USC–SIPI REPORT #436

FACIAL IDENTITY RECOGNITION AND ATTRIBUTE CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

By

Chun-Ting Huang

May 2017

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
USC Viterbi School of Engineering
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Suite 400
Los Angeles, CA 90089-2564 U.S.A.

FACIAL IDENTITY RECOGNITION AND ATTRIBUTE CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

by

Chun-Ting Huang

A Dissertation Presented to the
FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(ELECTRICAL ENGINEERING)

May 2017

Contents

List of Tables

List of Figures

Abstract

Robust face recognition plays a central role in biometric and surveillance applications. Although the subject has been studied for about four decades, there still exist quite a few technical challenges and system design issues in deploying it in a real-world video surveillance environment. Nowadays, the raw face images and their associated meta data are stored in a remote cloud storage system in a distributed face recognition platform. One key challenge in the overall system design is to ensure the security of stored data. In this research, we first conduct a survey on this technology and then, study the problems of cross-distance/environment face recognition and facial attribute classification with machine learning techniques.

The problem of long distance face recognition and attribute classification arising from surveillance applications impose major challenges. The captured face from the surveillance system can be low resolution and quality, which is further degraded by an uncontrolled outdoor environment such as long distance during daytime or nighttime. In addition, human age/gender inferred by face images are fundamental attributes in our social interactions. This research has many applications such as demographics analysis, commercial user management, visual surveillance, and even aging progression. Despite the rapid development in automatic face recognition, there is far less work on automatic age/gender classification in an unconstrained environment.

Research in this dissertation provides effective solutions to three topics: 1) cross-distance/environment face recognition, 2) cross-distance/spectral face recognition and 3) age/gender classification. For Topic 1, a two-stage alignment/enhancement filtering (TAEF) method is proposed to achieve the state-of-the-art performance. For Topic 2, a locally linear embedding (LLE) method is developed to restore low quality near-infrared (NIR) images so as to enhance the face recognition performance. For Topic 3, a new framework based on the convolutional neural network (CNN) is presented to achieve efficient age and gender classification using the information from full face and facial components.

Chapter 1

Introduction

1.1 Significance of the Research

With a rapid growth of security demand, authorities in many countries adopt video surveillance systems to enforce traffic management and monitor possible threats and criminal scenes. For instance, London has installed about half-million cameras to observe the public space, and major cities in the United States such as New York and Los Angeles are extending their surveillance networks largely in response to the 9/11 attack. As the result, the fast increasing volume of large video data has become more and more difficult to manage. Incidents like Boston Marathon bombing require heavy manual labor to examine low-quality face images within the camera's footage. The method to improve efficiency and accuracy of automatic face recognition in the context of video surveillance imposes a major challenge in computer vision research.

Although there have been progressive developments in the automatic face recognition, most state-of-the-art methods focus on faces with variant poses yet at a close distance and with sufficient quality. This situation leaves a noticeable gap between the clear image within datasets and the real footage obtained from surveillance systems. For example, face alignment plays an essential role if feature descriptors are applied for matching, thus the mainstream up-to-date researches on face alignment is dedicated on faces under unconstrained environment as known as Labeled Faces In-the-Wild (LFW) [66]. On the other hand, little research has been conducted to recognize low quality face images at a long distance. This issue has been defined as face recognition at a distance

(FRAD). As shown in various datasets, most surveillance footages taken from long distance can be easily distorted by illumination and polarization, which severely affect the system's performance.

Even if the FRAD problem is solved under visible light (VIS) environments, the ideal surveillance system should be able to operate round-the-clock service despite the environment. Currently cameras equipped with flash lights are served as the expediency for nighttime scenarios, but it is not appropriate to apply for long distance or convert surveillance. Therefore, we need to consider other options for nighttime face recognition. Methods like near infrared (NIR), shortwave infrared (SWIR), and thermal have been studied and applied in previous literatures. However, NIR has become more and more popular in recent years because of following reasons. First, NIR is not visible to human eyes, so it is natural to capture face expressions without interrupting subjects during collection. Second, environment factor gives less impact to NIR comparing to others. Third, NIR illuminator can penetrate glasses easily, which provides additional information if test subjects wear glasses. Fourth, NIR related equipments cost less than other image capturing devices, meaning they are more accessible to general public. Fifth, NIR images can be integrated into many established dataset such as mugshot or driver license dataset. Law enforcement departments can directly enforce cross-spectral matching, which makes NIR as a more feasible way to put into practice.

Furthermore, the age/gender attributes play an important role in human identity recognition. Automatic estimation of human age/gender can find many practical applications. Extensive research on this topic has been conducted for more than three decades. However, there was a big gap between traditional age/gender datasets and actual scenarios encountered in a real world environment. Many variations exist in human faces due to varying image quality, face poses, occlusion, and expression in practical applications.

Clearly, both the FRAD problem over daytime/nighttime and the age/gender classification problem in an unconstrained environment are challenging. They impose a high barrier on human face and attributes recognition systems. The following three solutions are proposed in this thesis research to address them.

- a long-distance daytime face recognition system using the two-stage alignment/enhancement filtering;
- an NIR image restoration using the locally linear embedding model;
- a convolutional neural network (CNN) based age/gender classification system.

We give a brief overview on each of them below.

1.1.1 Two-Stage Alignment/Enhancement Filtering

For the first topic, we provide a systematic solution called the Two-Stage Alignment/Enhancement Filtering (TAEF) system to address the FRAD problem during daytime. The TAEF incorporates multiple functions, from preprocessing like alignment and enhancement, to matching gallery set with long distance probe face images. Instead of only focusing on traditional feature extraction and comparison, the TAEF contains not only alignment structure and enhancement strategy in coarse-scale stage, but also face matching with refined candidate pool in fine-scale stage. In the coarse-scale stage, the alignment structure incorporates iterations of training using feature descriptors, where regressors learn estimations from the feedback provided by the Euclidean distance between prediction and ground truth, then regressors can provide parameters according to different levels. Moreover, the fine-scale stage of the TAEF is designed to optimize the recognition process in each level. The scheme utilizes voting mechanism with feature descriptors extracted from different face regions, and it carries top candidates with higher votes to the next level. Through this repetitive operation, the system can remove less possible candidates from calculated-weighted votes.

1.1.2 Image Restoration via Locally Linear Embedding

Although the TAEF system can manage the cross-distance issue with face images captured from VIS camera, it is difficult to apply the same rule to NIR images because of the spectrum difference. In other words, it does not work well for the FRAD problem during nighttime. For the second topic, we propose a restoration system that significantly improves the quality of NIR face images at a distance to bridge the gap between VIS and NIR. The restoration system is developed based on Locally Linear Embedding (LLE) that reconstructs image patches learned from two manifolds, and we also preserve image's local characteristic by constraining the reconstruction process within certain region, so that the recovered information will not be affected by other areas. For instance, if we aim to restore the areas around eyes, we will only refer to other eyes within the gallery set. The experiment is also conducted on LDHF dataset with NIR images taken at cross-distance during nighttime.

1.1.3 Age/Gender Attributes Classification

For the third topic, we present a CNN solution based on the Adience dataset [43], which was built recently for the age/gender classification in an unconstrained environment. The proposed Whole-Component cascaded CNN (WC-CNN) system consists of four building modules: 1) the face and facial components localization module, 2) the whole face network, 3) the facial component networks, and 4) the final classification module assisted by the confidence analysis. Each module is designed to serve a different purpose. The localization module takes care of all preprocessing tasks, such as face detection and facial landmark localization. It is used to localize the face and its component regions. The whole face network and the facial component networks are trained separately with extracted patches for age/gender classification. We use the whole face network as the primary classifier to yield the initial classification result. Finally, we use the confidence analysis to evaluate the confidence level of the initial decision. If the confidence level is high, we accept its decision. If the confidence level is low, the system will make a final decision by considering the outputs from the whole face network and the component networks jointly.

The above three-mentioned methods will be elaborated in Chapter 3, Chapter 4 and Chapter 5, respectively.

1.2 Review of Related Work

In this section, we provide a brief review on datasets and previous work that are closely related to our research.

1.2.1 Cross Distance/Spectral Face Datasets

There are few FRAD datasets accessible to the public. The UTK-LRHM dataset [176], which was built in 2008, contains 55 subjects in an indoor environment with distances ranging from 10 to 16 meters and 48 subjects in an outdoor environment with distances from 50 to 300 meters. Another dataset, built by Rara *et al.* [129] for stereo reconstruction in 2009, has 30 subjects with three distances (i.e., 3, 15, and 33 meters). Tome *et al.* [153] evaluated distance degradation using standard approaches and matchers on the ”Face still dataset” of the NIST Multiple Evaluation Grand Challenge (MBGC) [124]. Three cross-distance/spectral datasets have been released since 2011. The first one is called the Near-Infrared Face Recognition at a Distance (NFRAD) proposed by Maeng *et al.* [102]. The NFRAD dataset has 50 subjects taken VIS and NIR photos under controlled environment, and the captured distance includes 1 meter and 60 meter. This dataset provides some degree of pose change (frontal view, slight left and right face view angle), but the NIR illuminator causes halo like light pattern around the 60 meter subject, which sets the limitation for this dataset. The second dataset, proposed by Bourlai *et al.* [15], has 103 subjects in both indoor and outdoor environment. The NIR image's quality is slightly better than previous dataset, but the resolution is way lower than other two sets. At last, the LDHF dataset [103] gives 100 subjects with well-improved VIS and NIR capturing quality, and it is the only open-access dataset that is available to the public among those three datasets. Because of its high-resolution and completed distance set, we select the LDHF dataset to evaluate our system.

1.2.2 Cross Distance/Spectral Face Recognition

There are two recent works conducted on the LDHF dataset, Maeng *et al.* [102] and Kang *et al.* [78]. Maeng *et al.* proposed to apply Gaussian smoothing and histogram equalization as the preprocessing step, and then the Dense Scale Invariant Feature Transform (Dense-SIFT) [99] was extracted from 32 × 32 overlapping patches. Afterward, each patch is divided into 4 × 4 grids, where an 8-bin gradient orientation histogram was calculated to form a 128 dimension feature vector. The matching distance between the two feature vectors for VIS-to-VIS was the Euclidean distance, and Linear Discriminant Analysis (LDA) was added for the NIR-to-VIS matching. In addition, LDA projection matrices were learned from another dataset: CASIA HFB [93], so that feature vectors from LDHF's NIR images can be projected through the learned matrices. Kang *et al.* proposed a general framework to achieve cross-distance and cross-spectral matching. LLE was adopted in their work to map the long distance face image into short distance one. The learning was accomplished by collecting random sample patches from the whole face region, and then directly built the dictionary via high-quality and low-quality patches extracted from the same location. By referring to the dictionary, the input low-quality patch can be reconstructed through nearby high-quality neighbor patches. The result of Multiscale Local Binary Pattern (MLBP) [95] and SIFT extracted from preprocessing filters were applied for the score-level fusion in the final face matching step.

1.2.3 CNN-based Age/Gender Classification

Early age estimation work was conducted by extracting effective facial features according to their geometric distance [86], which was followed by difference models to estimate and classify target images into different age groups [128]. These methods depend on landmark localization to obtain results. However, these algorithms fail to provide accurate prediction because of unconstrained face images. Recently, there has been work applying the CNN to age estimation. Yi *et al.* [177] introduced the CNN to this problem using a subset of MORPH II to train a shallow network with only one convolutional layer. Wang *et al.* [168] treated the CNN as a feature extraction tool without fully utilizing its strength. Niu *et al.* [112] proposed multiple output CNNs with ordinal regression to achieve end-to-end learning for age estimation. However, all the reported results do not surpass Liu's result in [98] on the MORPH II dataset. There has been some progress in the development of new age and gender datasets in recent years. The Adience dataset [43] was proposed in 2014 aiming to narrow the gap between the dataset and the practical applications. It was built with raw smart-phone uploaded photos without further manual manipulation so that images inside the Adience dataset cover a wide range of postures, expressions, occlusions, and even quality variations.

1.3 Contributions of the Research

There are three main contributing chapters in this dissertation.

The contributions of Chapter 3 are given below.

- An automatic face alignment method is developed to handle alignment errors caused by the long distance effect, which is not covered by traditional face alignment papers with both probe and gallery images are both captured in a short distance under visible light.

- We propose the use of the MSRCR for face enhancement against harsh environments such as foggy and back-lighted conditions in the outdoor environment and demonstrate its effectiveness in comparison with several other enhancement methods. To the best of our knowledge, this is the first time for MSRCR to be introduced to the context of face alignment/recognition system.

- TAEF adopts a two-stage filtering mechanism: I) initial screening and II) iterative refinement. At the initial screening stage, face alignment is executed against the whole gallery image set to eliminate unlikely candidates at once for efficiency and only a few candidates are kept in the candidate pool. Then, at the iterative refinement stage, the alignment is conducted for every individual probe/gallery image pair for higher accuracy. The size of the candidate set is reduced one by one iteratively until only one candidate is left. With the two-stage filtering, TAEF strikes a balance between efficiency and accuracy.

The contributions of Chapter 4 are detailed as follows.

- We adopt the LLE system that improves the low-quality face images through mapping by structure and locality of extracted patches. The restored image possesses high-quality feature descriptors that show the applicability of the restoration system over cross-spectral and cross-distance environments.

- A grid-structured sampling strategy is adopted in the restoration system and, thus, the regional information can be preserved in the LLE mapping process. Furthermore, with overlapping face patches, each extracted grid covers the corresponding area with an inch-by-inch search, which helps restore face patches with all possible details from other subjects. The result of the proposed method outperforms some deep learning methods.

The contributions of Chapter 5 include the following.

- A novel age/gender classification system is proposed to handle faces in an unconstrained environment, which simulates real-world applications. The system locates facial components from the face region first, then a number of small networks are trained using the extracted patches. After that, a filtering mechanism based on the confidence analysis is adopted to conduct coarse-to-fine matching.
- We adopt a deep learning framework that applies both the whole face and the facial components networks to the same input image so as to maximize the integrated strength.
- Each CNN network requires only three or fewer convolutional layers, which is efficient to train and suitable for systems with limited resources.

1.4 Organization of the Dissertation

The rest of the dissertation is organized as follows. In Chapter 2, background and related literature of the secure data on cloud are reviewed and compared. In Chapter 3, we propose the TAEF system and evaluate its performance on LDHF dataset's VIS face images. The restoration system is presented and discussed in Chapter 4 in order to assist solving the FRAD problem during nighttime. Furthermore, the age and gender classification using component-based neural network is brought up in Chapter 5. Finally, concluding remarks and future works are given in Chapter 6.

Chapter 2

Survey on Secure Cloud Data Storage

Nowadays, the raw face images and their associated meta data are stored in a remote cloud storage system in a distributed face recognition platform. One key challenge in the overall system design is to ensure the security of stored data. We conduct a survey on this technology in this chapter.

Fast advances in broadband communication and high speed packet switching networks have made large file sharing much more effective during the last two decades. Consequently, the demand for rich media applications, such as multimedia mails, orchestrated presentations, high-quality audio and video sharing, collaborative documents, has grown tremendously. The amount of data and computing resources being used by those applications have also grown exponentially. As a result, the costs of IT service and support, such as investment in new hardware and software, staffing for installation and maintenance are rising consistently for both enterprises and individual users. Therefore, cloud computing has become an appealing new model of IT service provisioning and support driven by economic and productivity advantages. Instead of investing in new hardware and software, as well as maintaining those resources, users can use applications, infrastructures, servers, storage, network, and other computing resources that are available in the ’cloud’, which is a shared pool of computing resources that can be easily accessed through broadband network connections. This new IT service provisioning model offers users seemingly unlimited computing resources without up-front acquisition and/or sustaining maintenance costs. Moreover, it offers on-demand elasticity and flexibility in using computing resources. The utility pricing model allows users to pay for their actual usage only.

Storage, as one of the most influential and demanding computing resources in current digital era, is among the first being moved into the cloud. This type of cloud computing services, known as cloud storage, represents a business model in which the service provider rent spaces in their large scale storage infrastructure to organizations and individuals. It has always been one of the most prevalent services in cloud computing industry. As an extension of traditional data center or file hosting service into cloud, cloud storage has distinct characteristics including on-demand self-service, broadband network access, resource multiplexing, rapid elasticity and measured usage for utility billing. Besides the key advantages of cost saving, cloud storage can facilitate information sharing and task collaborating, promote portability and universal accessibility of data, as well as provide easy and convenient solutions to some other problems. For example, for disaster recovery purpose, organizations should maintain secondary off-premise data backups. Storage of sensitive data, such as financial, personal, or medical data are subject to more and more regulations and legal constraints. Cloud storage offered by a regulation-complied service provider can relieve data owners from the complicated process.

However, the promising new paradigm of cloud computing brings up unique challenges in terms of performance, availability, security, and scalability (known as PASS). Among these challenges, security issues have been reported as the biggest concern preventing enterprises and organizations from adopting cloud services according to recent researches [22]. Therefore, it is imperative to provide security strategies, tools, and mechanisms that meet user's requirements in the cloud. Security in cloud computing is a complex issue spanning across many aspects including physical security, infrastructure (distributed computers, servers and other hardware) security, data security, network security, software security. Moreover, it involves shared responsibilities and obligations among the constituents of the cloud service. Security enforcement would not be successful without agreement, trust, regulations and coordination among service providers and cloud users.

Since storage is one of the necessary core infrastructure in clouds, security of data in storage is one of the key concerns of any cloud computing systems, particularly in cloud storage services. The consequences of security breaches in cloud storage could be seriously damaging to both service providers and users. Without trust from users, the service provider could lose their customers. On the other hand, users whose valuable data lost, or sensitive information hacked could experience irrecoverable loss or damage. There have been many cases reported as threats of cloud storage security. Many leading service providers, including Amazon, Window and Google, encountered disconnections of their web-based cloud services due to different reasons such as power failure, hardware and software failures. For instance, Amazon Web Service's server was hit by lightning, causing destruction on power generator. Although Amazon successfully transferred data to a backup server, the service still stopped after their Uninterruptible Power Supply (UPS) went out. There was bulk email deletions in Gmail happened in 2006; numerous users found that they lost their emails and contact information without further notification from Google. Google was unable to restore the accounts after users responded the problem. Another incident happened recently in July, 2012. Because of a security loophole on the access control, Dropbox, a popular cloud storage service, was attacked by hackers. Some users reported that they received tons of spam emails, and some users' passwords were even leaked.

Although the security requirements for cloud storage vary with different applications and users, they share the same three basic objectives as any computer information systems [57]: integrity, confidentiality and availability. Many different tools have been developed to achieve these objectives, such as authentication, access control, encryption, certification, audition, digital signature. This chapter aims at providing a thorough study on recent data security mechanisms developed for the cloud storage. Based on the results of the study, we give our insights and suggestions on the future research directions in achieving each security objectives.

The rest of this chapter is organized as follows. The models of cloud storage systems and related security implications are firstly introduced in Section 2.1. A general conceptual system architecture model of the cloud storage is also proposed to address security issues in different layers. In Section 2.2, recent researches on data integrity protection such as proofs of retrievability and third party audition are reviewed and compared. In Section 2.3, data confidentiality and its related research work are discussed. A promising new encryption technique, fully homomorphic encryption, which allows algebraic operations performed on encrypted data, is examined in detail first, followed by discussion on access control and searchable encryption. In Section 2.4, we probe methods for ensuring data availability in distributed cloud storage systems, such as data synchronization, data recovery and information dispersal algorithms. At last, concluding remarks are given in Section 2.5.

2.1 Overview of Security in Cloud Storage

The scope and requirements for cloud security vary significantly with different cloud deployment model. National Institution of Standards and Technology (NIST) has defined [106] four deployment models of cloud computing. *Private cloud* is provisioned for exclusive use by a single organization comprising multiple users. *Community cloud* is provisioned for exclusive use by a specific community of users from multiple organizations that have shared concerns. *Public cloud* is provisioned for open use by the general public. *Hybrid cloud* is a combination of at least two of the above three. Apparently, the price of deployment decreases from private cloud to public cloud at the cost of increasing security concerns.

Cloud storage can be deployed in any of the four deployment models. Internet-based public cloud storage services are rapidly growing because they are able to provide users with biggest cost saving and most elasticity. Numerous storage service providers, including Amazon, IBM, Google, Microsoft, EMC, HP, Symantec, Rackspace, to name just a few, are competing in this enormous market. However, these public cloud storage services also face highest potential risks of security breaches because the shared infrastructure is open to the public. As a matter of fact, cloud storage systems deployed in other forms of multi-tenancy clouds, including hybrid clouds and community clouds, are also exposed to higher risks than private cloud. Even in the private cloud, it is highly likely that cloud storage is managed and operated by a service provider off premises, in order to take the most advantage of the cloud computing. In fact, some cloud storage usage, such as disaster recovery backup, requires off-premises storage. When data are no longer stored and managed by the data owner on its own premises, the data owner has less control over their data. Therefore, cloud storage security is challenging if the service providers are not trusted, regardless of the deployment model.

NIST has also defined three primary cloud service models. Software as a Service (SaaS) implies consumers utilize service provider's application running on cloud infrastructure, such as SalesForce CRM, YouTube, Google Apps (Gmail, Google Document).

Platform as a Service (PaaS), means the service provider builds an environment for consumers to establish acquired applications with programming languages, libraries and tools that are already supported in the platform. Famous PaaS includes Google App Engine, Microsoft Azure and Cloud Foundry from VMware. Infrastructure as a Service (IaaS), represents consumers can deploy and manage application, operating system with provided network and storage devices. Amazon's Elastic Compute Cloud (EC2) is a leading example, with other offerings like Rackspaces Mosso and GoGrid's ServePath [106]. From SaaS to PaaS, and to IaaS, users have progressively deeper control over the stack of cloud architecture, thus share more responsibility on security enforcement.

Basic cloud storage services are categorized as IaaS service model, although many cloud storage providers are offering value-added PaaS and SaaS services built upon their baseline IaaS services. As an IaaS, cloud storage allows users to strengthen the security measure using their own security protection mechanisms. For example, users can encrypt their data before moving them into the cloud storage using a private key managed by themselves. In this case, even if the data was accessed by unauthorized parties, the sensitive information would not be revealed without obtaining the key. However, users of SaaS services can only rely on the service provider's security measures.

The basic architecture of a cloud storage system is composed of a storage resource pool, including the distributed file system, the Service Level Agreements (SLA), and service interfaces [71, 182]. In order to conceptually understand the cloud storage systems, and how security protection mechanisms could be integrated and implemented in the system, we decompose the system architecture into a three layer reference model based on the logical function boundaries as shown in Fig. 2.1.

In physical storage infrastructure layer, there are distributed wired and wireless networks connecting a distributed storage device network. The second layer is storage management layer, which processes necessary operations, such as data placement, replication, and reduction, on the stored data in the first layer. By means of virtualization technology, this layer becomes the intelligent abstraction layer which hides the complexity of the underlying layer. The service interface layer provides the interface for users to access their data stored in the cloud storage. Basic cloud storage systems mostly provides either a client-side software or a web browser interface, or sometimes both. Client-side software has to be installed on the user's devices used to access the data, while a web browser interface allows access of data from any place without local installations. Some advanced cloud storage systems also provide an Application Programming Interface (API), which can be used to directly integrate access of stored data into other applications. Most of those applications belong to PaaS or SaaS based on the cloud storage infrastructure.

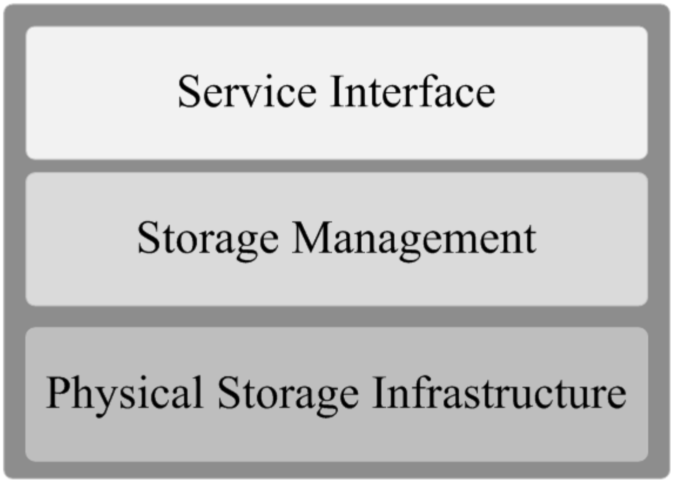

Figure 2.1: Cloud storage architecture

Since different layer has different functionalities, the security concerns in each layer have different emphasis. The physical storage infrastructure layer deals with physical and hardware security. The storage management layer should efficiently control the resource allocation and reliably perform data management. In the service interface layer, how to avoid the encroachment on rights of both clients and service providers using secure interfaces and APIs has been extensively discussed. In every layer, there could be risks, intrusions, and attacks against data integrity, confidentiality and/or availability. Therefore, storage security protection mechanisms should be integrated into every layer, and the security objectives can not be achieved without the collaborated efforts across all three layers. For example, to ensure data availability under any circumstances such as hardware failure or disasters, the physical storage infrastructure layer usually have duplicated data stored at different locations. In case data stored in one location was lost, the storage management layer should be able to locate the available data in another location and route it to the users upon their request. The service interface layer should be able to effectively receive incoming requests from anywhere and provide reliable access method to legitimate users.

Given the architecture overview of cloud storage and its security implications, we will discuss recent research efforts in achieving the three main security objectives, namely, data integrity, data confidentiality, and data availability, in the following three sections, respectively.

2.2 Data Integrity

Data integrity refers to the property that data has not been altered or destroyed in an unauthorized manner [172]. In cloud storage, since users no longer possess the physical storage of their data, how to efficiently verify the correctness of outsourced data stored in cloud server has become a challenging as well as a promising research topic for data storage security.

In traditional data communication networks, data integrity is usually threatened by malicious attackers only. Both the sender and the receiver of data are trusted and collaborated in detecting and protecting data integrity. However, in cloud storage, the cloud storage servers are not always trusted. The cloud storage service provider has motivations to elude the service users on stored data status. For instance, A service provider may remove the rarely accessed data in order to economize the storage usage, or hide the data loss incidents for maintaining its reputation. Moreover, a malicious server may change or replace the stored data. In order to prevent the above instances, it is more valuable to have data integrity verification process in place and regularly query the correctness of data in storage servers. An effective verification mechanism can also allow the user to detect the threats of data integrity in cloud storage sooner, and take necessary actions to minimize the damage or recover the lost caused.

There are three basic requirements for data integrity verification process, namely, efficiency, unbounded use, and self-protect mechanism. Efficiency implies minimal network bandwidth and client storage capacity are needed for the verification process. The client does not need to access the entire data for verification purpose. Unbounded use represents verification process should support unlimited number of queries. Self-protect mechanism means the process itself should be secure against malicious server that passes the integrity test without accessing the data.

A number of different techniques and mechanisms have been proposed and designed for cloud data integrity verification process. The mainstream of research in this field belongs to Proofs of Retrievability (POR) and Provable Data Possession (PDP), both were designed to the above three requirements. The two methods originally emerged with a similar concept but different approaches. Since then, each one had gone through further development along different directions such as dynamic data support, public verifiability, and privacy against verifiers. Dynamic data support allows a client to dynamically update their data partially after uploading the data. Public verifiability enables everyone, not just data owner or verifier, to perform verification process. Privacy against verifiers ensures that the verification process does not contain any private information of data owner. POR and PDP schemes with their developments will be discussed and compared in more detail later in this section.

Besides those two approaches, there are several methods studied to address the storage data integrity issue resulted from data insertion, modification and deletion at the block level. In 2010, Proof of Erasability (POE) scheme was proposed by Paul and Saxena [120]. POE addresses clients' need to ensure a comprehensive destruction of the stored data in the storage when they withdraw the data and disassociate with the storage provider. This model plays a role as probing engineering or destructor, which can ensure the stored data are shredded partially or fully based on the rules of data store. Nevertheless, this scheme only allows the data owner knowing the data are being destroyed. Another parallel scheme called Proofs of Secure Erasure (PoSE-s) also has a similar function on remote attestation [122]. Even though this scheme was proposed to replace hardware-based attestation, it is suitable for updating secure code and secure storage erasure for cloud.

In the following subsections, we will first introduce POR and PDP, followed by their developments to improve efficiency, dynamic data support and public verifiability.

2.2.1 Proofs of Retrievability and Provable Data Possession

As s widely studied mechanism to ensure data integrity, POR was firstly proposed by Juels and Kaliski in 2007 [73]. Fig. 2.2 depicts the general schematic of the proposed POR system, which ensures the server (prover) to a client (verifier) that the stored data are intact during the storing and retrieving process of the client. The client first encode a raw file F through an encoding algorithm into an encoded file F' and then stores it in the prover. A key generation algorithm produces a key K stored in the verifier, and it is used to encode. For checking process, the verifier can perform challenge-response process with prover in order to check if F can be retrieved.

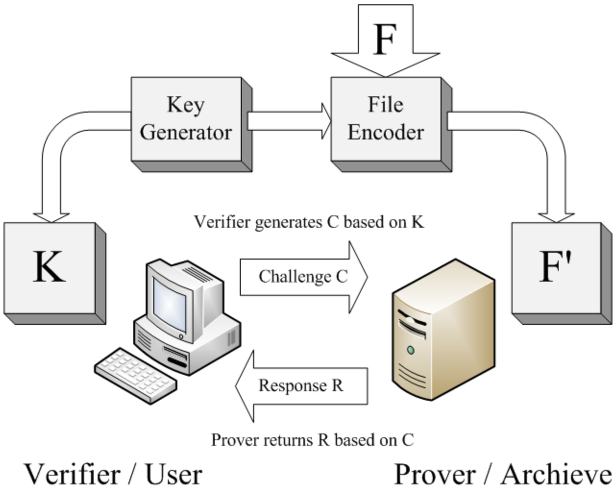

Figure 2.2: Schematic of a POR system

The first POR scheme introduced by Juels and Kaliski employed a sentinel scheme. POR protocol encrypts F and inserts randomly several *sentinels* into the other file data blocks after encryption. These sentinels play an crucial role for verification. The verifier can challenge the prover by pointing out the positions of a collection of sentinels, and the prover should return the values of the sentinels. If the values are different from the verifier's data, then it shows that prover has deleted or modified F. POR also includes error-correcting code to recover a small portion F if corrupted. However, this scheme requires pre-processing and encoding of F prior to store into the data storage, and it is bounded use - number of sentinels can be used up for limited queries. Therefore, Juels and Kaliski proposed another technique from Lillibridge *et al.* [96], Naor and Rothblum [110]. It stores the redundantly encoded data blocks with Message Authentication Code (MAC) to replace sentinels, and the MACs are stored together with data blocks. In this case, verification algorithm can examine the data integrity and ensures retrievability by requesting random number of block positions with their MACs. This approach resolves bounded use problem of the previous scheme, but at the cost of higher communication complexity of the audit.

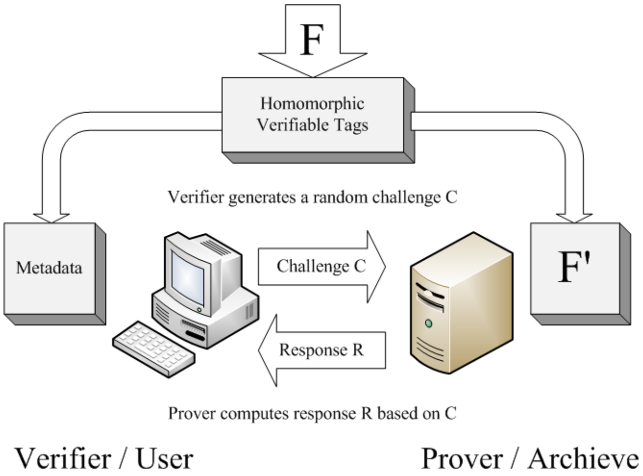
F
Homomorphic Verifiable Tags
Verifier generates a random challenge C
Metadata
Challenge C
Response R
F'
Prover computes response R based on C
Verifier / User
Prover / Archieve

Figure 2.3: Schematic of a PDP system

On the other hand, PDP came out concurrently with Juels-Kaliski's scheme. It was proposed by Ateniese et al. [5] and constructed based on symmetric key cryptography. PDP firstly chose RSA-based homomorphic verifiable tags [70] to combine multiple file blocks into a single value. A similar approach was also adopted later by Shacham and Waters [140]'s POR scheme in 2008. PDP scheme also provides data format independence, and it puts no restriction on the format of data. In other words, PDP allows any verifier (not only client) to query the server. POR and PDP both employed erasure code, which is a Forward Error Correction (FEC) for the binary erasure channel, helping recovery of the original message from slightly damaged data. The major difference between initial POR and PDP is that POR ensures not only data integrity at the server end but also retrievability, whereas PDP guarantees only data integrity at cloud data storage. Nevertheless, PDP is more efficient compared to Juels-Kaliski's POR, since it does not require any bulk encryption, and PDP requires smaller storage space on the client side and fewer bandwidths for challenges and responses. However, both schemes work on static data only, even though Ateniese *et al.* [6] proposed a dynamic version later in 2008, but it is restricted by number of queries and basic block operations.

2.2.2 Improvement on Public Verifiability

Since the Juels-Kaliski's original POR scheme was proposed without implementation of public verifiability, and its complexity was still high for communication and client storage, it became a popular topic for researchers to improve public verifiability and efficiency (discussed in the next subsection). In 2008, Shacham and Waters [140] proposed two new PORs system structures based on Juels-Kaliski's POR concept. Both solutions allow only one authentication value for the purpose of verification. The first one is privately verifiable using pseudorandom functions (PRFs); the second one is publicly verifiable, and it was built based on signature scheme of Boneh, Lynn and Shacham in a bilinear group [13]. Since the BLS signature was adopted, the public retrievability was achieved, and the proofs are reduced to a single authentication value, thus reduced communication complexity from $O(t)$ to $O(1)$, where t is the number of block positions. However, this scheme still works on static data only, without support of dynamic data update. Besides, the security parameter relies on Random Oracles, which means the client's challenge size grows up to $O(t^2)$.

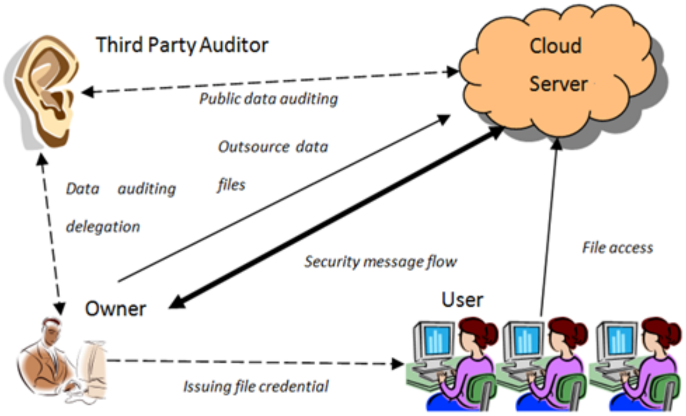

Figure 2.4: Third Party Auditor (TPA) structure

A new system model, as depicted in Fig. 2.4, which aimed at establishing a trustable mechanism between client and Cloud Storage Server (CSS) by introducing a *Third Party*

Auditor (TPA), was proposed in 2009 [164]. By using a privacy-preserving third-party auditing protocol, the TPA is trusted to monitor the stored data in cloud and transactions between the client and CSS, as well as assess and expose risks of the cloud services. This new scheme has been further developed based upon existing PORs and newly developed cryptographic primitives [161, 162, 163, 165, 166].

TPA typically adopts a public-key-based homomorphic authenticator with random masking to perform traffic auditing without a local copy of the data for integrity check. This public audit system can be constructed from the setup stage, which allows a user to initialize the secret parameters of the system, send the verification metadata to TPA, and audit the corresponding result. In this process, TPA will issue an audit message to the server for checking the user's data.

Homomorphic authenticators are used to verify metadata generated from individual data blocks while the aggregated authenticators can justify a linear combination of data blocks. As a paradigm, one can use a homomorphic token with distributed verification to check the integrity of erasure-coded data. The erasure-correcting codes play a vital role in preparing files for distribution so that the distributed files have redundancy parity vectors and the data dependability property. However, the linear combination of data blocks may potentially reveal users' privacy. With random masking, TPA cannot derive user's data content by building a correct group of linear equations.

The above model was further improved in [167] by integrating dynamic data support in 2011. Zhu *et al.* [193] also proposed a construction of dynamic audit services for untrusted and outsourced storage. It can detect abnormal behavior by using fragment structure, random sampling, and index-hash table.

Even though TPA-based schemes allows public verification of data integrity checking, They have a potential obstacle that requires an additional constituency, which is a third party auditor, added to the entire existing data storage scheme. The implementation of such schemes might be a burden for service providers because of additional costs. To address this concern, Han and Xin [60] proposed a new scheme offering the traditional TPA functions provided by CSP in a trustful manner. This scheme utilizes RSA and Bilinear Diffie-Hellman techniques, creates message header and mechanisms to achieve authentication process, while reducing complexity of cloud computing. Another work that provides public verifiability without help from the third party auditor was examined in [61] based on the work of Sebe *et al.* [139], and it has been proved to be secure from an untrusted server.

2.2.3 Improvement on Efficiency

Efficiency of POR and PDP has been improved from different aspects of the verification process. For instance, Curtmola *et al.* [35] showed how to integrate error-correcting codes with PDP and an adversarial error-correcting code construction similar to PORs. It also enabled PDP scheme to secure multiple replicas over distributed system without encoding each separate replica.

Besides, Dodis *et al.* [40] provided different solutions of optimized POR schemes under different constraints, such as bounded-use or unbounded-use, knowledge-soundness or information-soundness. They analyzed the tradeoffs in parameters and security between bounded and unbounded use schemes, and they also compared PORs under different circumstances in detail. It also improved the Shacham-Waters POR scheme by avoiding the usage of Random Oracles, which reduced the challenge size down to be linear in the security parameter, from $O(t^2)$ to $O(t)$.

In addition, a theoretical framework of PORs improvement was concurrently proposed by Bowers *et al.* [16]. The model offers an improvement over the protocols of Juels-Kaliski [73] and Shacham-Waters [141] by proposing a new variant to achieve lower storage overhead and tolerate higher error rates. The proposed POR scheme also decreased the challenge size to be linear of the security parameter. Another POR scheme was proposed by Kumar and Saxena in 2011 [147]. It targeted on simplification of Juels-Kaliski's sentinel scheme, making it suitable for limited computational power or small storage at verifier end. For PDP, Ateniese and Burns *et al.* concluded previous research developments and implementations of PDP in 2011 [4], and proposed two improved provably-secure PDP schemes with higher efficiency than previous ones.

2.2.4 Improvement on Dynamic Data Support

Supporting dynamic data update in data integrity verification schemes are especially challenging. Ateniese *et al.* [6] proposed the first partially dynamic PDP scheme in 2008. This scheme was more efficient in setup and verification phase compared to its previous version in [5], since it was only relied on symmetric-key cryptography. On the other hand, it only allowed a limited number of queries and basic block operations with limited functionality. For example, block insertion was not supported. Moreover, public verifiability was not supported either.

In 2009, Erway *et al.* proposed an improvement on PDP, referred as Dynamic Provable Data Possession (DPDP) [44]. In order to support provable updates on the stored data, this new model utilized authenticated directories based on rank information, and it defined the update as block insertion, modification or deletion to achieve dynamic PDP scheme. Nevertheless, this scheme maintains skip list [118] for tags and stores root metadata in clients side to prevent replay attack, so its computational and communication complexity can be up to $O(logt)$.

The dynamic data updates on POR were first considered in 2009. Wang *et al.* [164, 166] proposed the first scheme that achieved efficient data dynamics of the POR model by utilizing the homomorphic token with distributed verification of erasure-coded data, and manipulation of the Merkle Hash Tree (MHT) [107] respectively. The first scheme supported block update, delete and append operations only, while the second scheme provided both public verifiability and data dynamics for remote data integrity check, but the verification complexity increased to O(log n) from O(1) as a trade-off, and it achieved partially dynamic instead of fully dynamic. Both schemes also showed a new system model involving Third Party Auditor (TPA).

In addition, Zheng and Xu presented a new POR scheme with a fresh property, namely, fairness, to deal with dynamic data [189]. This property prevents unscrupulous clients from accusing a legitimate server about modifying their stored data. This issue arises because of the feature of dynamic data. POR for static data storage can solve this problem simply by asking the verifier to approve and sign digitally when the data has not been stored into the storage. The proposed Fair and Dynamic Proof Of Retrievability (FDPOR) was mainly composed of two parts, a new authenticated data structure: range-based 2-3 Tree (rb23Tree) and a new incremental signature scheme called hash-compress-and-sign. However, FDPOR did not support public verifiability, and complexity for both the verifier and the prover were higher than that of previous PORs.

2.2.5 Summary of Data Integrity

PORs and PDPs are the major remote data integrity checking protocols proposed in cloud storage systems. The original POR and PDP protocols differs in many aspects. PORs are considered to be more secure compared to PDPs, because it requires encryption of the original data and error correction coding to recover damaged data, while PDPs are known for higher efficiency and applicability to large scale public databases, such as digital libraries. With further improvement of each, the two schemes have been converging towards the same objectives. For example, although public verifiability and homomorphic verifiable tags were first known for PDPs, these characteristics are also

	Data Dynamic	Public Verifiability	Retrievability	Server Comp.	Verifier Comp.	Communication Comp.	TPA
2007 JK [73]	Static	No	Yes	$O(1)$	$O(1)$	$O(t)$	No
2008 SW [141]	Static	Yes	Yes	$O(1)$	$O(1)$	$O(1)$	No
2009 Wang [164, 166]	Partially dynamic	Yes	Yes	$O(logt)$	$O(logt)$	$O(logt)$	Yes
2009 Dodis [40]	Static	Yes	Yes	$O(1)$	$O(1)$	$O(1)$	No
2009 Bowers [16]	Static	Yes	Yes	$O(1)$	$O(1)$	$O(1)$	No
2010 Wang [159]	Static	Yes	Yes	$O(1)$	$O(1)$	$O(1)$	Yes
2011 Saxena [147]	Static	No	Yes	$O(1)$	$O(1)$	$O(1)$	No
2011 Zheng [189]	Partially dynamic	No	Yes	$O(logt)$	$O(logt)$	$O(logt)$	No
2007 Ateniese [5]	Static	Yes	No	$O(1)$	$O(1)$	$O(1)$	No
2008 Ateniese [6]	Partially dynamic	No	No	$O(1)$	$O(1)$	$O(1)$	No
2008 Curtmola [35]	Static	Yes	No	$O(1)$	$O(1)$	$O(1)$	No
2009 Erway [44]	Fully dynamic	Yes	No	$O(logt)$	$O(logt)$	$O(logt)$	No
2011 Ateniese [4]	Partially dynamic	Yes	No	$O(1)$	$O(1)$	$O(1)$	No
2011 Hao [61]	Fully dynamic	Yes	No	$O(logt)$	$O(logt)$	$O(logt)$	No

Table 2.1: Performance comparison for data integrity verification schemes.

applicable to PORs. On the other hand, some PDP variants may also adopt encryption and/or error correction coding tools to strengthen their security measurement. Therefore, it is all about making tradeoffs among security functionalities and efficiency.

In Table 2.1, we summarize the above reviewed POR and PDP schemes by a thorough comparison of their performances. It is noteworthy that schemes with dynamic data support suffers higher complexities compared to their counterparts. Future research directions include further improvements on efficiency and fully dynamic data support. To improve efficiency of those schemes, reducing communication cost and storage overhead are rightful considerations. However, fully dynamic data support is a challenging objective, because it increases complexity but reduces update information on server-end.

2.3 Data Confidentiality

Data confidentiality in cloud storage security refers to the property that information stored in the cloud storage is not made available or disclosed to unauthorized individuals, entities, or processes. Access control and data encryption have been widely deployed to protect data confidentiality in the traditional data communication networks. It is natural to extend their deployment in cloud storage systems. For instance, Secure Socket Layer (SSL) and AES-256 bit encryption are adopted in Dropbox to ensure data security.

However, data confidentiality in cloud storage systems faces new risks and challenges, thus calls for new techniques or improved mechanisms. In this section, we discuss new challenges faced by access control and data encryption mechanisms, as well as recent developments to meet those challenges of data confidentiality protection in cloud computing.

Although traditional encryption techniques can hide the information of data from the cloud server, it would not provide a satisfactory solution if users demand to compute on their stored data. Since the computing can not be functionally performed on the ciphertext, users would have to decrypt the data before performing any computation and re-encrypt after the computation. During this process, sensitive information could have been leaked to the curious server. Otherwise, user would be forced to compromise with the service provider by uploading plaintext and signing SLA, which exposes their data to higher risks. To solve this problem, there have been research attentions drawn to a newly proposed encryption primitive, namely, *Fully Homomorphic Encryption*, which allows ciphertext to be computed without affecting decipher process.

In the following of this section, we first examine new access control mechanisms with higher efficiency and fine-grain user control suitable for cloud storage. Then introduce some new concept of data encryption schemes, such as searchable encryption and FHE, and discuss their potential applications in protecting data confidentiality in cloud computing. Then, Other data confidentiality approaches are also briefly discussed. We provide our insights of the current research efforts and future directions in data confidentiality to summarize this topic.

2.3.1 Access Control

As mentioned above, access control has been one of the key mechanisms to protect data confidentiality in traditional data networks. It is designed to block unauthorized users and malicious hackers from accessing data. Although the objective of access control in cloud storage does not differ from that in traditional data network, the requirement does change. Traditional access control enforced by the service provider could not stop a curious cloud service provider accessing users' sensitive data, which was stored in the service provider's infrastructure and managed by the service provider. A curious cloud storage server trying to derive sensitive information from its stored data, or from data operations performed by data owner and authorized users, is a new threat model against data confidentiality in cloud storage service. Moreover, a malicious service provider could intentionally leak the data to unauthorized parties for profit, or a malicious attacker could compromise the service provider and get unauthorized access to the data.

To address this challenge, cryptographic access control schemes that shifted the access control agency from the service provider to the users have been proposed. Instead of relying on untrusted service provider to grant access control, users can enforce their own access control by selectively granting different decryption access to a certain part of encrypted data. By means of encryption, the owners of data, i.e., cloud storage users who lost their physical control over their own data could regain their control at the semantic level.

Plutus [74] and SiRiUS [67] are examples of using encryption to secure file sharing on remote untrusted storage. These schemes encrypted different files with different keys, thus changing the problem of access to files to the problem of key management. However, this approach is not scalable when applying to cloud storage, because the complexity of key management increases with the number of files and/or the number of users, which both could be enormous in a cloud storage system. As a large number of users are sharing the same infrastructure in a public cloud storage built upon a complicated network scale, it is crucial to have efficient, scalable and reliable access control mechanism in place.

In the following, we examine recent research on more efficient access control using encryption techniques developed for cloud storage systems.

Access Control using Attribute-Based Encryption

In attribute-based access control model, access is granted based on attributes of the user. When applied to cloud storage, access control is enforced on data encrypted using *attribute-based encryption* (ABE) schemes. In an ABE system, a user's keys and ciphertexts are labeled with sets of descriptive attributes. A particular key can decrypt a particular ciphertext only if there is a match between the attributes of the ciphertext and the user's key.

The concept of ABE was introduced by Sahai and Waters [135]. Their access control allowed for decryption when the number of overlapped attributes between a ciphertext and a private key exceeds a specified threshold k. The fuzzy nature of this scheme was originally designed for error-tolerant identity-based encryption scheme that could use biometric identities. However, with a threshold-based flat access structure, it could not be generalized to other applications. Two prominent ABE schemes with more general tree-access structures, namely, *Key-Policy Attribute-Based Encryption* (KP-ABE) [54] and *Ciphertext-Policy Attribute-Based Encryption* (CP-ABE) [9], were proposed in 2006 and 2007, respectively. Both algorithms associated a set of expressively descriptive attributes with a tree-access structures to enforce access control on the encrypted data, but they work in a reverse manner. In KP-ABE, each ciphertext was labeled with a set of attributes during encryption, while the users' private keys were associated with an access tree specifying which ciphertexts the key can decrypt. On the contrary, in CP-ABE, Users' private keys were based on a set of their attributes while ciphertexts are associated with an access tree over the attributes during encryption. As a result, in KP-ABE scheme, it is the key distributor (usually the service provider), who decides the access policy, while in CP-ABE scheme, it is the encryptor (usually the data owner) who controls the access over the encrypted data.

In the above mentioned ABE schemes, the access policy can only contain logical formula “and” and “or”, and threshold gates. A KP-ABE scheme was introduced in [114] which allows “negative” constraints to be represented in access policies. Additionally, many CP-ABE schemes were proposed such as [26, 90, 169] which either achieve chosen ciphertext attack (CCA) secure or are built on different security assumptions. Even though the KP-ABE and CP-ABE work in reverse manner, Goyal *et al.* [53] provided a generic approach to transform a KP-ABE scheme into a CP-ABE one. Malek and Miri combined the two ABE schemes into one system, and proposed a balanced access control that allows both service provider setting up system wide access policies and data owner setting up access structure to their own data [105]. Further research on ABE is also discussed in [180, 187]. In a dynamic system, access policies may differ from time to time, and user qualifications may also change. Therefore, the ability to revoke attributes from a user is desired in ABE systems. Several revocable ABE schemes [134, 184] were proposed where an ABE system is able to revoke users from accessing encrypted data to which they used to have access in the system.

When using the ABE in a system where there is a large number of attributes, assessing the qualification of users and generating decryption keys by a central authority becomes impractical. *Multi-Authority Attribute-Based Encryption* (MA-ABE) was first proposed to address this issue in 2007 [23]. In a MA-ABE scheme, attributes are divided into different sets, and each set can be managed by an independent attribute authority. Corresponding attribute keys for decryption are issued by multiple attribute authorities, and encryptors can specify an access policy that requires a user to obtain decryption keys for appropriate attributes from different authorities in order to decrypt a message. Subsequently, several other MA-ABE constructions were proposed in [24, 91].

Role-Based Access Control

Another access control model called *Role-Based Access Control* (RBAC) [2, 137], has also been commonly adopted in traditional storage system in order to simplify management of permissions. Its access policy is determined based on different roles assigned to users by the system, while the data owner can specify a set of permissions of their data to different roles. By separation the tasks of role assignment and permission assignment, RBAC is much more efficient and scalable compared to other access control based on individual users, because the number of roles are usually significantly less than the number of users. Furthermore, it makes dynamic access control easier. For example, in applications where permissions for roles change slowly, while users may enter, leave or change roles rapidly, the role manager can simply assign a new role to the user or revoke a role from the user. On the other hand, the data owner can also add permissions to a role or revoke permissions from a role. The authors of [111] suggested including RBAC in a new access control model for the health care system that can provide flexible access rights, because it can be modified dynamically while the task changed. However, one of the major criticisms of RBAC schemes is the complicated process when setting up the role structure. To make RBAC more efficient, roles can be structured hierarchically so that some roles inherit permissions from others.

To enforce role-based access control policies, one approach is to transform the access control problem into a key management problem. In the literature, there exist many hierarchical access control schemes [3, 39, 136] which have been constructed based on hierarchical key management (HKM) schemes. Because of the similarity in structures between hierarchical access control and RBAC, a hierarchical access control scheme can be easily used to enforce RBAC access policies in cloud environment. In 2010, a role-based encryption (RBE) scheme [191] was built directly on RBAC policies. The security of the hierarchical access control scheme relies on the correct execution of the key assignment process, while the security of the RBE is based on the security of the cryptographic algorithm. More specifically, when a user is assigned to a role in RBE, a decryption key is calculated through a cryptographic algorithm by taking as input of the secret value and the identity of user and role. In the hierarchical access control scheme, the key for the user is generated based on the access control policies of the whole system. In 2011, Zhu *et al.* [192] proposed a revocable RBE scheme which allows users to be granted or revoked role memberships dynamically.

In the above schemes for enforcing RBAC policies, user membership of each role and role hierarchy are managed by a central authority. However in large-scale RBAC systems which have hundred or even thousands of roles and hundreds of thousands of users and permissions, it is impractical to centralise the task of managing these users and permissions, and their relationships with the roles in a small team of security administrators. Zhou *et al.* [190] proposed a new RBE scheme using an *identity-based broadcast encryption* (IBBE) algorithm [37], which allows user memberships to be managed by individual roles. In the new RBE scheme, plaintext can be encrypted to a specified role, and only users in that role and its predecessor roles can decrypt the data with their role secrets and decryption keys. The employment of a broadcast encryption algorithm allows dynamically adding new users into a role without re-encryption, as well as revoking an existing user from a role without affecting any other existing users. In addition, this scheme has other features such as constant size keys and ciphertexts.

There have also been combined Attribute-Based Access control (ABAC) and RBAC schemes proposed in order to take advantage of both to provide effective access control for distributed and rapidly changing applications [83]. Hong *et al.* [64] implemented RBAC system for cloud storage via CP-ABE. In their work, permission assignments were handled by data owner while role assignments were handled by other users through propagation.

2.3.2 Searchable Encryption

With more and more data moving to the cloud storage, it becomes imperative to enable search over the huge amount of data for many user applications. To preserve data confidentiality and integrity, it is necessary to store encrypted data in the cloud storage servers. To perform searching over data, the user has to either store an index locally, or download all the encrypted data, decrypt it and search locally. Neither approach is efficient when the data size grows in the cloud. When users seek to search and download relevant files from a cloud storage system, it is often desirable for the Storage Service Provider (SSP) to host search service, because it can minimize the network traffic and reduce management complexity for the users.Therefore, how to perform searching on encrypted databases without the need of decryption has become an increasingly fascinating topic in cloud storage systems. Recently, there have been new cryptographic primitives, called *searchable encryption* schemes [12**?**], proposed to address this problem.

The basic idea of searchable encryption schemes is to encrypt a search index generated over a collection of data in such a way that its contents are hidden without appropriate tokens, which can only be generated with a secrete key. Given a token for a keyword, one can retrieve pointers to the encrypted data files that contain the keyword. During the retrieval process, there is no contents of either the data files or the keyword revealed, other than the fact that all the retrieved data files contain one keyword in common.

Searchable encryption schemes, including *Symmetric Search Encryption* (SSE) [146], *Asymmetric Search Encryption* (ASE) [12] and other improvements on both schemes are reviewed in [75]. SSE employs symmetric cryptographic algorithms, such as block cipher or hash function, therefore is suitable when the party that performs search over the data are also the one who generates it, whereas ASE employs asymmetric cryptographic algorithms such as elliptic curve, thus is also suitable when the party that performs search over the data are different than the one who generates it. Therefore, ASE has wider applications than SSE in cloud storage than SSE. Meanwhile, compared to SSE schemes, ASE can achieve more complex search queries, such as conjunctions of terms, but at the cost of higher complexity and weaker security guarantees. Efficient ASE, or ESE scheme was introduced in [8] to improve the efficiency when the keywords are hard to guess. However, it is more vulnerable to dictionary attacks.

Since SSE achieves higher efficiency and stronger security, it has been further developed recently. For example, dynamic SSE [76, 77] extended the inverted index approach [34] to allow update of the encrypted index and data files, and to achieve adaptive security against chosen-keyword attacks. Furthermore, Parallel and dynamic SSE [76] enables more efficient and scalable construction based on a keyword red-black tree-based multi-map data structure. On the other hand, SSE schemes with improved functionalities but compromised security have been proposed. Kuzu *et al.* [85] utilized locality sensitive hashing (LSH), which is widely used for fast similarity search in high dimensional spaces for plain data, and proposed a search scheme to enable fast similarity search in the context of encrypted data. Another approach, which was proposed by Wang and Cao *et al.* [160] to secure ranked keyword search in encrypted cloud data. This method utilized the *Order-Preserving Symmetric Encryption* (OPSE) [10, 11], which achieves both security and privacy-preserving by protecting sensitive weighted information.

For cloud storage that are accessible with multiple users, how to enforce privileges and access control while searching through cloud storage has attracted researchers' attentions. One approach was proposed by Singh and Srivatsa *et al.* [144] in 2009 which performs indexing in the trusted enterprise domain, and utilizes the resulting indices systematically with the *Access Control Barrel* (ACB) [143] primitives and concepts of user access hierarchy. This solution improves indexing efficiency and allows transferring the indices to the SSP for hosting, and it can be developed based on the integrity of search results returned by the SSP in the future.

Other than search algorithms on encrypted database, more general computation on encrypted database is a related topic. *Secure Computation ON an Encrypted Database* (SCONEDB) [171] was proposed to solve the *k-Nearest Neighbor* (kNN) computation in an encrypted database utilizing asymmetric scalar-product preserving encryption (ASPE). Besides, SCONEDB can incorporate other existing techniques, such as OPSE for the range query and homomorphic encryption for aggregate queries. CryptDB [126] implemented an integrated system that supports more general SQL query operations over encrypted database, by adapting a number of existing and new SQL-aware encryption primitives with different security properties and functionalities. CryptDB dynamically adjusts the encryption strategies using layered onion structures, where each data was dressed in increasingly stronger encryption, such that the outmost layer provides maximum security, whereas inner layers provide more functionality. A trusted proxy determines whether layers of encryption need to be removed when receiving a query from the user application.

2.3.3 Fully Homomorphic Encryption

Homomorphic encryption allows specific algebraic operations to be manipulated on a ciphertext, so it can produce the same encrypted result as the ciphertext of the result of the same (or different but known) operations performed on the plaintext. In other words, the operations to be performed on original data can now be performed on the encrypted ciphertext without knowing the original data. Homomorphic encryption can be categorized into two types: Partially Homomorphic Encryption (PHE) and Fully Homomorphic Encryption (FHE). PHE allows only one homomorphic operation, either addition (e.g. Paillier [116]) or multiplication (e.g., unpadded RSA), while FHE supports both addition and multiplication operations. Since the original unpadded RSA algorithm published in 1977, there have been many PHE algorithms developed. However, the partially homomorphic property of an encryption algorithm has rarely been considered advantageous, but rather vulnerable to adaptive chosen-ciphertext attacks. Therefore, PHE algorithms have been found useful only in limited security applications such as electronic voting systems. On the other hand, since the first FHE algorithm was announced in 2009 [50], it has been recognized as a huge breakthrough in the computing security field. Practical application of FHE cryptosystems will potentially enable development of computing programs, which runs on encrypted input data to generate encrypted output. These programs can thus be run by untrusted entities without revealing any sensitive information during the computing process.

A homomorphic cryptosystem ε consists of four algorithms, $KeyGen_{\varepsilon}$, $Encrypt_{\varepsilon}$, $Decrypt_{\varepsilon}$, and an $Evaluate_{\varepsilon}$ algorithm. The first three algorithms are defined the same as those in any public-key cryptosystems. The $KeyGen_{\varepsilon}(\lambda)$ produces key-pair (pk, sk) given a security parameter λ. The $Encrypt_{\varepsilon}$ algorithm takes pk and a plaintext π as input, and it outputs a ciphertext ϕ. The $Decrypt_{\varepsilon}$ takes sk and ϕ as input, and outputs the plaintext π. In addition, the $Evaluate_{\varepsilon}$ algorithm takes as input pk, a circuit C from a permitted set C_{ε}, and a set of ciphertexts $\varphi = (\phi_1, ...\phi_t)$, consequently outputs a ciphertext ϕ. The homomorphic cryptosystem ε is correct for C_{ε} if for any key-pair (pk, sk) generated by $KeyGen_{\varepsilon}(\lambda)$, any circuit $C \in C_{\varepsilon}$, any plaintexts $\pi_1, ..., \pi_t$, and any ciphertexts $\varphi = (\phi_1, ...\phi_t)$ with $\phi_i \rightarrow Encrypt_{\varepsilon}(pk, \pi_i)$, it is the case that

$$If \phi \leftarrow Evaluate_{\varepsilon}(pk, C, \varphi), \tag{2.1}$$

then

$$Decrypt_{\varepsilon}(sk, \phi) \rightarrow C(\pi_1, ..., \pi_t) \tag{2.2}$$

The computation complexity of all the above algorithms has to be polynomial in the size of C and security level parameter λ, which is defined as all known attacks against the scheme take time at least 2^{λ}. ε is fully homomorphic if it is homomorphic for all circuits[50].

A family of schemes $\varepsilon(d) : d \in Z^{+}$ is leveled fully homomorphic if they all use the same decryption circuit, $\varepsilon(d)$ is homomorphic for all circuits of depth at most d (that use some specified set of gates Γ), and the computational complexity of $\varepsilon(d)$'s algorithms is polynomial in λ, d, and (in the case of $Evaluate_{\varepsilon(d)}$) the size of C.

The first FHE scheme proposed by Craig Gentry in 2009 [50] applies lattice-based cryptography to construct the scheme, where lattice L was a set of points in the n-dimensional Euclidean space R^n with a strong periodicity property. The proposed scheme started from a somewhat homomorphic encryption scheme using ideal lattices, which is limited to "low-degree" polynomials evaluation on encrypted data due to the augment of noise in the ciphertext during evaluation. After this "*initial construction*" stage, a "*squash the decryption circuit*" technique was used to modify the scheme to make it "*bootstrappable*". The modified encryption scheme can evaluate its own decryption circuit, and effectively refresh the ciphertext to reduce the augmented noises, which eliminates the limitation on the depth of circuit evaluated over the ciphertext. In short, Craig Gentry slightly modified somewhat homomorphic encryption by recursive self-embedding. The resulting scheme can reduce the accumulated noise caused by multiple algebraic operations, thus make it possible to realize FHE in arbitrary depth.

However, this first FHE scheme is impractical since the computation complexity and ciphertext size are high-order polynomials in the security level parameter λ, which means they increase sharply in order to achieve a practically high enough security level.

This prohibit the practical application of the FHE, especially in the cloud computing context where high security level is crucial. Another major concern of this scheme is that its security was based on two relatively new assumptions, namely, the hardness of the worst-case Bounded Distance Decoding problem (BDD) on ideal lattice, and the hardness of the average-case Sparse Subset Sum Problem (SSSP) of the squashing step. Both are relatively untested cryptographic assumptions.

More recently, there have been growing research efforts made in searching practical FHE algorithms, which are more efficient and/or based on more reliable security assumptions. A second version of FHE scheme, known as DGHV, was proposed by Marten van Dijk, Craig Gentry, Shai Halevi, and Vinod Vaikuntanathan in 2010 [155]. DGHV uses Gentry's techniques with only elementary modular arithmetic over integers to convert a simple somewhat homomorphic encryption scheme to a bootstappable FHE scheme. This scheme achieved conceptual simplicity because all computations were performed over integers instead of ideal lattice. It also reduced the security assumption to the hardness of the Greatest Common Divisor (GCD) problem. However, the price of this tradeoff is the immense size of public key, which can be impractical for the current systems.

Stehle and Steinfeld [148] presented a faster homomorphic encryption in order to improve Gentry's scheme by a more aggressive analysis of the SSSP assumption, and introducing a probabilistic decryption algorithm implemented by an algebraic circuit of low multiplicative degree. With these two enhancements, this scheme obtains $O(\lambda^{3.5})$ bit complexity for refreshing a cipher text, whereas previous scheme claimed $O(\lambda^{6})$ for the same task, where λ is the security parameter. However, there is a non-zero probability of decryption error associated with this scheme.

Besides, Zvika Brakerski and Vinod Vaikuntanathan gave another improvement on Gentry's scheme [18] by changing the two security assumptions made in [50]. First, the somewhat homomorphic encryption was based on ring learning with errors (RLWE) assumption from Lyubashevsky, Peikert and Regev [101] instead of the ideal lattices BDD problem. Second, to make the somewhat homomorphic encryption scheme bootstrappalbe, it used a dimension-modulus reduction technique instead of Gentry's squashing technique, thus eliminating the assumption of SSSP. This new bootstrapping technique also shortened the ciphertext and reduced the complexity.

Based on the above improvement, Brakerski, Gentry, and Vaikuntanathan worked together to propose a new leveled FHE scheme without Gentry's bootstrapping procedure in 2011 [17]. By applying RLWE, this FHE scheme has $O(\lambda \cdot L^3)$ per-gate computation for L-level arithmetic circuits. As an optional approach, they also proposed a leveled FHE scheme using bootstrapping as optimization to further reduce the per-gate computation down to $O(\lambda^2)$, independent of L.

Following up in 2011, Coron *et al.* proposed an improvement of FHE over the integers described by van Dijk *et al.*. The proposed new scheme shortened the public key size from $O(\lambda^{10})$ to $O(\lambda^7)$ [32]. This procedure is done by using quadratic form instead of linear one in the public key elements, so that the full-length public key is compressed to a smaller subset of the original key. Instead of proposing any further improvement on FHE, Ron Rothblum manifests how to transform any additively homomorphic private-key encryption scheme into a public-key encryption scheme [132]. To construct this process, this scheme develops a theorem that any compact additively homomorphic with respect to addition modulo two can be transformed into a semantically scheme. In consequence, the public-key encryption scheme save one hop homomorphic with regard to the same set operations with private-key encryption, which are prior FHE schemes.

With all the theoretical development of different FHE algorithms, it is necessary to investigate their practical implementation. There were several implementations of Gentry's FHE in 2010, and the first attempt was made by Smart and Vercauteren [145].

	Solution on Ideal Lattices BDD	Solution on SSSP	Per-gate Comp.	Public Key Size	Asymptotic Comp.
2009 Gentry [50]	SVP	Availability of SVP Oracle	$O(\lambda^6)$		
2010 Stehle [148]		Refined Analysis	$O(\lambda^3)$		
2011 Brakerski [18]	RLWE		$O(\lambda^3)$		
2010 Dijk [155]	Replace Ideal Lattice	Choosing Large Enough θ	$O(\lambda^3)$	$O(\lambda^{10})$	
2011 Coron [32]		Refined Analysis	$O(\lambda^3)$	$O(\lambda^7)$	
2010 Vercauteren [145]			$O(\lambda^3)$		$O(n^{2.5})$
2011 Halevi [51]			$O(\lambda^3)$		$O(n^{1.5})$

Table 2.2: Performance comparison for FHE schemes.

They were able to implement the somewhat homomorphic scheme using ”*principle-ideal lattices*” of prime determinant, which can be implied by two integers only. However, they were not able to implement the bootstrapping functionality to obtain a fully homomorphic scheme. Bottleneck of this implementation was the failure to support a large amount of parameters.

Based on this work, in 2011, Gentry and Halevi developed a series of simplifications and optimizations that made bootstrapping implementation possible. As the result, the asymptotic complexity is reduced from Smart and Vercauteren's $O(n^2.5)$ to $O(n^1.5)$. The optimizations from this paper were also used in [32] in order to implement the fully homomorphic DGHV scheme under new variant. With the result of having similar performance, Coron *et al.* successfully showed that FHE can be implemented with a simple arithmetic scheme.

In Table 2.2, we provide a comparison of performance for different FHE schemes. BDD and SSSP are the problems that stated in the first FHE scheme.

2.3.4 Other Data Confidentiality Approaches

There are several other data confidentiality methods beside above. For instance, The application of cryptographic algorithms to data blocks in the cloud storage is a popular method used to ensure the confidentiality of stored data. A data confidentiality scheme in coreFS, which is a user-level network file system, was proposed in 2009 [181]. This scheme is constructed based on a new universal-hash stateful MAC. It has smaller computational overhead of cryptographic operations comparing to the MHT. Besides, it allows better communication capability. However, the choice of caching strategy, MAC tree update schedule, and the method to store the tree can affect the performance of this scheme.

Another data confidentiality scheme exploited the newly proposed Secure Provenance (SP) model based on the bilinear pairing techniques in 2010 [100]. This scheme records the ownership and the process history of data objects in the cloud storage in order to increase the trust from public users. The SP model consists of the following modules: system setup, key generation, anonymous authentication, authorized access, and provenance tracking. The provable security technique has been tested on this scheme under the standard SP model. It demands some practical considerations in real-world applications and further improvement under the current framework.

Different from above schemes, an compelling statement was proposed by Dijk and Juel in 2010 [156], which claimed that no cryptographic protocol, even including power primitives such as FHE, can enforce privacy requested by common cloud services alone. This paper also demonstrated that above demand can be achieved by other enforcements instead, such as tamperproof hardware, distributed computing, and complex trust ecosystems.

2.3.5 Summary of Data Confidentiality

Data confidentiality is one of the most critical issues for applications with sensitive data, such as personal information, customer's account information, financial and healthcare information. The new challenge of storing those data in clouds is how to prevent accidental or intentional data leakage to the cloud storage service provider, such that even if the service provider is compromised, the information can still be kept confidential. The fundamental solution to this problem is still data encryption. The user has to encrypt the data before they are moved to the cloud server, and keep it encrypted for the entire period during which the data are in the cloud. When the data needs to be accessed or processed by either the data owner or other legitimate users who have the key for decryption, it is not efficient to retrieve the encrypted data, decrypt it, process and re-encrypt it before sending back to the server. Therefore, new encryption mechanisms that allow for processing of the ciphertext directly without revealing the original information in the plaintext will have a significant potential in cloud storage of sensitive data. WIth encryption, data owners or users regain their control over their data that are not physically stored by themselves.

FHE is an ideal example of these encryption algorithms. However, the promising applications of current FHE algorithms are hindered by its computation complexity and other implementation difficulties. Improvements must be made before it can be put in practical applications. In addition, more implementations of various improvements are awaited to be evaluated on current platforms.

Unlike FHE, which has an ambitious aim at arbitrary computing on the ciphertext, other cloud encryption schemes aimed at specific type of control over encrypted data. For example, ABE allows access control being enforced on the encrypted data by incorporating attribute-based access structure into either the ciphertext or the decryption key. Searchable encryption schemes provide a way to search the ciphertext for a keyword token without revealing the real content of either data or the keyword.

2.4 Availability

As a different security measure, availability in cloud storage refers to that the data are accessible and usable when authorized users request them from any machine at any time. In an earlier stage of cloud computing, availability was of more security concern due to the lack of mature and reliable infrastructure. Many incidents of service unavailability occurred due to hardware failure and resulted in severe consequences. WIth better and more reliable infrastructures in place, the challenge facing the availability of cloud storage service is how to preserve the user's data in case of emergency, such as a natural disaster.

The most straightforward solution is to keep backup copies of data in multiple physical locations. Amazon EC2 and S3 provide a perfect example based on availability zone, which locates within divided geographic regions, for example, US-West and US-East. Each region contains several instances with same data. When accident occurs, Amazon EC2 and S3 can easily recover damaged or lose data from other availability zones within the same region to save power and time. However, this approach is not efficient in terms of storage resource utilization.

There have been backup storage management schemes, such as incremental backup and data deduplication, developed to improve the storage utilization. Incremental backup has been used widely in file backup services. It exploits the correlation between current files with previous backup version and only stores the differences. When incremental backup being deployed in a data block level or even data byte level, it becomes more efficient in storage utilization, but with higher processing overhead. Delta encoding is a famous incremental backup example applied by Dropbox. Data deduplication is a specialized data compression technique that identifies common data chunks within and across different files, and stores them only once to improve storage utilization. Unfortunately, data deduplication potentially undermining the data security in terms of both data integrity and data confidentiality. First, by definition, data deduplication alters the original data from the user and stores them in a different form in the cloud storage, thus results in concerns of data integrity. Second, data deduplication attempts to identify and exploit identical data chunks, while encryption algorithms usually try to randomize them to conceal the real contents. The encrypted ciphertext for the same plaintext will likely to be extremely different. To address these issues, efficient and secure data deduplication which allows data deduplication performed on encrypted ciphertext have been developed [149]. This technique utilized *convergent encryption*, in which the encryption key is generated using a hash function of the plaintext of the data chunk. Therefore, the same plaintext data chunk will be encrypted using the same key, no matter when and by whom it is encrypted. This results in the same ciphertext data chunk for the same plaintext. The scheme stores unique chunks of data or bytes during data analysis, and then compares other chunks to the stored data. If the compared result is matched, then the redundant part is replaced by a small pointer pointing to the location of the matched stored data.

Another proactive approach is to predict future availability failure occurrences so that actions could be taken earlier to avoid interruption of service. Guan *et al.* proposed two learning approaches to predict failure dynamics in cloud computing systems by using Bayesian methods and decision trees [55]. An initial stage is required for monitoring data, and then an ensemble Bayesian methods labels data that have anomalous behaviors. After all the anomalies are identified, the model can predict future failure occurrences based on decision tree classifiers.

Once the failure has occurred, data recovery schemes are necessary to reduce or eliminate the loss. Zhang [185] presented a data recovery method that examines the damage in a fine-grained cloud database and allows the cloud database owner to know and locate the damage precisely for the recovery purpose. Information dispersal algorithm [127] is used to enable greater availability of data when encountering physical failures and network outages.

Besides the above techniques, data recovery can also be achieved by new service framework. Chi-won Song *et al.* proposed Parity Cloud Service (PCS) in 2011 [170].

It generates virtual disk in user system for private backup and makes parity group of multiple users. The same data among those users in the parity group are stored at the server-end. Therefore, when users find out that original file requires recovery, they can request data from the server-end without violating privacy since private backup is stored at each user's virtual disk. This approach is simple and secure, but each user has to build up virtual disk, which costs additional overhead for users.

In summary, ensuring availability of users' data whenever users demands it is the basic and primary requirement in cloud storage. The main challenge arises when taking other performance and security concerns into consideration. Trade-offs between efficiency and reliability have to be made to balance the interests of service provider and the user. Furthermore, data integrity and data confidentiality should not be compromised by improved availability.

2.5 Summary

With the trend of rapid deployment of cloud storage and computing nowadays, it is essential for the cloud storage systems to be equipped with security solutions proven to be reliable and trustworthy. In this work, we conducted a survey on most recently developed or proposed primitives to ensure three of the most critical security measurements, namely, data integrity, data confidentiality, and availability, for the cloud storage systems. For each aspect, we identified the unique challenges that are different from those in traditional data network or file storage systems, summarized the existing development progress up to date, and provided insight into the future directions of research. Overall, we feel that the cloud storage security is still in its infancy and expect to see more salient breakthrough in the near future. For example, although the cloud storage security solutions have been developed rapidly in recent years, we have not yet seen a widely accepted model for the implementation. Besides the system design, the cloud storage security system should be flexible enough so that it can be improved by new cryptographic algorithms.

To sum up, this chapter provides a broad survey on data security for the cloud storage system, and since we often store the surveillance data in the cloud nowadays, biometric security is also as important as the data security. If there is no efficient and effective way to identify possible suspects through the stored surveillance data in the cloud, then extra efforts have to be done with the growing data to achieve the goal. Therefore, having the automatic face recognition system is as crucial as maintaining the data security on cloud.

Chapter 3

TAEF: A Cross-Distance/Environment Face Recognition Method

Although there have been progressive developments in automatic face recognition, most efforts focus on the situation where probe faces are located at a close distance with varying poses. Much less work has been conducted for Face Recognition at A Distance (FRAD), which is common in the video surveillance application. For convenience, long distance face images and short distance face images for matching are called *probe* and *gallery* images, respectively. The FRAD problem is challenging since the quality of probe images, which are often captured in an outdoor environment, is degraded by both distance and environmental effects while gallery images are typically captured in a controlled indoor environment. To address this cross-distance and cross-environment face matching challenge, we propose a solution called the Two-Stage Alignment/Enhancement Filtering (TAEF) method. The TAEF method consists of three main components: cross-distance face alignment, a cross-environment face enhancement and two-stage filtering.

Two-Stage Filtering. A coarse-to-fine two-stage filtering mechanism consisting of initial screening (or coarse-scale processing) and iterative refinement (or fine-scale processing) is proposed in this work. Given a probe image, the procedure of face alignment, enhancement and matching is executed against all gallery images to eliminate unlikely candidates at once at the first stage for efficiency. Then, the procedure is conducted for every individual probe/gallery image pair for higher accuracy and the least probable one

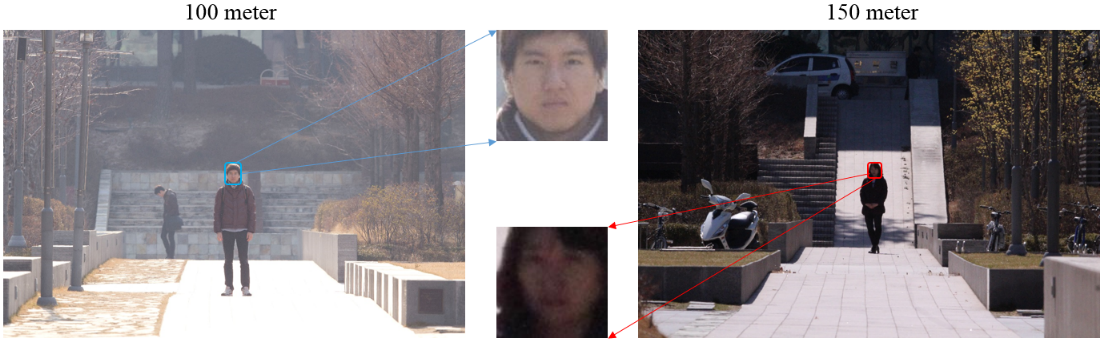

Figure 3.1: Two exemplary images in the LDHF database taken at a distance of 100 and 150 meters.

in the candidate pool is removed one by one until the last one at the second stage. With the two-stage filtering, TAEF strikes a balance between computational efficiency and matching accuracy.

The rest of this chapter is organized as follows. The background and related work are discussed in Section 3.1. The TAEF method is presented in Section 3.2. Experimental results are shown and evaluated in Section 3.3. Finally, concluding remarks are given in Section 3.4.

3.1 Background and Related Work

3.1.1 Face Alignment

Face alignment involves two steps: finding an initial rough face shape reference, and approaching the ground truth via iterative optimization. A well known technique is facial landmark localization that finds the coordinates of essential components iteratively. It can be further categorized into the model-based and regression-based approaches.

The model-based approach includes the Active Shape Model (ASM) [31] and the Active Appearance Model (AAM) [29]. They were derived based on the principal component analysis (PCA) of shape and appearance of landmarks. As an extension, a descriptor is used to capture the appearance for each landmark while these descriptors are constrained by a shape model in the AAM with a constrained local model (AAM/CLM) [33]. Saragih *et al.* [138] incorporated a mean-shift filter into AAM/CLM to achieve better matching capability. Cootes *et al.* [30] adopted a random forest method to compute accumulated votes to improve the alignment performance.

The regression-based approach utilizes local descriptors and regressors to reduce the matching error. The cascaded regression was introduced by Dollar *et al.* [41] for pose estimation in image sequences. Later, it was applied to face alignment. Cao *et al.* [20] proposed a regression method with two-stage training, where the cascaded regression was extended to the context of an affine transform. Xiong *et al.* [173] applied cascaded regression with the SIFT feature, examined the derived solution from a gradient descent view, and called it the Supervised Descent Method (SDM). Yan *et al.* [174] adopted a similar framework with the “learn-to-rank” and “learn-to-combine” modules placed in the front and the back of the main alignment module, respectively. Moreover, deep learning was introduced in [150, 152], which offers competitive performance.

Automatic face alignment techniques and systems have been extensively tested. For instance, Wagner *et al.* [158] used the spare representation in their alignment algorithm for the Multi-PIE database. Geng and Jiang [48] developed an automatic alignment system based on both holistic and local features and conducted experiments on the AR, GT and ORL databases. Deng *et al.* [38] proposed a Transform-Invariant PCA (TIPCA) method to achieve automatic alignment and tested its performance on the FERET dataset.

In this work, a face alignment method using cascaded regression is adopted. To address alignment distortion caused by the distance effect, we design a filter to mimic the long distance effect and generate facial landmarks accordingly.

3.1.2 Face Enhancement

Face image enhancement for FRAD received little attention before. However, it is much needed for FRAD as evidenced by the exemplary images shown in Figure 3.1. A system that incorporates wavelet decomposition, deblurring, denoising and linear stretching was proposed in [176] to recover quality loss due to the long distance. The reported recognition performance ranges from 50% to 70% due to low image resolution and quality. One face enhancement technique rooted in retinex theory was proposed by Land and McCann [88]. It examined relative lightness (instead of absolute lightness) in a local region to mimic the human visual experience. Later, Land [87] presented another approach to lightness computation. Based on this foundation, Jobson *et al.* [69] proposed a Single-Scale Retinex (SSR) method for the trade-off between rendition and dynamic range compression, and extended it to the Multi-Scale Retinex (MSR) method in [68]. More recently, Petro *et al.* [123] proposed a new method called MSR with Color Restoration (MSRCR).

In this work, the Multiscale Retinex with Color Restoration (MSRCR) method is adopted [123] for face enhancement to handle foggy and back-lighted conditions in the outdoor environment. To the best of our knowledge, this is the first time for MSRCR to be used in a face alignment/recognition system. We show that MSRCR can restore face quality in the LDHF database, where local contrast enhancement is used to overcome the back-lighted or foggy challenge while maintaining image color balance.

3.1.3 Long Distance Heterogeneous Face Database

As mentioned in Section 1.2, we select LDHF to evaluate our system. The LDHF database [78, 103] was released in 2012. It contains 1-meter indoor, 60-, 100- and 150-meter outdoor VIS and NIR images of 100 subjects.

In this work, we focus on visible-light images taken during daytime only. Two exemplary LDHF visible-light images are shown in Figure 3.1, which were taken at 100 and 150 meters, respectively. They have the same image size (i.e., 5184 × 3456 pixels) but different face sizes (i.e., 220 × 220 pixels for the 100-meter image and 120 × 120 pixels for the 150-meter image on average). The illumination in the LDHF database can be roughly categorized into three types, such as normal, foggy, and back-lighted. The two images in Figure 3.1 show how image quality can be affected by the foggy and the back-lighted environments. Apparently, both face alignment and enhancement techniques are needed before face matching. Furthermore, for cross-distance/environment facial matching, features contained in the interior face region could be too weak and additional information such as face contour and partial hair is helpful.

3.2 Proposed TAEF Method

The diagram of the TAEF system is shown in Figure 3.2. At the first stage, approximate facial landmarks are obtained using cascaded regression in the alignment step and MSRCR in the enhancement step. The objectives of this stage are two folds: filtering out unlikely candidates to reduce the size of the candidate pool and providing a better initialization for further processing. At the second stage, fine-scale alignment, enhancement and matching operations are performed iteratively to reduce the least probable candidate one by one until the last one is reached.

3.2.1 Coarse-scale Processing

Coarse-scale Alignment (C-Alignment)

Typically, the initial location of the face region is provided by face detection algorithms. It can be affected by the appearance variation such as poses, expressions and occlusion.

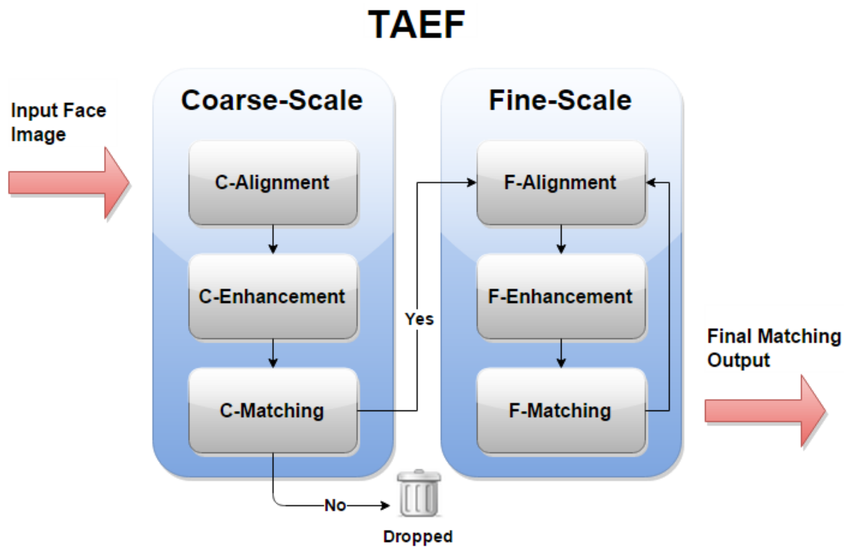

Figure 3.2: The system diagram of TAEF.

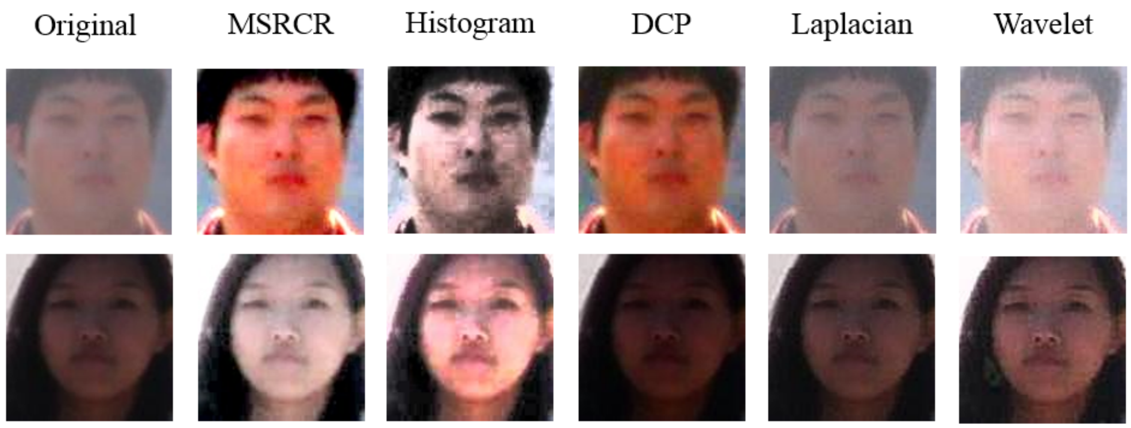

Figure 3.3: Two exemplary enhanced facial images (from left to right): original images and images enhanced by MSRCR, histogram equalization, dark channel prior, Laplacian sharpening and wavelet decomposition.

In this work, we focus on frontal faces with limited facial expression as provided by the LDHF database as the start point. With this simplified condition, we adopt the Viola-Jones [157] algorithm as the face detector and obtain an acceptable prediction outcome for the detected face region.

Most automatic alignment algorithms developed today do not work well in a long distance environment due to blurring and illumination distortions. Although the test image contains a long distance face, the ground truth is a short-distance face of higher resolution and better quality. Apparently, there is a mismatch between the training and testing data. To address the mismatch problem, we design a distortion filter that mimics the long distance effect so as to generate the facial landmarks of synthetic long distance images. The cascaded regression scheme has to be retrained using the distance-adjusted data. Besides, we need an initial location for each landmark in the cascaded regression. To achieve this goal, we conduct Procrustes analysis on all landmarks of images in the distance-adjusted training set to yield a face model formed by the averaged locations of landmarks. After the face region bounding box on the test image is generated and a face model is constructed based on the distance-adjusted training set, we map these landmarks to their corresponding locations in the test face region to generate initial landmark locations. Then, given an input face image, we apply the cascaded regression to reduce the distance between the estimated landmark position and that of the training data. This process can be written mathematically as follows.

To conduct the cascaded regression, we need initial facial shape and training set of multiple subjects. The initial facial shape is represented by the coordinates of N initial facial landmarks in form of $S^0 = [x_1, y_1, \cdots, x_N, y_N]$. The training set is denoted by $\{(I_k, \overline{S}_k)\}_{k=1}^{K}$, where I_k is the kth subject, $\overline{S}_k$ is the corresponding landmark-based facial representation, and K is the total number of training subjects. With these two inputs, cascaded regression generates a sequence of approximations $S^1, \cdots, S^t, \cdots, S^T$, where S^T is the converged output. The t-th facial shape is updated based on

$$S^t = S^{t-1} + R^t(I, S^{t-1}), \tag{3.1}$$

where

$$R^t = \arg\min_R \sum_{k=1}^{K} ||\overline{S}_k - [S_k^{t-1} + R(I_k, S_k^{t-1})]|| \tag{3.2}$$

is learned. R represents a regressor and S_k^{t-1} is the estimated shape produced in the previous stage. In this work, R is chosen to be a linear regressor since it can handle the desired task efficiently. To explain the above concept in words, we design a sequence of regressors, where each regressor is trained based on the difference between the estimated result from the previous stage, S_k^{t-1}, $k = 1, \cdots, K$, and the ground truth, $\overline{S}_k$. This iteration process stops when the training error converges. In our work, we adopt multi-scale HOG features [36] as the input descriptor for regressor's training. To be more specific, the large-scale HOG feature is extracted in the beginning stage while only a small-scale area around each estimated landmark is considered in the later stage.

Furthermore, we regulate the solution at the end of each iteration with two constraints. First, we adopt the sketch token feature from [97] as a reference for the face contour since it offers a reliable edge map using the mid-level feature. It is observed that the trained sketch token model offers an exceptional result on long distance faces in the designated region. Second, the estimated landmarks are constrained based on the face shape with the closest distance, so that no landmarks will deviate from the ground truth too much because of image quality degradation.

To sum up, the predicted facial landmarks for the kth subject at the end of the t-th iteration are obtained as the fusion of results from: 1) the predicted result from the t-th regressor, 2) the output obtained by imposing the sketch-token-based face contour constraint, and 3) the closest landmark model selected from the training set.

Coarse-scale Enhancement (C-Enhancement)

After getting landmarks from the C-alignment procedure, we attempt to restore the distorted facial color so as to allow robust cross-environment facial matching. We test several enhancement algorithms, including histogram equalization, dark channel prior [62], Laplacian sharpening, wavelet decomposition and MSRCR, and conclude that MSRCR

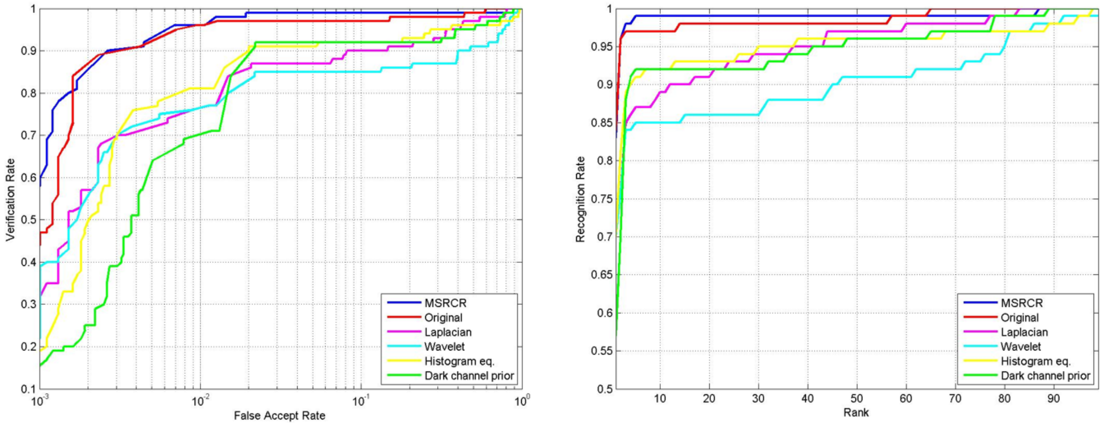

Figure 3.4: Comparison of the 150-to-1 meter face matching result with different enhancement methods: ROC (left) and CMC (right).

provides a superior performance on foggy and back-lighted images. Two examples are given in Figure 3.3 for subjective visual comparison. The top and bottom original images in Figure 3.3 are distorted by the foggy and back-lighted conditions, respectively. The goal of enhancement is to remove these environmental factors to allow cross-environment matching. We see that MSRCR does provide better results against the original ones.

For objective performance evaluation, we compare the Receiver Operating Characteristic (ROC) curve and the Cumulative Match Characteristic (CMC) curve of the matching result under different enhancement algorithms in Figure 3.4, where the matching result in this figure is generated using only the coarse-scale alignment/enhancement and will be detailed in the next subsection. We see from this figure that only MSRCR can improve the matching performance. It is worthwhile to point out that most image enhancement algorithms have been developed for white noise removal. The white noise model is however not suitable in characterizing outdoor distortions. The matching performance is actually worsen by these enhancement algorithms designed for other purposes. In contrast, MSRCR compensates the environment effect with a more suitable design and, as a result, it can offer better performance over the original one.

Facial Matching (C-Matching and F-Matching)

The facial matching component appears in both the coarse-scale and the fine-scale processing modules, where the same matching method is adopted and described below.

After proper alignment and enhancement, the first step is to extract feature descriptors (e.g. HOG and SIFT) and geometric features (i.e. facial landmarks) in polar coordinates. Polar coordinates are adopted because it can represent relative locations of aligned landmarks conveniently. Feature descriptors are extracted from two cropped face regions called the *interior face region* and the *bounded face region*, respectively. Two examples are illustrated in Figure 3.5. The interior face region includes major facial components up to eyebrows and down to part of chin without ears and hair. The bounded face region has the whole face including the face contour and partial hair such as bangs. Since the bounded face region is sensitive to background and hairstyle change, it is not used in traditional face recognition systems. However, for the cross-distance and cross-environment face recognition problem, the information contained in the interior face region could be too little. The additional information contained in the bounded face region can be helpful, and it is not proper to discard any relevant information due

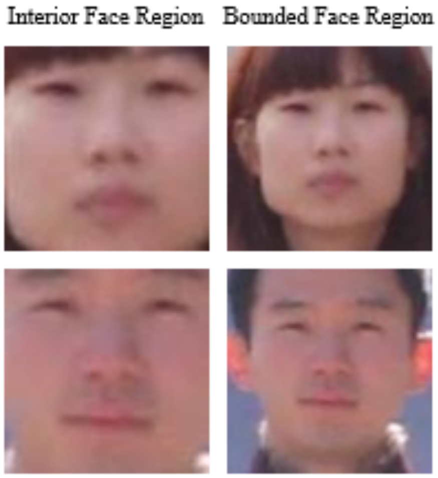

Figure 3.5: Illustration of interior and bounded face regions.

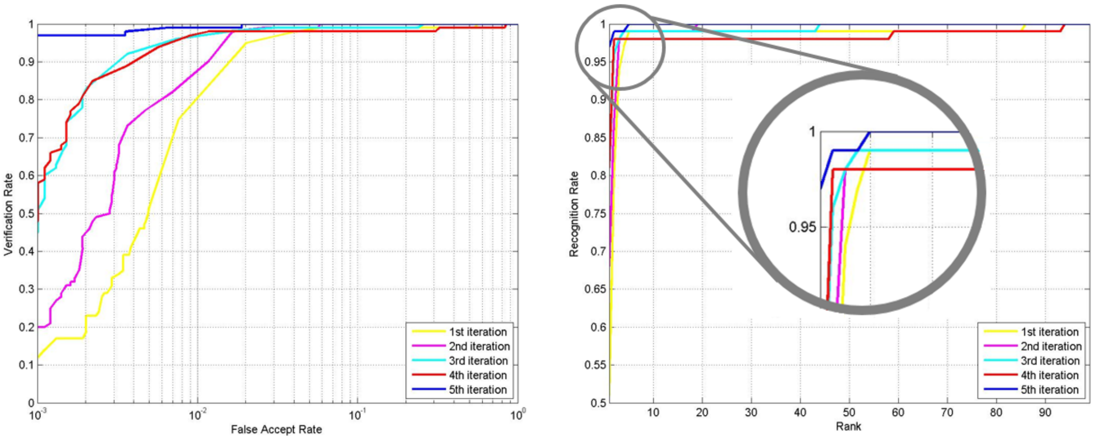

Figure 3.6: Performance of the 150-to-1 meter face matching result with different iterations: ROC (left) and CMC (right).

to aggressive cropping. Experiments show that the performance of using information from both interior and bounded face regions is better than that from only one of them. Thus, in our implementation, both HOG and SIFT features from the two regions are used separately as individual classifiers for further processing.

Moreover, both regions share the same Interpupillary Distance (IPD) as 40 pixels. The number 40 is chosen because it is the average IPD for 150 meter images, and the cutoff range is decided based on the total average face boundary. The reason to fix the IPD in both regions is to maintain the resolution alignment since we may generate distortions during resizing by allowing images of different scales.

After collecting all features from both interior and bounded regions of aligned and enhanced face images, we can measure the Euclidean distance of feature vectors and generate rank-order lists from all classifiers. Then, a weighted voting will be used to pile all classification results into one single rank-ordered list. By gradually eliminating less probable candidates in various stages, the TAEF system will provide the final ranked result.

3.2.2 Fine-scale Processing

An iterative alignment/enhancement filtering process is adopted in this stage. It means that, after eliminating the least possible candidate from the selection pool through alignment, enhancement and matching, features are extracted from re-normalized images based on remaining images in the pool at the next iteration. This process is described in detail below.

Sets of probe and gallery images are denoted by $P = \{P_1, \cdots, P_k, \cdots, P_{N_p}\}$ and $G = \{G_1, \cdots, G_k, \cdots, G_{N_g}\}$, where N_p and N_g are their sizes. Furthermore, $O_1, \cdots, O_k, \cdots, O_{N_p}$ are candidate pools for probe images $P_1, \cdots, P_k, \cdots, P_{N_p}$, respectively. The iterative filtering process consists of two steps at each iteration. First, probe image P_i is geometrically and photometrically normalized with subjects left in its candidate pool O_i so that the normalized probe image can be written as

$$P_i' = \Gamma(G_j, \Lambda(G_j, P_i)), \quad G_j \in O_i, \tag{3.3}$$

where Λ and Γ are the fine-scale alignment (F-alignment) and enhancement (F-enhancement) operations, respectively. Afterwards, the new candidate pool is expressed as

$$O_i' = \{O_i \mid V_j \geqslant N_v\}, \tag{3.4}$$

where

$$V_j = \sum_{l=1}^{N_c} w_l \cdot C_l(\Psi_l(P_i'), j) \tag{3.5}$$

is the vote collected from classifiers C_l, $l = 1, \cdots N_c$ using feature transform Ψ_l and weighting factor w_l, and N_v is a threshold of vote count for the pool.

The same cascaded regression in C-alignment is applied to the probe image in the F-alignment but with one major difference. That is, it is aligned with each individual gallery image G_j in the candidate pool O_i one by one, where the Procrustes analysis is conducted to derive the transform array. Translation, orthogonal rotation, reflection, and scale component are all calculated in this process. Furthermore, only reliable landmarks are selected to reduce the influence from inaccurate landmarks. For example, tips of eyes and the mouth often have higher steadiness and the nose rim location is difficult to determine in long distance. With these improvements, the F-alignment, denoted by Λ in Eq. (3.3) can reduce the estimation error based on the improvement in the last iteration.

The F-enhancement process, denoted by Γ in Eq. (3.3), is needed for photometric normalization, and it is achieved by region-based histogram matching. In contrast with the traditional histogram matching method that calculates the histogram of the whole face, we match histograms of the probe image and each individual gallery image in face sub-regions segmented by localized landmarks in the F-alignment step so as to differentiate images in a small candidate pool.

3.3 Experimental Results

3.3.1 Implementation Details

Since the LDHF database has a limited number of daytime gallery images to be used as training samples, we also include the MUCT database [108], which consists of 3755 images with 76 landmarks and is collected from 276 subjects. However, we only use 25 landmarks out of 76 in the C-alignment training process by focusing on visible landmarks such as center and edge tips of eyes and mouth in the long distance. The training set in C-alignment also contains gallery images from LDHF. Each face is manually labeled with 25 landmarks since the ground truth is not provided with the database. We simulate three long distance scenarios (contrast change due to long distance, fog and back-lighted) for each gallery image so that the size of the training images is tripled. This allows the regressor to learn in a cross-environment setting.

In the C-enhancement step, we apply MSRCR to the whole probe image, where its parameters are decided by the histogram distribution of each probe image. If the distribution contains a concentrated peak in the dark area, it should be under the back-lighted condition. If the distribution spans over a broad area, it should be under the foggy condition. Otherwise, it is under the normal condition. In our experiment, we set parameters $\alpha = 0.7$ and $c = 0.5$ in the Laplacian sharpening method and parameters $threshold = 50$ and $C = 2$ in the wavelet decomposition method. The enhancement performance is shown in Figure 3.4, where the verification rate under $FAR = 0.1\%$ is: 12% for dark channel prior, 18% for histogram equalization, 22% for wavelet decomposition, 32% for Laplacian sharpening, 44% for original and 58% for MSRCR.

In the fine-scale stage, TAEF collects votes from all classifiers to build up a candidate pool. We observe that HOG and SIFT features have the ability to select candidates of high similarity but with low first-rank accuracy. They can be used as the main features for both interior and bounded face regions, yet they need to be assisted with geometric features offered by landmarks. As a result, we have six classifiers based on the following feature sets: HOG and SIFT from interior and bounded face regions, landmark's angle and radius distributions (represented in polar coordinates). The voting mechanism collects votes from all six classifiers, and the top N candidates that receives most votes are placed in the initial candidate pool ($N = 5$ in the experiment). Then, one candidate is removed at each iteration until the final one is reached.

3.3.2 Performance Evaluation

We compare curves of ROC and CMC in Figure 3.6 to demonstrate the performance of the TAEF system. Note that we need a distance table among all candidates to draw ROC

Table 3.1: Comparison of ROC verification rates for 150-to-1 cross-distance face recognition.

Methods	0.1% FAR	1% FAR	10% FAR
Maeng [102]	93%	99%	100%
Kang [78]	75%	87%	99%
TAEF	**97%**	99%	100%

and CMC, where the distance table is built based on the received number of votes. A higher vote number means a closer distance and vice versa, and the distance is weighed by the iteration number.

We can see the performance improvement in ROC curve as the iteration number increases. For example, when $FAR = 0.001$, TAEF gives a verification rate of 12%, 20%, 45%, 48% and 97% at the 1st, 2nd, 3rd, 4th and 5th. The superior performance of the TAEF method is also demonstrated by the CMC plot. At the first iteration, the first rank recognition rate is 51% and it rapidly climbs up to 99% in rank 5. It follows the same aggregation pattern for later iterations. Its first rank recognition rates are 68%, 81%, 81% and 97% for iteration numbers 2, 3, 4 and 5, respectively.

For performance benchmarking, we selected the work of Maeng *et al.* [102] and that of Kang *et al.* [78]. Note that the former did not provide sufficient details on their alignment process while the latter relied on a commercial software called FaceVACS, and manually provided eye locations when the software failed to detect. For these reasons, we can only take the reported data from their papers for the comparison purpose. We list the ROC verification rates of three methods (TAEF and theirs) for 150-meter visible-light images in LDHF in Table 3.1. TAEF has the best performance among the three. Moreover, we test the TAEF method using 60 meter and 100 meter visible-light images in LDHF, and it gives 100% first rank recognition rate. Thus, TAEF offers the state-of-the-art performance for the FRAD problem at an outdoor setting.

We also compare the first rank recognition rate using features from only the interior or bounded face region or both under the same settings. The results for the 150-meter visible-light images at first rank are shown in Table 3.2. It is interesting to see that the performance of the bounded face region alone is better than that of the interior face region. Since the interior face region is blurred due to the long distance effect, its extracted features have limited discriminant power. The additional information from the bounded face region such as the face contour and hairstyle can play an important role although it is less robust. The TAEF system takes both into account and achieves the best performance.

Table 3.2: First rank recognition rates for different face regions.

Face region	Interior	Bounded	Both
1^{st} iteration	35%	46%	51%
2^{nd} iteration	49%	54%	68%
3^{rd} iteration	60%	69%	81%
4^{th} iteration	60%	72%	81%
5^{th} iteration	66%	86%	97%

3.3.3 Error Analysis

Among the 100 probe images located at the 150-meter distance, there are three failure cases for TAEF as shown in Figure 3.7. we show two intermediate processing results of probe images in the first two columns: the output after C-alignment in the first column and the final normalized result in the second column. Furthermore, their ground truth of the 1-meter gallery image is shown in the fourth column while their predicted match by TAEF is shown in the third column. The ground truth images of these three subjects from top to bottom rank as No. 2, No. 2, and No. 4, respectively.

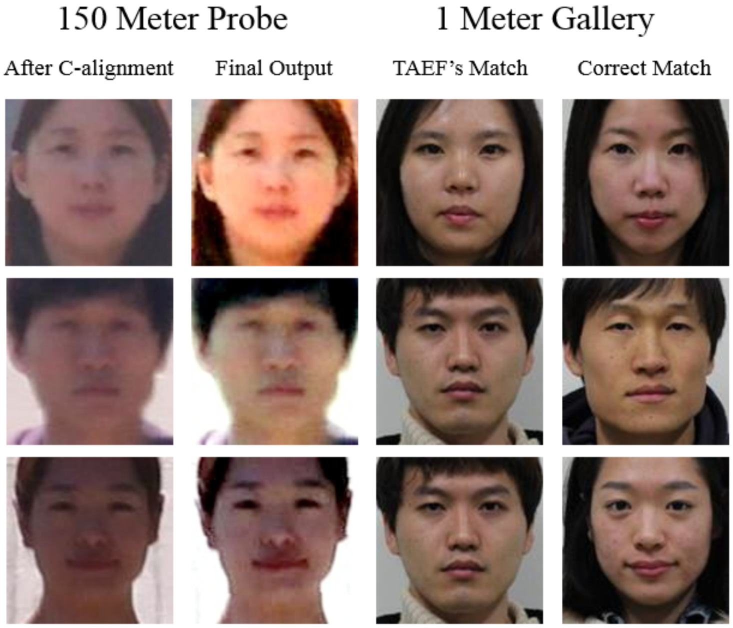

Figure 3.7: Error cases for 150 meter visible-light probe images.

One obvious reason for the error is attributed to the difference of the hair style between gallery and probe images. For example, for the case in the bottom row, the hairstyle of the 1-meter gallery ground truth is completely changed in her corresponding 150-meter probe image. Generally speaking, the hairstyle and the chin shape visible in the bounded face region do contribute positively to the recognition performance. This case happens to work against this policy.

Another reason is due to other environmental factors such as blurring, which is not yet considered in the TAEF system. For the first two rows, the interior face regions of the final output images from the TAEF system are still blurred. The loss of details in pupils and eye's shape can mislead HOG and SIFT descriptor classifiers. With these two blurred probe images, the TAEF system fails to choose the correct one in the last round.

3.4 Summary

In this chapter, we presented an interesting and practical face recognition problem where the probe images are located at a distance in an outdoor environment. We discussed several challenging issues existing in this problem and proposed a solution called the TAEF method to address them. The TAEF method offers a state-of-the-art solution to this cross-distance and cross-environment face matching problem during daytime, and we also offers a solution for a the cross-distance and cross-spectral matching during nighttime in the next chapter.

Chapter 4

Cross-Distance Near-Infrared Face Recognition

In this chapter, we continue to extend our work to a different environment setting, which is near-infrared (NIR) faces at long distance. An ideal surveillance system should operate around the clock, including both day and night time. Currently, cameras equipped with flash lights are used for night time to offer acceptable performance. However, they are not appropriate for long distance or convert surveillance. As a result, we need to consider other options for night time face recognition. Methods like near infrared (NIR), shortwave infrared (SWIR), and thermal infrared have been studied in the literatures. NIR has become popular in recent years for several reasons [194]. First, NIR is not visible to human eyes, and it is desired to capture face expressions without interrupting subjects in acquisition. Second, the environmental factor has less impact to NIR when compared with others. Third, the NIR illuminator can penetrate glasses easily, which provides additional information if the test subject wears glasses. Generally speaking, NIR offers a good choice for night time long distance face recognition.

On the other hand, if we desire to directly compare NIR images with photos captured in visual light (VIS), the comparison could be difficult due to highly distorted spectrum perception. Furthermore, the long distance environmental effects even corrupt the degraded image. In order to achieve NIR-to-VIS matching, where 1 meter VIS image is taken as the gallery set and cross-distance NIR images as the probe set, we need to develop an alternative approach instead of TAEF only, because it is almost infeasible to directly apply the same system when NIR sensing has different mechanism compared to VIS under different distances. As we can see from Figure 4.1, the image quality has different degradation via long distance under NIR or VIS spectrum.

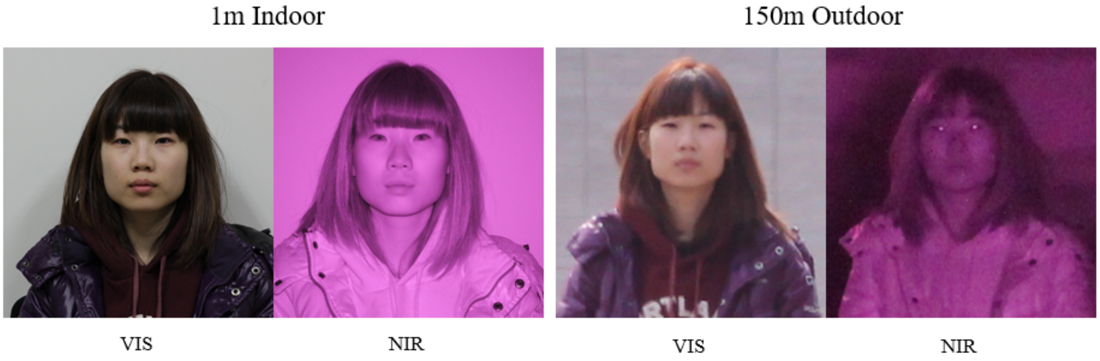

Figure 4.1: Comparison of VIS and NIR under 1 meter indoor and 150 meter outdoor environments.

Therefore, it is an urgent step to repair the low quality images before applying the matching process, because feature descriptors like HOG and SIFT give better performance while the testing input has closer image structure similarity to the gallery set. For example, Maeng *et al.* [103] proved there is obvious performance degradation of the matching score if we directly compare NIR with VIS face images without further enhancement. Under this observation, we propose a restoration scheme that adopts Locally Linear Embedding (LLE) [133], which reconstructs low-quality patches from the mapping between low-quality and high-quality patches. Moreover, we further segment image into overlapping grids, so that LLE can learn the local region characteristics within the grid-based structure. In consequence, the restored result helps the recognition system to extract better feature descriptors in the matching process, and it also demonstrates the strength of LLE's application on grid-based localized approach.

This chapter is organized as follows. The corresponding work and databases are introduced in Section 4.1, and experimental result and performance evaluation are presented in Section 4.3. Finally, we conclude this chapter in Section 4.4.

4.1 Near-Infrared Face Recognition

Previous works on NIR face recognition can be categorized into two major fields: NIR to NIR matching (known as intra-spectral matching), NIR to VIS matching (cross-spectral matching). NIR to NIR matching is commonly taken when NIR images are enrolled within the gallery dataset, and then the matching process is more focused on the feature degradation of long distance under the influence of NIR spectrum. However, the obvious drawback is the situation of having only VIS gallery images whereas the probe images are all NIR images. Pan *et al.* [117] applied spectral measurement on multiple facial tissue types, where subjects were taken at a close distance but with different pose and expression; Zhao *et al.* [188] utilized DCT and SVM for the cross-spectral matching, and other approaches like LBP, ELBP, DBC, Gabor, and Adaboost were adopted and obtained competitive results in this field [65, 92, 142, 183]. Bourlai [15] also evaluated cross-distance NIR to NIR matching with CSU face matcher.

For the second category, there are two major directions to handle NIR to VIS matching: correlation learning and face synthesis/reconstruction on pixel level. The correlation learning focuses on establishing the relation between NIR and VIS through learning strategy, like Yi *et al.* [179] proposed a canonical correlation analysis (CCA) based method with PCA; Liao *et al.* [94] provided another solution with Local Structure of Normalized Appearance (LSNA) and MB-LBP; Klare *et al.* [79, 80] presented a framework by solving nonlinear similarities between NIR and VIS images; Maeng [102] used DoG-SIFT to build up the relation; Yi *et al.* [178] also improved their work with Gabor filter and Restricted Boltzmann Machines (RBM). On the other hand, face synthesis/reconstruction on pixel level intends to restore low quality images via cross-spectral learning from two domains. For instance, Chen *et al.* [25] applied LLE with LBP for NIR to VIS restoration; Wang *et al.* [164] proposed the face analogy method incorporating LoG and checking facial texture patterns from the same region; Zhang *et*

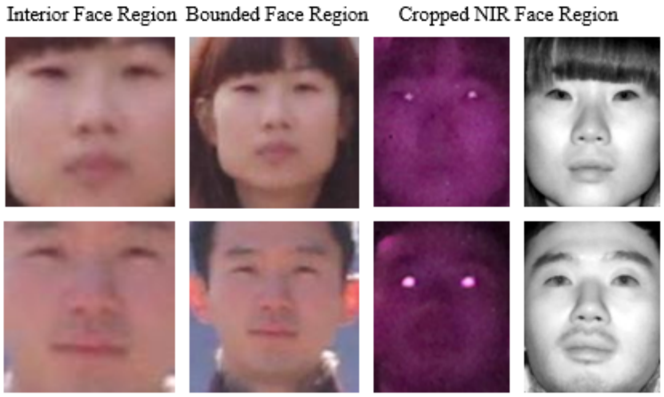

Figure 4.2: Illustration of different cropped facial regions (from left to right): interior VIS face region cropping, bounded VIS face region cropping, 150-meter and 1-meter NIR face region cropping regions.

al. [186] also presented the model using LBP and sparse representation; Kang *et al.* [78] utilized LLE with K-means and augmented heterogeneous face recognition (AHFR) to solve restoration for NIR face images at long distance. Furthermore, Xu *et al.* [72] proposed a cross-spectral joint l_0 minimization based dictionary learning with the same purpose from NIR to VIS.

4.2 Face Restoration From Near-Infrared Environment

4.2.1 Preprocessing

Due to the low signal-to-noise ratio (SNR) of NIR images, we adopt a tighter cropped facial region for NIR images as shown in Figure 4.2, where the IPD is set to 92 pixels. This choice not only rules out most background noise but also preserves needed facial texture. In this pre-processing step, all NIR face images are cropped into the same size (192 × 240 pixels), where eye locations are automatically detected and aligned using the technique described in previous chapter.

Because of very poor NIR image quality at 100 and 150 meters, the Viola-Jones detector is not as effective as being applied to VIS images. It gives 35 false positive and 5 false negative cases among one hundred 150-meter NIR images. We explore pupil's reflection from the NIR illuminator as an auxiliary tool to mark the eye location. With this extra procedure, we are able to detect all faces on NIR 150 meter images without any error. For image enhancement, we would like to maintain fidelity without removing too many details. The 3×3 median filter is applied to suppress high frequency noise, and a simple image contrast adjustment is adopted to enhance the image. The contrast adjustment is decided based on whole image's gray-scale histogram distribution. That is, since the NIR image does not receive sufficient light, we adjust the intensity values to meet the condition that 1% of data of the whole image is saturated at low and high intensities.

To partially recover the lost information caused by the long distance and the spectral difference partially, we need to bridge the gap between 1 meter and 150 meter NIR images by building their correspondence. Being inspired by the work in [21], we adopt the LLE method to achieve this goal, which is explained in detail in the next subsection.

4.2.2 Restoration System

The proposed restoration scheme is based on the framework in [78] but with additional features to boost up the performance. It consists of two stages: 1) the correspondence building stage and 2) the correspondence finding stage. In the first stage, we build the correspondence between high quality and low quality patches. In the second stage, we use the correspondence to reconstruct patches to restore the quality of the probe image.

The operation in the first stage is illustrated in Figure 4.3. We first partition 150-meter (low quality) and 1-meter (high quality) NIR images in the gallery set into multiple subregions, and extract a large number of patches from each subregion. The pixels of low quality and high quality patches are cascaded into a single vector. Then, we use the tree-structured vector quantization (TSVQ) to cluster vectors separately based on their source subregion. TSVQ clusters vectors into two groups at each level, and repeat the same procedure at each group recursively until the desired cluster number is reached. The centroid of each cluster is called a codeword, and the set of all codewords generated by the TSVQ is called a codebook (or a dictionary). Thus, we can associate a codebook with each subregion.

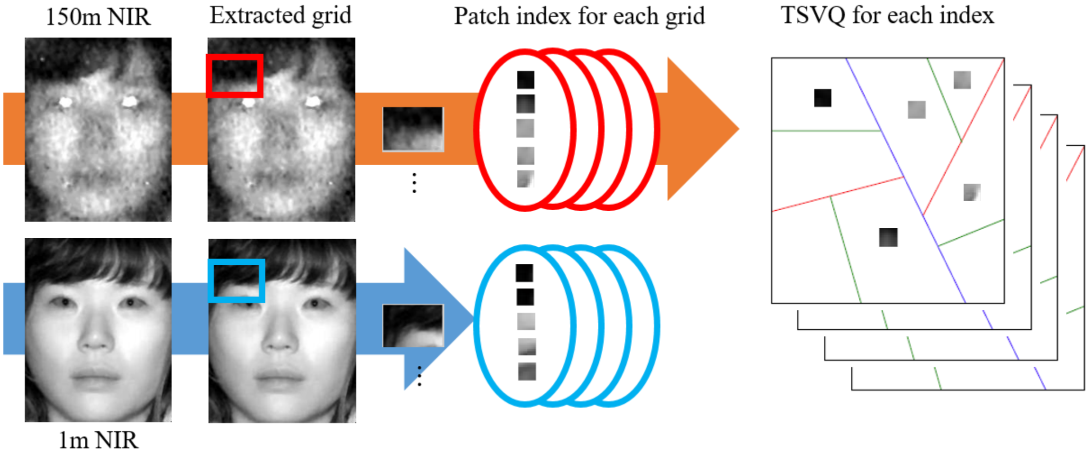

Figure 4.3: The construction of two corresponding codebooks using 150-meter and 1-meter NIR patches, where low quality (150-meter NIR) and high quality (1-meter NIR) patches are extracted from the same subregion of the same subject and cascaded into a vector. TSVQ is used to generate a codebook for each subregion

Mathematically, we use $\{(G_k^H, G_k^L)\}_{k=1}^K$ to denote the 1 meter and 150 meter NIR images from the same subject in the gallery set G, where K is the total number of subjects and superscripts L and H indicate low and high quality images, respectively. We divide each image into smaller subregions denoted by $\{D_t, t = 1, 2, ..., T\}$, and extract pairs of corresponding patches from the same location,

$$P_{D_t}^k = \{(\pi_i^H, \pi_i^L), \quad i = 1, 2, ..., n\},$$

where T is the total number of subregions in a facial image and n is the total number of patches in each subregion.

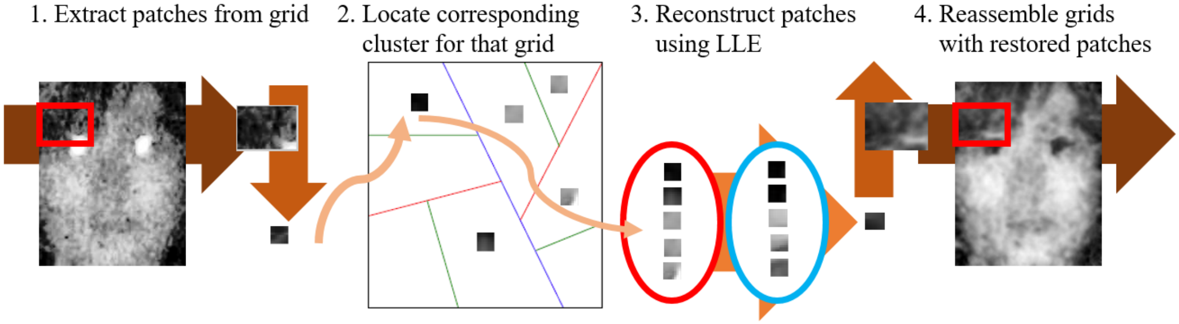
1. Extract patches from grid
2. Locate corresponding cluster for that grid
3. Reconstruct patches using LLE
4. Reassemble grids with restored patches

Figure 4.4: The process of restoring low quality patches using the LLE-based method. Patches are extracted from the designated subregion. Then, the system inqures the corresponding codebook, locates its neighbor patches, and reconstructs the target via LLE. Finally, the restored face image is resembled by all reconstructed patches.

From manifold learning, the correspondence between two manifolds can be learned when they possess similar local geometries. Here, we use LLE to learn the relationship between high quality (1 meter) and low quality (150 meter) patches in each cluster. Based on the learned relationship, we can reconstruct the 150 meter probe patches in each subregion and, then, restore the whole image accordingly.

Figure 4.4 demonstrates the restoration procedure. Once a probe image's patch is extracted, we can use it to locate the closest cluster, $\{C_o^L, o = 1, 2, ..., O\}$, in each subregion through the minimum Euclidean distance, where O indicates the total number of clusters in each subregion. For a low quality patch denoted by π_j^L, we select its S nearest neighbor patches of the same cluster and use them to calculate weights $\{w_s, s = 1, 2, ..., S\}$ that minimize the following error

$$\varepsilon_j = ||\pi_j^L - \sum_{\pi_s^L \in N_j^L} w_s \pi_s^L||, \tag{4.1}$$

where N_j^L denotes the nearest neighbor low quality patches from C_o^L, and $||.||$ is the Euclidean norm.

With weights w_s obtained from the above equation, we can use them to reconstruct the corresponding high quality patch

$$\pi_j^H = \sum_{\pi_s^H \in N_j^H} w_s \pi_s^H, \tag{4.2}$$

where N_j^H is the corresponding nearest neighbor high quality patches from cluster C_o^H. We should emphasize that C_o^H contains high quality patches in the same geometric location of C_o^L, except that they are extracted from 1 meter NIR image of the same subject. Furthermore, we take the regional background into consideration by averaging the restored image with the cluster mean image C_o^H in the subregion (e.g., one half from the restored image and the other half from the cluster mean image). Therefore, the restored result will be less sensitive if the nearest neighbor patches fail to represent the input probe patch. Finally, we reassemble patches back to the subregion to restore the whole facial image. We use overlapping subregions to reduce the blocking effect.

The major difference between VIS and NIR images is image quality. Furthermore, since there is no large-scale NIR face dataset with labeled landmark annotation, the performance of NIR facial landmark localization is limited. Here, we replace the C-alignment and the C-enhancement steps with “integration from face detection and eye location marking” and “LLE-based image restoration”, respectively. The TAEF system will output the final decision for NIR images right after C-matching. There is no processing needed in the fine-scale stage.

4.3 Experimental Result

4.3.1 Implementation Details

For the restoration mechanism, we select 48×64 as the subregion size and 8×8 as the patch size. They are determined by considering the trade off between restoration performance and system efficiency. A larger subregion offers higher efficiency but poor restoration capability, and vice versa. To generate more samples, we allow overlapping subregions and patches with a quarter of their boundary size. This also allows smoother transition across the boundaries of subregions and patches. As a result, there are 153 subregions per image and 609 patches per subregion. Since we adopt 10-fold cross-validation in the dictionary building stage, there are 54,810 patches per grid. We apply the 8-level TSVQ with 256 clusters per subregion. There are 214 patches per cluster on the average. In the patch reconstruction phase, we identify the associated cluster for a long-distance probe patch, choose its five nearest neighbors from the same cluster, and calculate their weights to approximate the probe one. Then, we use these weights and their corresponding 1 meter patches to reconstruct the 1 meter patch.

4.3.2 Performance Evaluation

Five pairs of pre-processed (or intermediate) and restored (or final) face images are shown in Figure 4.5, where intermediate results after the pre-processing step in Sec. 4.2.1 are shown in the first row and the final output images using the LLE-based restoration method are presented in the second row. Clearly, the restored ones give better visual quality than the pre-processed ones. They have less noise and bear higher similarity with the corresponding 1 meter NIR images. Moreover, we compare the visual appearances of the subregion and global-region restoration schemes in Figure 4.6. Note that there is no subregion decomposition in the global-region restoration scheme. All

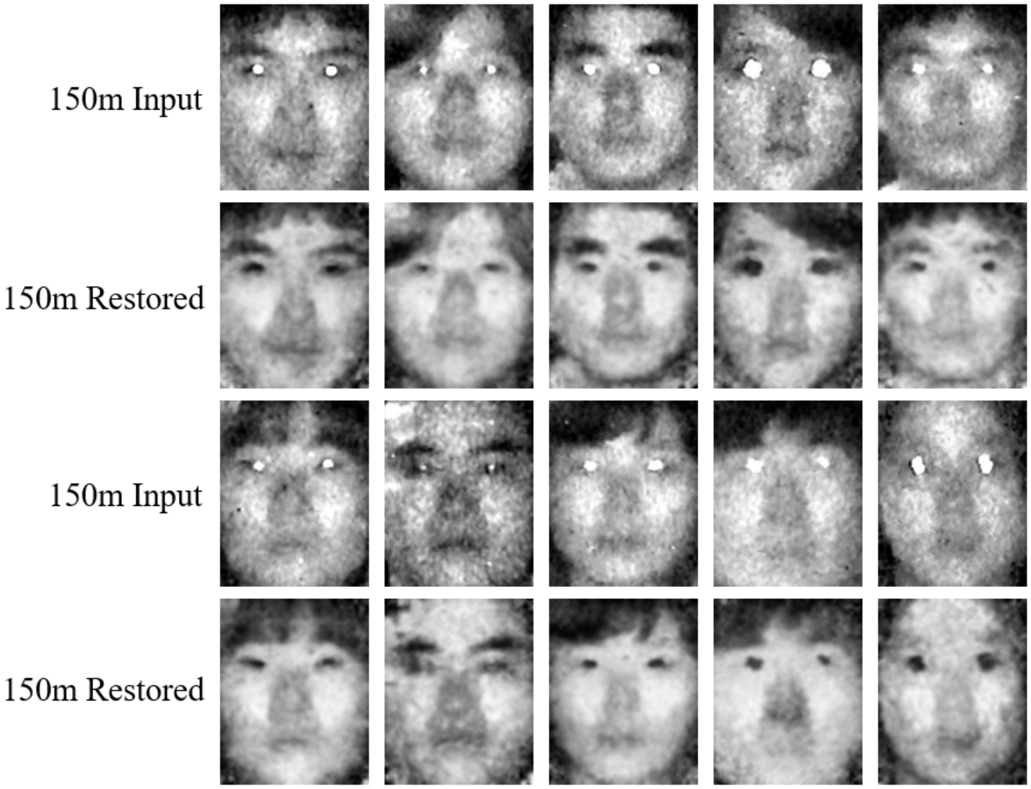

Figure 4.5: The restoration result of 150 meter NIR images. First and third rows show the original input images, and their corresponding outputs are listed in second and forth rows.

collected patches are restored by LLE in one single system. Its result is vulnerable to local variants. In contrast, the subregion restoration scheme gives significantly better results because of its ability to preserve local characteristics within each subregion.

Our restoration results are compared with those obtained by Kang *et al.* [78] in Figure 4.7. We see from the figure that the local facial textures are better preserved by our method. For instance, the top subject in Figure 4.7 has sharper and richer eyebrow shape as compared to the benchmark one. The bottom subject possesses distinct mouth characteristics such as the lip contour and corner.

With restoration, we are able to boost up the first rank accuracy rate on the 150 meter NIR images from 8% to 52% with the HOG feature, and 62% to 76% with the

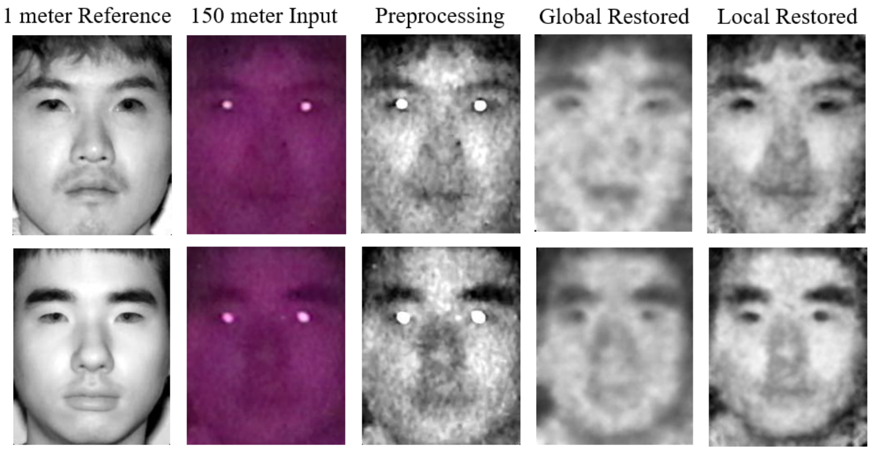

Figure 4.6: Comparison between the 1 meter NIR reference, the 150 meter NIR input, the input with preprocessing, the output of the global-region restoration, and the output using subregion restoration.

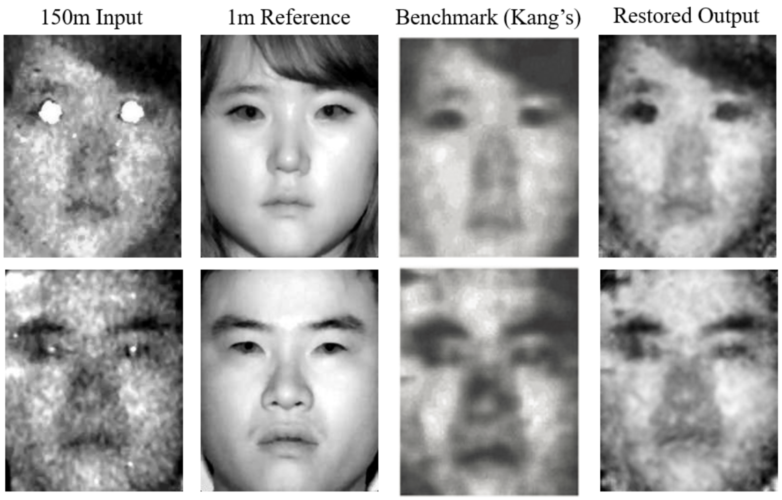

Figure 4.7: Comparison of restoration results obtained by our method (the last column) and by the method proposed by Kang *et al.* [78] (the 3rd column), where the first column shows the input 150 meter images and the second column shows the 1 meter reference images.

SIFT feature as illustrated in Figure 4.8. This performance gain demonstrates the ability of recovering some lost information by restoration. We further apply the C-matching step of TAEF to obtain weighted votes from HOG and SIFT extracted from the cropped NIR face region and the interior face region. The proposed TAEF with restoration can

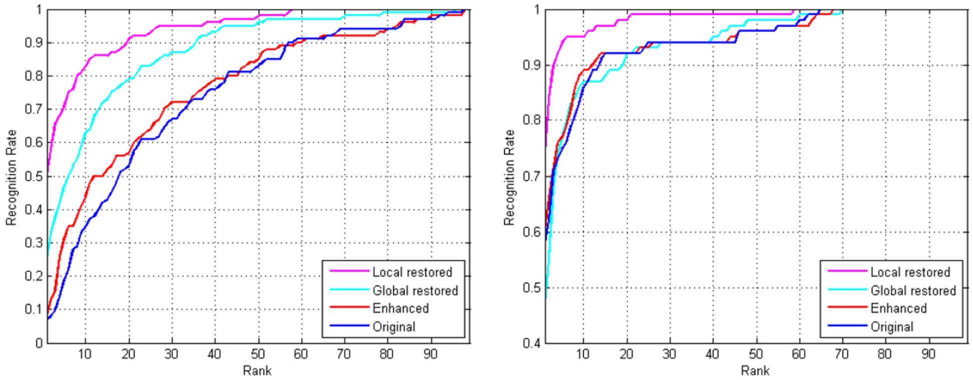

Figure 4.8: The CMC curves of the restored 150 meter NIR images. The left subfigure gives the comparison between subregion restoration, global-region restoration, pre-processed and original images (as illustrated in Figure 4.6) using the HOG feature while the right one uses the SIFT feature.

achieve 45% verification rate at 0.1% FAR for 150 meter NIR images, which outperforms Kang's system by 8% under the same distance. We show the comparison in Table 4.1 with other state-of-the-art methods.

4.3.3 Deep Learning

Deep learning is a popular tool in computer vision applications nowadays. However, due to the limited size of the LDHF dataset, it is difficult to train an effective deep network in our current context. We should emphasize that there is only one image for each subject under the same distance in this dataset, and the correspondence between images of each subject at various distances is fixed. The same data augmentation technique (e.g., rotation, mirroring or random sampling) has to be applied to both the input and output image pairs simultaneously. As a result, the technique does not offer more discriminant power among subjects.

On the other hand, several image processing (denoising, inpainting, and super-resolution) problems have been solved by the deep learning technology. In this subsection, we examine the application of deep learning to NIR image restoration.

Table 4.1: Comparison of ROC verification rates for 150-to-1 meter NIR-to-VIS face recognition.

Methods	0.1% FAR	1% FAR	10% FAR
Maeng [102]	6%	20%	56%
DWT-SVD [113]	29%	29%	64%
DWT-PCA [113]	36%	39%	68%
Kang [78]	37%	62%	92%
Local Restored	45%	72%	95%
SRCNN [42]	49%	65%	93%
Local Restored + SRCNN	49%	75%	97%

The super-resolution convolutional neural network (SRCNN) was proposed in [42] to learn the mapping between low- and high-resolution images. Here, we use it as another benchmark for image restoration. We adopt the same network architecture in [42] except changing the first layer's filter size from 9×9 to 7×7 to fit the input image size of our problem better. The restoration relation between low- and high-quality images is learned from training samples. Besides local restored LLE and SRCNN, we conduct experiments by cascading LLE and SRCNN (namely, applying SRCNN to the restored LLE's output). We compare the retored images obtained by the three methods in the left side of Figure 4.9. We see that images restored by the SRCNN are smoother and sometimes blurred with sufficient details. We also compare the CMC curves of the three methods in Figure 4.9. As shown in the figure, the CMC performance of the SRCNN is worst among the three. The local restored LLE method and the cascaded method have comparable performance. The cascaded method provides slightly better performance because the SRCNN method can improve the image quality based on restored images.

Finally, we compare the ROC verification rates of several methods in Table 4.1. Although the SRCNN method outperforms the local restored LLE method by 4% at the 0.1% FAR, it does not perform well on 1% FAR and 10% FAR. Overall, the cascaded method provides the best performance among all benchmarking methods in Figure 4.9 and Table 4.1.

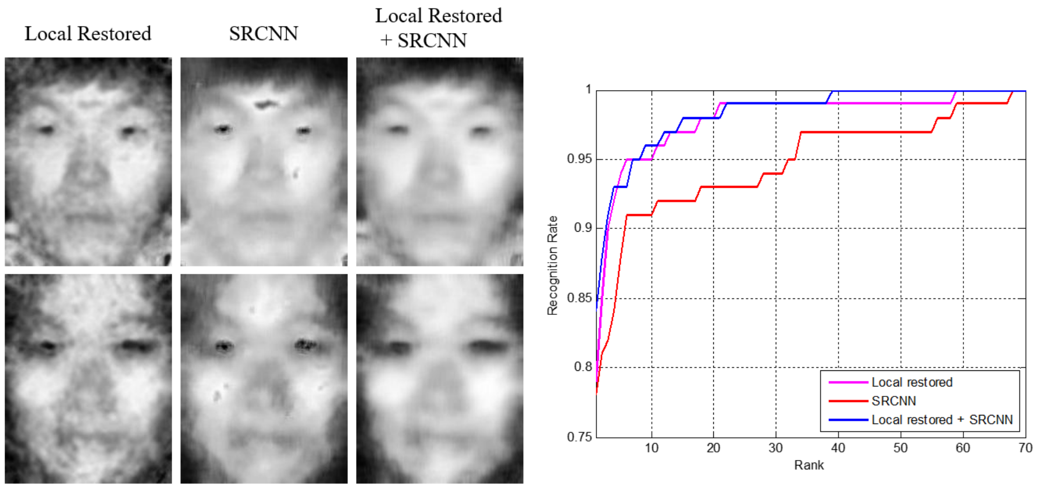

Figure 4.9: Left: Exemplary restored images obtained by the local restored LLE method, the SRCNN method, and the local restored LLE followed by the SRCNN. Right: The CMC curves of the three methods.

4.3.4 Error Analysis

We examine the error cases for NIR images in this subsection. We show several of them in Figure 4.10. The pre-processed image, the restored output, the predicted match and the ground truth are displayed in order along each row. The hair style difference between the gallery and the probe images plays an important factor. For instance, the subject in the first row has a smaller fringe in the probe image which is similar to the gallery one. Furthermore, blurred NIR images may lead to different gradients/contours that may confuse the classifier. As compared with the VIS case, NIR image quality degradation is much more significant and serious.

4.4 Summary

we presented a restoration system for cross-spectral face matching problem. The proposed restoration system can restore the NIR cross-distance probe image via learning the modality gap between VIS and NIR face images. In addition, given 1 meter and 150

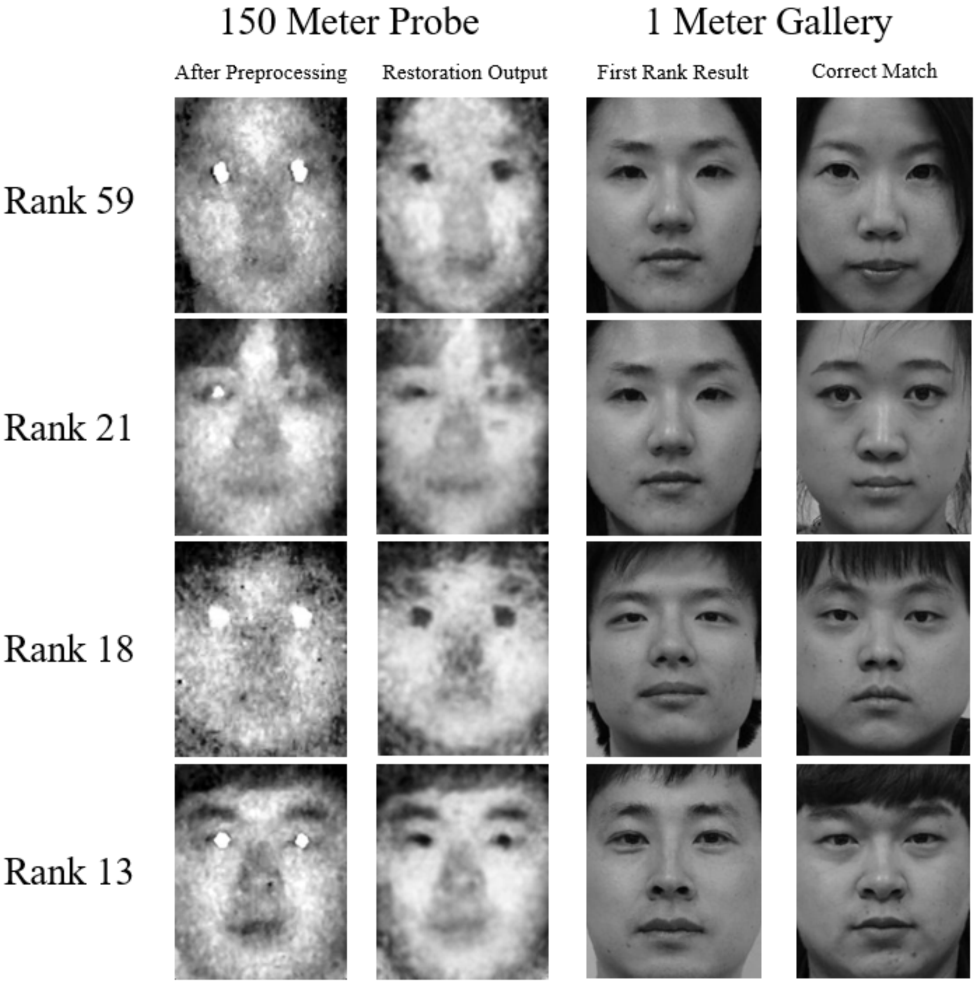

Figure 4.10: Error cases of 150 meter NIR probe images (from left to right): the original input after preprocessing, the restored output, the final matching result, and the ground truth.

meter corresponding NIR images, each subregion has its own LLE model to recover high-quality patches, which later become parts of the restored image. The effectiveness of the solution is demonstrated by higher accuracy rates.

The main issue in the FRAD problem is the lacking of a large long distance face dataset. It is critical to build such a dataset for further research advancement along this direction. Also, the convolutional neural network (CNN) has been tested in short-distance facial recognition problems and reported to have an impressive performance gain. It will be interesting to try the CNN solution if a large labeled long distance facial image dataset is available.

Chapter 5

Age/Gender Classification with Neural Networks

5.1 Introduction

The Convolutional Neural Network (CNN) is a powerful machine learning tool in computer vision. There have been numerous research activities on the application of the CNN to object detection, recognition and semantic segmentation, etc. Yet, recognizing human age/gender attributes using the CNN remains to be an interesting problem for further exploration. The automatic human age/gender classification technology finds many real world applications such as target advertisement, demographics analysis, visual surveillance, etc. Here, we explore the use of the CNN for age/gender classification based on human facial data.

Traditional facial image datasets with age/gender attributes were built in a controlled indoor environment. There exists a gap between the collected datasets and the uncontrolled environment in real world applications. In a practical scenario, facial variations exist due to image quality degradation, face poses, and occlusions. Thus, the attribute classification models trained by traditional datasets do not perform well in real world applications. There are new datasets built recently to narrow down the gap such as the Adience dataset [43]. The Adience dataset was built in 2014 for face-based age/gender classification in an unconstrained environment. It is one of the most challenging dataset in this field. We select the training/testing data from the Adience dataset in this work.

To achieve human age/gender classification from their facial image data, we present a novel CNN solution that contains two types of neural networks – the whole face network and the facial component networks, which provide the whole and component facial information, respectively. For this reason, we call the network solution as the Whole-Component CNN (WC-CNN) system. The proposed WC-CNN system consists of four building modules: 1) the face and facial components localization module, 2) the whole face network (or the global network), 3) the facial component networks (or the local networks), and 4) the final classification module using confidence analysis. Each module is designated with different functionalities.

The localization module takes care of all preprocessing tasks, such as face detection and facial landmark localization. The goal is to localize faces and their component regions. The whole face network and the facial component networks are trained separately with extracted patches for age/gender classification. We use the whole face network as the primary classifier to yield the initial classification result. Afterwards, confidence analysis is used to evaluate the confidence score of the initial decision. We accept its decision if the confidence score is high. Otherwise, the system will make a final decision by considering the outputs from both the global and local networks jointly. The proposed WC-CNN solution achieves the state-of-the-art performance for age/gender classification against the challenging Adience dataset.

The rest of this chapter is organized as follows. The related previous work is reviewed in Section 5.2. The proposed WC-CNN system is presented in Section 5.3. The experimental results of the WC-CNN system applied to the Adience dataset are shown in Section 5.4. Finally, concluding remarks are given in Section 5.5.

5.2 Review of Previous Work

Gender recognition is clearly a classification problem. On the other hand, age recognition can be formulated as either an estimation or a classification problem. That is, some researchers attempt to find the exact age via regression while others divide the entire age range into multiple intervals and treat it as a classification problem. Given human facial images, quite a few classic methods were proposed for age/gender recognition without the neural network technology. They were extensively discussed in several survey papers, *e.g.,* [46, 59, 104, 130]. We will have a brief review on this subject.

5.2.1 Age/Gender Datasets

There are a few popular age/gender datasets built under controlled indoor environment. Examples include the FERET [125], FG-NET [1], MORPH I, and MORPH II [131] datasets. The experiments in most recently published papers were conducted on these datasets. However, these datasets have limited variations in terms of facial poses, expressions and occlusions. They do not meet the need of practical real world applications.

The Adience dataset [43] was constructed to narrow the gap between experimental and practical applications. It was built from raw photos uploaded by smart phones without further processing, and the collected images cover a wide range of scenarios. A benchmark algorithm was proposed by Eidinger *et al.* in [43]. It is the dropout support vector machines (dropout-SVM), where the SVM classifier is trained using a dropout strategy to avoid overfitting. Levi and Hassner [89] demonstrated the effectiveness of the CNN solution with respect to the Adience dataset. Here, we propose a new CNN-based solution that outperforms the two previous methods described in [43] and [89] agaist the Adience dataset. Niu *et al.* [112] released a new dataset that focuses on Asian faces in 2016. It contains more than 160 thousand images. However, the collected faces are all near frontal. Thus, it does not fit our interest under the criterion of being captured in an uncontrolled environment.

5.2.2 Features-based Methods

Early age determination methods were built upon facial features extracted from various geometric distances [86]. Then, models were constructed using extracted features to estimate and classify subjects into different age groups [128]. These methods highly depend on landmark localization. However, since there was no mature alignment technique to provide accurate distance measure, the recognition performance was low at that time. The aging process was treated as a subspace or manifold in [49, 56] with an objective to learn the correspondence between different age stages of the same subject. Then, one may model a high dimensional image space with a low dimensional feature space. Nevertheless, this approach has a drawback; namely, the manifold structure of aging facial appearances may not be consistent among different subjects.

For gender classification, the neural-network-based solution was applied to a small set of frontal face images in [52]. The 3D head structure with image intensity values were also examined in [115]. Other gender classification methods directly applied the SVM [109] and AdaBoost [7] classifiers to image intensity values.

Generally speaking, for age/gender classification, one common strategy in traditional methods is to extract local features to represent the face image and then apply a machine learning technique for the classification task. Several popular features and classifiers have been combined to develop a total solution. Examples include the Gabor image features with the Fuzzy-LDA classifier in [47], the biologically insepired features (BIF) with manifold-learning in [175], the local binary patterns (LBP) with the support vector regression (SVR) in [27], etc. Some features are particularly powerful for gender classificion, *e.g.,* the Webers local texture features [154] and the shape and texture features [121]. More recently, Liu *et al.* [98] proposed a multistage learning system, called the Grouping Estimation Fusion (GEF), by fusing multiple decisions obtained from several age grouping methods and several global and local features such as BIF, HOG and LBP. This work obtained impressive age estimation results on the FG-NET and MORPH II datasets.

5.2.3 Components-based Methods

A comparison study between the whole-face-based and the component-based face recognition approaches was conducted in [63], and the whole-face-based approach was shown to give better performance. On the other hand, recent studies in [14, 45, 58] demonstrated that the component-based approach can provide an auxiliary tool to aid the whole-face-based approach in face detection, alignment, and recognition.

For age/gender classification, Han *et al.* [59] proposed a hierarchical age estimation method by applying the BIF features and the SVM classifier to facial components. However, the performance was limited by inaccurate facial localizations and weak feature discrimination. The component-based idea was also exploited by the CNN-based solution. For example, the DeepID [151] extracted features from different face regions to form complementary facial representations.

5.2.4 CNN-based Methods

Researchers applied the CNN to the age estimation problem in recent years. Yi *et al.* [177] used a subset of the MORPH II dataset to train a shallow CNN that consists of only one convolutional layer. Wang *et al.* [168] adopted the CNN as a feature extraction tool, yet did not utilize its strength fully. Furthermore, Niu *et al.* [112] proposed a multiple output CNN with ordinal regression to achieve end-to-end learning. The results reported

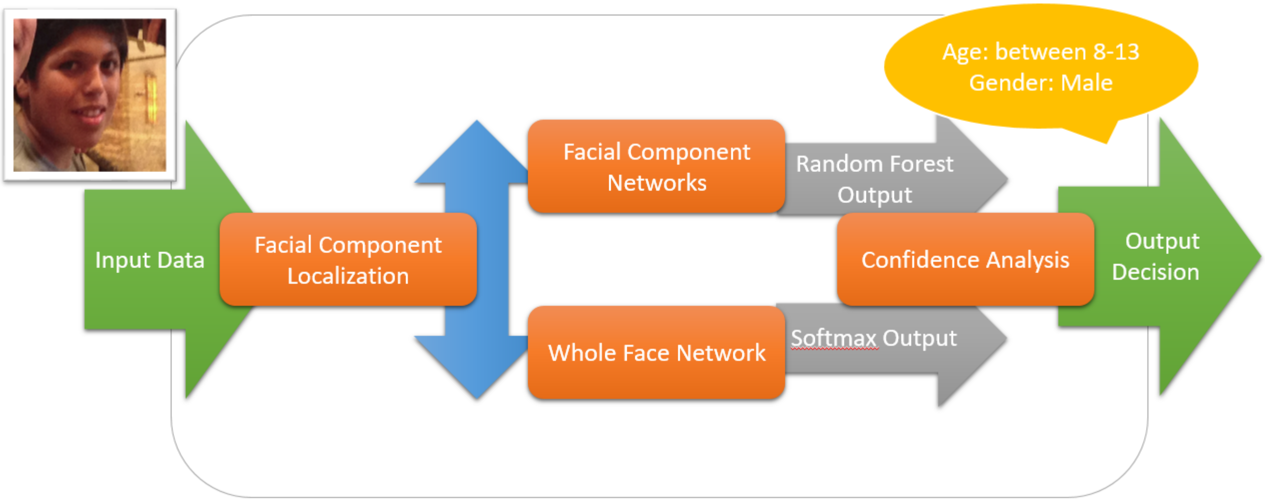

Figure 5.1: The diagram of the proposed WC-CNN system. Frist, the facial component localization module extract the whole face and local component regions from an input image. Then, we train the whole face network and the facial component networks with extracted whole face and local component regions, respectively. In the testing stage, we feed the corresponding regions into these two networks. Finally, if the decision from the whole-face network has a high confidence score, we accept this decision. Otherwise, the system will make a decision by combining results from both the whole face and the component networks.

in all these papers do not surpass the mean absolute errors (MAE) result reported in [98] on the MORPH II dataset.

5.3 Whole-Component CNN (WC-CNN) Method

The proposed whole-component CNN (WC-CNN) method is composed by four major modules: 1) the face and facial components localization module, 2) the whole face network, 3) the facial component networks, and 4) the final classification module. Those four modules are depicted in Fig. 5.1 below.

The WC-CNN method applies the face detection and facial landmark localization techniques to input images to extract the whole face and the facial component regions in the first module as discussed in Sec. 5.3.1. For the second and third modules, we train the whole face network and facial component networks using the extracted regions, respectively, in the training stage. We feed the extracted whole face and facial component regions into their respective networks in the testing stage. These two modules are detailed in Sec. 5.3.2. For the fourth module, we conduct confidence analysis on the initial decision obtained by the whole face network. This is achieved by analyzing the angle between a particular sample and the anchor vector of the network, which will be elaborated in Sec. 5.3.3. If the confidence score is high, we simply accept the decision. If the confidence score is low, the system will generate a final decision based on results from both the whole face and facial component networks.

5.3.1 Face and Facial Component Localization

.

This module serves two main purposes: 1) to locate face region and 2) to detect facial landmarks and identify facial component regions. Its outputs will be the whole face region and the facial component regions. In the implementation, we adopt the AlexNet [82] fine-tuned by the AFLW dataset [81] as the face detector for better face detection performance. The original AlexNet was trained for the classification problem. We change the fully-connected layers into the convolutional layers by reshaping layer parameters so that the network output is a heat map that marks potential face locations. We use the heat map followed by non-maximal suppression to detect faces. The detection accuracy of this method applied to the Adience dataset reaches 97.5%. For the undetected faces, we follow the default setting of the Adience dataset by claiming that the face is located at the image center.

Afterwards, the facial landmark localization algorithm is applied to the detected face region. To obtain accurate localization, we re-train the VGG Face in [119] with facial landmarks from the AFLW dataset. Generating accurate facial landmarks demands more detailed information on the face. This justifies the use of the VGG Face, which is a deep network and able to provide richer features. When faces are successfully detected, we can use facial landmarks to generate four facial component regions: left and right eyes, nose and mouth. The sizes of facial components can vary because of different postures or expressions. Thus, it is critical to apply a normalization procedure to these detected regions before feeding them to CNNs in the testing stage. We resize component regions according to their corresponding averaged sizes in each category. For instance, the left eye region will be normalized to the averaged size of all left eye regions in the training set.

5.3.2 Whole Face and Facial Component Networks

We present the architectures of the whole face network and the facial component networks and discuss their training in this subsection.

As shown in Fig. 5.2, the whole face network is composed by three convolutional and two fully-connected layers. Each convolutional layer is followed by the Rectified Linear Unit (ReLU), the max pooling operation, and the local response normalization (LRN) operation with an exception in the last convolutional layer which does not need the LRN. Its parameter setting is similar to that of the AlexNet. The number of layers and the number of filters per layer are determined by the classification numbers in the Adience dataset – eight age groups two gender categories with a total of 16 classes. Because of the low output dimensions, the small CNN fits our need while it is easier to train. Each fully-connected layer is followed by a ReLU and a dropout layer. Finally, a softmax operation is conducted and the training loss is calculated.

We consider four facial components: the left eye, the right eye, the nose and the mouth. They serve as the inputs to the facial component networks as shown in Fig. 5.3. The major difference between the whole face and the facial component networks

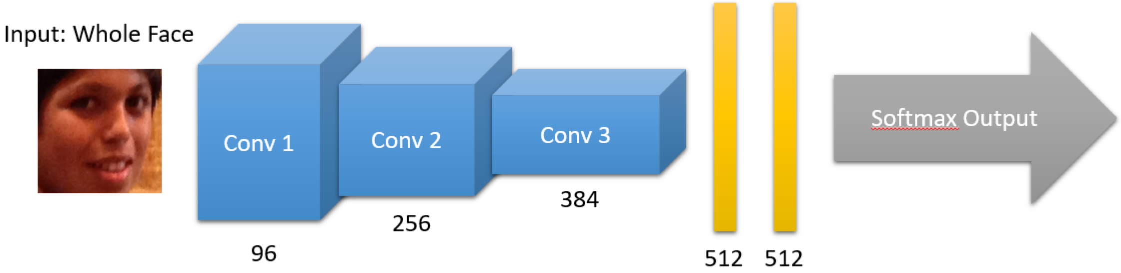

Figure 5.2: The architecture of the whole face network.

is the number of layers. The former requires three convolutional layers and two fully-connected layers while the latter only requires two convolutional layers and one fully-connected layer. Since the component region is smaller, it requires a simpler network. It will be shown in the section of experiments that the facial component network with two convolutional layers has almost the same verification accuracy as that with three convolutional layers.

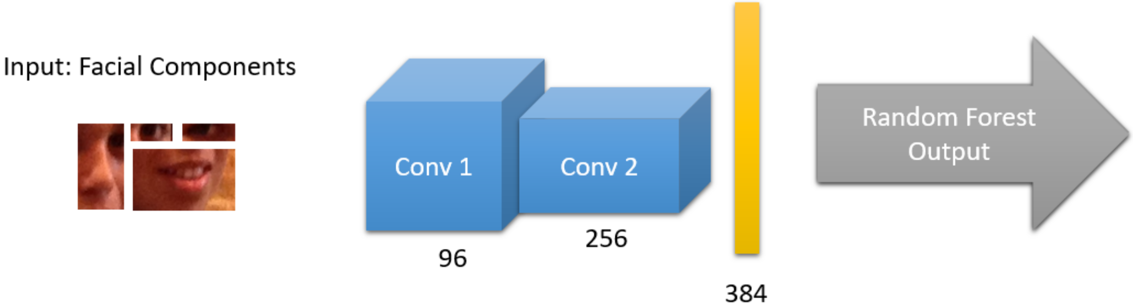

Figure 5.3: The architecture of the facial component network.

As to network initialization, we adopt the K-means clustering method to initialize the filter weights [28] to speed up the training time. If there is a cluster that contains few training samples, we can either reinitialize it with a random sample or just drop it. Both are proven to be effective.

5.3.3 Confidence Analysis

We first combine the whole-face and the facial component networks into one connected network as shown in Fig. 5.4, where all facial components networks and the whole face network is depicted from left to right. Each network has its own input data and layers. Their outputs are concatenated to form a softmax loss vector. Then, we can conduct the end-to-end training for the full system. However, the overall classification accuracy of this integrated system is better than the whole face network only by 1-2%. This small gain is probably due to the redundancy of the whole-face and the facial component networks. They share overlapping input regions while conflicting decisions cannot be properly resolved by the limited number of training samples.

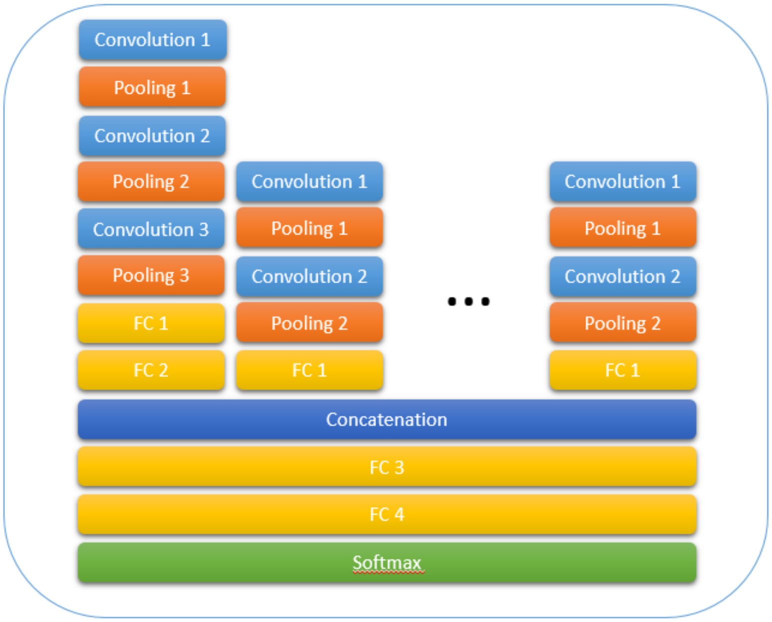

Figure 5.4: The integration of the whole-face and the facial component networks. Each network has its own input while the outputs of all networks are concatenated to form a softmax loss vector.

As an alternative, we use the whole face network as the primary classifier to yield an initial classification result. Confidence analysis is then conducted to evaluate the confidence score of the initial decision. We accept its decision if the confidence score is high. Otherwise, the system will make a final decision by considering the outputs from both the global and local networks jointly.

Our confidence analysis is built upon the “REctified-COrrelations on a Sphere” (RECOS) model proposed in [84]. The anchor vector refers to a set of filter weights associated with the same output node since it serves as a reference signal (or a visual pattern) in the testing stage. The convolution operation can be viewed as the signal correlation or projection onto a set of anchor vectors, and we can check the similarity between vectors by examining their correlation. A RECOS model is given in Fig. 5.5 to illustrate the confidence analysis. The RECOS model is a origin-centered unit sphere, and the dots on sphere’s surface at the last layer of a CNN are projected samples’ output vectors (anchors). If there are K decision classes, they form K clusters and each anchor points to the centroid location of each cluster. For a given decision, we calculate the distance between the decision vector and the anchor. If the distance is smaller (or larger), the decision is more (or less) confident.

In addition, in order to determine the confidence threshold within each decision class without handcrafting, we apply K-means clustering in each decision cluster. The cluster with the shortest geodesic distance between its centroid and the decision vector is labeled as the group with high confidence score. Other clusters’ decisions are then classified as the low confidence score group.

For low confidence decisions, we adopt random forest classifiers [19] to train features from the fully-connected layer in facial component networks, thus we can obtain more subtle information from component patches.

In the testing stage, we identify the confidence score from a test image. If the decision by the whole-face network is of low confidence, we will evaluate the test image by the facial component networks and then apply the trained random forest models to form weighted votes as the final result.

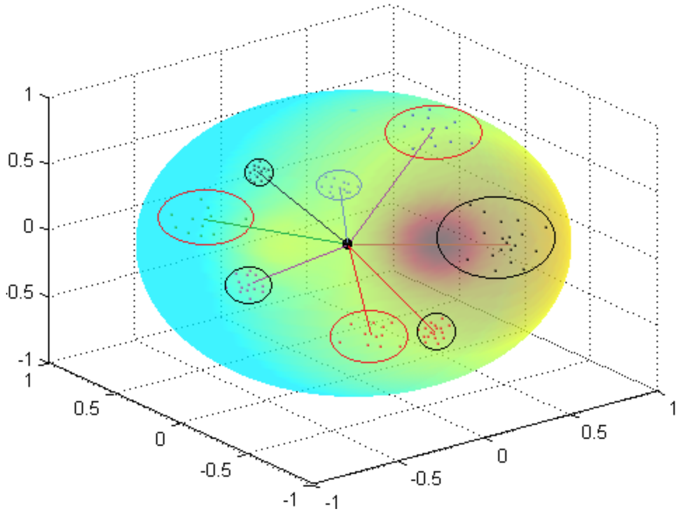

Figure 5.5: Illustration of confidence analysis using the anchor vector, where scattered dots on the sphere represent projected decision vectors from network's outputs.

5.4 Experimental Results

In this section, we first describe implementation details and then compare the performance the proposed WC-CNN with that of other age/gender classification methods.

5.4.1 Implementation Details

The experiment is conducted on the Adience dataset [43], which is designed for age/gender classification in an unconstrained environment. The Adience dataset contains 19,487 images of 2,284 subjects with 8 age groups: 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53 and 60-. Most age groups have around one to two thousand images except for two senior groups (only around eight hundred images each), and the 25-32 group (about five thousand images).

For the network parameter, we adopt the same setting as the Alex-Net in the kernel size, the filter number and the stride number. For the whole face network, the input data dimension is 250×250 with three channels. The input data size for the facial component networks are listed are: eyes - $150 \times 50 \times 3$, nose - $65 \times 125 \times 3$, and mouth - $150 \times 75 \times 3$.

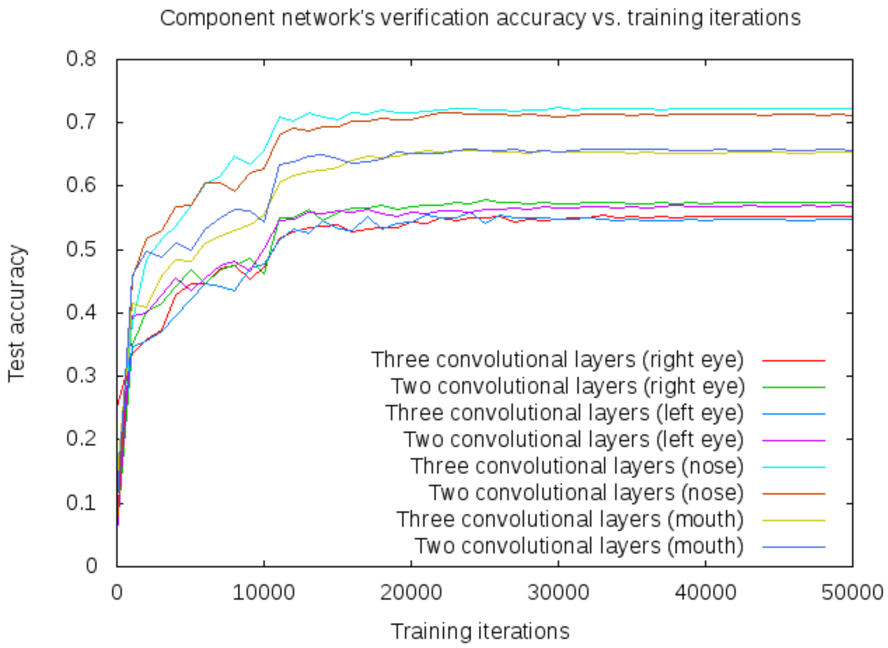

Figure 5.6: The verification accuracy versus training iterations for the facial component network, which shows that network with two convolutional layers can achieve a similar accuracy rate with the network with three convolutional layers on all facial components.

Fig. 5.6 shows that we only need two convolutional layers to reach similar performance with that with three. This figure is measured and compared based on the verification set from one test fold of the Adience dataset.

For the training, the batch size is set to 50, the momentum and weight decay are set to 0.9 and 0.0005, respectively. The learning rate starts at 0.001 and, then, it is decreased by a factor of 10 every 40 epochs. In total, it decreases three times before the learning stops after 200 epochs. It takes approximately two hours on GPU Titan X for the whole face network and one hour or less for the facial component networks.

5.4.2 Performance Evaluation

We follow the predefined five-fold subject-exclusive cross-validation protocol of the Adience dataset [43] to evaluate the performance of age/gender classification. Table 5.1

lists the accuracy rates of all five testing folds for age classification categorized by the confidence score. The rows of the table represent five folds from the testing set. The first and third columns show the percentage of testing samples are classified as high confidence and low confidence, respectively. Since the whole face network is confident about its result in the high confidence decisions, it is reasonable to have higher accuracy rate. Fourth and fifth columns are used to demonstrate the performance gap of having weighted votes from the facial component networks. We observe that there are different degree of performance improvement in test folds depending on their difficult cases.

Since both confidence groups have different testing samples in different folds, which affects the final decision's accuracy rate after we combine them together as the final result. We compare the proposed WC-CNN system with other methods against the Adience dataset in Table 5.2. The first column represents the age classification result with the first rank accuracy rate, and the second column shows the accuracy rate covering the adjacent age groups with one group distance, namely "1-off" prediction. The third column demonstrates that our proposed system also works well on gender classification problem, obtaining similar performance gain with age classification. As shown in this table, the WC-CNN offers the state-of-the-art performance in both age and gender classification.

Table 5.1: The age classification results on the five-fold cross validation testing set. Testing set samples are divided into two groups according to their confidence score. Each group has its sample number percentage and the corresponding accuracy rate.

	High Confidence Score Group		Low Confidence Score Group		
	Percentage of Testing Set	Accuracy Rate	Percentage of Testing Set	Accuracy Rate without Component	Accuracy Rate with Component
Test Fold 0	48.87%	72.60%	51.13%	48.31%	56.37%
Test Fold 1	40.22%	50.03%	59.78%	38.38%	39.84%
Test Fold 2	49.56%	69.67%	50.44%	43.55%	46.97%
Test Fold 3	40.72%	57.21%	59.28%	37.65%	38.23%
Test Fold 4	43.17%	64.15%	56.83%	41.06%	47.64%

We compare the proposed WC-CNN system with other methods against the Adience dataset in Table 5.2. The first column represents the age classification result with

the first rank accuracy rate, and the second column shows the accuracy rate covering the adjacent age groups with one group distance, namely "1-off" prediction. The third column demonstrates that our proposed system also works well on gender classification problem, obtaining similar performance gain with age classification. As shown in this table, the WC-CNN offers the state-of-the-art performance in both age and gender classification.

Table 5.2: Age/gender classification result on Adience dataset. We compare our work with other works under three categories: age classification with exact match to the ground truth, age group matching up to one adjancent age group (could be one group younger or older), and the exact match of gender classification.

Methods	Age (Exact)	Age (1-off)	Gender
Eidlinger [43]	45.1 ± 2.6	79.5 ± 1.4	77.8 ± 1.3
Levi [89]	50.7 ± 5.1	84.7 ± 2.2	86.8 ± 1.4
Ours	**54.3 ± 3.5**	**87.6 ± 1.9**	**89.6 ± 1.3**

5.4.3 Error Analysis

In order to find out the performance bottleneck of our solution, we dig into the misclassification cases for error analysis. There are two major reasons: 1) misalignment in facial component localization, and 2) data constraints in the Adience dataset such as low image resolution, occlusion or heavy makeup. In addition, the data characteristics has a great impact on the age/gender classification. For instance, the performance of gender classification is poor on baby images, since it is even a challenging task for humans to distinguish the gender of a baby. For age classification, groups of higher variations are more difficult for WC-CNN to classify. Age groups such as teenagers or people in their 40s have lower accuracy rates as compared with other age groups.

5.5 Conclusion and Future Work

The WC-CNN method, which incorporates the whole face and the facial component networks, was proposed for age/gender classification. The proposed scheme contains relatively simple CNN architectures which is easy to train and test. It is suitable for the implementation in embedded systems with limited resources. Even with the proposed solution, age/gender classification in an unconstrained environment remains to be an open and challenging problem. There is still room for further improvement as the future work. For example, it is desired to take care those difficult cases as mentioned at the end of Sec. 5.4.3.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Three major tasks were accomplished in this thesis research.

First, we examined and verified the performance of the TAEF system on LDHF dataset with cross-distance and cross-spectral conditions. The TAEF is composed by two stages: coarse-scale processing and fine-scale processing. In the coarse-scale stage, a quick scan-through procedure of face alignment, enhancement and matching is executed on all gallery images, so a new candidate pool is established for eliminating most unlikely candidates. Therefore, the overall efficiency is improved by checking the confined list within the pool only. In the fine-scale stage, the procedure is further conducted into more detailed alignment and enhancement by pair-wise reference. For example, each individual pair in the candidate pool has its own alignment/enhancement process, thus the comparison between different candidates at the matching step will have the maximum inter-distance. The results demonstrated TAEF's superior verification rate and its ability to adapt cross-distance VIS environment.

Next, since the TAEF system cannot provide competitive results with the NIR image as the input, we presented a restoration system for the cross-spectral matching problem. The proposed restoration system can restore the NIR cross-distance probe image via learning the modality gap between VIS and NIR face images. In addition, given 1 meter and 150 meter corresponding NIR images, each grid has its own LLE to recover high-quality patches, which later become parts of the restored image. We show the effectiveness of the system by proving the improvement of the accuracy rate over images after preprocessing. The proposed TAEF with the restoration system was evaluated under cross-distance and cross-spectral face matching, and our experimental results demonstrated the proposed method achieves competitive performance compared to other baselines.

Finally, we proposed a CNN-based system to solve the human age/gender classification problem. The system consists of four building modules: 1) the face and facial components localization module, 2) the whole face network, 3) the facial component networks, and 4) the final classification module assisted by the confidence analysis. The localization module is used to localize the face and its component regions. The whole face network is used as the primary classifier to yield the initial classification result. An confidence analysis is conducted to evaluate the confidence level of the initial decision. If the confidence level is high, we accept its decision. If the confidence level is low, the system will make a final decision by considering the outputs from the whole face network and the component networks jointly.

6.2 Future Work

There are several possible extensions of the research presented in this dissertation.

6.2.1 Building Larger Face Datasets

As compared with other face recognition datasets, the size and the variety of the existing long distance face recognition dataset are still very limited. For instance, the LDHF dataset contains only one hundred subjects with cross-spectral and cross-distance images, which is relatively small comparing to other state-of-the-art face datasets like well-known Labeled Faces in the Wild (LFW) or newly built Labeled Wikipedia Faces (LWF) datasets. LFW contains 13233 images with 5749 subjects, and LWF has 8500 images from 1500 identities. Comparing to those large volume datasets, it is important to build a larger dataset in the research community to facilitate future research and development work along this line. Furthermore, it is worthwhile to build a novel dataset contained with cross-distance videos instead of images. In this way, the recorded subjects are no longer bounded to certain distances, which increases the flexibility and applicability of the experiment. The number of subjects in this dataset should be larger than LDHF, providing a competitive choice among other collections. Furthermore, to reduce the distance between our experiment and practical usage, it is necessary to consider some degree of expression, pose, and occlusion into further improvement. Similar to LFW, the established dataset should be inclusive of all unconstrained environment scenarios. In this way, it can significantly bridge the gap between experiments and practical situations and fill the hole of current major researches in unconstrained face recognition, which have not taken cross-distance and cross-spectral into consideration.

6.2.2 Novel Face Recognition Techniques Based on Features and Raw Data

There are two main face recognition approaches nowadays; namely, the traditional feature-based approach and the modern CNN-based approach. The former can be used if the number of training samples is small. The latter can be used if the number of training samples is large. For the traditional feature-based approach, one can develop face feature descriptors such as SIFT or HOG from a small set of training samples for the matching purpose. However, these descriptors perform poorly when the image quality is degraded due to cross-distance and cross-spectral effects. For the modern CNN-based approach, we need a large number of labeled training samples, which may impose some challenges when the size of the training dataset is small. It will be interesting to take the strengths of both approaches and consider a joint approach. A recognition system can first learn from a small training dataset using features. Then, when the size of the training data grows, we can gradually switch to the data-driven learning method such as the CNN-based methodology.

Bibliography

[1] (2009). The fg-net aging database. http://www.fgnet.rsunit.com/. [Online].

[2] Ahn, G.-J. and Sandhu, R. (2000). Role-based authorization constraints specification. *ACM Trans. Inf. Syst. Secur.*, 3:207–226.

[3] Atallah, M. J., Blanton, M., Fazio, N., and Frikken, K. B. (2009). Dynamic and efficient key management for access hierarchies. *ACM Trans. Inf. Syst. Secur.*, 12(3):18:1–18:43.

[4] Ateniese, G., Burns, R., Curtmola, R., Herring, J., Khan, O., Kissner, L., Peterson, Z., and Song, D. (2011). Remote data checking using provable data possession. *ACM Trans. Inf. Syst. Secur.*, 14(1):12:1–12:34.

[5] Ateniese, G., Burns, R., Curtmola, R., Herring, J., Kissner, L., Peterson, Z., and Song, D. (2007). Provable data possession at untrusted stores. In *Proceedings of the 14th ACM conference on Computer and communications security*, CCS '07, pages 598–609, New York, NY, USA. ACM.

[6] Ateniese, G., Di Pietro, R., Mancini, L. V., and Tsudik, G. (2008). Scalable and efficient provable data possession. In *Proceedings of the 4th international conference on Security and privacy in communication netowrks*, SecureComm '08, pages 9:1–9:10, New York, NY, USA. ACM.

[7] Baluja, S. and Rowley, H. A. (2007). Boosting sex identification performance. *International Journal of computer vision*, 71(1):111–119.

[8] Bellare, M., Boldyreva, A., and O'Neill, A. (2007). Deterministic and efficiently searchable encryption. In *Proceedings of the 27th annual international cryptology conference on Advances in cryptology*, CRYPTO'07, pages 535–552, Berlin, Heidelberg. Springer-Verlag.

[9] Bethencourt, J., Sahai, A., and Waters, B. (2007). Ciphertext-policy attribute-based encryption. In *Proceedings of the 2007 IEEE Symposium on Security and Privacy*, SP '07, pages 321–334, Washington, DC, USA. IEEE Computer Society.

[10] Boldyreva, A., Chenette, N., Lee, Y., and O'Neill, A. (2009). Order-preserving symmetric encryption. In *Proceedings of the 28th Annual International Conference on Advances in Cryptology: the Theory and Applications of Cryptographic Techniques*, EUROCRYPT '09, pages 224–241, Berlin, Heidelberg. Springer-Verlag.

[11] Boldyreva, A., Chenette, N., and O'Neill, A. (2011). Order-preserving encryption revisited: Improved security analysis and alternative solutions. In *Proceedings of the 31st Annual Conference on Advances in Cryptology*, CRYPTO'11, pages 578–595, Berlin, Heidelberg. Springer-Verlag.

[12] Boneh, D., Di Crescenzo, G., Ostrovsky, R., and Persiano, G. (2004). Public key encryption with keyword search. In Cachin, C. and Camenisch, J., editors, *Advances in Cryptology - EUROCRYPT 2004*, volume 3027 of *Lecture Notes in Computer Science*, pages 506–522. Springer Berlin / Heidelberg.

[13] Boneh, D., Lynn, B., and Shacham, H. (2001). Short signatures from the weil pairing. In Boyd, C., editor, *Advances in Cryptology ASIACRYPT 2001*, volume 2248 of *Lecture Notes in Computer Science*, pages 514–532. Springer Berlin / Heidelberg.

[14] Bonnen, K., Klare, B. F., and Jain, A. K. (2013). Component-based representation in automated face recognition. *IEEE transactions on information forensics and security*, 8(1):239–253.

[15] Bourlai, T., Von Dollen, J., Mavridis, N., and Kolanko, C. (2012). Evaluating the efficiency of a night-time, middle-range infrared sensor for applications in human detection and recognition. In *SPIE Defense, Security, and Sensing*, pages 83551B–83551B. International Society for Optics and Photonics.

[16] Bowers, K. D., Juels, A., and Oprea, A. (2009). Proofs of retrievability: theory and implementation. In *Proceedings of the 2009 ACM workshop on Cloud computing security*, CCSW '09, pages 43–54, New York, NY, USA. ACM.

[17] Brakerski, Z., Gentry, C., and Vaikuntanathan, V. (2011). Fully homomorphic encryption without bootstrapping. Cryptology ePrint Archive, Report 2011/277.

[18] Brakerski, Z. and Vaikuntanathan, V. (2011). Fully homomorphic encryption from ring-lwe and security for key dependent messages. In *Proceedings of the 31st annual conference on Advances in cryptology*, CRYPTO'11, pages 505–524, Berlin, Heidelberg. Springer-Verlag.

[19] Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

[20] Cao, X., Wei, Y., Wen, F., and Sun, J. (2014). Face alignment by explicit shape regression. *IJCV*, 107(2):177–190.

[21] Chang, H., Yeung, D.-Y., and Xiong, Y. (2004). Super-resolution through neighbor embedding. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE.

[22] Chantry, D. (2009). Mapping applications to the cloud. Technical report.

[23] Chase, M. (2007). Multi-authority attribute based encryption. In *Theory of Cryptography, 4th Theory of Cryptography Conference*, volume 4392 of *Lecture Notes in Computer Science*, pages 515–534. Springer.

[24] Chase, M. and Chow, S. S. M. (2009). Improving privacy and security in multi-authority attribute-based encryption. In *Proceedings of the 2009 ACM Conference on Computer and Communications Security*, pages 121–130. ACM.

[25] Chen, J., Yi, D., Yang, J., Zhao, G., Li, S. Z., and Pietikainen, M. (2009). Learning mappings for face synthesis from near infrared to visual light images. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 156–163. IEEE.

[26] Cheung, L. and Newport, C. (2007). Provably secure ciphertext policy abe. In *Proceedings of the 14th ACM conference on Computer and communications security*, CCS '07, pages 456–465, New York, NY, USA. ACM.

[27] Choi, S. E., Lee, Y. J., Lee, S. J., Park, K. R., and Kim, J. (2011). Age estimation using a hierarchical classifier based on global and local facial features. *Pattern Recognition*, 44(6):1262–1281.

[28] Coates, A. and Ng, A. Y. (2012). Learning feature representations with k-means. In *Neural Networks: Tricks of the Trade*, pages 561–580. Springer.

[29] Cootes, T. F., Edwards, G. J., and Taylor, C. J. (1998). Active appearance models. In *ECCV*, pages 484–498. Springer.

[30] Cootes, T. F., Ionita, M. C., Lindner, C., and Sauer, P. (2012). Robust and accurate shape model fitting using random forest regression voting. In *ECCV*, pages 278–291. Springer.

[31] Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59.

[32] Coron, J.-S., Mandal, A., Naccache, D., and Tibouchi, M. (2011). Fully homomorphic encryption over the integers with shorter public keys. In *Proceedings of the 31st annual conference on Advances in cryptology*, CRYPTO'11, pages 487–504, Berlin, Heidelberg. Springer-Verlag.

[33] Cristinacce, D. and Cootes, T. (2008). Automatic feature localisation with constrained local models. *Pattern Recognition*, 41(10):3054–3067.

[34] Curtmola, R., Garay, J., Kamara, S., and Ostrovsky, R. (2011). Searchable symmetric encryption: Improved definitions and efficient constructions. *J. of Computer Security*, 19(5):895–934.

[35] Curtmola, R., Khan, O., and Burns, R. (2008). Robust remote data checking. In *Proceedings of the 4th ACM international workshop on Storage security and survivability*, StorageSS '08, pages 63–68, New York, NY, USA. ACM.

[36] Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *CVPR*, volume 1, pages 886–893. IEEE.

[37] Delerablée, C. (2007). Identity-based broadcast encryption with constant size ciphertexts and private keys. In *Proceedings of the Advances in Crypotology 13th international conference on Theory and application of cryptology and information security*, ASIACRYPT'07, pages 200–215, Berlin, Heidelberg. Springer-Verlag.

[38] Deng, W., Hu, J., Lu, J., and Guo, J. (2014). Transform-invariant pca: A unified approach to fully automatic facealignment, representation, and recognition. *TPAMI*, 36(6):1275–1284.

[39] di Vimercati, S. D. C., Foresti, S., Jajodia, S., Paraboschi, S., and Samarati, P. (2010). Encryption policies for regulating access to outsourced data. *ACM Trans. Database Syst.*, 35(2).

[40] Dodis, Y., Vadhan, S., and Wichs, D. (2009). Proofs of retrievability via hardness amplification. In *Proceedings of the 6th Theory of Cryptography Conference on Theory of Cryptography*, TCC '09, pages 109–127, Berlin, Heidelberg. Springer-Verlag.

[41] Dollar, P., Welinder, P., and Perona, P. (2010). Cascaded pose regression. In *CVPR*, pages 1078–1085.

[42] Dong, C., Loy, C. C., He, K., and Tang, X. (2015). Image super-resolution using deep convolutional networks.

[43] Eidinger, E., Enbar, R., and Hassner, T. (2014). Age and gender estimation of unfiltered faces. *IEEE Transactions on Information Forensics and Security*, 9(12):2170–2179.

[44] Erway, C., Küpçü, A., Papamanthou, C., and Tamassia, R. (2009). Dynamic provable data possession. In *Proceedings of the 16th ACM conference on Computer and communications security*, CCS '09, pages 213–222, New York, NY, USA. ACM.

[45] Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2010). Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 32(9):1627–1645.

[46] Fu, Y., Guo, G., and Huang, T. S. (2010). Age synthesis and estimation via faces: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 32(11):1955–1976.

[47] Gao, F. and Ai, H. (2009). Face age classification on consumer images with gabor feature and fuzzy lda method. In *International Conference on Biometrics*, pages 132–141. Springer.

[48] Geng, C. and Jiang, X. (2013). Fully automatic face recognition framework based on local and global features. *Machine vision and applications*, 24(3):537–549.

[49] Geng, X., Zhou, Z.-H., and Smith-Miles, K. (2007). Automatic age estimation based on facial aging patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 29(12):2234–2240.

[50] Gentry, C. (2009). Fully homomorphic encryption using ideal lattices. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, STOC '09, pages 169–178, New York, NY, USA. ACM.

[51] Gentry, C. and Halevi, S. (2011). Implementing gentry's fully-homomorphic encryption scheme. In *Proceedings of the 30th Annual international conference on Theory and applications of cryptographic techniques: advances in cryptology*, EUROCRYPT'11, pages 129–148, Berlin, Heidelberg. Springer-Verlag.

[52] Golomb, B. A., Lawrence, D. T., and Sejnowski, T. J. (1990). Sexnet: A neural network identifies sex from human faces. In *NIPS*, volume 1, page 2.

[53] Goyal, V., Jain, A., Pandey, O., and Sahai, A. (2008). Bounded ciphertext policy attribute based encryption. In *Automata, Languages and Programming, 35th International Colloquium, 2008*, volume 5126 of *Lecture Notes in Computer Science*, pages 579–591. Springer.

[54] Goyal, V., Pandey, O., Sahai, A., and Waters, B. (2006). Attribute-based encryption for fine-grained access control of encrypted data. In *Proceedings of the 13th ACM conference on Computer and communications security*, CCS '06, pages 89–98, New York, NY, USA. ACM.

[55] Guan, Q., Zhang, Z., and Fu, S. (2011). Proactive failure management by integrated unsupervised and semi-supervised learning for dependable cloud systems. In *Proceedings of the 2011 Sixth International Conference on Availability, Reliability and Security*, ARES '11, pages 83–90, Washington, DC, USA. IEEE Computer Society.

[56] Guo, G., Fu, Y., Dyer, C. R., and Huang, T. S. (2008). Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Transactions on Image Processing*, 17(7):1178–1188.

[57] Guttman, B. and Roback, E. A. (1995). Sp 800-12. an introduction to computer security: the nist handbook. Technical report, Gaithersburg, MD, United States.

[58] Han, H., Klare, B. F., Bonnen, K., and Jain, A. K. (2013a). Matching composite sketches to face photos: A component-based approach. *IEEE Transactions on Information Forensics and Security*, 8(1):191–204.

[59] Han, H., Otto, C., and Jain, A. K. (2013b). Age estimation from face images: Human vs. machine performance. In *2013 International Conference on Biometrics (ICB)*, pages 1–8. IEEE.

[60] Han, S. and Xing, J. (2011). Ensuring data storage security through a novel third party auditor scheme in cloud computing. In *Cloud Computing and Intelligence Systems (CCIS), 2011 IEEE International Conference on*, pages 264 –268.

[61] Hao, Z., Zhong, S., and Yu, N. (2011). A privacy-preserving remote data integrity checking protocol with data dynamics and public verifiability. *Knowledge and Data Engineering, IEEE Transactions on*, 23(9):1432 –1437.

[62] He, K., Sun, J., and Tang, X. (2011). Single image haze removal using dark channel prior. *TPAMI*, 33(12):2341–2353.

[63] Heisele, B., Ho, P., and Poggio, T. (2001). Face recognition with support vector machines: Global versus component-based approach. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 688–694. IEEE.

[64] Hong, C., lv, Z., Zhang, M., and Feng, D. (2011). A secure and efficient role-based access policy towards cryptographic cloud storage. In *Proceedings of the 12th international conference on Web-age information management*, WAIM'11, pages 264–276, Berlin, Heidelberg. Springer-Verlag.

[65] Huang, D., Wang, Y., and Wang, Y. (2007a). A robust method for near infrared face recognition based on extended local binary pattern. In *Advances in Visual Computing*, pages 437–446. Springer.

[66] Huang, G. B., Ramesh, M., Berg, T., and Learned-Miller, E. (2007b). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst.

[67] jin Goh, E., Shacham, H., Modadugu, N., and Boneh, D. (2003). Sirius: Securing remote untrusted storage. In *in Proc. Network and Distributed Systems Security (NDSS) Symposium 2003*, pages 131–145.

[68] Jobson, D. J., Rahman, Z.-U., and Woodell, G. A. (1997a). A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Transactions on Image Processing*, 6(7):965–976.

[69] Jobson, D. J., Rahman, Z.-U., and Woodell, G. A. (1997b). Properties and performance of a center/surround retinex. *Transactions on Image Processing*, 6(3):451–462.

[70] Johnson, R., Molnar, D., Song, D., and Wagner, D. (2002). Homomorphic signature schemes. In Preneel, B., editor, *Topics in Cryptology CT-RSA 2002*, volume 2271 of *Lecture Notes in Computer Science*, pages 204–245. Springer Berlin / Heidelberg.

[71] Jones, M. T. (2010). Anatomy of a cloud storage infrastructure. Technical report, IBM.

[72] Juefei-Xu, F., Pal, D. K., and Savvides, M. (2015). Nir-vis heterogeneous face recognition via cross-spectral joint dictionary learning and reconstruction. *CVPRW, June*.

[73] Juels, A. and Kaliski, Jr., B. S. (2007). Pors: proofs of retrievability for large files. In *Proceedings of the 14th ACM conference on Computer and communications security*, CCS '07, pages 584–597, New York, NY, USA. ACM.

[74] Kallahalla, M., Riedel, E., Swaminathan, R., Wang, Q., and Fu, K. (2003). Plutus: Scalable secure file sharing on untrusted storage. In *Proceedings of the 2nd USENIX Conference on File and Storage Technologies*, pages 29–42, Berkeley, CA, USA. USENIX Association.

[75] Kamara, S. and Lauter, K. (2010). Cryptographic cloud storage. In Sion, R., Curtmola, R., Dietrich, S., Kiayias, A., Miret, J., Sako, K., and Seb?, F., editors, *Financial Cryptography and Data Security*, volume 6054 of *Lecture Notes in Computer Science*, pages 136–149. Springer Berlin / Heidelberg.

[76] Kamara, S. and Papamanthou, C. (2013). Parallel and dynamic searchable symmetric encryption. In *Financial Cryptography*, pages 258–274.

[77] Kamara, S., Papamanthou, C., and Roeder, T. (2012). Dynamic searchable symmetric encryption. In *ACM Conference on Computer and Communications Security*, pages 965–976.

[78] Kang, D., Han, H., Jain, A. K., and Lee, S.-W. (2014). Nighttime face recognition at large standoff: Cross-distance and cross-spectral matching. *Pattern Recognition*, 47(12):3750–3766.

[79] Klare, B. and Jain, A. K. (2010). Heterogeneous face recognition: Matching nir to visible light images. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 1513–1516. IEEE.

[80] Klare, B. F. and Jain, A. K. (2013). Heterogeneous face recognition using kernel prototype similarities. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(6):1410–1422.

[81] Köstinger, M., Wohlhart, P., Roth, P. M., and Bischof, H. (2011). Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 2144–2151. IEEE.

[82] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

[83] Kuhn, D. R., Coyne, E. J., and Weil, T. R. (2010). Adding attributes to role-based access control. *Computer*, 43(6):79–81.

[84] Kuo, C.-C. J. (2016). Understanding convolutional neural networks with a mathematical model. *Journal of Visual Communication and Image Representation*, 41:406–413.

[85] Kuzu, M., Islam, M. S., and Kantarcioglu, M. (2012). Efficient similarity search over encrypted data. In *Proceedings of the 2012 IEEE 28th International Conference on Data Engineering*, ICDE '12, pages 1156–1167, Washington, DC, USA. IEEE Computer Society.

[86] Kwon, Y. H. and da Vitoria Lobo, N. (1999). Age classification from facial images. *Computer Vision and Image Understanding*, 74(1):1–21.

[87] Land, E. H. (1983). Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image. *Proceedings of the National Academy of Sciences of the United States of America*, 80(16):5163.

[88] Land, E. H. and McCann, J. (1971). Lightness and retinex theory. *JOSA*, 61(1):1–11.

[89] Levi, G. and Hassner, T. (2015). Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 34–42.

[90] Lewko, A. B., Okamoto, T., Sahai, A., Takashima, K., and Waters, B. (2010). Fully secure functional encryption: Attribute-based encryption and (hierarchical) inner product encryption. In Gilbert, H., editor, *Advances in Cryptology EUROCRYPT 2010*, volume 6110 of *Lecture Notes in Computer Science*, pages 62–91. Springer Berlin / Heidelberg.

[91] Lewko, A. B. and Waters, B. (2011). Decentralizing attribute-based encryption. In *Advances in Cryptology - EUROCRYPT 2011*, volume 6632 of *Lecture Notes in Computer Science*, pages 568–588. Springer.

[92] Li, S. Z., Chu, S. R., Liao, S., and Zhang, L. (2007). Illumination invariant face recognition using near-infrared images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(4):627–639.

[93] Li, S. Z., Lei, Z., and Ao, M. (2009). The hfb face database for heterogeneous face biometrics research. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 1–8. IEEE.

[94] Liao, S., Yi, D., Lei, Z., Qin, R., and Li, S. Z. (2009). Heterogeneous face recognition from local structures of normalized appearance. In *Advances in Biometrics*, pages 209–218. Springer.

[95] Liao, S., Zhu, X., Lei, Z., Zhang, L., and Li, S. Z. (2007). Learning multi-scale block local binary patterns for face recognition. In *Advances in Biometrics*, pages 828–837. Springer.

[96] Lillibridge, M., Elnikety, S., Birrell, A., Burrows, M., and Isard, M. (2003). A cooperative internet backup scheme. In *Proceedings of the annual conference on USENIX Annual Technical Conference*, ATEC '03, pages 3–3, Berkeley, CA, USA. USENIX Association.

[97] Lim, J., Zitnick, C. L., and Dollár, P. (2013). Sketch tokens: A learned mid-level representation for contour and object detection. In *CVPR*.

[98] Liu, K.-H., Yan, S., and Kuo, C.-C. J. (2015). Age estimation via grouping and decision fusion. *IEEE Transactions on Information Forensics and Security*, 10(11):2408–2423.

[99] Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee.

[100] Lu, R., Lin, X., Liang, X., and Shen, X. S. (2010). *Secure provenance: the essential of bread and butter of data forensics in cloud computing*, pages 282–292. ACM.

[101] Lyubashevsky, V., Peikert, C., and Regev, O. (2013). On ideal lattices and learning with errors over rings. *J. ACM*, 60(6):43:1–43:35.

[102] Maeng, H., Choi, H.-C., Park, U., Lee, S.-W., and Jain, A. K. (2011). Nfrad: Near-infrared face recognition at a distance. In *Biometrics (IJCB)*, pages 1–7. IEEE.

[103] Maeng, H., Liao, S., Kang, D., Lee, S.-W., and Jain, A. K. (2013). Nighttime face recognition at long distance: Cross-distance and cross-spectral matching. In *ACCV*, pages 708–721. Springer.

[104] Makinen, E. and Raisamo, R. (2008). Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):541–547.

[105] Malek, B. and Miri, A. (2009). Combining attribute-based and access systems. In *Computational Science and Engineering, 2009. CSE '09. International Conference on*, volume 3, pages 305 –312.

[106] Mell, P. and Grance, T. (2009). The nist definition of cloud computing. Technical report.

[107] Merkle, R. C. (1980). *Protocols for Public Key Cryptosystems*, pages 122–134. IEEE Computer Society Press.

[108] Milborrow, S., Morkel, J., and Nicolls, F. (2010). The muct landmarked face database. *Pattern Recognition Association of South Africa*.

[109] Moghaddam, B. and Yang, M.-H. (2002). Learning gender with support faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):707–711.

[110] Naor, M. and Rothblum, G. (2006). The complexity of online memory checking. Cryptology ePrint Archive, Report 2006/091.

[111] Narayanan, H. and Gunes, M. (2011). Ensuring access control in cloud provisioned healthcare systems. In *Consumer Communications and Networking Conference (CCNC), 2011 IEEE*, pages 247 –251.

[112] Niu, Z., Zhou, M., Wang, L., Gao, X., and Hua, G. (2016). Ordinal regression with multiple output cnn for age estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4920–4928.

[113] Omri, F., Foufou, S., and Abidi, M. (2014). Nir and visible image fusion for improving face recognition at long distance. In *Image and Signal Processing*, pages 549–557. Springer.

[114] Ostrovsky, R., Sahai, A., and Waters, B. (2007). Attribute-based encryption with non-monotonic access structures. In *Proceedings of the 14th ACM conference on Computer and communications security*, CCS '07, pages 195–203, New York, NY, USA. ACM.

[115] O'toole, A. J., Vetter, T., Troje, N. F., and Bülthoff, H. H. (1997). Sex classification is better with three-dimensional head structure than with image intensity information. *Perception*, 26(1):75–84.

[116] Paillier, P. (1999). Public-key cryptosystems based on composite degree residuosity classes. In Stern, J., editor, *Advances in Cryptology EUROCRYPT 99*, volume 1592 of *Lecture Notes in Computer Science*, pages 223–238. Springer Berlin / Heidelberg.

[117] Pan, Z., Healey, G., Prasad, M., and Tromberg, B. (2003). Face recognition in hyperspectral images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(12):1552–1560.

[118] Papamanthou, C., Tamassia, R., and Triandopoulos, N. (2008). Authenticated hash tables. In *Proceedings of the 15th ACM conference on Computer and communications security*, CCS '08, pages 437–448, New York, NY, USA. ACM.

[119] Parkhi, O. M., Vedaldi, A., and Zisserman, A. (2015). Deep face recognition. In *British Machine Vision Conference*, volume 1, page 6.

[120] Paul, M. and Saxena, A. (2010). Proof of erasability for ensuring comprehensive data deletion in cloud computing. In Meghanathan, N., Boumerdassi, S., Chaki, N., and Nagamalai, D., editors, *Recent Trends in Network Security and Applications*, volume 89 of *Communications in Computer and Information Science*, pages 340–348. Springer Berlin Heidelberg.

[121] Perez, C., Tapia, J., Estévez, P., and Held, C. (2012). Gender classification from face images using mutual information and feature fusion. *International Journal of Optomechatronics*, 6(1):92–119.

[122] Perito, D. and Tsudik, G. (2010). Secure code update for embedded devices via proofs of secure erasure. In *Proceedings of the 15th European conference on Research in computer security*, ESORICS'10, pages 643–662, Berlin, Heidelberg. Springer-Verlag.

[123] Petro, A. B., Sbert, C., and Morel, J.-M. (2014). Multiscale retinex. *Image Processing On Line*, pages 71–88.

[124] Phillips, P. J., Flynn, P. J., Beveridge, J. R., Scruggs, W. T., Otoole, A. J., Bolme, D., Bowyer, K. W., Draper, B. A., Givens, G. H., Lui, Y. M., et al. (2009). Overview

of the multiple biometrics grand challenge. In *Advances in Biometrics*, pages 705–714. Springer.

[125] Phillips, P. J., Wechsler, H., Huang, J., and Rauss, P. J. (1998). The feret database and evaluation procedure for face-recognition algorithms. *Image and vision computing*, 16(5):295–306.

[126] Popa, R. A., Redfield, C. M. S., Zeldovich, N., and Balakrishnan, H. (2012). Cryptdb: Processing queries on an encrypted database. *Commun. ACM*, 55(9):103–111.

[127] Rabin, M. O. (1989). Efficient dispersal of information for security, load balancing, and fault tolerance. *J. ACM*, 36(2):335–348.

[128] Ramanathan, N. and Chellappa, R. (2006). Modeling age progression in young faces. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 387–394. IEEE.

[129] Rara, H., Elhabian, S., Ali, A., Miller, M., Starr, T., and Farag, A. (2009). Face recognition at-a-distance based on sparse-stereo reconstruction. In *CVPR Workshops*, pages 27–32. IEEE.

[130] Reid, D., Samangooei, S., Chen, C., Nixon, M., and Ross, A. (2013). Soft biometrics for surveillance: an overview. *Machine learning: theory and applications. Elsevier*, pages 327–352.

[131] Ricanek, K. and Tesafaye, T. (2006). Morph: A longitudinal image database of normal adult age-progression. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 341–345. IEEE.

[132] Rothblum, R. (2011). Homomorphic encryption: From private-key to public-key. In Ishai, Y., editor, *Theory of Cryptography*, volume 6597 of *Lecture Notes in Computer Science*, pages 219–234. Springer Berlin / Heidelberg.

[133] Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326.

[134] Sahai, A., Seyalioglu, H., and Waters, B. (2012). Dynamic credentials and ciphertext delegation for attribute-based encryption. In *Advances in Cryptology - CRYPTO 2012*, volume 7417 of *Lecture Notes in Computer Science*, pages 199–217. Springer.

[135] Sahai, A. and Waters, B. (2005). Fuzzy identity-based encryption. In *Advances in Cryptology - EUROCRYPT 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005, Proceedings*, volume 3494 of *Lecture Notes in Computer Science*, pages 457–473. Springer.

[136] Samarati, P. and di Vimercati, S. D. C. (2010). Data protection in outsourcing scenarios: issues and directions. In *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, 2010*, pages 1–14. ACM.

[137] Sandhu, R., Coyne, E., Feinstein, H., and Youman, C. (1996). Role-based access control models. *Computer*, 29(2):38 –47.

[138] Saragih, J. M., Lucey, S., and Cohn, J. F. (2011). Deformable model fitting by regularized landmark mean-shift. *IJCV*, 91(2):200–215.

[139] Sebé, F., Domingo-Ferrer, J., Martinez-Balleste, A., Deswarte, Y., and Quisquater, J.-J. (2008). Efficient remote data possession checking in critical information infrastructures. *IEEE Trans. on Knowl. and Data Eng.*, 20:1034–1038.

[140] Shacham, H. and Waters, B. (2008). Compact proofs of retrievability. In *Proceedings of the 14th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology*, ASIACRYPT '08, pages 90–107, Berlin, Heidelberg. Springer-Verlag.

[141] Shacham, H. and Waters, B. (2013). Compact proofs of retrievability. *J. Cryptology*, 26(3):442–83.

[142] Shen, L., He, J., Wu, S., and Zheng, S. (2012). Face recognition from visible and near-infrared images using boosted directional binary code. In *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*, pages 404–411. Springer.

[143] Singh, A., Srivatsa, M., and Liu, L. (2007). Efficient and secure search of enterprise file systems. In *Web Services, 2007. ICWS 2007. IEEE International Conference on*, pages 18 –25.

[144] Singh, A., Srivatsa, M., and Liu, L. (2009). Search-as-a-service: Outsourced search over outsourced storage. *ACM Trans. Web*, 3:13:1–13:33.

[145] Smart, N. and Vercauteren, F. (2010). Fully homomorphic encryption with relatively small key and ciphertext sizes. In Nguyen, P. and Pointcheval, D., editors, *Public Key Cryptography PKC 2010*, volume 6056 of *Lecture Notes in Computer Science*, pages 420–443. Springer Berlin / Heidelberg.

[146] Song, D. X., Wagner, D., and Perrig, A. (2000). Practical techniques for searches on encrypted data. In *Security and Privacy, 2000. S P 2000. Proceedings. 2000 IEEE Symposium on*, pages 44 –55.

[147] Sravan Kumar, R. and Saxena, A. (2011). Data integrity proofs in cloud storage. In *Communication Systems and Networks (COMSNETS), 2011 Third International Conference on*, pages 1 –4.

[148] Stehle, D. and Steinfeld, R. (2010). Faster fully homomorphic encryption. In Abe, M., editor, *Advances in Cryptology - ASIACRYPT 2010*, volume 6477 of *Lecture Notes in Computer Science*, pages 377–394. Springer Berlin / Heidelberg.

[149] Storer, M. W., Greenan, K., Long, D. D., and Miller, E. L. (2008). Secure data deduplication. In *Proceedings of the 4th ACM international workshop on Storage security and survivability*, StorageSS '08, pages 1–10, New York, NY, USA. ACM.

[150] Sun, Y., Wang, X., and Tang, X. (2013). Deep convolutional network cascade for facial point detection. In *CVPR*, pages 3476–3483.

[151] Sun, Y., Wang, X., and Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1891–1898.

[152] Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *CVPR*, pages 1701–1708. IEEE.

[153] Tome, P., Fierrez, J., Alonso-Fernandez, F., and Ortega-Garcia, J. (2010). Scenario-based score fusion for face recognition at a distance. In *CVPR Workshops*, pages 67–73. IEEE.

[154] Ullah, I., Hussain, M., Muhammad, G., Aboalsamh, H., Bebis, G., and Mirza, A. M. (2012). Gender recognition from face images with local wld descriptor. In *2012 19th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 417–420. IEEE.

[155] Van Dijk, M., Gentry, C., Halevi, S., and Vaikuntanathan, V. (2010). Fully homomorphic encryption over the integers. In Gilbert, H., editor, *Advances in Cryptology EUROCRYPT 2010*, volume 6110 of *Lecture Notes in Computer Science*, pages 24–43. Springer Berlin / Heidelberg.

[156] Van Dijk, M. and Juels, A. (2010). On the impossibility of cryptography alone for privacy-preserving cloud computing. In *Proceedings of the 5th USENIX conference on Hot topics in security*, HotSec'10, pages 1–8, Berkeley, CA, USA. USENIX Association.

[157] Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *IJCV*, 57(2):137–154.

[158] Wagner, A., Wright, J., Ganesh, A., Zhou, Z., Mobahi, H., and Ma, Y. (2012). Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *TPAMI*, 34(2):372–386.

[159] Wang, C., Cao, N., Li, J., Ren, K., and Lou, W. (2010a). Secure ranked keyword search over encrypted cloud data. In *Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference on*, pages 253 –262.

[160] Wang, C., Cao, N., Ren, K., and Lou, W. (2012a). Enabling secure and efficient ranked keyword search over outsourced cloud data. *IEEE Transactions on Parallel and Distributed Systems*, 23(8):1467–1479.

[161] Wang, C., Chow, S. S., Wang, Q., Ren, K., and Lou, W. (2013). Privacy-preserving public auditing for secure cloud storage. *IEEE Transactions on Computers*, 62(2):362–375.

[162] Wang, C., Ren, K., Lou, W., and Li, J. (2010b). Toward publicly auditable secure cloud data storage services. *Network, IEEE*, 24(4):19 –24.

[163] Wang, C., Wang, Q., Ren, K., Cao, N., and Lou, W. (2012b). Toward secure and dependable storage services in cloud computing. *IEEE Trans. Serv. Comput.*, 5(2):220–232.

[164] Wang, C., Wang, Q., Ren, K., and Lou, W. (2009a). Ensuring data storage security in cloud computing. In *Quality of Service, 2009. IWQoS. 17th International Workshop on*, pages 1 –9.

[165] Wang, C., Wang, Q., Ren, K., and Lou, W. (2010c). Privacy-preserving public auditing for data storage security in cloud computing. In *Proceedings of the 29th conference on Information communications*, INFOCOM'10, pages 525–533, Piscataway, NJ, USA. IEEE Press.

[166] Wang, Q., Wang, C., Li, J., Ren, K., and Lou, W. (2009b). Enabling public verifiability and data dynamics for storage security in cloud computing. In *Proceedings of the 14th European conference on Research in computer security*, ESORICS'09, pages 355–370, Berlin, Heidelberg. Springer-Verlag.

[167] Wang, Q., Wang, C., Ren, K., Lou, W., and Li, J. (2011). Enabling public auditability and data dynamics for storage security in cloud computing. *Parallel and Distributed Systems, IEEE Transactions on*, 22(5):847 –859.

[168] Wang, X., Guo, R., and Kambhamettu, C. (2015). Deeply-learned feature for age estimation. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 534–541. IEEE.

[169] Waters, B. (2011). Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization. In *Public Key Cryptography PKC 2010*, volume 6571 of *Lecture Notes in Computer Science*, pages 53–70. Springer Berlin / Heidelberg.

[170] won Song, C., Park, S., wook Kim, D., and Kang, S. (2011). Parity cloud service: A privacy-protected personal data recovery service. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on*, pages 812 –817.

[171] Wong, W. K., Cheung, D. W.-l., Kao, B., and Mamoulis, N. (2009). Secure knn computation on encrypted databases. In *Proceedings of the 35th SIGMOD international conference on Management of data*, SIGMOD '09, pages 139–152, New York, NY, USA. ACM.

[172] X.800, C. R. (1991). Security architecture for open systems interconnection for ccitt applications. Technical report.

[173] Xiong, X. and De la Torre, F. (2013). Supervised descent method and its applications to face alignment. In *CVPR*, pages 532–539.

[174] Yan, J., Lei, Z., Yi, D., and Li, S. (2013). Learn to combine multiple hypotheses for accurate face alignment. In *ICCV Workshops*, pages 392–396.

[175] Yan, S., Wang, H., Tang, X., and Huang, T. S. (2007). Learning auto-structured regressor from uncertain nonnegative labels. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE.

[176] Yao, Y., Abidi, B. R., Kalka, N. D., Schmid, N. A., and Abidi, M. A. (2008). Improving long range and high magnification face recognition: Database acquisition, evaluation, and enhancement. *Computer Vision and Image Understanding*, 111(2):111–125.

[177] Yi, D., Lei, Z., and Li, S. Z. (2014a). Age estimation by multi-scale convolutional network. In *Asian Conference on Computer Vision*, pages 144–158. Springer.

[178] Yi, D., Lei, Z., Liao, S., and Li, S. Z. (2014b). Shared representation learning for heterogeneous face recognition. *arXiv preprint arXiv:1406.1247*.

[179] Yi, D., Liu, R., Chu, R., Lei, Z., and Li, S. Z. (2007). Face matching between near infrared and visible light images. In *Advances in Biometrics*, pages 523–530. Springer.

[180] Yu, S., Wang, C., Ren, K., and Lou, W. (2010). Achieving secure, scalable, and fine-grained data access control in cloud computing. In *INFOCOM, 2010 Proceedings IEEE*, pages 1 –9.

[181] Yun, A., Shi, C., and Kim, Y. (2009). On protecting integrity and confidentiality of cryptographic file system for outsourced storage. In *Proceedings of the 2009 ACM workshop on Cloud computing security*, CCSW '09, pages 67–76, New York, NY, USA. ACM.

[182] Zeng, W., Zhao, Y., Ou, K., and Song, W. (2009). Research on cloud storage architecture and key technologies. In *Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, ICIS '09, pages 1044–1048, New York, NY, USA. ACM.

[183] Zhang, B., Zhang, L., Zhang, D., and Shen, L. (2010a). Directional binary code with application to polyu near-infrared face database. *Pattern Recognition Letters*, 31(14):2337–2344.

[184] Zhang, F., Li, Q., and Xiong, H. (2012). Efficient revocable key-policy attribute based encryption with full security. In *Eighth International Conference on Computational Intelligence and Security 2012*, pages 477–481. IEEE.

[185] Zhang, M., Cai, K., and Feng, D. (2010b). Fine-grained cloud db damage examination based on bloom filters. In *Proceedings of the 11th international conference on Web-age information management*, WAIM'10, pages 157–168, Berlin, Heidelberg. Springer-Verlag.

[186] Zhang, Z., Wang, Y., and Zhang, Z. (2011). Face synthesis from near-infrared to visual light via sparse representation. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–6. IEEE.

[187] Zhao, F., Nishide, T., and Sakurai, K. (2011). Realizing fine-grained and flexible access control to outsourced data with attribute-based cryptosystems. In *Proceedings of the 7th international conference on Information security practice and experience*, ISPEC'11, pages 83–97, Berlin, Heidelberg. Springer-Verlag.

[188] Zhao, S. and Grigat, R.-R. (2005). An automatic face recognition system in the near infrared spectrum. In *Machine Learning and Data Mining in Pattern Recognition*, pages 437–444. Springer.

[189] Zheng, Q. and Xu, S. (2011). Fair and dynamic proofs of retrievability. In *Proceedings of the first ACM conference on Data and application security and privacy*, CODASPY '11, pages 237–248, New York, NY, USA. ACM.

[190] Zhou, L., Varadharajan, V., and Hitchens, M. (2011). Enforcing role-based access control for secure data storage in the cloud. *Comput. J.*, 54(10):1675–1687.

[191] Zhu, Y., Ahn, G.-J., Hu, H., and Wang, H. (2010). Cryptographic role-based security mechanisms based on role-key hierarchy. In *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, 2010*, pages 314–319. ACM.

[192] Zhu, Y., Hu, H., Ahn, G.-J., Wang, H., and Wang, S.-B. (2011a). Provably secure role-based encryption with revocation mechanism. *J. Comput. Sci. Technol.*, 26(4):697–710.

[193] Zhu, Y., Wang, H., Hu, Z., Ahn, G.-J., Hu, H., and Yau, S. S. (2011b). Dynamic audit services for integrity verification of outsourced storages in clouds. In *Proceedings of the 2011 ACM Symposium on Applied Computing*, SAC '11, pages 1550–1557, New York, NY, USA. ACM.

[194] Zou, X., Kittler, J., and Messer, K. (2007). Illumination invariant face recognition: A survey. In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pages 1–8. IEEE.