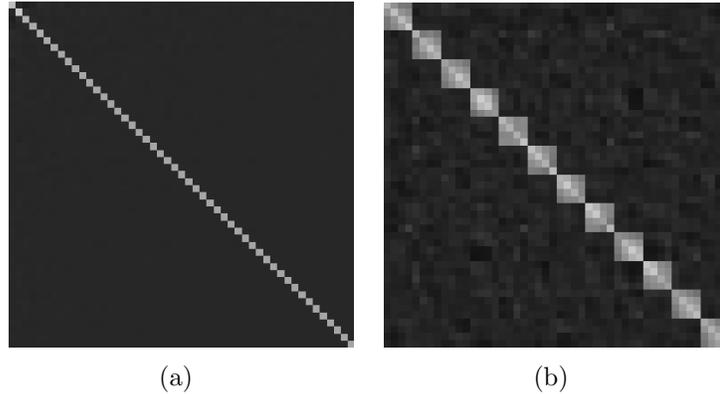


Figure 4.1: Visualization of covariance matrices (black color represents zero values and the elements with higher values are illustrated brighter). (a) Covariance between features ($\tilde{X}^* \tilde{X}^{*T}$). (b) Covariance between samples ($\tilde{X}^{*T} \tilde{X}^*$).

values and the elements with higher values are brighter. The results show that the correlation between different features in an individual feature set is minimal. On the other hand, Fig. 4.1(b) shows the covariance between samples in a transformed feature set ($\tilde{X}^{*T} \tilde{X}^*$). Being a block diagonal matrix, Fig. 4.1(b) clearly shows that the samples have higher correlation with only the ones in the same class.

Similar to the CCA method, feature level fusion can be performed either by concatenation or summation of the transformed feature vectors, as shown in Eqs. (3.9) and (3.10). However, the summation method has the advantage of lower number of dimensions, while the change in recognition results is very small. In our experiments, we use the summation method, shown in Eq. (3.10), for both CCA and DCA approaches.

4.2.2 Multiset Discriminant Correlation Analysis

Multiset Discriminant Correlation Analysis (MDCA) generalizes DCA to be applicable to more than two sets of features. Here, we assume that we have m sets

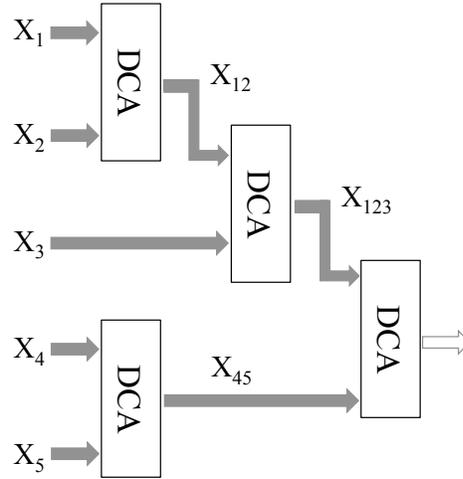


Figure 4.2: Multiset discriminant correlation analysis techniques for five sample sets with $rank(X_1) > rank(X_2) > rank(X_3) > rank(X_4) = rank(X_5)$.

of features, $X_i \in \mathbb{R}^{p_i \times n}$, $i = 1, 2, \dots, m$, which are sorted by their rank, that is $rank(X_1) \geq rank(X_2) \geq \dots \geq rank(X_m)$. MDCA applies DCA on two sets of features at a time. Based on the approach presented in the previous section, the maximum length of the fused feature vector is $\min(c - 1, rank(X_i), rank(X_j))$. In order to maintain the maximum possible length of the fused feature vector, in each step, the two feature sets with the highest ranks will be fused together. For example, in the first step, X_1 and X_2 , which have the highest ranks, will be fused. The result of the fusion of X_1 and X_2 will be fused with the next highest rank feature set, *i.e.*, X_3 , and so on. If there exists feature sets with equal ranks, they can be fused together at any time. We choose the length of the fused feature vector, r , to be equal to $\min(c - 1, rank(X_i), rank(X_j))$.

Fig. 4.2 shows an example framework of MDCA for five feature sets with $rank(X_1) > rank(X_2) > rank(X_3) > rank(X_4) = rank(X_5)$. In the first step of MDCA, we fuse X_1 and X_2 , which have the highest ranks. X_4 and X_5 , which have equal ranks, will be also fused together. The length of the X_{12} is expected to be greater than the

length of the X_{45} . Therefore, in the next step, X_3 is fused with X_{12} . In this way, we keep the maximum possible length for the fused feature vector in every step. The expected, possibly shorter, feature vector length can be determined in the final step, $r \leq \min(c - 1, \text{rank}(X_{123}), \text{rank}(X_{45}))$.

4.3 Experiments and Analysis

In this chapter, we present several sets of experiments to demonstrate the performance of our proposed feature level fusion technique. We devise experiments for combining different features extracted from a single modality as well as combining feature vectors extracted from different biometric modalities. Section 4.3.1 shows experimental results for combining different feature vectors extracted from a single modality. Additionally, Sections 4.3.2, 4.3.3, and 4.3.5 present experiments on the fusion of different biometric modalities. In Section 4.3.2, experiments are performed on fusing features from frontal/near-frontal face, profile/near-profile face, and ear modalities extracted from West Virginia University (WVU) database [95]. Similarly, in Section 4.3.3, experiments are conducted on fingerprint and iris modalities from Multimodal Biometric Dataset Collection, BIOMDATA [178]. Section 4.3.5 presents experiments on fusing information from weak biometric modalities, *i.e.*, periocular, mouth, and nose regions, extracted from face images in AR face database [179]. Section 4.3.6 evaluates the scalability of the proposed DCA method in dealing with new subjects that are not seen during the training. Finally, as an example of the applicability of the proposed approach to other applications, Section 4.3.7 shows how the proposed method helps to improve the accuracy of sketch to mugshot matching.

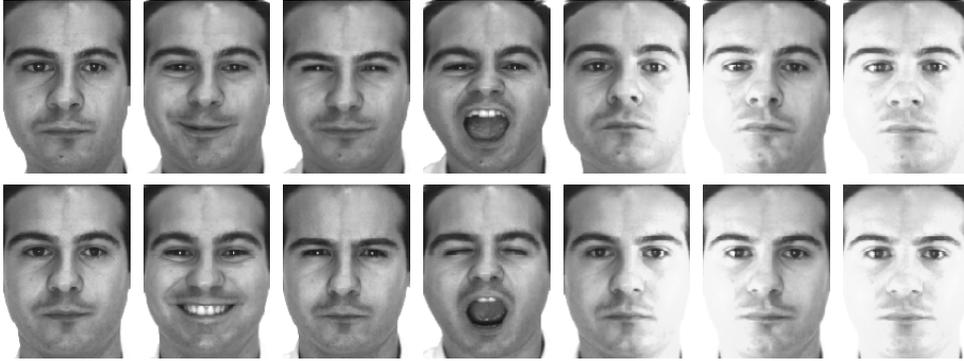


Figure 4.3: Sample face images of a subject in AR database (Top row: first session; Bottom row: second session).

Table 4.1: Maximum rank-1 recognition rates using individual and fused feature vectors (H: HOG, S: SURF, G: Gabor).

Method	H	S	G	HG	SG	HSG
Serial	85.14	81.00	79.57	86.57	81.86	86.57
Parallel	85.14	81.00	79.57	89.43	82.43	-
CCA/MCCA	85.14	81.00	79.57	94.43	91.57	95.14
DCA/MDCA	85.14	81.00	79.57	98.00	95.71	98.71

4.3.1 Unimodal Multi-Feature Fusion

In this section, we present experiments to show the effectiveness of the proposed method in combining feature sets extracted from a single modality. We evaluated our algorithms on a set of 100 subjects from AR face database [179, 180]. The AR face database consists of frontal face images with varying facial expressions and illumination. Fig. 4.3 shows sample images of one subject in the AR database. The face images are captured in two sessions. In this experiment, seven images of each subject from the first session are used for training and seven images from the second session are used for testing.

Three different features are extracted from these images. These features include Gabor wavelet features [130], Histogram of Oriented Gradients (HOG) [98], and Speeded-Up Robust Features (SURF) [181]. We employ forty Gabor filters in five scales and eight orientations. Since the adjacent pixels in an image are usually cor-

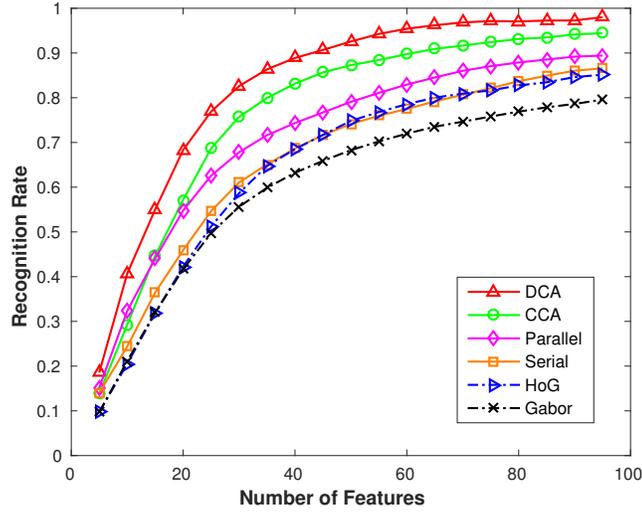


Figure 4.4: Accuracy of the unimodal biometric systems using Gabor and HOG features on AR face database.

related, the information redundancy can be reduced by downsampling the feature images that result from Gabor filters [72, 130]. In our experiments, the feature images are downsampled by a factor of five. HOG features, on the other hand, are extracted in 5×5 cells for nine orientations. We use the UOCTTI variant for the HOG presented in [131]. UOCTTI variant computes both directed and undirected gradients as well as a four dimensional texture-energy feature, but projects the result down to 31 dimensions (27 dimensions corresponding to different orientation channels, 9 contrast insensitive and 18 contrast sensitive, and 4 dimensions capturing the overall gradient energy in square blocks of four adjacent cells)³. Finally, we extract SURF features from 68 keypoints in every image. These points are the facial landmarks detected by fitting an Active Appearance Model (AAM) to the face images. A 64-dimensional feature vector is extracted from each point and the final feature vector is constructed by concatenating the feature vectors of all keypoints. A simple

³VLFeat open source library is used to extract the HOG features [132].

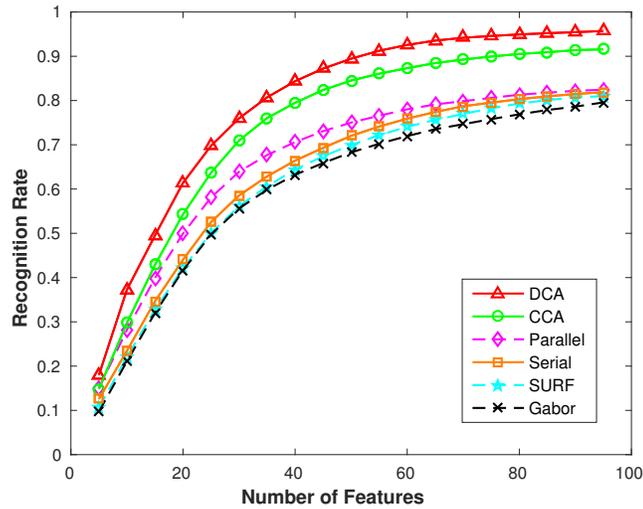


Figure 4.5: Accuracy of the unimodal biometric systems using Gabor and SURF features on AR face database.

minimum distance classifier is used for classification, in which one minus the sample linear correlation between observations is used as the distance.

Figs. 4.4, 4.5, and 4.6 show the experimental results for combining different feature vectors. Table 4.1 shows the maximum rank-1 recognition rates, over the number of features, obtained using individual and fused feature vectors. As mentioned before, the goal is to combine relevant information from the two input feature vectors into a single vector, which is expected to be more discriminative than any of the input feature vectors. Therefore, a fusion method that decreases the correlations between features will be more effective.

As it is clearly seen from the results, serial feature fusion [47] is not always successful in this regard, and in some cases, the fused feature vector has even less discriminative power than the input feature vector. Parallel feature fusion [48, 49] does not show a more discriminative feature either and in case of Gabor-SURF fusion, the fused feature vector works almost similar to the SURF feature vector. Note that the

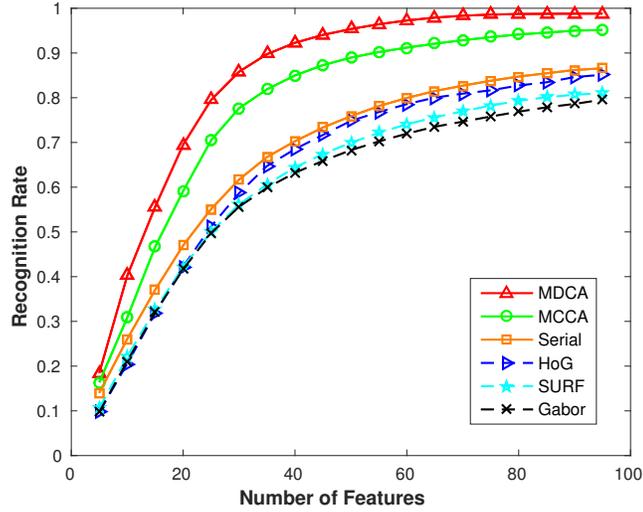


Figure 4.6: Accuracy of the unimodal biometric systems using Gabor, SURF and HOG features on AR face database.

parallel feature fusion method cannot be applied on more than two sets of variables; therefore, it is excluded in the third experiment. For the cases of more than two feature sets, here, we use Multiset Canonical Correlation Analysis (MCCA) [182] and MDCA methods.

The CCA-based feature fusion [50] and the proposed DCA feature fusion methods, on the other hand, work very well in combining different feature vectors. The reason might be the fact that these methods reduce the redundant information between two input feature vectors. Incorporating the class associations in its analysis, DCA provides a more powerful feature vector than CCA for the recognition purposes. The experimental results verify the effectiveness of our proposed method in comparison with serial, parallel and CCA-based feature fusion techniques. As mentioned in Section 4.1, the JSRC [170] and SMDL [171] methods does not combine feature vectors extracted from multiple modalities into a single fused feature vector that can be used by any classifier. Therefore, these methods are not included in this experiment;

however, they will be evaluated in the other experiments presented in Sections 4.3.2, 4.3.3, and 4.3.5.

4.3.2 WVU Database

In this set of experiments, we evaluate the performance of the proposed algorithm in combining feature vectors extracted from different biometric modalities on the WVU database [95]. This database consists of almost 110 seconds long video clips with rates of thirty frames per second, captured with a camera that rotates around the face. There are 402 subjects in the database. This database has 55 subjects with eyeglasses, 42 subjects with earrings, 38 subjects with partially occluded ears, and 2 subjects with fully occluded ears [183]. For subjects #239, #302, the ears are fully occluded with the hair, and for subject #308, just small portions of the ears are visible. Therefore, we exclude these three subjects and use the remaining 399 subjects in our experiments.

The video clips are captured by rotating a camera around the face; it starts from the left profile image of the face and ends at the right profile image. If we assume that the rotation for the left profile image is 0° and the rotation for the right profile image is 180° , the frontal image of the face is in the middle of the clip, *i.e.*, 90° of rotation. For our experiments, we choose frames that are five degrees of rotation apart. Figure 4.7 shows a sample of these frames in the range of 0° to 90° . We extract three different biometric modalities (frontal/near-frontal face, ear, and profile/near-profile face) from the above-mentioned frames. The best exposure of the profile face and the ear is at 0° while the best exposure of the frontal face is at 90° . For each modality, we choose ten images with up to 45° of rotation from their best exposure.



Figure 4.7: Different frames of the subject #1 from profile to frontal equally distanced by 5° .

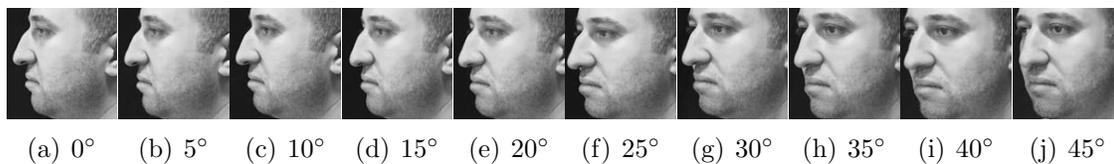


Figure 4.8: Profile/near-profile face images detected in different frames.

The face detection method proposed in [127] is used to automatically extract frontal and profile faces in each frame. For each subject, we extract ten profile and near-profile faces spanning between 0° and 45° , and ten frontal and near-frontal face images spanning from 45° to 90° degrees of rotation. Figures 4.8 and 4.9 show the sample profile/near-profile face and frontal/near-frontal face images extracted from the corresponding frames shown in Fig. 4.7.

On the other hand, the ear detection method proposed in [184] is used to automatically extract the ear regions. The ear detection method uses the deformable part model to find 17 landmarks on the ear helix and anti-helix. Figure 4.10(a) shows these

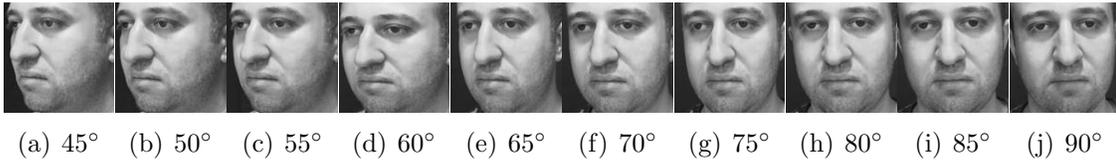


Figure 4.9: Frontal/near-frontal face images detected in different frames.

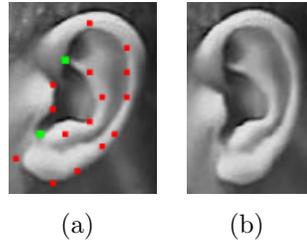


Figure 4.10: Ear normalization for in-plane rotations. (a) Detected Landmarks. (b) Normalized Ear.

landmarks on a sample ear image. We use the two green landmarks, the Triangular Fossa and Incisure Intertragica, to normalize the ear for in-plane pose variations. The normalized ear is shown in Fig. 4.10(b). For each subject, we extract ten ear images spanning between 0° and 45° . Figure 4.11 shows the sample ear images extracted from the corresponding frames shown in Fig. 4.7.

In our experiments, all the face images are normalized to 120×120 pixels and all ear images are normalized to 120×80 pixels. For feature extraction, Gabor features are extracted in five scales and eight orientations, and similar to the setting described in Section 4.3.1, the feature images are downsampled by a factor of five. The most important advantage of Gabor filters is their invariance to rotation, scale, and translation. Furthermore, they are robust against photometric disturbances, such as illumination change and image noise [73, 133].

We perform three multimodal experiments using WVU database. These experiments include the fusion of (a) frontal face and ear, (b) profile face and ear, and (c) all three modalities. For the first experiment, ten face images of each subject are ran-

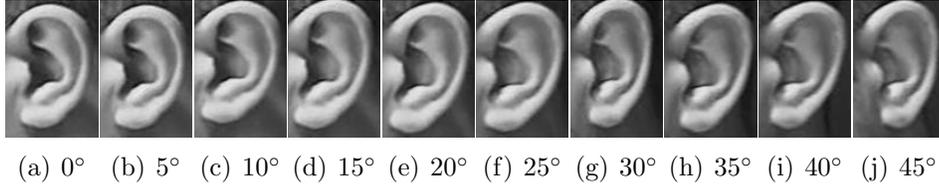


Figure 4.11: Ear images detected in different frames.

domly paired with ten ear images of the same subject to create a multimodal dataset of face-ear pairs. Five randomly chosen pairs are used for training and the remaining five are used for testing. In order to validate the robustness of the experiments, repeated random sub-sampling validation is applied and the results are averaged over 10 iterations. The same setting is used for the second and third experiments using ear-profile pairs and face-ear-profile trios, respectively.

4.3.2.1 Comparison of Methods

The performance of the proposed feature level fusion algorithm is compared with that of several state-of-the-art feature level, matching score level and decision level fusion algorithms. The feature level fusion techniques include the serial feature fusion [47], the parallel feature fusion [49], the CCA-based feature fusion [50, 182], and the most recently published JSRC [170] and SMDL [171] methods. In order to prevent the small sample size problem in the CCA-based approach, dimensionality reductions based on Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are applied [51, 74]. PCA and LDA are also used for dimensionality reduction and discriminant analysis of the results of the serial and parallel methods. Except for the JSRC and SMDL methods, which are restricted to work with a sparse representation classifier, all other feature level techniques use a simple KNN classifier with $K = 1$, *i.e.*, a minimum distance classifier, for classification. Here, one minus

Table 4.2: Rank-1 recognition rates obtained by a KNN classifier (K=1) using individual modalities in WVU database.

Modality	Face	Ear	Profile Face
Recognition Rate	82.59 ± 2.21	79.66 ± 2.43	81.71 ± 2.01

the sample linear correlation between observations is used as the distance. Note that in case of more than two modalities (Face+Ear+Profile), the parallel feature fusion method cannot be applied and Multiset-CCA [182] and Multiset-DCA are used.

For matching score level fusion and decision level fusion, we use Sparse Logistic Regression (SLR) [185] and SVM [186] techniques. For matching score level fusion, the probability outputs for each modality of the query samples are added together to produce the final score values, which are used for classification. For decision level fusion, on the other hand, the subject chosen by the maximum number of modalities was taken to be from the correct class. Following the notation in [170] and [171], we denote the score level fusion of these methods as SLR-Sum and SVM-Sum, and the decision level fusion as SLR-Major and SVM-Major. Moreover, we compare with the multiclass implementation of the Multiple Kernel Learning (MKL) algorithm [187].

Table 4.2 shows the rank-1 recognition rate for the individual modalities of face, ear and profile face, and Table 4.3 shows the multimodal fusion results. It is clear that the proposed DCA technique outperforms the other fusion methods. It is also shown that the combination of LDA + CCA is not effective for separating the classes due to the fact that the transformation applied by the CCA does not preserve the properties achieved by the LDA.

The complexity of the above-mentioned feature level fusion algorithms are compared using their run-time values. Table 4.4 shows the average computation time for each algorithm. Note that the run-time values are for recognition of one multimodal face-ear pair in WVU database averaged over multiple runs. Note that the

Table 4.3: Rank-1 recognition rates for multimodal fusion of face, ear and profile face biometrics in WVU database.

Modality Method	Face+Ear	Ear+Profile	Face+Ear +Profile
SVM-Major	85.09 ± 2.02	85.31 ± 2.12	87.59 ± 1.20
SVM-Sum	94.18 ± 1.54	94.42 ± 1.22	95.12 ± 1.02
SLR-Major	85.92 ± 1.87	85.85 ± 1.55	88.12 ± 1.24
SLR-Sum	94.37 ± 1.14	94.63 ± 1.20	95.57 ± 0.93
MKL	92.51 ± 1.21	92.97 ± 1.17	94.46 ± 0.98
Serial + PCA + KNN	89.14 ± 1.15	89.46 ± 1.13	92.28 ± 1.11
Serial + LDA + KNN	94.23 ± 1.02	95.14 ± 1.20	95.14 ± 1.04
Parallel + PCA + KNN	90.71 ± 2.05	90.61 ± 1.86	-
Parallel + LDA + KNN	93.38 ± 1.66	93.13 ± 1.67	-
PCA + CCA/MCCA + KNN	94.10 ± 0.87	94.34 ± 0.57	97.74 ± 0.54
LDA + CCA/MCCA + KNN	94.44 ± 0.88	94.89 ± 0.54	97.86 ± 0.49
JSRC	96.20 ± 0.52	97.74 ± 0.42	98.74 ± 0.32
SMDL	97.24 ± 0.48	97.97 ± 0.42	99.20 ± 0.24
DCA/MDCA + KNN	98.56 ± 0.15	99.38 ± 0.08	99.85 ± 0.03

Table 4.4: Average run-time values of different feature level fusion techniques for recognition of one multimodal face-ear pair in WVU database.

Method	Run Time (in milliseconds)
Serial + PCA + KNN	19
Serial + LDA + KNN	24
Parallel + PCA + KNN	39
Parallel + LDA + KNN	42
PCA + CCA + KNN	19
LDA + CCA + KNN	21
JSRC	8406
SMDL	7882
DCA + KNN	19

serial, parallel, CCA and DCA algorithms are very fast because they only apply the transformations obtained from the training process. Parallel feature fusion method is slightly more time consuming because it deals with complex feature vectors. JSRC and SMDL algorithms, on the other hand, are very time consuming and cannot be used in real-time applications.

Table 4.5: BIOMDATA multimodal biometric dataset.

Biometric Modality	# of Subjects	# of Samples
Iris	231	3043
Fingerprint	270	7136
Palm	263	673
Hand	219	2837
Voice	240	640
Face	205	1170

4.3.3 Multimodal Biometric Dataset, BIOMDATA

In this set of experiments, we use the multimodal biometric dataset (BIOMDATA) collected in West Virginia University [178]. This dataset is a comprehensive collection of image and sound files for six biometric modalities: iris, face, voice, fingerprint, hand geometry, and palm print, from subjects of different ethnicity, gender, and age. It is a challenging data set, as many of the samples suffer from various artifacts such as blur, occlusion, shadows, and sensor noise, as shown in Fig. 4.12. Table 4.5 shows the number of subjects and samples available in each modality. Due to privacy issues related to identifying individuals, face data is not made available in combination with other modalities; therefore, it cannot be used in a multimodal experiment. Following the experimental setting in [170] and [171], we chose iris and fingerprint modalities for our experiments. All the evaluations are performed on a subset of 219 subjects having samples in both modalities. In total, there are two iris (left and right eye) and four fingerprint modalities (thumb and index fingers from both hands).

Fig. 4.13 shows the preprocessing steps for a sample iris image. We segmented the iris images using the method proposed in [188]. As shown in Fig. 4.13(b), the non-iris areas in the segmented region are removed as noise. Following the segmentation step, iris regions are normalized and 25×240 bit-wise iris templates are generated by extracting log-Gabor features using the publicly available source code of Masek

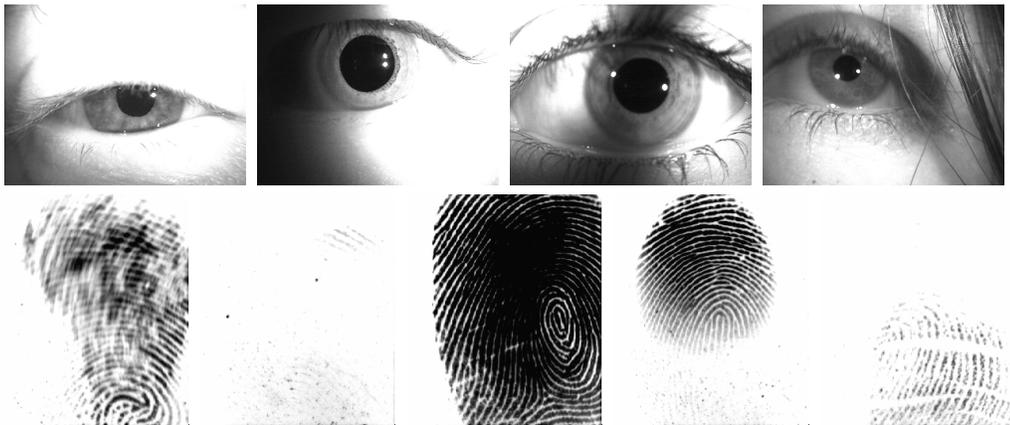


Figure 4.12: Examples of challenging samples in BIOMDATA database. The images are corrupted with blur, occlusion, shadows, and sensor noise.

Table 4.6: Rank-1 recognition rates obtained by a minimum distance classifier using individual modalities on BIOMDATA database.

Modality	Recognition rate
Iris (Left)	51.29 ± 2.12
Iris (Right)	57.33 ± 1.80
Fingerprint (Left thumb)	78.22 ± 1.39
Fingerprint (Left index)	90.10 ± 1.11
Fingerprint (Right thumb)	79.60 ± 1.22
Fingerprint (Right index)	91.29 ± 0.92

and Kovesi [189]. On the other hand, we enhanced the fingerprint images using the filtering methods described in [1]. Following the image enhancement step, the core points of the fingerprints are detected [77] and Gabor features in eight orientations are extracted around each detected core point. Fig. 4.14 shows the preprocessing steps for a sample fingerprint image.

Four samples randomly chosen from each modality are used for training and the remaining samples are used for testing. The recognition results are averaged over five runs. As before, all experiments, except for the JSRC method, use a minimum distance classifier. One minus the sample linear correlation between observations is used as the distance.

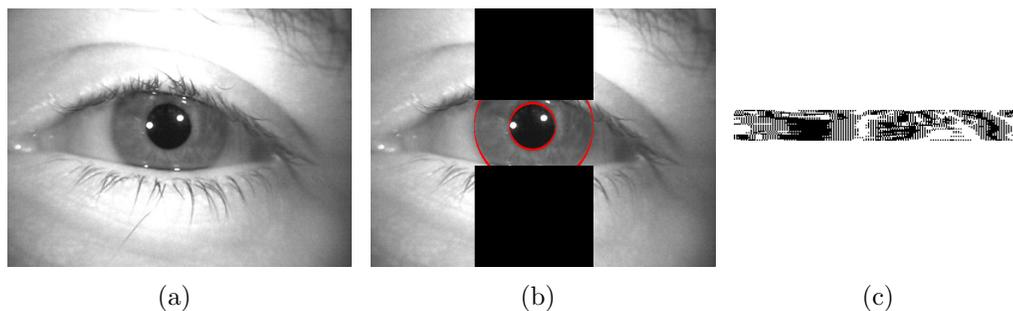


Figure 4.13: Preprocessing for iris images. (a) Original iris image from BIOMDATA database. (b) Segmented iris area. (c) 25×240 binary iris template.

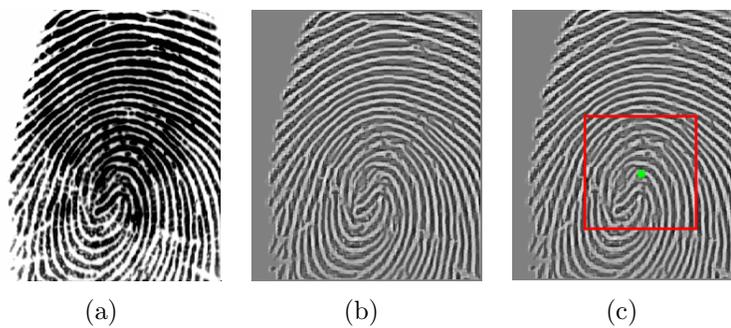


Figure 4.14: Preprocessing for fingerprint images. (a) Original fingerprint image from BIOMDATA database. (b) Enhanced image using the method in [1]. (c) Core point of the fingerprint and the region of interest around it.

Table 4.6 shows the rank-1 recognition rate for the individual iris and fingerprint modalities, and Table 4.7 shows the multimodal fusion results. We compare the proposed feature level fusion technique with several state-of-the-art feature level, matching score level and decision level fusion algorithms mentioned in Section 4.3.2.1. Experimental results clearly show that the proposed DCA technique outperforms the other fusion methods.

Table 4.7: Rank-1 recognition rates for multimodal fusion of iris and fingerprint biometrics in BIOM-DATA database.

Method \ Modality	2 Irises	4 Fingerprints	All 6 Modalities
SVM-Major	62.30 ± 2.35	90.14 ± 1.23	92.24 ± 1.11
SVM-Sum	71.03 ± 2.12	93.43 ± 0.96	97.51 ± 0.71
SLR-Major	61.73 ± 1.86	89.23 ± 1.03	91.18 ± 1.27
SLR-Sum	69.43 ± 1.88	93.67 ± 0.76	97.09 ± 0.80
MKL	68.23 ± 1.52	93.28 ± 0.84	95.96 ± 0.65
Serial + PCA + KNN	62.48 ± 1.24	94.46 ± 1.14	94.85 ± 0.87
Serial + LDA + KNN	70.31 ± 1.30	96.22 ± 0.98	96.22 ± 0.88
Parallel + PCA + KNN	68.22 ± 2.11	-	-
Parallel + LDA + KNN	72.25 ± 1.93	-	-
PCA + CCA/MCCA + KNN	78.51 ± 0.90	96.32 ± 0.74	97.20 ± 0.56
LDA + CCA/MCCA + KNN	78.90 ± 0.90	96.40 ± 0.72	97.51 ± 0.57
JSRC	78.20 ± 0.87	97.60 ± 0.42	98.60 ± 0.32
SMDL	83.77 ± 0.89	97.56 ± 0.41	99.10 ± 0.30
DCA/MDCA + KNN	84.16 ± 0.79	98.71 ± 0.42	99.60 ± 0.28

Table 4.8: Rank-1 recognition rates on the virtual multimodal database of face and ear images.

Ear	Face	Serial	Parallel	CCA	DCA
84.55 ± 0.91	80.42 ± 0.87	87.89 ± 0.87	90.60 ± 1.02	95.37 ± 0.74	99.86 ± 0.08

4.3.4 Multimodal Biometrics: AR Face Database and USTB

Ear Database

In this set of experiments, we use a virtual multimodal database generated by combining AR face database [179] and USTB ear database [190]. Based on the assumption that different biometric traits of the same person are independent, we can construct the multimodal databases by pairing a subject from face database with a subject from ear database [191].

The USTB ear database III consist of ear images of 79 subjects, 77 of whom have at least 10 profile images. We automatically extract the ear regions using the ear detection method proposed in [184]. Fig. 4.15 shows ear images of one subject extracted from USTB database.

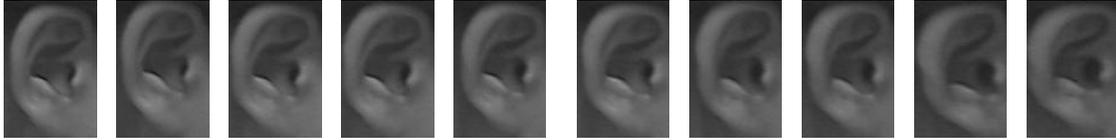


Figure 4.15: Sample ear images of a subject cropped from USTB database III.

The AR face database consists of frontal face images with varying facial expressions and illumination. Since we have 77 subjects in the ear database, we use the first 77 subjects from the AR face database. Fig. 4.3 shows sample images of one subject in the AR database. The face images are captured in two sessions [180]. In order to populate the training set, five face images of each subject, randomly selected from the first session, are paired with five ear images randomly chosen from a subject in USTB database. Similarly, five face images of each subject, randomly selected from the second session are paired with the remaining five ear samples of the subject to create the testing set.

Similar to the experiments in the previous section, Gabor wavelets are used for feature extraction and the minimum Euclidean distance classifier is used for classification. Table 4.8 shows the rank-1 recognition rate for the individual modalities of ear and face as well as the multimodal fusion results using serial, parallel, CCA and DCA methods. The recognition result are averaged over five runs. As seen from the results, the proposed DCA feature fusion algorithm outperforms the other methods and reaches the impressive recognition rate of 99.86%.

4.3.5 AR Face Database

In this set of experiments, we show the applicability of the proposed MDCA algorithm in fusing information from weak biometric modalities extracted from face

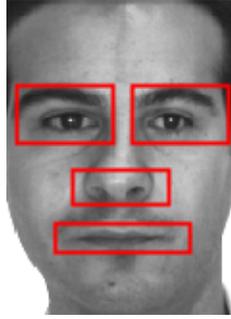


Figure 4.16: Face mask used to crop out different modalities.

Table 4.9: Rank-1 recognition rates obtained by a KNN classifier using individual modalities in AR database. Modalities include 1. left periocular, 2. right periocular, 3. nose, 4. mouth, and 5. face.

Modality	1	2	3	4	5
Recognition Rate	84.14	84.29	73.57	74.29	90.57

Table 4.10: Rank-1 recognition rates for multimodal fusion of different modalities in AR database. Modalities include 1. left periocular, 2. right periocular, 3. nose, 4. mouth, and 5. face.

Method \ Modality	{1,2}	{1,2,3}	{1,2,3,4}	{1,2,3,4,5}
Serial + PCA + KNN	85.57	88.71	90.42	90.71
Serial + LDA + KNN	89.43	92.14	92.86	93.57
PCA+CCA/MCCA+KNN	90.57	92.86	94.43	96.57
LDA+CCA/MCCA+KNN	91.28	92.57	93.71	97.00
JSRC	92.14	92.86	94.43	98.57
SMDL	92.29	92.86	95.14	98.85
DCA/MDCA + KNN	92.71	93.28	97.43	99.14

images. These modalities include left and right periocular, mouth, and nose regions, as shown in Fig. 4.16. It was shown that the periocular regions, nose and mouth can be considered as useful biometrics [192–194]; however, they are not as discriminative as the whole face [170].

We evaluated our algorithms on a set of 100 subjects from AR face database [179, 180] described in Section 4.3.1. Similar to the setup in [170], seven images of each subject from the first session are used for training and seven images from the second session are used for testing. Gabor features in five scales and eight orientations are extracted from all modalities.

Table 4.11: Rank-1 recognition rates for multimodal fusion of all modalities in AR database.

SVM-Major	SVM-Sum	SLR-Major	SLR-Sum	MKL	Serial+LDA	LDA+MCCA	JSRC	SMDL	MDCA
85.71	92.85	86.85	93.71	93.00	93.57	97.00	98.57	98.85	99.14

Table 4.9 shows the rank-1 recognition rates for the individual modalities. The major challenge here is to be able to fuse weak modalities with a strong modality based on the whole face without deteriorating the accuracy performance with respect to that of the strong modality [195].

Table 4.10 shows the recognition rates for different feature level fusion methods using combinations of different modalities. The results of fusing all five modalities with other methods including matching score level and decision level fusion techniques are presented in Table 4.11. It is obvious that the proposed method has a higher recognition rate than the other feature level fusion techniques. Moreover, the results show that adding more modalities increases the accuracy of the multimodal system over the performance of all the individual modalities.

4.3.6 Scalability of DCA

In this section, we evaluate the scalability of the proposed method in dealing with new subjects that were not used for training. The goal is to examine if DCA is trained on a separate a population of subjects whether the transformation matrices will still perform well on new subjects. We use a population of subjects to train DCA and obtain the transformation matrices. Another population of subjects, which which is not used for training, is used for evaluating the recognition performance.

For this purpose, we use the WVU database [95] with 399 subjects, introduced in Section 4.3.2. Similar to the experiment in Section 4.3.2, three different biometric modalities, *i.e.*, face (from frames between 45 and 90 degrees), ear and profile face

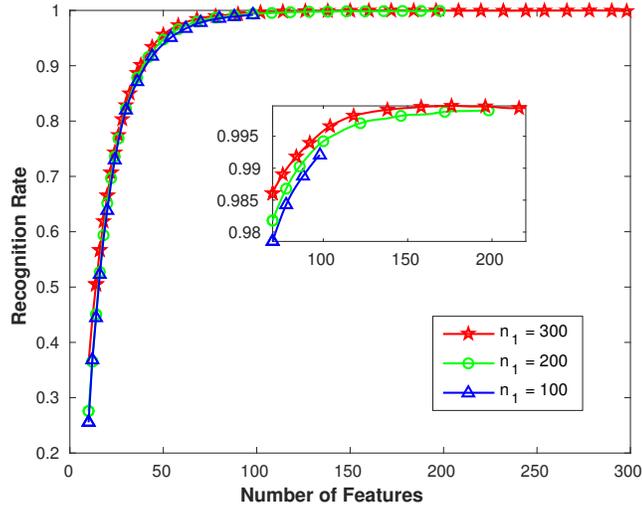


Figure 4.17: Scalability of the proposed DCA algorithm using different number of training subjects and testing on unseen populations.

(from frames between 0 and 45 degrees), are extracted from these frames. Each time we repeat the experiment, we randomly select a frame from the specified range for each modality for each subject to create the multimodal samples. A multimodal sample is a trio of a face, an ear, and a profile face image of a subject. Here, we have ten multimodal (face-ear-profile) samples per subject.

We divide the database into two populations with n_1 subjects for training the DCA and n_2 subjects for testing the performance. Five randomly selected multimodal samples from the first population, *i.e.*, training set, are used to calculate the transformation matrices of the DCA. The obtained transformation matrices are used to transform and fuse the feature sets of the second population, *i.e.*, testing set. We divide the second population into gallery and probe sets, which are used for the evaluation. Five randomly chosen multimodal (face-ear-profile) samples, for each subject, are used as gallery samples and the remaining five samples are used as probe.

Table 4.12: Maximum rank-1 recognition rates over the number of features in Fig. 4.17.

n_1	100	200	300
Recognition Rate	99.32 ± 0.085	99.89 ± 0.051	99.98 ± 0.012

Each time we repeat the experiment, we separate 99 randomly selected subjects from the database for the test population. Then, using the remaining subjects, we conduct three experiments with different number of training subjects in the first populations, $n_1 = 100, 200, 300$. In order to validate the robustness of the experiments, repeated random sub-sampling validation is applied and the results are averaged over 100 iterations. Fig. 4.17 shows the rank-1 recognition rate of the system with different number of training subjects n_1 . Table 4.12 shows the maximum recognition rate over the number of features in each case. The results show that the proposed algorithm is robust and it still performs well on new unseen subjects.

Since the maximum number of features is limited to $c - 1$, c being the number of training subjects, the three diagrams shown in Fig. 4.17 have different domains. In case of $n_1 = 100$, we are only limited to 99 features and the maximum recognition rate achieved by these features is 99.32%. The other cases use more subjects for training; therefore, not only the training becomes more robust, but also the number of features increases, *i.e.*, 199 and 299. This helps achieve higher recognition accuracies, 99.89% and 99.98%. This phenomenon is clearly shown in the magnified part of Fig. 4.17.

4.3.7 Sketch to Mugshot Matching

In this section, we present an experiment that shows the applicability of DCA in improving the accuracy of a sketch to mugshot matching technique. Matching sketches to facial photographs is a challenging face recognition problem, which assists

law enforcement to determine the identity of criminals [196]. Due to the large differences between sketches and photos and the unknown mechanism of sketch generation, it is difficult to match photos and sketches because they represent two different modalities. One way to solve this problem is to first transform a query sketch into a photo image and then match the synthesized photo with real photos in the gallery [197].

In this experiment, we use the publicly available Chinese University of Hong Kong (CUHK) face photo-sketch dataset [197]. It includes 188 faces where for each face, there is a sketch drawn by an artist and a photo taken in frontal pose under normal lighting condition and neutral expression. In this database, 88 faces are preselected for training and the remaining 100 faces are selected for testing. There is no identity overlap between the training and testing sets. Given a face sketch, we synthesize a pseudo-photo using a multiscale Markov Random Fields (MRF) model, which learns the face structure across different scales [197]. The MRF model is obtained using the training set of the CUHK database with 88 photo-sketch pairs. Pseudo-photos are synthesized for the remaining 100 sketch images in the testing set of the CUHK database⁴. Fig. 4.18 shows a sample face photo-sketch pair and the synthesized pseudo-photo.

The projection matrices of DCA are obtained using the training set of the CUHK database with 88 photo-sketch pairs. The remaining 100 real photos and the synthesized pseudo-photos are used as gallery and probe sets, respectively. Similar to the setting in Section 4.3.1, we extract Gabor and HOG features from these images and fuse them using DCA. A simple minimum distance classifier is used for recognition. Table 4.13 shows the rank-1 recognition rate and compares the performance with that

⁴For synthesizing, we used the open-source code available from [198].

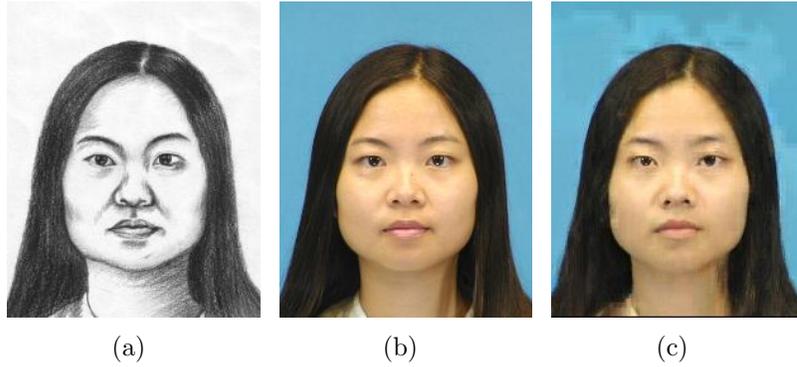


Figure 4.18: Photo synthesis result: (a) sketch drawn by the artist; (b) real photo; and (c) pseudo-photo synthesized from the sketch.

Table 4.13: Rank-1 recognition rate for photo-sketch matching in CUHK database.

Method	Ref. [197]	Ref. [199]	DCA
Recognition Rate	96.3	96	100

of [197] and the most recently published work [199]. The results show the advantages in fusing different features using DCA, as it significantly improves the sketch to photo matching accuracy.

4.4 Conclusions and Future works

In this chapter, we presented a feature fusion technique based on correlation analysis of the feature sets. Our proposed method, called Discriminant Correlation Analysis, uses the class associations of the samples in its analysis. It aims to find transformations that maximize the pair-wise correlations across the two feature sets and at the same time, separate the classes within each set. These characteristics make DCA an effective feature fusion tool for pattern recognition applications. Moreover, DCA is computationally efficient and can be employed in real-time applications. Extensive experiments on various multimodal biometric databases demonstrated the efficacy of

our proposed approach in the fusion of multimodal feature sets or different feature sets extracted from a single modality. In future, we will apply the proposed method to low-resolution face recognition problem. We aim to find transformations between high resolution and low resolution face images using the proposed DCA algorithm. This will be used in recognition of low resolution faces which are

CHAPTER 5

Low Resolution Face Recognition Using Discriminant Correlation Analysis

5.1 Motivation

Facial recognition systems have been evolving since the 1960s to automatically identify a person from a digital image, mug shot, or video. However, most of the attention, in face recognition research, focused on large enough frontal faces, which contain sufficient information for recognition. The expanding installation of surveillance cameras increases the demand for face recognition technology for surveillance applications, ranging from small-scale stand-alone camera applications in banks and supermarkets to large-scale multiple networked closed-circuit televisions in law enforcement applications in public streets. Due to the large distance between the surveillance camera and the subject, the captured images usually present very Low Resolution (LR) face regions. The discriminant properties of the face are usually degraded in the LR images, which significantly decreases the accuracy of traditional face recognition algorithms developed for High Resolution (HR) images [55]. Another problem with the LR face images is the difference in resolution of the probe and gallery images, which makes the recognition process more challenging [58].

As shown in Fig. 5.1, there are two standard approaches to address the LR face recognition problem. In the first approach, the gallery images are downsampled to the resolution of the probe images and then recognition is applied. However, in this approach, the discriminating information available in the HR gallery images is wasted. In the second approach, on the other hand, the probe images are upsampled to the resolution of the gallery images using super-resolution techniques [59–65]. In this method, the discriminating information of the HR images is preserved. However, the main goal of most super-resolution techniques is to obtain a visually appealing reconstruction, not to enhance recognition performance [66]. Therefore, not only there is no discriminating information added to the LR images, but also the redundant information may actually reduce the recognition accuracy. Recently, there have been attempts to apply super-resolution and recognition simultaneously [55, 57]. But in these approaches, given a probe image, an optimization needs to be repeated for each gallery image in the database, which results in significant computational cost, especially in big databases.

In this paper, we propose a novel approach for matching LR probe face images with HR gallery face images using *Discriminant Correlation Analysis (DCA)*. The proposed method projects the LR and HR feature vectors in a common domain in which not only the problem of feature vector length mismatch is solved, but also the correlation between corresponding features in LR and HR domain is maximized. In addition, the process of calculating the projection matrices considers the class structure of the data and it aims to separate the classes in the new domain, which is very beneficial from the recognition perspective. It is worth mentioning that the proposed method has a very low computational complexity and it can be employed

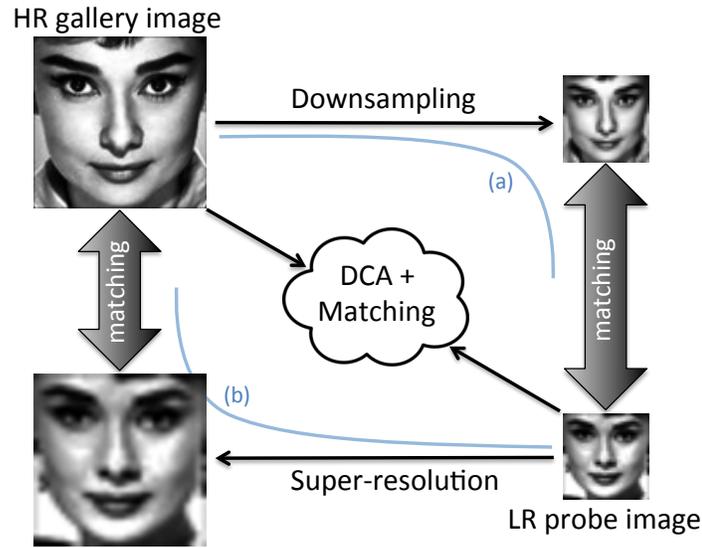


Figure 5.1: Standard approaches to matching a LR probe to a HR gallery. (a) Downsampling the gallery and then matching. (b) Upsampling the probe (interpolation or super-resolution) and then matching. In the proposed approach, we project LR and HR images to a common linear space for matching.

for processing several faces that appear in a crowded image in real-time. Extensive experiments performed on low-resolution surveillance images from the Surveillance Cameras Face (SCface) database [2] verify the effectiveness of our proposed method, which outperforms the state-of-the-art LR face recognition techniques.

The rest of the paper is organized as follows: Section 5.2 describes the proposed method for low-resolution face recognition and provides the explanation of the discriminant correlation analysis. The implementation details and experimental results on several databases are presented in Section 5.3. Finally, Section 5.4 concludes the paper.

5.2 Proposed Method

In this section, we describe the proposed method for low-resolution face recognition. Opposed to the standard approaches shown in Fig. 5.1, we do not upsample

or downsample the probe or gallery images for matching. However, we seek two linear transformation matrices W_h and W_l to project high- and low-resolution images to a common low-dimensional space, respectively. Wang *et al.* [200] use Canonical Correlation Analysis (CCA) to find such transformation matrices. In this section, we will briefly describe CCA and the major concerns about it. Then, we will describe the proposed Discriminant Correlation Analysis (DCA), which overcomes the shortcomings of CCA.

5.2.1 Canonical Correlation Analysis (CCA)

Canonical correlation analysis (CCA) is one of the valuable multi-data processing methods, which has been widely used to analyze the mutual relationships between two sets of variables. Suppose that $H \in \mathbb{R}^{p \times n}$ and $L \in \mathbb{R}^{q \times n}$ denote two matrices, each containing n training feature vectors extracted from high- and low-resolution images. Let $S_{hh} \in \mathbb{R}^{p \times p}$ and $S_{ll} \in \mathbb{R}^{q \times q}$ denote the within-sets covariance matrices of H and L and $S_{hl} \in \mathbb{R}^{p \times q}$ denote the between-set covariance matrix (note that $S_{lh} = S_{hl}^T$). The overall $(p + q) \times (p + q)$ covariance matrix, S , contains all the information on associations between pairs of features:

$$S = \begin{pmatrix} \text{cov}(h) & \text{cov}(h, l) \\ \text{cov}(l, h) & \text{cov}(l) \end{pmatrix} = \begin{pmatrix} S_{hh} & S_{hl} \\ S_{lh} & S_{ll} \end{pmatrix}. \quad (5.1)$$

However, the correlation between these two sets of feature vectors may not follow a consistent pattern, and thus, understanding the relationships between these two sets of feature vectors from this matrix is difficult [134]. CCA aims to find the linear

combinations, $H^* = W_h^T H$ and $L^* = W_l^T L$, that maximize the pair-wise correlations across the two data sets:

$$\text{corr}(H^*, L^*) = \frac{\text{cov}(H^*, L^*)}{\text{var}(H^*) \cdot \text{var}(L^*)}, \quad (5.2)$$

where $\text{cov}(H^*, L^*) = W_h^T S_{hl} W_l$, $\text{var}(H^*) = W_h^T S_{hh} W_h$ and $\text{var}(L^*) = W_l^T S_{ll} W_l$. Optimization is performed using Lagrange multipliers by maximizing the covariance between H^* and L^* subject to the constraints $\text{var}(H^*) = \text{var}(L^*) = 1$. The transformation matrices, W_h and W_l , are then found by solving the eigenvalue equations [134]:

$$\begin{cases} S_{hh}^{-1} S_{hl} S_{ll}^{-1} S_{lh} \hat{W}_h = R^2 \hat{W}_h \\ S_{ll}^{-1} S_{lh} S_{hh}^{-1} S_{hl} \hat{W}_l = R^2 \hat{W}_l \end{cases}, \quad (5.3)$$

where \hat{W}_h and \hat{W}_l are the eigenvectors and R^2 is the diagonal matrix of eigenvalues or squares of the *canonical correlations*. The number of non-zero eigenvalues in each equation is $d = \text{rank}(S_{hl}) \leq \min(n, p, q)$, which will be sorted in decreasing order, $r_1 \geq r_2 \geq \dots \geq r_d$. The transformation matrices, W_h and W_l , consist of the sorted eigenvectors corresponding to the non-zero eigenvalues. $H^*, L^* \in \mathbb{R}^{d \times n}$ are known as canonical variates. For the transformed data, the sample covariance matrix defined in Eq. (5.1) will be of the form:

$$S^* = \left(\begin{array}{cccc|cccc} 1 & 0 & \dots & 0 & r_1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & r_2 & \dots & 0 \\ \vdots & & \ddots & & \vdots & & \ddots & \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & r_d \\ \hline r_1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & r_2 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & & \ddots & & \vdots & & \ddots & \\ 0 & 0 & \dots & r_d & 0 & 0 & \dots & 1 \end{array} \right).$$

The above matrix shows that the canonical variates have nonzero correlation only on their corresponding indices. The identity matrices in the upper left and lower right corners show that the canonical variates are uncorrelated within each data set.

Two major concerns arise when using CCA. The first issue is encountered in case of a small sample size problem. In many real world applications, the number of samples is usually less than the number of features ($n < p$ or $n < q$). This makes the covariance matrices singular and non-invertible. Therefore, we will face a major problem in inverting the S_{hh} and S_{ll} matrices used in Eq. (5.3). A solution to overcome this issue is to reduce the dimensionality of the feature vectors before applying CCA. Therefore, a two stage PCA + CCA approach can be considered [97].

The second issue in CCA-based approaches is their negligence of the class structure among samples. CCA decorrelates the features, but in pattern recognition problems, we are also interested in separating the classes. The dimensionality reduction approaches based on Linear Discriminant Analysis (LDA) [174] consider this matter by

finding projections that best separate the classes. However, a *two stage* LDA + CCA will not be an effective solution due to the fact that the transformation applied by the second stage, *i.e.*, CCA, will not preserve the properties achieved by the first stage, *i.e.*, LDA. Therefore, we need transformations that not only maximize the pair-wise correlations between LR and HR features, but also *simultaneously* separate the classes within each set of features. In the next section, we present a solution to achieve this goal.

5.2.2 Discriminant Correlation Analysis (DCA)

DCA incorporates the class structure, *i.e.*, memberships of the samples in classes, into the correlation analysis, which helps in highlighting the differences between classes and at the same time maximizing the pair-wise correlations between LR and HR features.

Let's assume that the samples in the data matrix are collected from c separate classes. Accordingly, the n columns of the data matrix are divided into c separate groups, where n_i columns belong to the i^{th} class ($n = \sum_{i=1}^c n_i$). Let $h_{ij} \in H$ denote the feature vector corresponding to the j^{th} sample in the i^{th} class. \bar{h}_i and \bar{h} denote the means of the h_{ij} vectors in the i^{th} class and the whole data set, respectively. That is, $\bar{h}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} h_{ij}$ and $\bar{h} = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n_i} h_{ij} = \frac{1}{n} \sum_{i=1}^c n_i \bar{h}_i$. The between-class scatter matrix is defined as

$$S_{bh_{(p \times p)}} = \sum_{i=1}^c n_i (\bar{h}_i - \bar{h})(\bar{h}_i - \bar{h})^T = \Phi_{bh} \Phi_{bh}^T, \quad (5.4)$$

where

$$\Phi_{bh_{(p \times c)}} = [\sqrt{n_1}(\bar{h}_1 - \bar{h}), \sqrt{n_2}(\bar{h}_2 - \bar{h}), \dots, \sqrt{n_c}(\bar{h}_c - \bar{h})]. \quad (5.5)$$

If the number of features is higher than the number of classes ($p \gg c$), it is computationally easier to calculate the covariance matrix as $(\Phi_{bh}^T \Phi_{bh})_{c \times c}$ rather than $(\Phi_{bh} \Phi_{bh}^T)_{p \times p}$. As presented in [74], the most significant eigenvectors of $\Phi_{bh} \Phi_{bh}^T$ can be efficiently obtained by mapping the eigenvectors of $\Phi_{bh}^T \Phi_{bh}$. Therefore, we only need to find the eigenvectors of the $c \times c$ covariance matrix $\Phi_{bh}^T \Phi_{bh}$.

If the classes were well-separated, $\Phi_{bh}^T \Phi_{bh}$ would be a diagonal matrix. Since $\Phi_{bh}^T \Phi_{bh}$ is symmetric positive semi-definite, we can find transformations that diagonalize it:

$$P^T (\Phi_{bh}^T \Phi_{bh}) P = \hat{\Lambda}, \quad (5.6)$$

where P is the matrix of orthogonal eigenvectors and $\hat{\Lambda}$ is the diagonal matrix of real and non-negative eigenvalues sorted in decreasing order.

Let $Q_{(c \times r)}$ consist of the first r eigenvectors, which correspond to the r largest non-zero eigenvalues, from matrix P . We have:

$$Q^T (\Phi_{bh}^T \Phi_{bh}) Q = \Lambda_{(r \times r)}. \quad (5.7)$$

The r most significant eigenvectors of S_{bh} can be obtained with the mapping: $Q \rightarrow \Phi_{bh} Q$ [74]:

$$(\Phi_{bh} Q)^T S_{bh} (\Phi_{bh} Q) = \Lambda_{(r \times r)}. \quad (5.8)$$

$W_{bh} = \Phi_{bh} Q \Lambda^{-1/2}$ is the transformation that unitizes S_{bh} and reduces the dimensionality of the data matrix, H , from p to r . That is:

$$W_{bh}^T S_{bh} W_{bh} = I, \quad (5.9)$$

$$H'_{(r \times n)} = W_{bh}^T H_{(p \times n)}. \quad (5.10)$$

H' is the projection of H in a new space, where the between-class scatter matrix is I and the classes are separated. Note that there are at most $c - 1$ nonzero generalized

eigenvalues; therefore, an upper bound on r is $c - 1$ [177]. Other upper bounds for r are the ranks of the data matrices, *i.e.*, $r \leq \min(c - 1, \text{rank}(H), \text{rank}(L))$.

Similar to the above approach we solve for the data set of low resolution images, L , and find a transformation matrix W_{bl} , which unitizes the between-class scatter matrix for the low resolution images, S_{bl} and reduces the dimensionality of L from q to r :

$$W_{bl}^T S_{bl} W_{bl} = I, \quad (5.11)$$

$$L'_{(r \times n)} = W_{bl(r \times q)}^T L_{(q \times n)}. \quad (5.12)$$

The updated Φ'_{bh} and Φ'_{bl} are non-square $r \times c$ orthonormal matrices. Although $S'_{bh} = S'_{bl} = I$, the matrices $\Phi'_{bh}{}^T \Phi'_{bh}$ and $\Phi'_{bl}{}^T \Phi'_{bl}$ are strict diagonally dominant matrices ($\forall i: |a_{ii}| > \sum_{j \neq i} |a_{ij}|$), where the diagonal elements are close to one and the non-diagonal elements are close to zero. This makes the centroids of the classes have minimal correlation with each other, and thus, the classes are separated.

Now that we have transformed H and L to H' and L' , where the between-class scatter matrices are unitized, we need to make the features in one set have nonzero correlation only with their corresponding features in the other set. To achieve this, we need to diagonalize the between-set covariance matrix of the transformed data sets, $S'_{hl} = H' L'^T$. We use singular value decomposition (SVD) to diagonalize S'_{hl} :

$$S'_{hl(r \times r)} = U \Sigma V^T \quad \Rightarrow \quad U^T S'_{hl} V = \Sigma. \quad (5.13)$$

Note that H' and L' are of rank r and $S'_{hl(r \times r)}$ is nondegenerate. Therefore, Σ is a diagonal matrix whose main diagonal elements are non-zero. Let $W_{ch} = U \Sigma^{-1/2}$ and $W_{cl} = V \Sigma^{-1/2}$, we have:

$$(U \Sigma^{-1/2})^T S'_{hl} (V \Sigma^{-1/2}) = I, \quad (5.14)$$

which unitizes the between-set covariance matrix, S'_{hl} . Now, we transform the data sets as follows:

$$H^* = W_{ch}^T H' = \underbrace{W_{ch}^T W_{bh}^T}_{W_h} H = W_h H, \quad (5.15)$$

$$L^* = W_{cl}^T L' = \underbrace{W_{cl}^T W_{bl}^T}_{W_l} L = W_l L. \quad (5.16)$$

where $W_h = W_{ch}^T W_{bh}^T$ and $W_l = W_{cl}^T W_{bl}^T$ are the final transformation matrices for H and L , respectively.

It can be easily shown that the between-class scatter matrices of the transformed data sets are still diagonal; hence, the classes are separated. The between-class scatter matrix for H^* is calculated as:

$$S_{bh}^* = W_{ch}^T \underbrace{W_{bh}^T S_{bh} W_{bh}}_I W_{ch}. \quad (5.17)$$

From Eq. (5.9), $W_{bh}^T S_{bh} W_{bh} = I$ and since U is an orthogonal matrix, we have:

$$S_{bh}^* = (U\Sigma^{-\frac{1}{2}})^T (U\Sigma^{-\frac{1}{2}}) = \Sigma^{-1}. \quad (5.18)$$

Similarly, we can show that $S_{bl}^* = \Sigma^{-1}$, which is diagonal.

Fig. 5.2(a) shows the covariance between features in a transformed data set ($H^* H^{*T}$), which is a strict diagonally dominant matrix. Black color represents zero values and the elements with higher values are brighter. Fig. 5.2(a) shows that the correlation between the features of a data set is minimal. On the other hand, Fig. 5.2(b) shows the covariance between samples in a transformed data set ($H^{*T} H^*$). Being a block diagonal matrix, Fig. 5.2(b) clearly shows that the samples have higher correlation with only the ones in the same class.

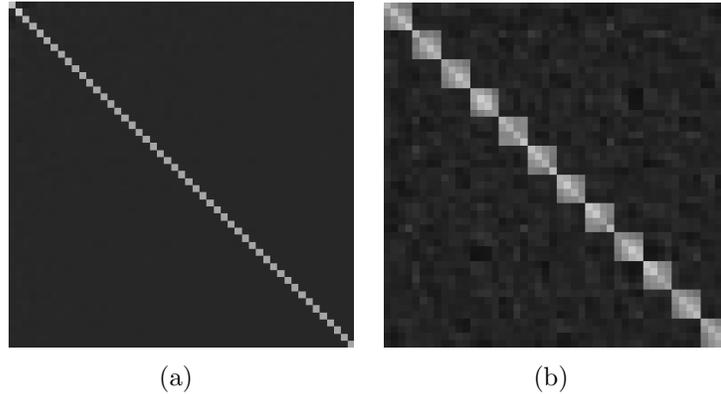


Figure 5.2: Visualization of covariance matrices (black color represents zero values and the elements with higher values are illustrated brighter). (a) Covariance between features (H^*H^{*T}). (b) Covariance between samples ($H^{*T}H^*$).

5.3 Experiments and Analysis

In this section, we present experiments to demonstrate the performance of our proposed LR face recognition technique. The experiments are performed on low-resolution surveillance images from the Surveillance Cameras Face (SCface) database [2]. SCface database has challenging face variations, which result in very low benchmark performance ranging between 0.7% and 7.7% as reported in [2]. Fig. 5.3 shows sample HR gallery and LR probe images of a few sample subjects. Most of the probe images in the database have even more extreme pose variations than the ones in these examples; however, those images could not be shown according to the database release agreement.

SCface contains images of 130 subjects taken in an uncontrolled environment using five video surveillance cameras of different qualities. Similar to a typical commercial surveillance system, the cameras are placed slightly above the subject’s head and the individuals are not looking at a fixed point during the recordings. Images are taken at three distinct distances from the cameras with the outdoor sunlight as a source

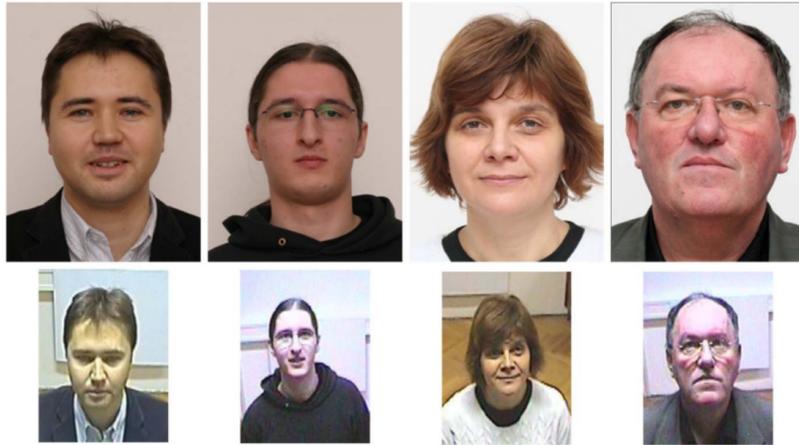


Figure 5.3: Sample images from SCface database. Top row: high-resolution gallery images. Bottom row: low-resolution probe images.

of uncontrolled illumination. Fig. 5.4 shows the camera setting and the distances at which the images are captured. The face resolution in distance 2 is obviously less than the one in distance 3, and the images captured in distance 1 are with the lowest resolution. Using the eye coordinates provided in the database, all images are scaled and rotated in a way that the distance between the two eyes is always half of the face width and the eyes lie on a straight line. We resize all the face images in distance 1, 2 and 3 to 16×16 , 32×32 , and 64×64 , respectively. The size of the HR face image is also set to 128×128 . Fig. 5.5 shows sample images of a subject captured at different distances. Please note that, the HR gallery images are captured using a high-quality digital camera.

We choose Gabor wavelets to extract features from HR and LR face images. The most important advantage of Gabor wavelets is their invariance to rotation, scale, and translation. Furthermore, they are robust against photometric disturbances, such as illumination change and image noise [73, 133]. Here, we employ forty Gabor filters in five scales and eight orientations. Since the adjacent pixels in an image are usually correlated, the information redundancy can be reduced by downsampling the feature

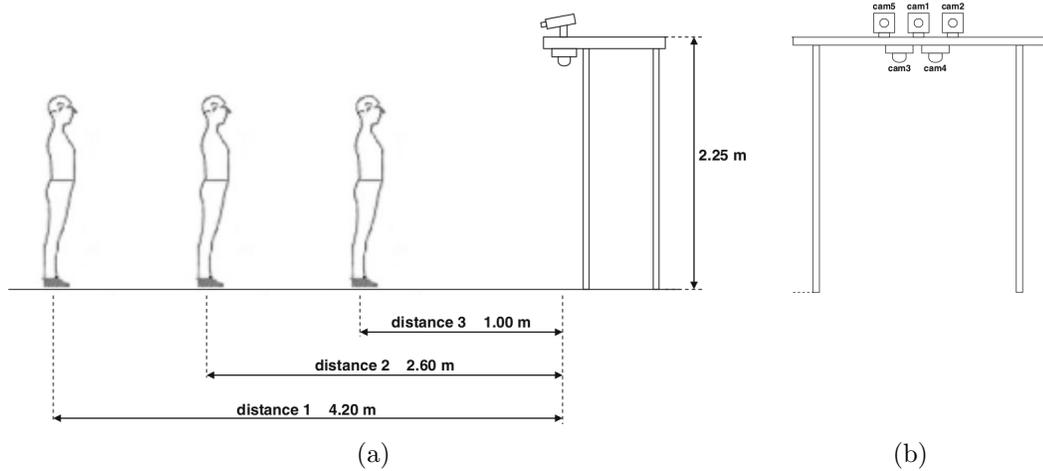


Figure 5.4: Setting of the surveillance cameras in SCface database. (a) Surveillance cameras capturing distances. (b) Set-up of the five cameras. The images are from [2].

images that result from Gabor filters [130]. Here, we downsample the feature images by a factor of four.

The images of 50 randomly chosen subjects are used for training and the images of the remaining 80 subjects are used for testing. That is, there is no identity overlap between the training and testing sets. The training set is used to obtain the projection matrices of DCA, W_h and W_l , and the testing set is used to evaluate the performance. A simple minimum distance classifier is used for classification. In order to validate the robustness of the experiments, repeated random sub-sampling validation is applied and the results are averaged over 10 iterations.

The performance of the proposed algorithm is compared with that of several state-of-the-art methods. These methods include the base eigenface approach as presented in the SCface report [2], two well-known and recently published super-resolution methods¹, Sparse representation-based Super-Resolution (SSR) [64] and Cascade Sparse

¹We have used the codes and the pretrained models available from the author's website [201].

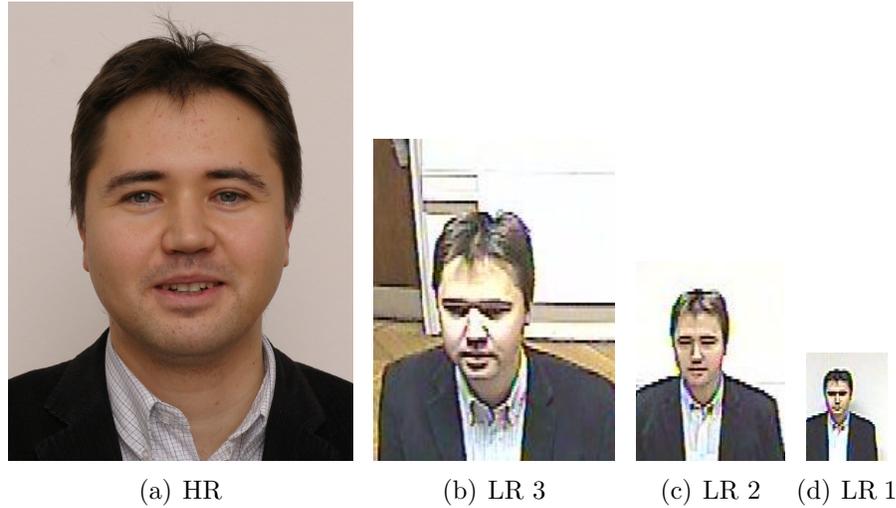


Figure 5.5: Sample images from SCface database. (a) HR gallery image (face size: 128×128). (b) LR probe image in distance 3 (64×64). (c) LR probe image in distance 2 (32×32). (d) LR probe image in distance 1 (16×16).

Coding based Deep Network (CSCDN) [65], and a recently published LR face recognition method based on CCA [200].

Table 5.1 shows the rank-1 recognition rate for fifteen sets of probe images captured by different cameras and at different distances. As mentioned before, the face images at distance 1, 2 and 3 are of sizes 16×16 , 32×32 , and 64×64 , respectively. In order to make the results easier to perceive, we present the average accuracies of all five cameras at different distances and the total average on all 650 images in Table 5.2. It is obvious that the proposed DCA method outperforms all the compared techniques. Another interesting result is that the super-resolution based approaches have similar performances in very low-resolution cases, *i.e.*, 16×16 . That is probably because neither of the super-resolution techniques adds additional information to the image that can be useful for the recognition purpose, regardless of how appealing the super-resolved image looks to human eyes.

Table 5.1: Rank-1 recognition rates for different camera/distance in SCface database.

Camera/Distance	SCface [2]	SSR [64]	CSCDN [65]	CCA [200]	Proposed DCA
cam1_1	2.3	6.50	6.21	9.15	10.37
cam1_2	7.7	14.01	15.13	16.25	21.60
cam1_3	5.4	17.50	18.13	22.12	31.04
cam2_1	3.1	7.62	7.63	11.89	12.99
cam2_2	7.7	16.88	16.62	18.80	21.31
cam2_3	3.9	18.31	19.87	19.74	25.12
cam3_1	1.5	7.38	7.50	9.43	12.20
cam3_2	3.9	9.75	10.13	11.23	16.96
cam3_3	7.7	19.63	20.12	23.54	25.54
cam4_1	0.7	7.44	7.38	9.77	12.57
cam4_2	3.9	15.25	15.62	16.24	16.52
cam4_3	8.5	18.50	18.88	19.33	23.78
cam5_1	1.5	6.25	6.25	8.72	12.81
cam5_2	7.7	10.12	10.38	11.71	15.81
cam5_3	5.4	16.50	16.87	18.74	22.15

Table 5.2: Rank-1 recognition rates for different LR image resolutions in SCface database.

LR	HR	SCface [2]	SSR [64]	CSCDN [65]	CCA [200]	DCA
16 × 16	128 × 128	1.82	7.04	6.99	9.79	12.19
32 × 32	128 × 128	6.18	13.20	13.58	14.85	18.44
64 × 64	128 × 128	6.18	18.09	18.97	20.69	25.53
total	128 × 128	4.73	12.78	13.18	15.11	18.72

5.4 Conclusions

In this paper, we presented a low-resolution face recognition technique based on Discriminant Correlation Analysis. DCA analyzes the correlation of the features in high- and low-resolution images and aims to find projections that maximize the pairwise correlations between the two feature sets and at the same time, separate the classes within each set. This makes it possible to project the features extracted from high- and low-resolution images into a common space, in which we can apply the

matching. The proposed method is computationally efficient and can be applied to challenging real-time applications such as recognition of several faces appearing in a crowded frame of a surveillance video. Extensive experiments performed on low-resolution surveillance images from the SCface database demonstrated the efficacy of our proposed approach in recognition of low-resolution faces, which outperformed other state-of-the-art techniques.

CHAPTER 6

Conclusions

In this study, four main challenges in biometrics were discussed: 1) feasibility of a secure and privacy-preserving biometric identification system in untrusted public cloud servers, 2) single sample face recognition in unconstrained environments, 3) feature-level fusion in multimodal biometric systems, and 4) low resolution face recognition in surveillance systems.

In Chapter 2, we presented a privacy-preserving cloud-based and cross-enterprise biometric identification solution. The proposed system called CloudID is the first cloud-based biometric identification system with a proven zero data disclosure possibility. In this approach, all biometric information are encrypted and the identification process is performed in the encrypted domain without decrypting the biometric data.

Unlike other privacy-preserving biometric identification methods, CloudID is not limited to any special type of biometric data and it can work with any biometric trait and feature extraction techniques. Moreover, it does not use distance-based matching, which is proven to have information leakage. However, it applies a conjunctive range query over encrypted gallery samples, which returns *true* response only if all the features of the gallery sample fall in certain ranges defined by predicates created using the query sample. This makes CloudID secure against *center search attack* in

which the attacker can recover the biometric template even if it is stored encrypted. We proposed a k-d tree structure to quantize the biometric feature vectors and define the range predicates. This structure also allows the system to handle variations in the biometric data.

CloudID enables clients to securely store their confidential information in the cloud and facilitates remote biometric-based identification by enterprises that are granted access to these confidential records. It provides a solution to the concerns about the security and confidentiality of personal information stored in the cloud through the use of biometrics, while guarding against identity theft.

We implemented a working prototype of the CloudID and evaluated it using a face biometric database. Our experimental results show that CloudID can be used in practice for biometric identification with a proven zero data disclosure. The main weakness of the CloudID is its complexity and the size of the ciphertext. In order to perform real-time identification in case of large databases, more computing resources need to be allocated by the cloud provider, which might be costly.

In Chapter 3, we proposed a single sample face recognition system for real-world applications in unconstrained environments. The potential application of this system is in many realistic scenarios like passport identification and video surveillance. The proposed system is fully automatic and robust to pose and illumination variations in face images. The system synthesizes the frontal views using a piece-wise affine warping. The warping is applied to triangular pieces determined by an enhanced AAM. The accuracy of the model fitting has direct influence on the quality of the synthesized frontal view, especially in case of semi-profile view images. In order to enhance the fitting accuracy, we initialize the AAM using estimates of the facial

landmark locations obtained by a method based on flexible mixture of parts. The fitting accuracy is further improved by training the AAM with in-the-wild images and using powerful optimization technique.

Experimental results demonstrated the efficacy of our proposed fitting approach. Histograms of oriented gradients and Gabor wavelet features are extracted from the synthesized frontal views. We use canonical correlation analysis to fuse these two features of a face image into a single but more discriminative feature vector. In contrast with other state-of-the-art methods, our system is fully automatic and does not require any manual intervention. It also handles a wide and continuous range of poses, *i.e.*, it is not restricted to any predetermined pose angles. Furthermore, it is capable of recognizing subjects from non-frontal view images and from images with different illumination using only a single image as the gallery. Experimental results performed on FERET, CMU-PIE and LFW databases demonstrated the effectiveness of our proposed method, which outperforms the state-of-the-art algorithms.

In Chapter 4, we presented a feature fusion technique based on correlation analysis of the feature sets. Our proposed method, called Discriminant Correlation Analysis, contemplates the class associations of the samples in its analysis. It aims to find transformations that maximize the pair-wise correlations across the two feature sets and at the same time, separate the classes within each set. These characteristics make DCA an effective feature fusion tool for pattern recognition applications. Moreover, DCA is computationally efficient and can be employed in real-time applications. Extensive experiments on various multimodal biometric databases demonstrated the efficacy of our proposed approach in the fusion of multimodal feature sets or different feature sets extracted from a single modality.

In Chapter 5, we applied the proposed DCA method to low-resolution face recognition problem. DCA analyzes the correlation of the features in high- and low-resolution images and aims to find projections that maximize the pair-wise correlations between the two feature sets and at the same time, separate the classes within each set. This makes it possible to project the features extracted from high- and low-resolution images into a common space, in which we can apply the matching. The proposed method is computationally efficient and can be applied to challenging real-time applications such as recognition of several faces appearing in a crowded frame of a surveillance video. Extensive experiments performed on low-resolution surveillance images from the SCface database demonstrated the efficacy of our proposed approach in recognition of low-resolution faces, which outperformed other state-of-the-art techniques.

APPENDIX

Proof of the Security of the CloudID

We review the proof of the security of CloudID's searchable encryption scheme presented in [80]. Let's define a security game in which an adversary is given a number of tokens and is required to distinguish two encrypted messages. The i^{th} experiment in the game proceeds as follows:

- **Setup** - The challenger generates the public and secret keys and PK is passed to the adversary.

$$PK \leftarrow (PK_1, PK_2, \dots, PK_t)$$

$$SK \leftarrow (SK_1, SK_2, \dots, SK_t)$$

- **Query Phase I** - The adversary adaptively requests for the tokens of the predicates $P_1, P_2, \dots, P_{q'} \in \Phi$, and the challenger responds with the corresponding tokens

$$TK_j \leftarrow GenToken(SK, P_j).$$

- **Challenge** - The adversary chooses two data-biometric pairs (M_0, B_0) and (M_1, B_1) subject to the following restrictions:

- $P_j(B_0) = P_j(B_1)$ for all $j = 1, \dots, q'$.
- If $M_0 \neq M_1$, then $P_j(B_0) = P_j(B_1) = 0$ for all $j = 1, \dots, q'$.

In i^{th} experiment, the challenger constructs the following ciphertexts:

$$C_j \leftarrow \begin{cases} Encrypt(PK_j, M_0) & \text{if } P_j(B_0) = 1 \quad \text{and } j \geq i \\ Encrypt(PK_j, M_1) & \text{if } P_j(B_1) = 1 \quad \text{and } j < i \\ Encrypt(PK_j, \perp) & \text{o/w,} \end{cases}$$

and returns $C \leftarrow (C_1, C_2, \dots, C_t)$.

- **Query Phase II** - The adversary can request more tokens for predicates $P_{q'+1}, \dots, P_q \in \Phi$ as long as they adhere to the above restrictions.
- **Guess** - The challenger flips a coin $\beta \in \{0, 1\}$ and gives $C_* = \text{Encrypt}(PK_{B_\beta}, M_\beta)$ to the adversary, who returns a guess $\beta' \in \{0, 1\}$ of β . The advantage of adversary in attacking the system is defined as

$$\text{Adv} = |\Pr(\beta = \beta') - \frac{1}{2}|.$$

If Exp^i is the probability that the adversary guesses $\beta' = 1$ in experiment i , in a chain of $t + 1$ experiments, the adversary's advantage can be calculated by the differences in the outer experiments

$$\text{Adv} = |\text{Exp}^1 - \text{Exp}^{t+1}| \leq \sum_{i=1}^t |\text{Exp}^i - \text{Exp}^{i+1}|.$$

Since the public key system is semantically secure, $|\text{Exp}^i - \text{Exp}^{i+1}|$ and consequently adversary's advantage are negligible, which make the Φ -searchable system secure.

Bibliography

- [1] S. Chikkerur, C. Wu, and V. Govindaraju, “A systematic approach for feature extraction in fingerprint images,” in *Biometric Authentication*. Springer, 2004, pp. 344–350.
- [2] M. Grgic, K. Delac, and S. Grgic, “SCface—surveillance cameras face database,” *Multimedia tools and applications*, vol. 51, no. 3, pp. 863–879, 2011.
- [3] A. K. Jain, A. Ross, and S. Pankanti, “Biometrics: a tool for information security,” *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 125–143, 2006.
- [4] A. Jain, L. Hong, and S. Pankanti, “Biometric identification,” *Commun. ACM*, vol. 43, no. 2, pp. 90–98, 2000.
- [5] Javelin Strategy & Research, “Smartphones, tablets, and fraud: When apathy meets security,” 2014, accessed: 2014-11-18. [Online]. Available: <https://www.noknok.com/sites/default/files/whitepapers/smartphonetabletsandfraudnoknokfinal.pdf>
- [6] Z. Erkin, M. Franz, J. Guajardo, S. Katzenbeisser, I. Lagendijk, and T. Toft, “Privacy-preserving face recognition,” in *Privacy Enhancing Technologies*. Springer, 2009, pp. 235–253.
- [7] A. Sadeghi, T. Schneider, and I. Wehrenberg, “Efficient privacy-preserving face recognition,” *Information, Security and Cryptology—ICISC 2009*, pp. 229–244, 2009.
- [8] M. Barni, T. Bianchi, D. Catalano, M. Di Raimondo, R. Donida Labati, P. Failla, D. Fiore, R. Lazzeretti, V. Piuri, F. Scotti *et al.*, “Privacy-preserving fingeocode authentication,” in *Proceedings of the 12th ACM workshop on Multimedia and security*, 2010, pp. 231–240.
- [9] Y. Huang, L. Malka, D. Evans, and J. Katz, “Efficient privacy-preserving biometric identification,” in *Network and Distributed System Security Symposium*, 2011, pp. 1–14.

- [10] M. Blanton and P. Gasti, “Secure and efficient protocols for iris and fingerprint identification,” in *Computer Security–ESORICS*. Springer, 2011, pp. 190–209.
- [11] M. Osadchy, B. Pinkas, A. Jarrous, and B. Moskovich, “SCiFI - a system for secure face identification,” in *IEEE Symposium on Security and Privacy (SP)*, 2010, pp. 239–254.
- [12] M. Osadchy, B. Pinkas, A. Jarrous, and B. Moskovich, “System for secure face identification (SCIFI) and methods useful in conjunction therewith,” Sep. 24 2013, uS Patent 8,542,886.
- [13] J. Bringer, H. Chabanne, and B. Kindarji, “Error-tolerant searchable encryption,” in *IEEE International Conference on Communications (ICC)*, 2009, pp. 1–6.
- [14] J. Bringer, H. Chabanne, and B. Kindarji, “Identification with encrypted biometric data,” *Security and Communication Networks*, vol. 4, no. 5, pp. 548–562, 2011.
- [15] J. Bringer, H. Chabanne, and A. Patey, “Practical identification with encrypted biometric data using oblivious ram,” in *International Conference on Biometrics (ICB)*. IEEE, 2013, pp. 1–8.
- [16] J. Bringer, H. Chabanne, and A. Patey, “SHADE: Secure HAMming DistancE computation from oblivious transfer,” in *Financial Cryptography and Data Security*. Springer, 2013, pp. 164–176.
- [17] J. Bringer, H. Chabanne, M. Favre, A. Patey, T. Schneider, and M. Zohner, “GSHADE: faster privacy-preserving distance computation and biometric identification,” in *Proceedings of the 2nd ACM workshop on Information hiding and multimedia security*. ACM, 2014, pp. 187–198.
- [18] E. Pagnin, C. Dimitrakakis, A. Abidin, and A. Mitrokotsa, “On the leakage of information in biometric authentication,” in *Progress in Cryptology–INDOCRYPT 2014*. Springer, 2014, pp. 265–280.
- [19] X. Zhang and Y. Gao, “Face recognition across pose: A review,” *Pattern Recognition*, vol. 42, no. 11, pp. 2876–2896, 2009.
- [20] Y. Moses, Y. Adini, and S. Ullman, “Face recognition: The problem of compensating for changes in illumination direction,” in *European Conference on Computer Vision (ECCV)*. Springer, 1994, pp. 286–296.
- [21] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” *ACM Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

- [22] J.-Y. Guillemaut, J. Kittler, M. T. Sadeghi, and W. J. Christmas, “General pose face recognition using frontal face model,” in *Progress in Pattern Recognition, Image Analysis and Applications*. Springer, 2006, pp. 79–88.
- [23] J. Heo and M. Savvides, “Face recognition across pose using view based active appearance models (VBAAMs) on CMU Multi-PIE dataset,” in *Computer Vision Systems*. Springer, 2008, pp. 527–535.
- [24] H. Gao, H. K. Ekenel, and R. Stiefelhagen, “Pose normalization for local appearance-based face recognition,” in *Advances in Biometrics*. Springer, 2009, pp. 32–41.
- [25] X. Chai, S. Shan, X. Chen, and W. Gao, “Locally linear regression for pose-invariant face recognition,” *IEEE Transactions on Image Processing*, vol. 16, no. 7, pp. 1716–1725, 2007.
- [26] A. Li, S. Shan, X. Chen, and W. Gao, “Maximizing intra-individual correlations for face recognition across pose differences,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 605–611.
- [27] S. Du and R. Ward, “Component-wise pose normalization for pose-invariant face recognition,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009, pp. 873–876.
- [28] V. Blanz and T. Vetter, “Face recognition based on fitting a 3D morphable model,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.
- [29] P. Breuer, K.-I. Kim, W. Kienzle, B. Scholkopf, and V. Blanz, “Automatic 3D face reconstruction from single images or video,” in *8th IEEE International Conference on Automatic Face & Gesture Recognition*, 2008, pp. 1–8.
- [30] C. D. Castillo and D. W. Jacobs, “Using stereo matching with general epipolar geometry for 2D face recognition across pose,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2298–2304, 2009.
- [31] M. S. Sarfraz and O. Hellwich, “Probabilistic learning for fully automatic face recognition across pose,” *Image and Vision Computing*, vol. 28, no. 5, pp. 744–753, 2010.
- [32] A. Asthana, C. Sanderson, T. D. Gedeon, and R. Goecke, “Learning-based face synthesis for pose-robust recognition from single image.” in *BMVC*, 2009, pp. 1–10.
- [33] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith, “Fully automatic pose-invariant face recognition via 3D pose normalization,” in *IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 937–944.

- [34] E. Mostafa, A. Ali, N. Alajlan, and A. Farag, "Pose invariant approach for face recognition at distance," in *European Conference on Computer Vision (ECCV)*. Springer, 2012, pp. 15–28.
- [35] E. A. Mostafa and A. A. Farag, "Dynamic weighting of facial features for automatic pose-invariant face recognition," in *9th Conference on Computer and Robot Vision (CRV)*. IEEE, 2012, pp. 411–416.
- [36] A. Sharma, M. Al Haj, J. Choi, L. S. Davis, and D. W. Jacobs, "Robust pose invariant face recognition using coupled latent space discriminant analysis," *Computer Vision and Image Understanding*, vol. 116, no. 11, pp. 1095–1110, 2012.
- [37] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler, "Robust face recognition for uncontrolled pose and illumination changes," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 43, no. 1, pp. 149–163, 2013.
- [38] H. T. Ho and R. Chellappa, "Pose-invariant face recognition using markov random fields," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1573–1584, 2013.
- [39] D. Yi, Z. Lei, and S. Z. Li, "Towards pose robust face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 3539–3545.
- [40] Y. Guo, X. Ding, and J.-H. Xue, "MiLDA: A graph embedding approach to multi-view face recognition," *Neurocomputing*, vol. 151, pp. 1255–1261, 2015.
- [41] S. Gao, Y. Zhang, K. Jia, J. Lu, and Y. Zhang, "Single sample face recognition via learning deep supervised autoencoders," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 10, pp. 2108–2118, 2015.
- [42] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [43] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *5th IEEE International Conference on Automatic Face and Gesture Recognition, 2002. Proceedings*, 2002, pp. 46–51.
- [44] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.
- [45] A. Ross and A. Jain, "Multimodal biometrics: An overview," in *12th European Signal Processing Conference (EUSIPCO)*, 2004, pp. 1221–1224.

- [46] X. Xu and Z. Mu, "Feature fusion method based on KCCA for ear and profile face based multimodal recognition," in *IEEE International Conference on Automation and Logistics (ICAL)*, 2007, pp. 620–623.
- [47] C. Liu and H. Wechsler, "A shape-and texture-based enhanced fisher classifier for face recognition," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 598–608, 2001.
- [48] J. Yang and J.-y. Yang, "Generalized K–L transform based combined feature extraction," *Pattern Recognition*, vol. 35, no. 1, pp. 295–297, 2002.
- [49] J. Yang, J.-y. Yang, D. Zhang, and J.-f. Lu, "Feature fusion: Parallel strategy vs. serial strategy," *Pattern Recognition*, vol. 36, no. 6, pp. 1369–1381, 2003.
- [50] Q.-S. Sun, S.-G. Zeng, Y. Liu, P.-A. Heng, and D.-S. Xia, "A new method of feature fusion and its application in image recognition," *Pattern Recognition*, vol. 38, no. 12, pp. 2437–2448, 2005.
- [51] N. M. Correa, T. Adali, Y.-O. Li, and V. D. Calhoun, "Canonical correlation analysis for data fusion and group inferences," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 39–50, 2010.
- [52] J. Yang and X. Zhang, "Feature-level fusion of fingerprint and finger-vein for personal identification," *Pattern Recognition Letters*, vol. 33, no. 5, pp. 623–628, 2012.
- [53] K.-H. Pong and K.-M. Lam, "Multi-resolution feature fusion for face recognition," *Pattern Recognition*, vol. 47, no. 2, pp. 556–567, 2014.
- [54] W.-P. Li, J. Yang, and J.-P. Zhang, "Uncertain canonical correlation analysis for multi-view feature extraction from uncertain data streams," *Neurocomputing*, vol. 149, pp. 1337–1347, 2015.
- [55] P. H. Hennings-Yeomans, S. Baker, and B. Kumar, "Simultaneous super-resolution and feature extraction for recognition of low-resolution faces," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.
- [56] Y. M. Lui, D. Bolme, B. Draper, J. R. Beveridge, G. Givens, P. J. Phillips *et al.*, "A meta-analysis of face recognition covariates," in *3rd International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*. IEEE, 2009, pp. 1–8.
- [57] W. W. Zou and P. C. Yuen, "Very low resolution face recognition problem," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 327–340, 2012.

- [58] S. Biswas, K. W. Bowyer, and P. J. Flynn, “Multidimensional scaling for matching low-resolution face images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 2019–2030, 2012.
- [59] K. Jia and S. Gong, “Generalized face super-resolution,” *IEEE Transactions on Image Processing*, vol. 17, no. 6, pp. 873–886, 2008.
- [60] C. Liu, H.-Y. Shum, and W. T. Freeman, “Face hallucination: Theory and practice,” *International Journal of Computer Vision (IJCV)*, vol. 75, no. 1, pp. 115–134, 2007.
- [61] X. Ma, J. Zhang, and C. Qi, “Hallucinating face by position-patch,” *Pattern Recognition*, vol. 43, no. 6, pp. 2224–2236, 2010.
- [62] J.-S. Park and S.-W. Lee, “An example-based face hallucination method for single-frame, low-resolution facial images,” *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1806–1816, 2008.
- [63] W. Zhang and W.-K. Cham, “Learning-based face hallucination in DCT domain,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.
- [64] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.
- [65] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, “Deep networks for image super-resolution with sparse prior,” in *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 370–378.
- [66] S. Biswas, G. Aggarwal, P. J. Flynn, and K. W. Bowyer, “Pose-robust recognition of low-resolution face images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 3037–3049, 2013.
- [67] R. Padilha and F. Pedone, “Confidentiality in the cloud,” *IEEE Security & Privacy*, no. 1, pp. 57–60, 2015.
- [68] D. A. Fernandes, L. F. Soares, J. V. Gomes, M. M. Freire, and P. R. Inácio, “Security issues in cloud environments: a survey,” *International Journal of Information Security*, vol. 13, no. 2, pp. 113–170, 2014.
- [69] K. Ren, C. Wang, Q. Wang *et al.*, “Security challenges for the public cloud,” *IEEE Internet Computing*, vol. 16, no. 1, pp. 69–73, 2012.
- [70] H. Takabi, J. B. Joshi, and G.-J. Ahn, “Security and privacy challenges in cloud computing environments.” *IEEE Security & Privacy*, vol. 8, no. 6, pp. 24–31, 2010.

- [71] B. R. Kandukuri, V. R. Paturi, and A. Rakshit, "Cloud security issues," in *IEEE International Conference on Services Computing (SCC)*, 2009, pp. 517–520.
- [72] M. Haghghat, S. Zonouz, and M. Abdel-Mottaleb, "Identification using encrypted biometrics," in *Computer Analysis of Images and Patterns (CAIP)*. Springer, 2013, pp. 440–448.
- [73] M. Haghghat, S. Zonouz, and M. Abdel-Mottaleb, "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7905–7916, 2015.
- [74] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [75] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.
- [76] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Advances in Cryptology (EUROCRYPT99)*. Springer, 1999, pp. 223–238.
- [77] A. Jain, S. Prabhakar, L. Hong, and S. Pankanti, "Filterbank-based fingerprint matching," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 846–859, 2000.
- [78] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148–1161, 1993.
- [79] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *IEEE Symposium on Security and Privacy*, 2000, pp. 44–55.
- [80] D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in *Proceedings of the 4th conference on Theory of cryptography*. Springer-Verlag, 2007, pp. 535–554.
- [81] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [82] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

- [83] L. Shen, L. Bai, and M. Fairhurst, "Gabor wavelets and general discriminant analysis for face identification and verification," *Image Vision Comput.*, vol. 25, no. 5, pp. 553–563, 2007.
- [84] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, vol. 11, pp. 467–476, 2002.
- [85] S. Meshgini, A. Aghagolzadeh, and H. Seyedarabi, "Face recognition using gabor-based direct linear discriminant analysis and support vector machine," *Computers & Electrical Engineering*, vol. 39, no. 3, pp. 727–745, 2013.
- [86] N. Nabizadeh and M. Kubat, "Brain tumors detection and segmentation in mr images: Gabor wavelet vs. statistical features," *Computers & Electrical Engineering*, vol. 45, pp. 286–301, 2015.
- [87] N. Nabizadeh, M. Kubat, N. John, and C. Wright, "Efficacy of gabor-wavelet versus statistical features for brain tumor classification in MRI: A comparative study," in *Int'l Conf. IP, Comp. Vision, and Pattern Recognition (IPCV)*, 2013, pp. 242–247.
- [88] J. K. Kamarainen, V. Kyrki, and H. Kalviainen, "Invariance properties of gabor filter-based features-overview and applications," *IEEE Transactions on Image Processing*, vol. 15, no. 5, pp. 1088–1099, 2006.
- [89] M. B. A. Haghghat and E. Namjoo, "Evaluating the informativity of features in dimensionality reduction methods," in *5th International Conference on Application of Information and Communication Technologies (AICT)*. IEEE, 2011, pp. 1–5.
- [90] P. N. Belhumeur, J. a. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [91] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Comput.*, vol. 12, no. 10, pp. 2385–2404, 2000.
- [92] University of California Corporate Compliance Policies and Procedures, "Legal medical record standards, policy no. 9420," pp. 1–19, 2008. [Online]. Available: <http://policy.ucop.edu/doc/1100168/LegalMedicalRecord>
- [93] University of Miami, "Pegasus HPC System," 2015, accessed: 2015-04-18. [Online]. Available: http://ccs.miami.edu/hpc/?page_id=4872
- [94] S. Rane, Y. Wang, S. Draper, and P. Ishwar, "Secure biometrics: Concepts, authentication architectures, and challenges," *IEEE Signal Processing Magazine*, vol. 30, no. 5, pp. 51–64, 2013.

- [95] G. Fahmy, A. El-Sherbeeney, S. Mandala, M. Abdel-Mottaleb, and H. Ammar, "The effect of lighting direction/condition on the performance of face recognition algorithms," in *SPIE Conference on Biometrics for Human Identification*, 2006, pp. 188–200.
- [96] Q. V. Le, "Building high-level features using large scale unsupervised learning," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 8595–8598.
- [97] M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi, "Fully automatic face normalization and single sample face recognition in unconstrained environments," *Expert Systems with Applications*, vol. 47, pp. 23–34, 2016.
- [98] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2005, pp. 886–893.
- [99] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *European Conference on Computer Vision (ECCV)*. Springer, 1998, pp. 484–498.
- [100] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.
- [101] A. Lanitis, C. J. Taylor, and T. F. Cootes, "A unified approach to coding and interpreting face images," in *IEEE International Conference on Computer Vision (ICCV)*, 1995, pp. 368–373.
- [102] G. J. Edwards, T. F. Cootes, and C. J. Taylor, "Face recognition using active appearance models," in *European Conference on Computer Vision (ECCV)*. Springer, 1998, pp. 581–595.
- [103] R. S. Ghiass, O. Arandjelovic, H. Bendada, and X. Maldague, "Vesselness features and the inverse compositional aam for robust face recognition using thermal ir." in *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, 2013.
- [104] M. Hasan, S. N. H. S. Abdullaha, and Z. A. Othman, "Efficient face recognition technique with aid of active appearance model," in *Intelligent Robotics Systems: Inspiring the NEXT*. Springer, 2013, pp. 101–110.
- [105] H. Van Kuilenburg, M. Wiering, and M. Den Uyl, "A model based method for automatic facial expression recognition," in *Proceedings of the 16th European Conference on Machine Learning*, ser. ECML'05. Springer, 2005, pp. 194–205.

- [106] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010, pp. 94–101.
- [107] C. Martin, U. Werner, and H.-M. Gross, "A real-time facial expression recognition system based on active appearance models using gray images and edge images," in *8th IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2008, pp. 1–6.
- [108] F. Tang and B. Deng, "Facial expression recognition using aam and local facial features," in *3rd International Conference on Natural Computation (ICNC)*, vol. 2. IEEE, 2007, pp. 632–635.
- [109] L. C. Trutoiu, J. K. Hodgins, and J. F. Cohn, "The temporal connection between smiles and blinks," in *10th IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2013, pp. 1–6.
- [110] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [111] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *European Conference on Computer Vision (ECCV)*. Springer, 2008, pp. 504–513.
- [112] H. Wang, S. Z. Li, Y. Wang, and J. Zhang, "Self quotient image for face recognition," in *IEEE International Conference on Image Processing (ICIP)*, vol. 2, 2004, pp. 1397–1400.
- [113] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision." in *IJCAI*, vol. 81, 1981, pp. 674–679.
- [114] A. B. Ashraf, S. Lucey, and T. Chen, "Fast image alignment in the fourier domain," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 2480–2487.
- [115] N. Komodakis and G. Tziritas, "Image completion using efficient belief propagation via priority scheduling and dynamic pruning," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2649–2661, 2007.
- [116] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, 2007.

- [117] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “DeepFace: Closing the gap to human-level performance in face verification,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1701–1708.
- [118] Y. Sun, Y. Chen, X. Wang, and X. Tang, “Deep learning face representation by joint identification-verification,” in *Advances in Neural Information Processing Systems*, 2014, pp. 1988–1996.
- [119] Y. Sun, X. Wang, and X. Tang, “Deeply learned face representations are sparse, selective, and robust,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 2892–2900.
- [120] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, “Bayesian face revisited: A joint formulation,” in *European Conference on Computer Vision (ECCV)*. Springer, 2012, pp. 566–579.
- [121] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823.
- [122] I. Matthews and S. Baker, “Active appearance models revisited,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.
- [123] R. Gross, I. Matthews, and S. Baker, “Generic vs. person specific active appearance models,” *Image and Vision Computing*, vol. 23, no. 12, pp. 1080–1093, 2005.
- [124] S. Baker, R. Gross, and I. Matthews, “Lucas-kanade 20 years on: A unifying framework: Part 3,” Robotics Institute, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-03-35, November 2003.
- [125] G. Tzimiropoulos and M. Pantic, “Optimization problems for fast AAM fitting in-the-wild,” in *IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 593–600.
- [126] Y. Yang and D. Ramanan, “Articulated pose estimation with flexible mixtures-of-parts,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1385–1392.
- [127] X. Zhu and D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2879–2886.
- [128] P. F. Felzenszwalb and D. P. Huttenlocher, “Pictorial structures for object recognition,” *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, 2005.

- [129] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.
- [130] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image processing*, vol. 11, no. 4, pp. 467–476, 2002.
- [131] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.
- [132] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," <http://www.vlfeat.org/>, 2008.
- [133] J.-K. Kamarainen, V. Kyrki, and H. Kalviainen, "Invariance properties of gabor filter-based features-overview and applications," *IEEE Transactions on Image Processing*, vol. 15, no. 5, pp. 1088–1099, 2006.
- [134] W. J. Krzanowski, *Principles of multivariate analysis: a user's perspective*. Oxford University Press, Inc., 1988.
- [135] M. Haghghat, M. Abdel-Mottaleb, and W. Alhalabi, "Computationally efficient statistical face model in the feature space," in *IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, 2014, pp. 126–131.
- [136] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 Faces in-the-Wild Challenge: The first facial landmark localization challenge," in *IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2013, pp. 397–403.
- [137] P. N. Belhumeur, D. W. Jacobs, D. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 545–552.
- [138] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," in *European Conference on Computer Vision (ECCV)*. Springer, 2012, pp. 679–692.
- [139] M. B. A. Haghghat, A. Aghagolzadeh, and H. Seyedarabi, "Multi-focus image fusion for visual sensor networks in DCT domain," *Computers & Electrical Engineering*, vol. 37, no. 5, pp. 789–797, 2011.
- [140] M. B. A. Haghghat, A. Aghagolzadeh, and H. Seyedarabi, "Real-time fusion of multi-focus images for visual sensor networks," in *6th Iranian Conference on Machine Vision and Image Processing (MVIP)*, 2010, pp. 1–6.

- [141] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, “Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition,” in *IEEE International Conference on Computer Vision (ICCV)*, vol. 1, 2005, pp. 786–791.
- [142] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic, “Face frontalization for alignment and recognition,” *arXiv preprint arXiv:1502.00852*, 2015.
- [143] D. Cox and N. Pinto, “Beyond simple features: A large-scale feature search approach to unconstrained face recognition,” in *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG)*, 2011, pp. 8–15.
- [144] S. U. Hussain, T. Napoléon, and F. Jurie, “Face recognition using local quantized patterns,” in *British Machine Vision Conference*, 2012, pp. 11–pages.
- [145] L. Wolf, T. Hassner, and Y. Taigman, “Similarity scores based on background samples,” in *Asian Conference on Computer Vision (ACCV)*. Springer, 2010, pp. 88–97.
- [146] X. Cao, D. Wipf, F. Wen, G. Duan, and J. Sun, “A practical transfer learning algorithm for face verification,” in *IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 3208–3215.
- [147] O. Barkan, J. Weill, L. Wolf, and H. Aronowitz, “Fast high dimensional vector multiplication face recognition,” in *IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 1960–1967.
- [148] J. Hu, J. Lu, and Y.-P. Tan, “Discriminative deep metric learning for face verification in the wild,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1875–1882.
- [149] J. Hu, J. Lu, J. Yuan, and Y.-P. Tan, “Large margin multi-metric learning for face and kinship verification in the wild,” in *Asian Conference on Computer Vision (ACCV)*. Springer, 2014, pp. 252–267.
- [150] X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li, “High-fidelity pose and expression normalization for face recognition in the wild,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 787–796.
- [151] L. Zheng, K. Idrissi, C. Garcia, S. Duffner, and A. Baskurt, “Triangular similarity metric learning for face verification,” in *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 1, 2015, pp. 1–7.
- [152] J. Lu, V. Erin Liong, and J. Zhou, “Simultaneous local binary feature learning and encoding for face recognition,” in *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 3721–3729.

- [153] A. Ouamane, M. Bengherabi, A. Hadid, and M. Cheriet, “Side-information based exponential discriminant analysis for face verification in the wild,” in *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 2, 2015, pp. 1–6.
- [154] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, “Face recognition by independent component analysis,” *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1450–1464, 2002.
- [155] T.-K. Kim, K.-Y. K. Wong, B. Stenger, J. Kittler, and R. Cipolla, “Incremental linear discriminant analysis using sufficient spanning set approximations,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–8.
- [156] G. Guo, S. Z. Li, and K. L. Chan, “Face recognition by support vector machines,” in *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 196–201.
- [157] Z. Guo, D. Zhang, and X. Mou, “Hierarchical multiscale LBP for face and palmprint recognition,” in *IEEE International Conference on Image Processing (ICIP)*, 2010, pp. 4521–4524.
- [158] S. Liao, Z. Lei, D. Yi, and S. Z. Li, “A benchmark study of large-scale unconstrained face recognition,” in *IEEE International Joint Conference on Biometrics (IJCB)*, 2014, pp. 1–8.
- [159] D. Chen, X. Cao, F. Wen, and J. Sun, “Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 3025–3032.
- [160] S. Liao, “Benchmark of large-scale unconstrained face recognition(BLUFR),” <http://www.cbsr.ia.ac.cn/users/scliao/projects/blufr/>, accessed: 2016-03-18.
- [161] K. Q. Weinberger and L. K. Saul, “Distance metric learning for large margin nearest neighbor classification,” *The Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.
- [162] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, “Large scale metric learning from equivalence constraints,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2288–2295.
- [163] Z. Lei, D. Yi, and S. Z. Li, “Learning stacked image descriptor for face recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, in press, 2016.
- [164] J. R. Kettenring, “Canonical analysis of several sets of variables,” *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.

- [165] A. A. Nielsen, "Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 293–305, 2002.
- [166] Y.-H. Yuan, Q.-S. Sun, Q. Zhou, and D.-S. Xia, "A novel multiset integrated canonical correlation analysis framework and its application in feature fusion," *Pattern Recognition*, vol. 44, no. 5, pp. 1031–1040, 2011.
- [167] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [168] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 372–386, 2012.
- [169] A. Ross and N. Poh, "Multibiometric systems: Overview, case studies, and open issues," in *Handbook of Remote Biometrics*. Springer, 2009, pp. 273–292.
- [170] S. Shekhar, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Joint sparse representation for robust multimodal biometrics recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 113–126, 2014.
- [171] S. Bahrapour, N. M. Nasrabadi, A. Ray, and W. K. Jenkins, "Multimodal task-driven dictionary learning for image classification," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 24–38, 2016.
- [172] M. Haghghat, M. Abdel-Mottaleb, and W. Alhalabi, "Discriminant correlation analysis for feature level fusion with application to multimodal biometrics," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 1866–1870.
- [173] M. Haghghat, M. Abdel-Mottaleb, and W. Alhalabi, "Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition," *IEEE Transactions on Information Forensics and Security*, 2016, <http://dx.doi.org/10.1109/TIFS.2016.2569061>.
- [174] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [175] T.-K. Kim, J. Kittler, and R. Cipolla, "Discriminative learning and recognition of image set classes using canonical correlations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1005–1018, 2007.

- [176] Y. Ma, S. Lao, E. Takikawa, and M. Kawade, "Discriminant analysis in correlation similarity measure space," in *Proceedings of the 24th international conference on Machine learning (ICML)*. ACM, 2007, pp. 577–584.
- [177] R. O. Duda and P. E. Hart, *Pattern classification and scene analysis*. Wiley New York, 1973.
- [178] S. Crihalmeanu, A. Ross, S. Schuckers, and L. Hornak, "A protocol for multibiometric data acquisition, storage and dissemination," *Technical Report, WVU, Lane Department of Computer Science and Electrical Engineering*, 2007.
- [179] A. M. Martinez and R. Benavente, "The AR face database," *CVC Technical Report*, vol. 24, 1998.
- [180] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.
- [181] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [182] Y.-O. Li, T. Adali, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, 2009.
- [183] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, "A survey on ear biometrics," *ACM Computing Surveys (CSUR)*, vol. 45, no. 2, p. 22, 2013.
- [184] J. Lei, J. Zhou, and M. Abdel-Mottaleb, "Gender classification using automatically detected and aligned 3D ear range data," in *International Conference on Biometrics (ICB)*. IEEE, 2013, pp. 1–7.
- [185] B. Krishnapuram, L. Carin, M. A. Figueiredo, and A. J. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 957–968, 2005.
- [186] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.
- [187] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.
- [188] S. J. Pundlik, D. L. Woodard, and S. T. Birchfield, "Non-ideal iris segmentation using graph cuts," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2008, pp. 1–6.

- [189] L. Masek and P. Kovesi, "Matlab source code for a biometric identification system based on iris patterns," *The School of Computer Science and Software Engineering, The University of Western Australia*, vol. 26, 2003.
- [190] Ear Recognition Laboratory at University of Science & Technology Beijing (USTB), "USTB ear image database," <http://www1.ustb.edu.cn/resb/en/index.htm>, accessed: 2015-03-18.
- [191] Z. Huang, Y. Liu, C. Li, M. Yang, and L. Chen, "A robust face and ear based multimodal biometric system using sparse representation," *Pattern Recognition*, vol. 46, no. 8, pp. 2156–2168, 2013.
- [192] U. Park, R. Jillela, A. Ross, and A. K. Jain, "Periocular biometrics in the visible spectrum," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 96–106, 2011.
- [193] A. Moorhouse, A. N. Evans, G. Atkinson, J. Sunf, and M. Smith, "The nose on your face may not be so plain: Using the nose as a biometric," pp. 1–6, 2009.
- [194] M. Balasubramanian, S. Palanivel, and V. Ramalingam, "Real time face and mouth recognition using radial basis function neural networks," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6879–6888, 2009.
- [195] H. Li, K.-A. Toh, and L. Li, *Advanced topics in biometrics*. World Scientific, 2012.
- [196] H. Han, B. F. Klare, K. Bonnen, and A. K. Jain, "Matching composite sketches to face photos: A component-based approach," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 191–204, 2013.
- [197] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, 2009.
- [198] J. Xie, "MATLAB implementation of converting face to sketch and vice versa," <https://github.com/ClaireXie/face2sketch>, accessed: 2016-02-24.
- [199] R. Srinivasan and A. Roy-Chowdhury, "Robust face recognition based on saliency maps of sigma sets," in *IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2015, pp. 1–6.
- [200] Z. Wang, W. Yang, and X. Ben, "Low-resolution degradation face recognition over long distance based on CCA," *Neural Computing and Applications*, vol. 26, no. 7, pp. 1645–1652, 2015.
- [201] J. Yang, "Jianchao Yang's home page," <http://www.ifp.illinois.edu/~jyang29/>, accessed: 2016-03-17.