THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學
Pao Yue-kong Library
包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

Pao Yue-kong Library, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

http://www.lib.polyu.edu.hk

The Hong Kong Polytechnic University

Department of Computing

Use of Intelligent System Techniques for Storage and Retrieval of Biometrics Data with Application to Personal Identification

CHEUNG King Hong

A thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

August 2005

Pao Yue-kong Library
PolyU · Hong Kong

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

(Signed)

CHEUNG King Hong (Name of Student)

Abstract

Biometrics (or biometric recognition) refers to the technology recognizing individuals based on their physiological and/or behavioral characteristics. It has advantages over token-based and knowledge-based personal recognition technologies. The important biometric systems including face, fingerprint, iris and palmprint are in fact image-based recognition systems. For identification purpose, these systems can be regarded as Content-Based Image Retrieval (CBIR) systems.

The storage and retrieval requirements of image-based biometric systems, nevertheless, make them to be a special type of CBIR systems. First, personal physiological characteristics cannot be properly described by traditional text-based methods. Second, there is no generally accepted indexing and classification methods being used in common biometrics systems. Third, the captured biometric images always contain noise and variations due to change in capture environments and user habit. Fourth, in real world applications, they should be scalable for large databases and should provide relevant security mechanisms. Finally, they may make its own decision or provide help to make decision on whether a claim is accepted or rejected.

In this research, we focus on the storage and retrieval of some image-based biometric systems supported by personal physiological characteristics. We have considered the storage and retrieval of biometric templates as an application of target and category search in narrow domain. We have incorporated some primitive visual-based image features including texture, lines and points, as content descriptors of biometric images. Various intelligent system techniques, e.g. machine learning and fuzzy, have been used to extract and represent the visual-based image features of palmprint and face images. We have developed methods that can compactly

represent and effectively retrieve palmprint and face templates and we are the first to consider retrieval in large palmprint databases. We originally identify and study three design issues of cancelable biometrics, which is a security and privacy enhancement method proposed for template protection. We are among the first to consider the three issues integrally when designing and evaluating cancelable biometrics.

Publications arising from the thesis

Referred Journal Article

1. Kong, A.; Cheung, K.-H.; Zhang, D.; Kamel, M.; You, J., 2005. An Analysis of BioHashing and Its Variants. Pattern Recognition **39**(7), pp. 1359-1368.
2. You, J.; Kong, W.-K.; Zhang, D.; Cheung, K.-H., 2004. On hierarchical palmprint coding with multi-features for personal identification in large databases. IEEE Transactions on Circuit Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics **14**(2), pp. 234-243.

Working Journal Article

3. Cheung, K.-H.; Kong A.; Zhang, D.; Kamel, M.; You J., 2005. On Cancelable Biometrics for Personal Authentication. To be submitted.
4. Cheung, K.-H.; Kong A.; You J.; Zhang, D.; Kamel, M., 2006. Evaluating Eigenpalm from an Application Perspective. Submitted to Pattern Recognition.

Book Chapter

5. Cheung, K.-H.; Kong, A.; Zhang, D.; Kamel, M.; You, J., 2006. Revealing the secret of FaceHashing. Lecture Notes in Computer Science **3832**, Springer-Verlag GmbH, pp.106-112.
6. You, J.; Li, Q.; Cheung, K.H.; Bhattacharya, P., 2005. An Integration of Biometrics and Mobile Computing for Personal Identification. Lecture Notes in Computer Science **3687**, Springer-Verlag GmbH, pp. 226-235.
7. Cheung, K.-H.; Kong, A.; Zhang, D.; Kamel, M.; You, J.; Lam, H.-W.T., 2005. An Analysis on Accuracy of Cancelable Biometrics based on BioHashing.

Lecture Notes in Computer Science **3683**, Springer-Verlag GmbH, pp.1168-1172.

8. Cheung, K.-H.; You, J.; Liu, J.; Ao Ieong, W.H.T., 2004. Appearance-based Face Recognition using Aggregated 2D Gabor Features. Lecture Notes in Computer Science **3214**, Springer-Verlag GmbH, pp. 572-579.
9. You, J.; Kong, W.K.; Zhang, D.; Cheung, K.-H., 2004. A new approach to personal identification in large database by hierarchical palmprint coding with multi-features. Lecture Notes in Computer Science **3072**, Springer-Verlag GmbH, pp. 739-745.

Referred Conference Article

10. Cheung, K.H.; You, J.; Li, Q.; Bhattacharya, P., 2005. A new approach to appearance-based face recognition. Proc. of the IEEE International Conference on Systems, Man and Cybernetics (IEEE SMC'2005) **2**, pp. 1686-1691.
11. Cheung, K.H.; You, J.; Li, Q.; Bhattacharya, P., 2005. On Aggregated 2D Gabor Features for Appearance-based Face Recognition. Proc. of the 2005 International Conference on Imaging Science, Systems, and Technology: Computer Graphics (CISST'2005), pp. 33-39.
12. Cheung, K.-H.; Kong, A.; You, J.; Zhang, D., 2005. An Analysis on Invertibility of Cancelable Biometrics based on BioHashing. Proc. of the 2005 International Conference on Imaging Science, Systems, and Technology: Computer Graphics (CISST'2005), pp. 40-45.
13. Cheung, K.H.; You, J.; Li, Q.; Bhattacharya, P., 2005. Appearance-based face recognition using aggregated 2D Gabor features. Proc. of the Workshop on Pattern Recognition in Information Systems (PRIS'2005), pp. 81-93.

14. Cheung, K.H.; You, J.; Kong, W.K.; Zhang, D., 2004. A Study of Aggregated 2D Gabor Features on Appearance-based Face Recognition. Proc. of the Third International Conference on Image and Graphics (ICIG 2004), pp. 310-313.
15. Cheung, K.H.; Kong, W.K.; You, J.; Zhang, D., 2003. An integration of principal component analysis and self-organizing map for effective palmprint retrieval. Proc. of the 16th International Conference on Computer Applications in Industry and Engineering (CAINE-2003), pp. 101-104.
16. You, J.; Kong, W.K.; Zhang, D.; Cheung, K.H., 2003. On hierarchical palmprint coding with multi-features for personal identification in large databases. Proc. of the 16th International Conference on Computer Applications in Industry and Engineering (CAINE-2003), pp. 105-107.
17. Cheung, K.H.; Kong, W.K.; You, J.; Zhang, D., 2003. On Effective Palmprint Retrieval for Personal Identification. Proc. of 2003 International Conference on Imaging Science, Systems, and Technology (CISST'2003), pp. 111-117.
18. You, J.; Cheung, K.H.; Liu, J., 2003. A hierarchical approach to content-based image retrieval. Proc. of 2003 International Conference on Imaging Science, Systems, and Technology (CISST'2003), pp. 127-133.
19. Cheung, K.H.; Dillon, T.; You, J.; Liu, J., 2001. ColorGuide: A wavelet-based data warehousing approach to hierarchical image retrieval. Proc. of International Conference on Image Science, Systems and Technology (CISST'2001), pp. 386-392.

Acknowledgements

I would like to express my sincere gratitude to Professor Tharam S. DILLON, my initial chief supervisor and currently my co-supervisor, for taking me into the world of research, and to Dr. Jane Jia YOU, my current chief supervisor, for her supervision over the period of my study. I would like to express my thanks to Dr. Allan Kang-Ying WONG, my co-supervisor, for many fruitful discussions with and advices from him. I would like to thank Professor David Da-Peng ZHANG, the founder and director of the Biometrics Research Centre, for granting me access to their databases and providing me opportunities of exposure.

I would like to express my greatest gratitude to Dr. Korris Fu-Lai CHUNG, the Chairman of the Board of Examiners, to Professor Cheng-Qi. ZHANG, the External Examiner, and to Professor Pong-Chi YUEN, the External Examiner, for their involvement in my PhD examination.

I would like to show my utmost gratefulness to Mr. Adams Wai-Kin KONG, my research partner, for bringing me up and growing with me in the world of research. I would like to show my thankfulness to Dr. John Pui-fai SUM for his advices, suggestions and many precious discussions between us.

I would like to say thank you to my former roommates Mr. Vincent Chi-hung LI, Ms. Connie POON, Ms. Tiffany Ya TANG, Mr. Ray Ying-Wai LAM, Mr. Cody Ka-pang WONG, Mr. Jacky Koon-hang SHIU and my current roommates Mr. Wilfred Wan-Kei LIN, Mr. Richard Sui-Lun WU, Mr. Siu-Nam CHAUNG and Mr. Tak-Chung FU for accompanying me through the period of my study.

I would like to thank the personnel of Research Office and Department of Computing for providing me various kinds of help throughout the period of my study.

I would also like to thank all my friends in Department of Computing for providing me everything I need during my study.

Last but not least, I would like to express my greatest thankfulness to my family. Had I had eternal support from my family, I would not be able to finish my studies.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

> *Biometrics offers greater security and convenience than traditional methods of personal recognition. In some applications, biometrics can replace or supplement the existing technology. In others, it is the only viable approach.*
>
> (Prabhakar et al. 2003)

Biometrics (or biometric recognition) refers to the technology recognizing individuals based on their physiological and/or behavioral characteristics. It has advantages over token-based and knowledge-based personal recognition technologies.

Biometric data is fundamentally one form of multimedia data. The nature and volume of biometric data (and information) are beyond the reach of traditional data management approaches. Traditional data management approaches (databases) are designed for *structured* data, i.e. precise data type (Narasimhalu et al. 1997): numerical or alphanumerical attributes. Precise query result (*exact match*) is expected, in addition. Owing to the nature of biometric data, which is *unstructured* and *multidimensional* and, the demands of *non-exact match* queries, traditional database technology is unable to keep up with these ever demanding requirements. Thus, there is a surge of effective and yet efficient methods for the management of ever growing biometric data. Storage and retrieval is the core of biometric data management and is a multidisciplinary subject including but not limited to pattern recognition, computer vision, signal processing, machine learning, database management and information retrieval.

Content-based approach is a promising way to multimedia data storage and retrieval. For content-based retrieval, there are two types of methods proposed for

this subject, text-based and audiovisual-based which complement each other. Text-based methods support the retrieval, as in Information Retrieval, using text attributes that are high-level (logical) content descriptors. Multimedia data stored for retrieval are usually annotated manually. Audiovisual-based methods support the use of high- and middle- level (logical) as well as low-level (primitive) content descriptors for retrieval. Content descriptors are automatically or semi-automatically extracted, from multimedia data, for retrieval. Image (Smeulders et al. 2000; Chen 1998) and video processing (Chen 1998) together with audio (Lu 2001) and speech processing (Juang and Chen 1998; Rudnicky et al. 1997) provides cues to automatic extraction of information/features in multimedia data. For a large volume of multimedia data, it is inefficient to manually annotate all data items. Although manual annotation is usually done by professions, it is subjective yet inconsistent and is hardly a complete set of descriptions that answer a wide range of queries. Audiovisual based methods, on the contrary, are comparatively consistent and usually designed to be automated. Thus, they are the more viable way to handle large volume of multimedia data automatically and semi-automatically. (Gorsky and Mehrotra 1989; Gudivada and Raghavan 1995; Gupta and Jain 1997; Chang et al. 1997; Smeulders et al. 2000) Much research has been conducted on various forms of indexing trees in attempting to support the storage and retrieval of multimedia data. (Chávez et al. 2001; Böhm et al. 2001) As an alterative approach, intelligent system techniques operating on one or multiple features have been studied.

The important biometric systems including face, fingerprint, iris and palmprint are in fact image-based recognition systems. These systems for identification purpose can be regarded as Content-Based Image Retrieval (CBIR) systems.

The storage and retrieval requirements of image-based biometric systems, nonetheless, are substantially different from traditional CBIR systems. First, they are visual-based CBIR systems because personal physiological characteristics are inappropriate to be described using text-based methods. Second, they have to manage noise and variations of input images because an input image of a person, which is presented at a different time and perhaps in a different environment, can be considerably different from the analogous templates stored in the database of image-based biometric systems. Third, in real world application, they should be scalable for large databases and should provide relevant security mechanisms. Finally, they may make its own decision or provide help to make decision on whether a claim is accepted or rejected.

In the course of study, we focus in the storage and retrieval of some image-based biometric systems that support personal physiological characteristics. We have considered the storage and retrieval of biometric image templates as an application of target and category search in narrow image domain retrieval. We have incorporated some primitive visual-based image features, texture, lines and points, as content descriptors of biometric images. Various intelligent system techniques, e.g. machine learning and fuzzy, have been used to extract and represent the visual-based image features of palmprint and face images. We have developed methods that can compactly represent and effectively retrieve palmprint and face image templates and we are the first to consider retrieval in large palmprint databases. We originally identify and study three design issues of cancelable biometrics, which is a security and privacy enhancement method proposed for the protection of user templates in a biometric system, and their relations. We are among the first to propose the

consideration of the three issues integrally when designing and evaluating cancelable biometrics.

The rest of this thesis is organized as follows. Chapter 2 provides the background and a review of the literature of image retrieval, especially the content-based approach and its relating issues. Chapter 3 introduces some tools that will be come across in the course of study. Chapter 4 and 5 give studies on biometrics image databases, palmprint and face correspondingly. Chapter 6 discusses some design issues regarding the storage security evolved from biometrics image databases. Chapter 7 presents an analysis of BioHashing and its variants that raises the issue of using “safe” tokens and it supplements the discussion in Chapter 6. Finally, Chapter 8 gives our conclusion, contributions and some future research directions.

Chapter 2 Background Research

2.1 Introduction to Biometrics

In every day's life, there are questions about someone's identity. These questions become more acute in today's networked society as any operations and processes have gone online, operate remotely and performed in an unattended environment. Traditional recognition approaches: *token*-based and *knowledge*-based cannot meet the ever demanding requirements of the applications that require personal authentication. Biometrics is an alternative that can replace or supplement the traditional recognition approaches. (Jain et al. 2004; Prabhakar et al. 2003)

To qualify as a measurement for personal recognition, a biometric should posses the following characteristics. (Jain et al. 2004; Prabhakar et al. 2003)

Universality: each person should have the characteristic;

Distinctiveness: any two persons should be sufficiently different in terms of the characteristic;

Permanence: the characteristic should be sufficiently invariant (with respect to the matching criterion) over a period of time;

Collectability: the characteristic can be measured quantitatively.

Biometric systems operate in one of the two modes: *verification* or *identification*. In verification mode, the system validates a person's identity by comparing the acquired biometric data against that person's biometric data stored in the database of the biometric system, so-called 1-to-1 matching. In identification mode, the system searches the entire database of the biometric system for a match, so-called 1-to-many matching. A practical system should consider, in a whole, the following issues. (Jain et al. 2004; Prabhakar et al. 2003)

Performance: the achievable recognition accuracy and speed, the resources required to achieve the desired recognition accuracy and speed, as well as the operational and environmental factors that affect the accuracy and speed;

Acceptability: the extent to which people are willing to accept the use of a particular biometric identifier (characteristic) in their daily lives;

Circumvention: how easily the system can be fooled using fraudulent methods.

Every biometric system commits two types of errors. (Jain et al. 2004; Prabhakar et al. 2003)

False Match/Accept Error (Type I): Mistaking biometric measurements from two different persons to be from the same person;

False Non-Match/Reject Error (Type II): Mistaking two biometric measurements from the same person to be from two different persons.

An operational biometric system regulates the trade-offs of these two types of recognition errors using a threshold. (see Chapter 3.3 also and Figure 3.5 for an illustration) Besides the two recognition error, there are two system errors, namely *Failure To Capture* (FTC) and *Failure To Enroll* (FTE). FTC occurs when an automatic biometric data capture device fails to capture a valid sample from a user while FTE occurs when a user fails to enroll a biometric system.

Some popular biometrics include DeoxyriboNucleic Acid (DNA), face, fingerprint, hand geometry, iris, palmprint and voice. A detailed introduction is given by Jain et al. (2004) and a discussion on security and privacy concerns is given by Prabhakar et al. (2003).

2.2 Content-Based Image Retrieval (CBIR)

> *Images are being generated at an ever-increasing rate by sources such as defense and civilian satellites, military reconnaissance and surveillance flights, fingerprinting and mug shot capturing devices, scientific experiments, biomedical imaging, and home entertainment systems.* (Gudivada and Raghavan 1995)

Accompanying the growth of the World Wide Web (WWW) and multimedia information, boost in hardware, software and communication technology, and many activities, e.g. advertising and promotion, have gone online, there is also a demand for online accessible image retrieval or database systems.

An effective and yet efficient image retrieval or database system is needed to manage this huge amount of image data is needed. There are five types of image data (Gorsky and Mehrotra 1989): Iconic data, Image-related data, Information extracted from images, Image-world relationships and World-related data. Iconic data is raw data in an encoding format used for storage. Image-related data are generally content independent data and are also known as metadata. The remaining three types of data are content dependent and in ascending order with respect to level of abstraction. Using content dependent data to index or search for a similar image in an image database is generally referred to Content-based Image Retrieval.

Image retrieval has since the 1970s been a focus of interest in two research communities, computer vision and database management. (Gorsky and Mehrotra 1989; Gudivada and Raghavan 1995; Gupta and Jain 1997) They look at image retrieval, nonetheless, from two different angles. The computer vision community has taken a visual-based approach while the database management community has used a text-based approach. The visual-based approach identifies visual

characteristics through series of computations and searches images based on these characteristics. The text-based approach first annotates images manually and then searches images using text-based attributes in database management systems (Rui et al. 1999). Following the success of Information Retrieval (IR) and Natural Language Processing (NLP) techniques, they are introduced and incorporated in order to build better image retrieval systems. It is worth to point out that visual- and text-based approaches are complementing each other instead of opposing. It is expected, moreover, a better integration of the two approaches to bridge the semantic gap (Smeulders et al. 2000).

Content-based Image Retrieval (CBIR), in general, refers the use of content dependent data to index or search for a similar image in an image database (Gorsky and Mehrotra 1989). CBIR, which incorporates and text-based visual-based features, consists of three processes: image feature extraction and representation, similarity measure and, access or retrieval. Some well known CBIR systems are QBIC (Flickner et al. 1995), Photobook (Pentland et al. 1996), PicHunter (Cox et al. 2000), and for more, please refer to Rui et al. (1999).

Image feature extraction and representation

> *In developing CBIR systems, there is an inherent tradeoff between the degree of automation desired for feature extraction and the level of domain independence realized in the system.* (Gudivada and Raghavan 1995)

An ambitious approach, dynamic feature extraction, to build CBIR systems is to develop an automatic and dynamic feature extraction mechanism that computes required features (determined a priori or not) dynamically and synthesizes those computed features to logical ones under guidance of domain expert(s). This favours

applications deal with comparatively small image collections and retrieval of images is performed by domain experts.

CBIR systems that are not dynamic in feature extraction, i.e. a priori feature extraction, provide reasonable degree of domain independence. Predefined set of low-level features are extracted from an image and logical features are determined at the time the image is inserted into the database.

Similarity Measures

Since the revolution in retrieval bought about by IR, researchers begin to think of using best match instead of exact match to fulfill the search for something similar and provide some relevant entries in case no indexed entries match exactly. Use of similarity measures distinguishes multimedia retrieval, or one of its specifications: image retrieval, from traditional database retrieval.

A ranked-order list of entries is returned based on the similarity measured between the query and entries in the databases according to a measurement, e.g. a metric. Various metrics may be used to define a similarity measurement of various image features.

Access or retrieval

Access or retrieval of images in a CBIR system is generally conducted through specifying a query. A query can be specified by values, such as text (e.g. keyword), numerical values for features, or by examples, i.e. providing an image in a query. Query by examples differs from Query by values that, instead of providing values of image features directly, they are extracted from the query image by the feature extraction mechanism of that system.

Relevant entries, based on user preferences and similarity measure results, are returned and formatted to display. Some relevance feedback mechanisms can be

introduced to improve the effectiveness of the image retrieval systems. (Gudivada and Raghavan 1995; Rui et al. 1997a; Rui et al. 1997b)

2.2.1 User aims in CBIR

User aims specified are boarder than the presence or absence of visual characteristics. Three broad categories of user aims (Cox et al. 2000) are identified and summarized by (Smeulders et al. 2000). They are briefly described and a comprehensive comparison is shown in Table 2.1.

Target search. User intends to search for a specific image, such as searching a catalogue, finding other images containing objects desired within the image(s) at hand. Systems providing this search method are suitable for catalogue searching of art and industrial items.

Category search. User intends to retrieve an arbitrary representative of a particular grouping. Example(s) of a certain group is/are provided for searching other members of the same group. Systems providing this search method usually have domain dependent similarity measure.

Association search. User does not have any intention at first but finds something interesting. Iteratively refining the search intention is expected.

From viewpoints other than the functional aspects of CBIR, user aims may cover a wider scope (Smeulders et al. 2000). It is worth to point out that the query interface, e.g. linguistic (text) vs. visual (Enser 1995) can constraint the specification of user aims and thus their category.

Table 2.1 The three user aims in image retrieval: Target-, Category- and Association-search.

	Target	**Category**	**Association**
Object code	1 specific object	An arbitrary object	Not defined at start
Query by example	1…N objects	1…N objects with class labels	N objects plus association
Similarity	Feature-based	Class driven	Session-specific
Events in F-space	Proximity to query	Class membership	Clusters
Feedback	Rank ordered on proximity	Likelihood on class membership	Relevance feedback on association value
Interactive update of **Images of query** **Features of query** **Similarity measure**	— Refine on the way —	Expand query Refine on the way Adapt to group	Refine on the way Alter on the way Reshape to goal
Builds on	Pattern matching in computer vision Object recognition	Object recognition Statistical pattern recognition	Feature sets and similarity function of computer vision
Nature	—	—	Iterative Interactive Explicative
Challenges	Huge amount of objects to search among Incomplete query specification Incomplete image description Variability of sensing conditions and object states	Interactive manipulation of results Usually very large number of object classes Absence of an explicit training phase for feature and classifier tuning	Semantic gap Understandable display and relevance feedback
Examples	Search catalogue, e.g. Cox et al. 2000	Search catalogue of variety, e.g. Huet and Hancock 1999 Jain and Vailaya 1998	Kato et al. 1992; With relevance feedback, e.g. Hiroike et al. 1999 Frederix et al. 2000

2.2.2 The image domain in CBIR

The image domain of image retrieval systems can be thought of as a one-dimensional space with the narrow domain at one end while the board domain at the other end. The narrow domain and board domain are defined as follows. (Smeulders et al. 1998)

A narrow domain has a limited and predictable variability in all relevant aspects of its appearance.

A board domain has an unlimited and unpredictable variability in its appearance even for the same semantic meaning.

In narrow domain, content of images are of limited variability and the recording circumstances are usually the same within the same domain. The semantic descriptions about images are generally well defined and unique. Criminal photograph database is an example of this domain.

In board domain, images can be interpreted in various contexts, usually not in full. A generic photograph archive is one typical example of this domain; the World Wide Web is perhaps the broadest one available.

Table 2.2 Narrow versus Broad domain in image retrieval

	Narrow	**Broad**
Variance of content	Low	High
Sources of knowledge	Specific	Generic
Semantics	Homogeneous	Heterogeneous
Ground truth	Likely	Unlikely
Content description	Objective	Subjective
Scene and sensor	Possibly controlled	Unknown
Aimed application	Specific	Generic
Type of application	Professional	Public
Tools	Model-driven Specific invariants	Perceptual Cultural General invariants
Interactivity	Limited	Pervasive Iterative
Evaluation	Quantitative	Qualitative
System architecture	Tailored database-driven	Modular interaction-driven
Size	Medium	Large to very large
A source of inspiration	Object recognition	Information retrieval

The semantic gap (see Chapter 2.2.3 for the definition) between features and semantic interpretations is small for the narrow domain but large for the broad domain. Characteristics of the narrow and the broad domain in CBIR are listed in Table 2.2.

2.2.3 Semantic gap and sensory gap in CBIR

The semantic gap and sensory gap is defined as follows (Smeulders et al. 2000).

> *The semantic gap is the lack of coincidence between the information that one can extract form the visual data and the interpretation that the same data have for a user in a given situation.*
>
> *The sensory gap is the gap between the object in the world and the information in a (computational) description derived from a recording of that scene.*

The pivotal point in content-based retrieval is that the user seeks semantic similarity, but the database can only provide similarity by data processing. Associating a complete semantic system with image data can solve the general object recognition problem from a single image but it does not succeed till now. Research is, therefore, concentrated in linking semantics to sets of data values. (Smeulders et al. 2000)

The sensory gap poses problems for describing the content of images, especially when recording conditions are not known. Referencing a mixture of interpretations of an image may help in eliminating disambiguation. (Smeulders et al. 2000)

2.3 Image Processing Fundamentals

CBIR is realized through the use of image processing techniques. We are going to look at some of the concepts and techniques in the following subsections.

2.3.1 A general process flow of image processing

Gonzalez and Woods (Gonzalez and Woods 1992) have outlined the fundamental steps in digital image processing (Figure 2.1).

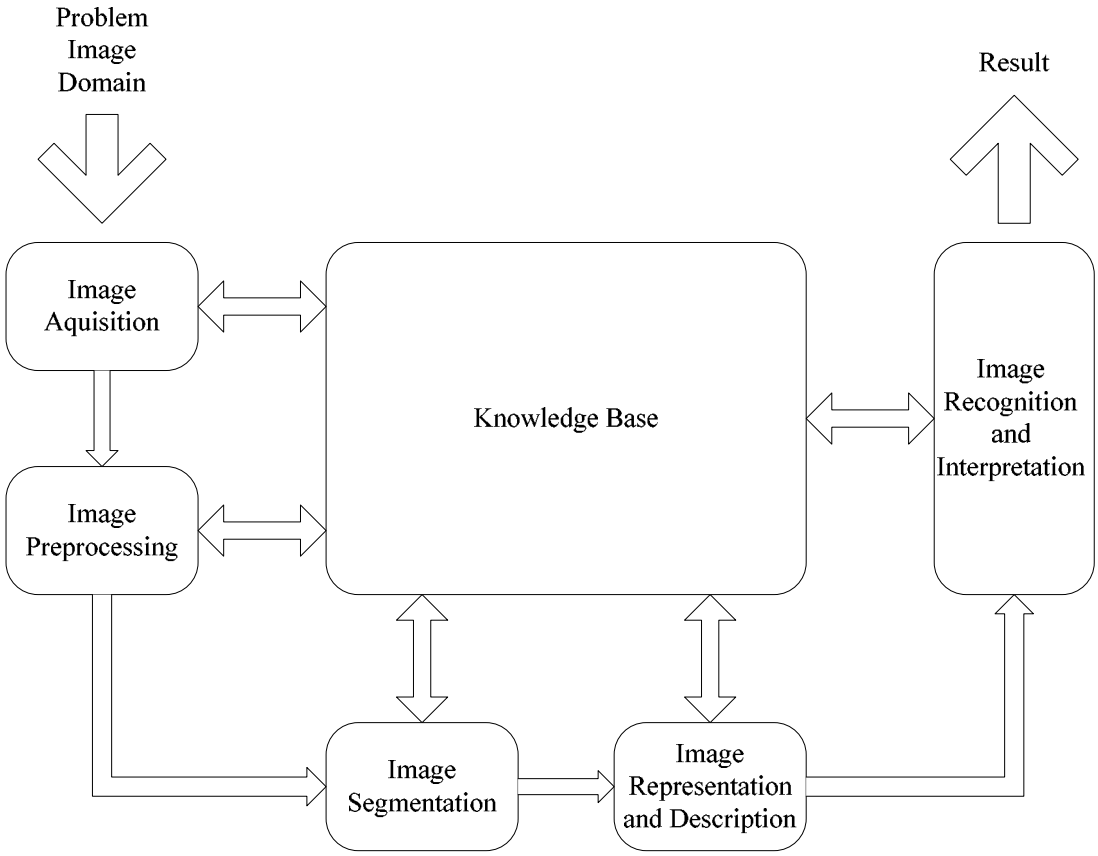


Figure 2.1 Fundamental Steps in Image Processing

Image acquisition

The first step in image processing is acquisition of the target image. Two physical devices are required for image acquisition. One is a sensing device which is sensitive to certain band(s) in the electromagnetic energy spectrum and produces electrical signal output proportional to the level of energy sensed, e.g. a Charged Coupled Device (CCD). Another device is a digitizer which is used to convert the electrical signal output of the sensing device to digital output. More information may be found in Chen (1998).

Image preprocessing

Once a target image is obtained, it may not in the desired form for further processing because of some practical constraints such as hardware limitations and environment settings during acquisition. So the target image needs to be preprocessed in

facilitating the latter steps. Image enhancements, restoration, compression and denoising are some common preprocessing techniques.

Image Segmentation

Image segmentation divides the target image into its constituent parts. This is one of the most difficult tasks in image processing.

Image Representation and Description

Image representation is a process of transforming raw data into a form suitable for subsequent computer processing, that is to encode the raw data. The purpose of image description, also known as feature extraction and/or selection, is to select a set of features that can be used to discriminate images.

Image Recognition and Interpretation

Image recognition labels a part of the image based on its descriptor information. Image interpretation is to assigns meanings to recognized objects which are viewed as a whole, integrated part.

Knowledge Base

Knowledge base is a place that stores prior or domain specific knowledge.

2.4 Image Features: the content descriptors

Image features generally refer to the meaningful and discriminant information extracted, either manually, semi-automatically, or automatically from images. It is worth to note that there is not any generally accepted (logical) categorization of image features. Here we briefly introduce some common image features and the categorization presented below may be named by their nature, by the object to be represented or by the extraction method.

Color

Colour plays a key role in two ways: one, colour provides important evidence about surfaces and materials, and two, colour vision is an integral part of the higher vertebrate foveal visual system. (McCabe 1997)

Until now, nonetheless, many processes of our visual system are still mysteries but the physical nature of color can be specified in some standard way, color models/spaces. (Gonzalez and Woods 1992; McCabe 1997)

RGB model. Each color appears to be a combination of its primary spectral components, i.e. the three primary colors: Red, Green and Blue. This is the color model used in color monitors and many video cameras. (Gonzalez and Woods 1992)

CMY(K) model. The primary colors of pigments, also known as secondary colors: Cyan, Magenta, Yellow and (Black) are used instead to serve as bases to represent colors. This model is used when hardcopy output is of interest. (Gonzalez and Woods 1992; Rubner 1999)

YIQ model. YIQ stands for Luminance, Inphase and Quadrature respectively and IQ are decoupled color information. This model is used in commercial color TV broadcasting. (Gonzalez and Woods 1992)

HSI model. HSI stands for Hue, Saturation and Intensity respectively. Hue and Saturation together form Chromaticity. (Gonzalez and Woods 1992)

Munsell HSV model. HSV stands for Hue, Saturation and Value (Gonzalez and Woods 1992). Hue is invariant under orientation of the object with respect to illumination and camera direction and hence more suited for object retrieval (Smeulders et al. 2000). It is mainly used in User Interface designs which provide designers a uniform way to easily manipulate colors.

CIE XYZ space. Luminance (Y) is represented independently from Chrominance (X, Z). A CIE xy chromaticity diagram can be derived from normalized values but the resultant diagram shows an uneven distribution of colors. (McCabe 1997)

CIE Y*uv* and Y*u'v'* space. *Yuv* partially solve the uneven distribution of colors in chromaticity diagram while *Yu'v'* is standardized to be a uniform chromaticity space.(McCabe 1997)

CIE Lab space. Lab space models the human perception in Euclidean distance. (Smeulders et al. 2000)

Color, which is an image content descriptor by itself, is often a criterion in the determination of other image features, e.g. edge, salient point, structure, histogram, etc.

Edge, Boundary and Region, Surface

Image edges(/lines) are generally regarded as discontinuities, which can arise from texture and color, between neighbourhoods. Related edges form a boundary and the area bounded is called region in a two dimensional plane (2D) and is called surface in a three dimensional space (3D). Regions and surfaces conforming to some geometrical constraints are known as shapes.

Previous works have been done on edge detection (Wu and Li 1997) in specific of straight lines (Burns et al. 1986), on 2D (Gonzalez and Woods 1992), on 3D (Wallace and Mitchell 1981; Zucker and Hummel 1981), on color (Trémeau and Colantoni 2000; Mojsilovic et al. 2000) and on texture (Ma and Manjunath 2000).

Point/Spot

A point or spot within an image is a small area. Point or spot whose magnitude differs significantly from its neighborhood catches visual focus. Therefore, they can

be used to characterize an image. (Sebe et al. 2000) This is regarded as an extreme form of weak segmentation (Smeulders et al. 2000).

Description of Structure and Layout (Spatial and Color)

Spatial relationships are the spatial ordering among spatial objects such as points, lines, regions and objects. Spatial relationships can be classified into three aspects: topological, directional and distance/metric relationships. Topological relationships include disjoint, adjacent, overlap, contain, etc. Directional relationships can be expressed as 4–(above, below, left and right) or 8–direction (north, northeast, east, southeast, south, southwest, west and northwest) or an even more complex description model. Distance/metric relationships are those physical measurements of the space between two entities. (Borowski et al. 2000; Vazirgiannis et al. 1998)

Color can add a new constraint on the discrimination of objects or by itself forms another layout, e.g. Trémeau and Colantoni (2000) and Garcia and Tziritas (1999).

Texture

There is no formal definition of textures. In computer vision, texture is defined as all what is left after color and local shape have been considered or it is defined by such terms as structure and randomness (Smeulders et al. 2000). Textures refer to the visual patterns that have properties of homogeneity that do not result from the presence of only a single color or intensity (Smith and Chang 1996). Texture descriptors provide measures of properties such as smoothness, coarseness, and regularity. It contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment (Haralick et al. 1973). Please see Chapter 2.4.2 for more discussion on different methods and their categorization.

Transform Coefficients

Transform Coefficients are the resultant information of image content that has undergone image transforms. Transform coefficients provide us with a different view of the image content such that some manipulations can be performed more effective and efficient than in the input space. One of the well-known applications is in image compression that high frequency components, which are insensitive to human visual perception, are removed while image quality can be retained. Another example is the use of Fourier Transform in texture feature extraction.

Histogram and Moments

In image processing and retrieval, the underlying frequency distribution of image pixels is what we are interested in. A histogram is a visualization of a frequency distribution in the form of a bar graph. The computational complexity of histograms is relatively low but that of the similarity measure between histograms is high because of the usually high dimensionality of histograms.

Moments have been broadly used to describe geometric properties of an object or a distribution (curve) (You and Bhattacharya 2000). It has been suggested that the computational complexity of measuring similarity of histograms can be reduced by modeling the histograms using their first few moments (Stricker and Orengo 1995).

Histogram and Moments bear the low sensitivity to camera and object motion because distribution of image content is invariant to image rotation and changes gradually under image translation (Mandal et al. 1996; Sebe et al. 2000). Nevertheless, they lack spatial distribution information about color (Ma and Zhang 1998) as they capture the global information of the image.

Concepts, Domain Knowledge

(Domain) Knowledge is modeled, in (Yoshitaka et al. 1994; Chua et al. 1994), as a set of concepts that contain descriptions, relationships with other concepts and rules

for their recognition in the contents of nodes. Actually, this is the reconsideration of the knowledge representation, conceptual/cognitive modeling.

The source of general knowledge has been identified in Smeulders et al. (2000) to be the following six classes.

Literal. Laws of syntactic equality and similarity define the relation between image pixels or image features regardless of its physical or perceptual causes.

Perceptual. Laws describing the human perception of equality and similarity are that they define equality on the same basis as the user experiences it.

Physical. Physical laws describing equality and difference of images under differences in sensing and object surface properties.

Geometric. Geometric and topological rules describe equality and differences of patterns in space. Although two objects are geometrically equal, the physical properties of their surfaces or the physical conditions of the sensing may be different.

Categorical. Category-based rules encode the characteristics common to class *z* of the space of all notions *Z*. Categories are almost exclusively used in narrow domains

Cultural. Man-made customs or man-related patterns introduce rules of culture-based equality and difference. Language is collected under culture.

Text Annotation (Keyword, Description)

Text appears more natural to people when formulating a query. It can represent our concepts, feelings, ideas, etc. So this is one way, and the most direct one, that we can think of, to describe an image and formulate a query. There are several ways to index images based on text (Chua et al. 1994): one is to use keywords; another is to use descriptions, e.g. captions; third is to use free-text.

Several image feature classification schemes are given in 2.4.1. The image processing method of image features are selectively described in 2.4.2.

2.4.1 Classification of image features

Gong (1998) classifies image features into five classes: Pixel-level, Global, Textural, Object and Conceptual. More generally, we can classify features based on their level of abstraction (Ahuja and Schachter 1981), complexity (Gudivada and Raghavan 1995) and level of segmentation (Smeulders et al. 2000).

Level of abstraction

We can categorize features into high-, middle- and low-level features. High- and middle-level features and, low-level features are equivalent to logical features and primitive features in Gudivada and Raghavan (1995).

High-level (**Logical**). High-level features are semantically rich information that is close to human perception. They are usually extracted manually or, highly intervened through the synthesis of low-level and/or middle-level features. Consequently, domain experts are required to accomplish the task. Domain concepts and text annotation are categorized as high-level features.

Middle-level (**Logical**). Middle-level features are one step further compared to low-level features while less domain knowledge but some general knowledge is required. Shape matching and object recognition are members of this type.

Low-level (**Primitive**). Low-level features are usually referred to features extracted at the signal processing level. They are totally machine extractable or minimum human intervention is required. Therefore, they can be processed automatically or semi-automatically. Color content, point, texture and transform coefficients belong to this class. Text information extracted from images (Jung et al. 2004) is also in this category.

Gong's classification (Gong 1998) is in ascending order of level of abstraction. Pixel-level, Global and Textural features are low-level features while Conceptual features are high-level features. Object features are in between high-level and low-level features, so we may call them middle-level ones.

Chang et al. (1997) classifies features in four level of abstraction, in ascending order: feature-, object-, syntax- and semantic-level. Feature-level corresponds to low-level while semantic-level corresponds to high-level. Object- and syntax-level are middle-level features.

Complexity

Features taxonomy can be considered in another dimension, complexity, i.e. features can be divided into complex and generic features (Gudivada and Raghavan 1995).

Complex. The only feature considered to be complex is (domain) concepts because it is the only class that can be synthesized from any of the compositions of generic features.

Generic. Generic features include all features except (domain) concepts. Some of the features may be generated based fully or partially from some other features but not any compositions of generic features.

Level of Segmentation

Segmentation is vital to image retrieval because several features, e.g. edge and shape, salient and, layout, depend on good segmentation. The following feature classification scheme is given in Smeulders et al. (2000).

Strong segmentation: (**segmented**) **object features**. Strong segmentation is a division of the image into regions T that contain exactly the pixels of the silhouette of objects O in the real world, such that $T = O$. It should be noted that object segmentation for broad domains of general images is not likely to

succeed, with a possible exception for sophisticated techniques in very narrow domains.

Weak segmentation: **salient features** (**region**). Weak segmentation is a grouping of the image data in conspicuous regions *T*, which are internally homogenous according to criterion, hopefully $T \subset O$. Since $T \subset$ O, it cannot be guaranteed that *T* fully covers *O*. Homogeneity criterion can be color (Forsyth and Fleck 1999; Pauwels and Frederix 2000), color and texture (Carson et al. 1997; Mirmehdi and Petrou 2000). Isolated Points (Schmid and Mohr 1997; Gevers and Smueulders 2000), which are the limit case of weak segmentation, do not require homogeneity but the effectiveness is subject to selection of the points.

Sign location: **sign probabilities**. Localizing signs is finding an object with a fixed shape and semantic meaning, $T = x_{center}$. Signs are helpful in content-based retrieval as they deliver an immediate and unique semantic interpretation.

Partitioning: **global features**. A partitioning is a division of the data array regardless of the data, $T \neq O$. *T* may equal to the whole image or fixed partitioning, e.g. dividing image into equal size tile and summarizing dominant feature values from each tile (Picard and Minka 1995).

Semantic and Perceptual

Image content can be classified as semantic content and perceptual content. Objects, events, and their relations are classified as semantic content. Color/Intensity, shape and texture are classified as perceptual content. (Jung et al. 2004; Kim 1996)

2.4.2 Image processing methods

Since our focus lies in primitive image features, we will review two widely studied primitive image features, color and texture (often represented by various transforms). Choosing an image processing method or a collection of them depends heavily on the characteristics of images in target application domain. Methods that work well in one domain can perform disappointingly in others (Tuceryan and Jain 1998).

Color

Colors in an image are represented using a color model, one of those reviewed above in Chapter 2.4. Some of the color models imitate the human perception, e.g. Munsell HSV and CIE Lab, that can be manipulated directly to represent color differences using proper distance measure. Color histogram is a popular method representing an image for retrieval that aggregates directly from the raw representation of an image or just part of an image. (Rui et al. 1999; Smeulders et al. 2000; Stricker and Orengo 1995; Swain and Ballard 1991)

Two physiological characteristics of human vision, color opponency and constancy have been considered. Color opponency (McCabe 1997; Smeulders et al. 2000), which can be derived from RGB model, is useful in saving storage and transmission bandwidth as well as in CBIR. Color constancy (McCabe 1997; Smeulders et al. 2000), which can also be derived from RGB model, is useful in eliminating the effects of illumination variation on an apparent color.

Invariant color representation has been explored for shape and object recognition.

Texture

In earlier times, textures are modeled in a *statistical* sense or a *structural* sense (Haralick 1979). Ahuja and Schachter (1981) later suggested, instead of using Haralick's (1979) classification. Two low-level texture models: pixel-based and

Region-based. Tuceryan and Jain (1998) refine Haralick's (1979) classification of texture models to four categories: *statistical*, *geometrical*, *model-based*, *signal processing methods*.

Statistical methods measure the spatial distribution of pixels in an image. One well-known method is co-occurrence matrices. Several statistics can be obtained from co-occurrence matrix including *energy*, *entropy*, *contrast*, *homogeneity* and *correlation* to describe the texture in an image. Another recognized method is autocoorelation that can measure the regularity and fineness/coarseness of the texture in an image.

Geometrical methods include the previously known structural methods modeling structural primitives and placement rules. Local properties and shape are examples of structural primitives. Spatial relationships are common illustration of placement rules.

Model-based methods include Random Field models (e.g. Markov, MRF) and Fractals are based on the building of an image model such that their parameters represent some characteristics of texture.

Signal processing methods are adopted for texture analysis because human perform frequency analysis on images in the brain. Spatial domain filters and spectral (or Fourier) domain filters have been extensively studied. Recently advances use multiresolution methods, such as wavelets and Gabor filters, to extract texture features. Please see also Chapter 3.1 for more details.

2.4.3 Similarity measure

Similarity of images is a measure of how much one image resembles or differs from another. Each feature type has its own characteristics and thus different similarity measures fit for different feature types. It is worth to notice that hierarchically

ordered features may help staying away from problems arisen from segmentation. (Smeulders et al. 2000) We will state some well-known similarity measures.

Similarity measures can be a *metric* or *non-metric*. Metric-based (dis)similarity measures have been widely used in computer vision, pattern recognition communities. Metric-based techniques satisfy the following constraints.

Minimality: $D(a, b) \geq D(a, a) = D(b, b) = 0$ (2.1)

Symmetry: $D(a, b) = D(b, a)$ (2.2)

The Triangle Inequality: $D(a, b) + D(b, c) \geq D(a, c)$ for all b (2.3)

Distance measures are one of the metric-based techniques. One family of distance measures, L_p-norm or L_p-metric, is generally known as Minkowski distance (Eq. 2.4). L_1-norm is generally known as Manhattan distance or City Block distance (Eq. 2.5) and L_2-norm is generally known as Euclidean distance (Eq. 2.6).

$$L_p(a,b) = \left(\sum_{i=1}^{n} | a_i - b_i |^p \right)^{1/p} \tag{2.4}$$

$$L_1(a,b) = \sum_{i=1}^{n} | a_i - b_i | \tag{2.5}$$

$$L_2(a,b) = \sqrt{\sum_{i=1}^{n} | a_i - b_i |^2} \tag{2.6}$$

Another often used distance measure is angular or angle-based distance (Eq. 2.7).

$$A(a,b) = -\cos(a,b) = -\frac{\sum_{i=1}^{n} a_i b_i}{\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2} \tag{2.7}$$

2.5 Recapitulation

In this chapter, we have first made an introduction to biometrics. We have reviewed, especially the content-based domain, some of the basic concepts and issues in image retrieval in image processing and in image features.

Chapter 3 Tools for Content-Based Retrieval

3.1 Transform-based Methods

Transform-based methods are some mathematical transformations that have been introduced because of some of their nice properties. We briefly describe the mathematical formulation of Fourier Transform, Wavelets and Gabor Filters in the following subsections.

3.1.1 Fourier Transform (FT)

Fourier Transform, also known as Fourier analysis, breaks down a signal into constituent sinusoids (of unlimited duration) of different frequencies (Misiti et al. 1996). In mathematical sense, it is a technique that transforms a signal into a time-based one to a frequency-based one. It is a mathematical tool introduced to the signal processing field and afterwards, introduced to image processing in its discrete, 2-D form. In the following equations, $j = \sqrt{-1}$.

1D forward and inverse Continuous Fourier Transform are defined as follows.

$$F(u) = \int_{-\infty}^{\infty} f(x) e^{-j2\pi ux} dx \tag{3.1}$$

$$f(x) = \int_{-\infty}^{\infty} F(u) e^{j2\pi ux} du \tag{3.2}$$

2D forward and inverse Continuous Fourier Transform are defined as follows.

$$F(u,v) = \int_{-\infty}^{\infty} \int f(x,y) e^{-j2\pi(ux+vy)} dx\, dy \tag{3.3}$$

$$f(x,y) = \int_{-\infty}^{\infty} \int F(u,v) e^{j2\pi(ux+vy)} du\, dv \tag{3.4}$$

1D forward and inverse Discrete Fourier Transform is defined as follows.

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) e^{-j2\pi ux/N} \tag{3.5}$$

where u = 0, 1, 2, …, N-1

$$f(x) = \sum_{u=0}^{N-1} F(u) e^{j2\pi ux/N} \tag{3.6}$$

where x = 0, 1, 2, …, N-1

2D forward and inverse Discrete Fourier Transform is defined as follows.

$$F(u,v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi(ux/M + vy/N)} \tag{3.7}$$

where u = 0, 1, 2, …, M-1 and v = 0, 1, 2, …, N-1

$$f(x,y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v) e^{j2\pi(ux/M + vy/N)} \tag{3.8}$$

where x = 0, 1, 2, …, M-1 and y = 0, 1, 2, …, N-1

FT has a serious drawback in that time information is lost during the transformation from the time domain to the frequency domain. When it is extended to image processing, the spatial information is lost during the transformation from the spatial domain to the frequency domain.

Fourier Mellin Transform (FMT)

Fourier Mellin Transform (Casasent and Psaltis 1976; Sheng and Arsenault 1986) is translation, in-plane rotation (referred as *Z*-plane rotation in Lai et al. 2001) and scale invariant transformation. FMT was first described in Casasent and Psaltis (1976) for optical systems. It was introduced into pattern recognition in Sheng and Arsenault (1986), into image registration in Srinivasa Reddy and Chatterji (1996) and into face recognition in Lai et al. (2001).

By the properties of FT, the magnitude of FT, $|F(u, v)|$, is invariant to translations in spatial domain but it is variant to in-plane rotation and scale in the spatial domain. A polar (coordinate) transform on the centralized magnitude of FT can separate the effect of in-plane rotation and scale in spatial domain $|F(r, \theta)|$. The Mellin Transform is applied on r and FT is applied on θ. In implementation, it is a logarithmic scaling of r followed by a 2D FT. (Casasent and Psaltis 1976) A process flow diagram of FMT is depicted in Figure 3.1.

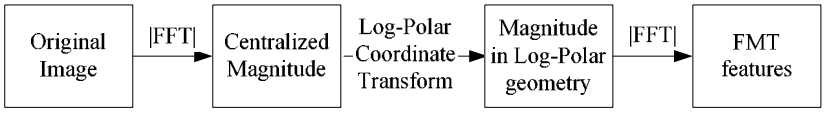


Figure 3.1 Process flow of Fourier Mellin Transform

A detailed discussion of the invariant properties of FMT can be found in Casasent and Psaltis (1976), Srinivasa Reddy and Chatterji (1996) and, Lai et al. (2001). The implementation of FMT has been considered in Srinivasa Reddy and Chatterji (1996).

Holistic Fourier Invariant Features (HFIF)

Holistic Fourier Invariant Features (Lai et al. 2001) are features extracted through an integration of wavelet transform (WT, see Chapter 3.1.2 for details) and FMT (see above for details). FMT is applied on the low frequency subband that results from WT. The process flow is shown in Figure 3.2.

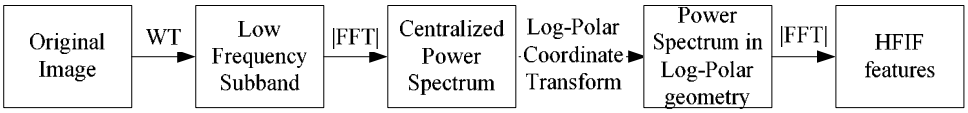


Figure 3.2 Process flow of Holistic Fourier Invariant Features extraction

3.1.2 Wavelets and Wavelet Transform (WT)

Wavelet Transform, also known as Wavelet analysis, produces a time-scale view of a signal (Misiti et al. 1996). It offers the promise of compact representation and efficient detection of image components that match the wave-shape of the chosen (mother) wavelet. A wavelet is a waveform of limited support, i.e. effectively limited duration, and has an average value of zero.

Wavelets are a new way to see and represent signal. It is an alternative to the Fourier Transform but not a replacement (Strang and Nguyen 1996). Wavelet analysis and synthesis have been applied in signal processing to deal with audio, image and video processing and there is an increasing interest in how to utilize this tool.

Much research on image compression, storage and retrieval has been done with the use of wavelets in image for (Albanesi and Bertoluzza 1995; Albanesi et al. 1999; Liang and Kuo 1999; Saha and Vemuri 1999; Strang and Nguyen 1996; Yang and Mitra 1998; Zamora et al. 2000). Some research, instead of working in the spatial domain, has worked on the wavelet coefficients of images after being decomposed using wavelet transform, (Albuz et al. 1999; Liang and Kuo 1997; Liang and Kuo 1999; Servetto et al. 1997; You and Bhattacharya 2000). This research, usually, extracted features from the wavelet coefficients and features extracted were indexed for fast searching based on similarity. Similarly, some researchers utilized the properties of wavelet analysis to perform shape/region detection, e.g. finding faces in images (Garcia and Tziritas 1999; Karlekar and Desai 2000).

Wavelet transform is the process that connects a function *f(t)* and its wavelet coefficients. The function *f(t)* can be represented by the combination of basis functions as follows.

$$b_{JK} = \int_{-\infty}^{\infty} f(t)\omega_{JK}(t)dt \tag{3.9}$$

$$f(t) = \sum_{j.k} b_{jk}\omega_{jk}(t) \tag{3.10}$$

Eq. 3.9 represents the wavelet analysis while Eq. 3.10 represents the wavelet synthesis. There is connection between filter bank and wavelets, “lowpass filter” leads to scaling functions, Eq. 3.11, while “highpass filter” leads to wavelets Eq. 3.12.

$$\phi(t) = \sqrt{2}\sum_{k=0}^{N} c(k)\phi(2t-k) \tag{3.11}$$

$$\omega(t) = \sqrt{2}\sum d(k)\phi(2t-k) \tag{3.12}$$

Figure 3.3 illustrates the 1D wavelet decomposition and reconstruction of an input signal X_0 that uses a two-channel filter bank. Input signal X_0 goes through the highpass filter and the lowpass filter for decomposition and then are downsampled,

the resultant (shaded area) is wavelet coefficients. They can be reconstructed, i.e. get back to original, by being upsampled and then going through the synthesis highpass and lowpass filters.

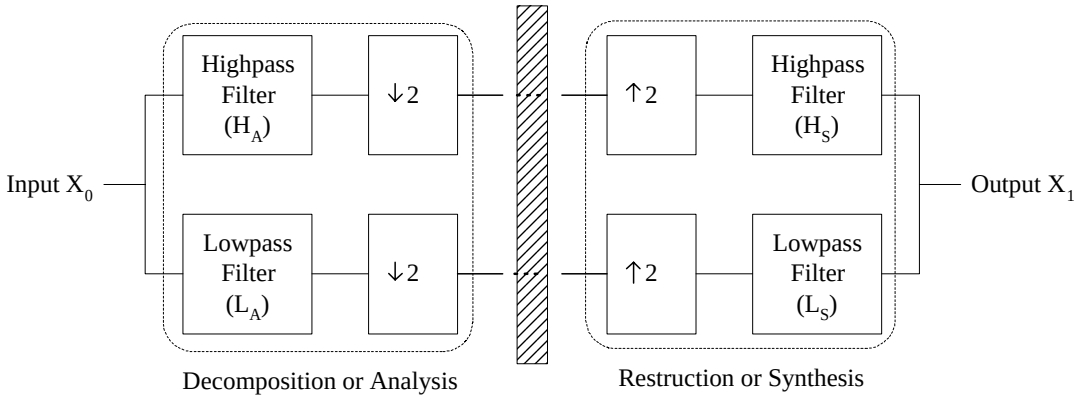


Figure 3.3 Wavelet analysis and synthesis: Two-channel filter banks

Multiresolution

The simultaneous appearance of multiple scales is known as multiresolution. Multiresolution can be presented as follows; subspace V_{j-1} and its complementary subspace W_{j-1} together form the higher or finer space V_j.

$$V_j = V_{j-1} \oplus W_{j-1} \tag{3.13}$$

In other words, $V_0 \subset V_1 \subset V_2 \subset \cdots$. A realization of the direct sum of coarser subspaces to become finer subspace is shown as follows.

$$\sum_k a_{jk}\phi_{jk}(t) = \sum_k a_{j-1,k}\phi_{j-1,k}(t) + \sum_k b_{j-1,k}\omega_{j-1,k}(t) \tag{3.14}$$

Figure 3.4 presents the decomposition of an input signal X_0 in the multiresolution manner implemented as two-channel filter bank.

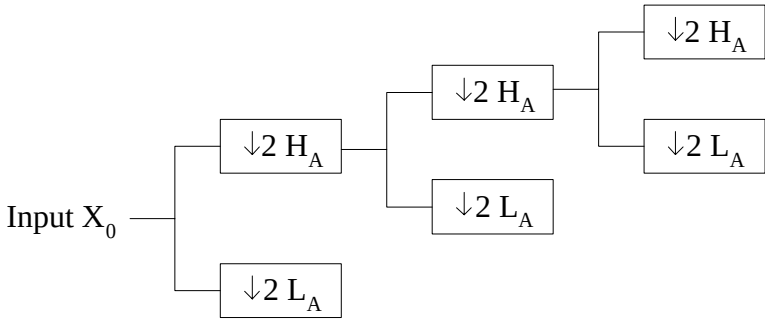


Figure 3.4 Wavelet multiresolution decomposition: Two-channel filter banks

3.1.3 Gabor Filters

2D Gabor filters proposed by Daugman (1985) extended from the original (generally referred to as 1D version) proposed by D. Gabor (1946) in communication engineering has shown its capability in modeling the receptive-field profiles of simple cells in mammalian visual cortex. Simple cells in mammalian visual cortex are organized roughly in the polar Orientation-Frequency plane to form a receptive-field such that with their specific 2D locations in visual space, preferred orientations and spatial frequencies, localized 2D spectral information is captured. (Daugman 1985) A set of 2D Gabor filters, which are of various spatial dimensions and, spatial-frequency and orientation bandwidths to imitate those empirical characteristics of receptive-field, can be used to extract different kinds of information from images.

2D Gabor filters have been shown to be effective for texture analysis on monochromatic images. They optimally achieve joint resolution in space and spatial frequency; their orientations, radial frequency bandwidths and center frequencies are all tunable. Thereby, it is adopted to perform texture analysis. (Bovic et al. 1990; Liu and Wechsler 2003)

2D Gabor filters, in general, are of the following functional form:

$$h(x, y) = g(x', y') \exp[2\pi i(Ux + Vy)] \tag{3.15}$$

where $(x', y') = (x\cos\phi + y\sin\phi, -x\sin\phi + y\cos\phi)$ and

$$g(x, y) = \left(\frac{1}{2\pi\lambda\sigma^2}\right)\exp\left[-\frac{(x/\lambda)^2 + y^2}{2\sigma^2}\right] \tag{3.16}$$

where $\sigma = \sigma_x = \sigma_y$ is the standard deviation of the Gaussian envelope

The corresponding Fourier transform is

$$H(u, v) = \exp\{-2\pi^2\sigma^2[(u' - U')^2\lambda^2 + (v' - V')^2]\} \tag{3.17}$$

where $(u', v') = (u\cos\phi + v\sin\phi, -u\sin\phi + v\cos\phi)$ and

$(U', V') = (U\cos\phi + V\sin\phi, -U\sin\phi + V\cos\phi)$

Circular Gabor Filters

If modulating Gaussians of 2D Gabor filters have the same orientation as the complex sine grating, i.e. $\phi = \theta$ and aspect ratio is 1 (i.e. $\lambda = 1$), a family of circular 2D Gabor filters is of the following form in the spatial domain:

$$h(x, y) = g(x', y') \exp(2\pi i F x') \tag{3.18}$$

The corresponding Fourier transform of Eq. 3.18 is

$$H(u, v) = \exp\{-2\pi^2\sigma^2 [(u' - F)^2\lambda^2 + (v')^2]\} \tag{3.19}$$

with radial central frequency $F = \sqrt{U^2 + V^2}$ cycles/image and orientation, $\theta = \tan^{-1}(V/U)$ degrees or radians (measured from u-axis).

The expanded form of the family of circular Gabor filters in spatial domain is

$$h(x, y, \theta, F, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left[\frac{x^2 + y^2}{-2\sigma^2} + 2\pi i F (x\cos\theta + y\sin\theta)\right] \tag{3.20}$$

Elliptical Gabor Filters

The expanded form of the family of elliptical Gabor filters in spatial domain is

$$h(x, y, \omega, \theta) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-\omega}{8\kappa^2}(4x'^2 + y'^2)\right)\left[\exp(i\omega x') - \exp(\frac{\kappa^2}{2})\right] \tag{3.21}$$

for $\kappa = \sqrt{2\ln 2}\left(\frac{2^\delta + 1}{2^\delta - 1}\right)$ and $\omega = \frac{\kappa}{\sigma}$, i $= \sqrt{-1}$, u is the sinusoidal wave frequency in spatial domain, θ is the orientation of the filter and δ is the half-amplitude bandwidth of the frequency response (Lee 1996).

Applying Gabor filter to a monochromatic texture image $t(x, y)$ yields

$$k(x, y) = k_c(x, y) + i k_s(x, y) \tag{3.22}$$

where

$$k_c(x, y) = \mathrm{Re}\{k(x, y)\} = h_c(x, y) * t(x, y), \tag{3.23}$$
$$k_s(x, y) = \mathrm{Im}\{k(x, y)\} = h_s(x, y) * t(x, y) \tag{3.24}$$

The amplitude and phase (envelopes) of $k(x, y)$ are, respectively,

$$m(x, y) = [k_c^2(x, y) + k_s^2(x, y)]^{1/2}, \tag{3.25}$$
$$\psi(x, y) = \tan^{-1}[k_s(x, y)/k_c(x, y)] \tag{3.26}$$

3.2 Learning Methods

In this subsection, we will briefly describe two types of learning methods, linear subspace methods and self organizing map. They are widely used in many applications across various disciplines (Haykin 1999; Kohonen et al. 1996)

3.2.1 Linear subspace methods

Linear subspace methods are used to transform the original sample space into its subspace through a linear transformation with certain objective in order that the original sample space can be better separated and represented more compactly.

Principal Component Analysis (PCA) is an unsupervised learning method while Linear Discriminant Analysis (LDA) is a supervised learning method. Before we go into the details, we would like to define three scatter/covariance matrices, with-in class S_W, between-class S_B and total S_T, that will be used below. Suppose there are L (pattern) classes and c_i samples in class C_i where $i = 1 \ldots L$ and, altogether M samples Moreover, μ is the mean of all samples while μ_i is the mean of all samples of class C_i,

i.e. $\mu = \frac{1}{M}\sum_{i=1}^{M} X_i$ and $\mu_i = \frac{1}{c_i}\sum_{X \in C_i} X$.

$$S_W = \frac{1}{M}\sum_{i=1}^{L}\sum_{X \in C_i} c_i (X - \mu_i)(X - \mu_i)^{\mathrm{T}} \tag{3.27}$$

$$S_B = \frac{1}{M}\sum_{i=1}^{L} c_i (\mu_i - \mu)(\mu_i - \mu)^{\mathrm{T}} \tag{3.28}$$

$$S_T = S_W + S_B = \sum_{X} (X - \mu)(X - \mu)^{\mathrm{T}} \tag{3.29}$$

The three scatter matrices are symmetric.

Principal Component Analysis (PCA)

Principal Component Analysis (Gonzalez and Woods 1992; Halici and Ongun 1996; Haykin 1999), which is also known as (discrete) Karhunen-Loève Transform or Hotelling Transform, is a statistical method that linearly maps the data space

(original distribution) to feature space (usually a subspace of the original) with minimal mean square (approximation) error. Transforming from original space (Analysis), data can be effectively represented by a subspace of fewer dimensions, i.e. Principal Components, with the essential information retained such that mean-squared error is optimized and is equal to the sum of variances of truncated elements. It is well known for feature extraction/selection in pattern recognition, noise reduction in signal processing and de-correlation.

PCA can be formulated as the maximization of the criterion in Eq. 3.30 and the maximization of the criterion corresponds to an eigenvalue problem given in Eq. 3.31(Yang et al. 2004).

$$J(w) = w^T S_T w \tag{3.30}$$

To solve the eigenvalue problem, Eigenvalue Decomposition (EVD) is applied on the covariance matrix of (training) samples. Eigenvectors of the covariance matrix, which are orthonormal bases, are found and sorted in descending order according to their importance, i.e. the magnitude of corresponding eigenvalues. The method is as follows (Gonzalez and Woods 1992; Haykin 1999).

Suppose there are M real valued vectors $\{X_j \in \boldsymbol{R}^n | X_j = [x_{1,} x_{2,\ldots,} x_{n,}]\}$, where $j = 1 \ldots M$. The covariance matrix, also known as the total scatter matrix S_T, is calculated as in Eq. 3.29.

$$S_T W = W \lambda \tag{3.31}$$

where W is a $n \times n$ matrix containing eigenvectors of S_T such that $W^T W = \boldsymbol{I}$, i.e. each vector is orthogonal to the others and is normalized, and λ is a $n \times n$ diagonal matrix, $\{\lambda = \text{diag}\ [\lambda_1,\ldots, \lambda_n] \mid \lambda_1 = \lambda_{max} > \lambda_2 > \ldots > \lambda_n\}$, containing eigenvalues as diagonal elements.

Since the columns of W are ordered in descending order of the magnitude of their eigenvalues, by truncating $(n - m)$ columns of W, the columns of the

resulting matrix P (of m dimension) are known as the Principal Components and the space spanned by P is known as the Principal Subspace, i.e. the feature space. Through the use of Principal Components, the Principal Subspace can effectively represent the data space. Thus the important features are selected or the dimension of data space is reduced.

Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis is a widely used classification method (Webb 1999). It is also known as Fisher Discriminant Analysis (FDA), Fisher Linear Discriminant (FLD) (Liu and Wechsler 2002). It is to maximize the *Fisher's Criterion* $J(w)$, that is the ratio of the variance of between-class samples to the variance of within-class samples, as shown below in Eq. 3.33. This maximization, however, does not directly link to the classification error, i.e. usually the system performance index (Lu et al. 2003).

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \tag{3.33}$$

The following criterion is sometimes used as an alternative to Eq. 3.33 (Yang et al. 2003; Lu et al. 2003)

$$J(w) = \frac{w^T S_B w}{w^T S_T w} \tag{3.34}$$

The maximization of the Fisher's Criterion $J(w)$, which is a *Rayleigh Quotient*, corresponds to a generalized eigenvalue problem (Melzer 2004). For the mathematical derivations, please refer to Melzer (2004).

$$S_B W = S_W W \Lambda \tag{3.35}$$

There are two generally used LDA implementations, Foley-Sammon LDA (FSLDA, Foley and Sammon Jr. 1975; Duchene and Leclercq 1988) and Uncorrelated LDA (ULDA, Jin et al. 2001a, 2001b). They differ by satisfying different orthogonality constraints during the maximization of the Fisher's Criterion.

FSLDA satisfies the constraint in Eq. 3.36 while ULDA satisfies the constraint in Eq. 3.37 (Jin et al. 2001a, 2001b; Yang et al. 2002)

$$w_i^T w_j = 0 \ \forall i \neq j \tag{3.36}$$
$$w_i^T S_T w_j = 0 \ \forall i \neq j \tag{3.37}$$

The discriminant vectors of FSLDA need to be determined one by one. The discriminant vectors of ULDA, in contrast, can be determined in a whole. In general, LDA uses the following algorithm that produces uncorrelated discriminant vectors similar to ULDA.

Suppose we have the scatter matrices, S_W and S_B, of training samples in hand.

We first apply EVD on S_W

$$S_W = V D V^T \tag{3.38}$$

Discarding all zero eigenvalues d_i (= 0) of D (in implementation, eigenvalues of extremely small magnitude) and their associating eigenvectors v_i of V, a whitening matrix B is formed by

$$B = (V')(D')^{-1/2} \tag{3.39}$$

where $D' = \{\text{diag}(d_j) > 0\}$ and $V' = \{v_j\}$. S_B is transformed into S'_B by

$$S'_B = B S_B B^T \tag{3.40}$$

We then apply EVD on S_B

$$S'_B = U \Lambda U^T \tag{3.41}$$

The associated eigenvectors U' of the m biggest eigenvalues $\lambda_1, \cdots, \lambda_m$ of Λ are selected to form the transformation matrix W.

$$W = B (U') \tag{3.42}$$

In case the with-in class scatter matrix S_W becomes singular owing to the high feature dimensionality and the small sample size (SSS) problem (Lu et al. 2003; Yang et al. 2005), PCA has been applied to the input and feature space respectively before performing LDA (Swets and Weng 1996; Belhumeur et al. 1997; Liu and Wechsler 2002; Lu et al. 2003) which is known as PCA plus LDA approach (PCA+LDA). Yang and Yang (2003) give theoretical support to the PCA+LDA approach.

3.2.2 Self-Organizing Map (SOM)

The Self Organizing Map or Self-Organizing Feature Map (Haykin 1999; Kohonen 1997; Mitra and Pal 1994), proposed by T. Kohonen, is a well-known unsupervised learning algorithms in the field of artificial neural networks. It models the neurobiological behaviour of the human brain. SOM is basically a kind of competitive learning that only one neuron will fire after mutual competition of neurons, i.e. winner-takes-all. Its aim is to adaptively generate, based on the input patterns (of arbitrary dimension), a lower dimensional (usually one- or two-dimension) topologically ordered discrete map. Although, in general, it can be extended to higher dimension, one- or two-dimensional SOM is commonly adopted because of its simplicity and expressiveness.

SOM has been used in industrial monitoring and analysis, statistical pattern recognition including texture analysis and classification and other areas such as image compression and encoding, robotics and telecommunication (Haykin 1999; Kohonen 1997; Kohonen et al. 1996). It is capable of clustering the training data without any pre-classification of the training data. Nevertheless, for practical reasons, primary data is seldom used directly in the application of artificial neural network (e.g. SOM); thus, feature extraction is usually performed before applying artificial neural network, e.g. clustering (Kohonen et al. 1996).

There are two stages of operation in SOM: Formation of SOM and Calibration of SOM. Formation of SOM has four phases: first is initialization (of synaptic weights), second is competition, third is cooperation and the final stage is synaptic adaptation. (Haykin 1999)

Suppose an SOM consisting of, in total, l^2 neurons. Let the input space is of m dimension. An input vector randomly chosen from the input space is denoted by

$$\boldsymbol{x} = [x_1, x_2, \ldots, x_m]^T \quad (3.43)$$

The major steps of the generation of the self-organizing map are summarized as follows:

Step 1 Initialization (**of synaptic weights**). The synaptic weight vector ($\boldsymbol{w_i}$) of each neuron is of the same dimension as the input space. Each synaptic weight (w_{ij}) can be initialized randomly within the range of the domain or by picking small values from a random number generator.

$$\boldsymbol{w_i} = [w_{i1}, w_{i2}, \ldots, w_{im}]^T \text{ for } i = 1, 2, \ldots, l^2 \quad (3.44)$$

Step 2 Competition. A discriminant function, $d(\boldsymbol{x}, \boldsymbol{w}_i)$, set the basis for neurons' competition and each neuron computes its resulting value using the discriminant function. The neuron with the most distinct value is chosen as the winner. Since only one winning neuron is selected for each input pattern, if more than one neuron have same distinct value, one of them will be selected randomly to be the winner. The winning neuron n_w is defined as

$$n_w = arg\ min_i\ d(\boldsymbol{x}, \boldsymbol{w}_i) \text{ for } i = 1, 2, \ldots, l^2 \quad (3.45)$$

Step 3 Cooperation. The winning neuron n_w becomes the center for determining the spatial position of topological neighboring neurons through a neighborhood function $N(n_w, n_i, t)$ that defines neighborhood members with respect to the central element, based on the distance $h(n_w, n_i)$ between the center (i.e. winning neuron, n_w) and surrounding elements (n_i) and, training time (t).

$$h(n_w, n_i) = \| c(n_i) - c(n_w) \| \text{ for } i = 1, 2, \ldots, l^2 \quad (3.46)$$

where $\|\cdot\|$ denotes the Euclidean norm and $c(n_i)$ determine the spatial location, i.e. coordinate, of neuron n_i in the topographic map.

A neighborhood function can be stated in discrete form or continuous form. Neighborhood function, $N(n_w, n_i, t)$, can be represented as a step function for discrete case, while it may be expressed, in continuous case, as a Guassian

function. For 2D topology, rectangular, 2D Guassian (Eq. 3.47) or Mexican hat is commonly chosen to define neighborhood. Both the winning neuron and its neighboring neurons, unlike in general competitive learning, learn from the input pattern; nearer neurons are adjusted more. The size of neighborhood decreases with training time.

$$N(n_w, n_i, t) = \exp(\frac{h(n_w, n_i)}{2\sigma^2(t)}) \quad (3.47)$$

Step 4 Synaptic Adaptation. The synaptic weight of neuron *j* being adjusted in relation to the input vector $\boldsymbol{x}$ at time *t* can be expressed as follows. This adjustment is applied to the winning neuron n_w and its neighboring neurons determined by $N(n_w, n_i, t)$. There are two phases of Synaptic Adaptation, Ordering and Convergence/Tuning. The size of neighborhood $N(n_w, n_i, t)$ and learning rate $\eta(t)$ of SOM at two phases are different. At the early ordering phase, we would like the whole SOM to learn quickly about the input patterns, so the neighborhood may include all neurons and the learning rate is relatively large, e.g. 0.1. The two parameters are expected to decrease gradually with time in the ordering phase. At the convergence/tuning phase, we would like to fine tune the feature map so as to provide an accurate statistical quantification of the input space, so the neighborhood may only include the nearest ones and the learning rate is small, e.g. 0.01, but, to avoid the occurrence of metastable state, not zero. The update of the synaptic weight vector is defined as

$$\boldsymbol{w_i}(t+1) = \boldsymbol{w_i}(t) + \eta(t)\, N(n_w, n_i, t)\, (\boldsymbol{x} - \boldsymbol{w_i}(t)) \text{ for } i = 1, 2, \ldots, l^2 \quad (3.48)$$

where $\eta(t)$ is the learning rate

More specifically the formation of SOM includes the following basic operations (Haykin 1999; Halici and Ongun 1996):

Step 1 Initialization. Randomly choose values to initialize weight vectors $\boldsymbol{w}_i(0)$ for $i = 1, 2,\ldots, l^2$; or randomly select from the available input vectors as weight vectors $\boldsymbol{w}_i(0)$.

Step 2 Sampling. Randomly draw one from the available input vectors as $\boldsymbol{x}$

Step 3 Similarity matching. Apply $d(\boldsymbol{x}, \boldsymbol{w}_i)$ on all neurons and determine n_w

Step 4 Updating. Adjust $\boldsymbol{w}_i(t)$ to $\boldsymbol{w}_i(t+1)$ as described above

Step 5 Continuation. Continue with steps 2 to 5 until no observable changes in the feature map

Calibration of SOM (Mitra and Pal 1994) is actually labeling the training samples/input patterns with a corresponding (winning) class/node number that is computed using the same discriminant function, $d(\boldsymbol{x}, \boldsymbol{w}_i)$, in the Formation stage of SOM. This can provide some qualitative information about the topological ordering between the input and output space.

3.3 Benchmarking Methods

Retrieval performance is a vital index that helps the design of CBIR, including the choice of image features, similarity measures and retrieval methods. In this section, we would like to review some benchmarking methods that have been used to evaluate the retrieval performance of CBIR. (Ma and Zhang 1998; Smeulders et al. 2000)

Accuracy and Error

Accuracy is one of the most generally used and accepted performance indexes for retrieval systems. It is often used in reporting the performance of retrievals in top few matches, e.g. *k*-NN. Suppose *C* corrected matches out of *Q* queries is measured from a retrieval system. *Accuracy* and *Error*, which may be reported in percentage, are defined by

$$\text{Accuracy} = \frac{C}{Q} \tag{3.49}$$

$$\text{Error} = 1 - \frac{C}{Q} \tag{3.50}$$

Precision and Recall

Precision and *Recall* are two performance indexes introduced from information retrieval community (Narasimhalu et al. 1997; Smeulders et al. 2000). Suppose, with respect to a query, *R* relevant images are in the database and, *r* relevant images and *N* images are returned. Precision and recall are defined as follows.

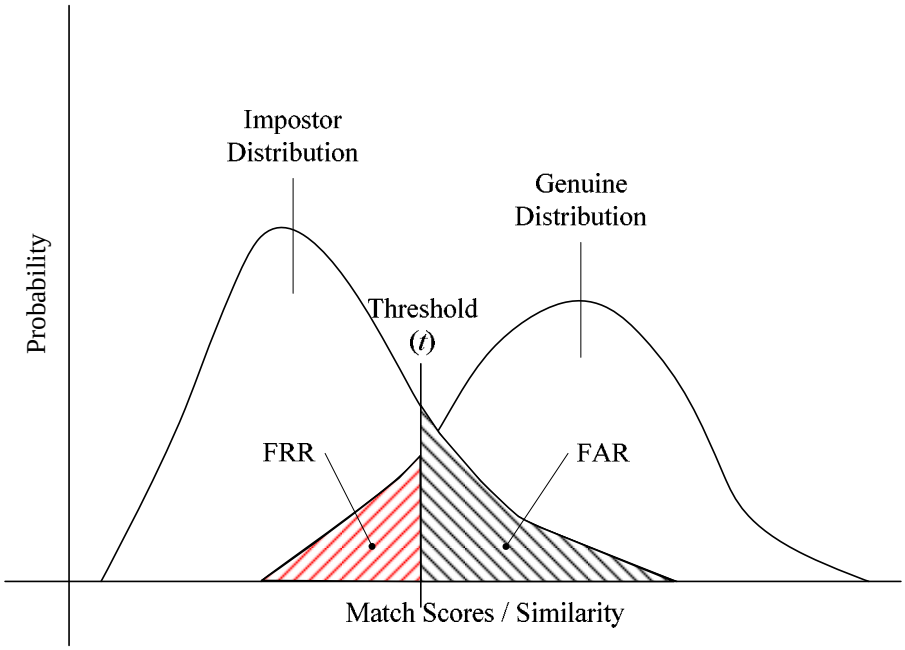
$$\text{Prescision} = \frac{r}{N} \tag{3.50}$$

$$\text{Recall} = \frac{r}{R} \tag{3.51}$$

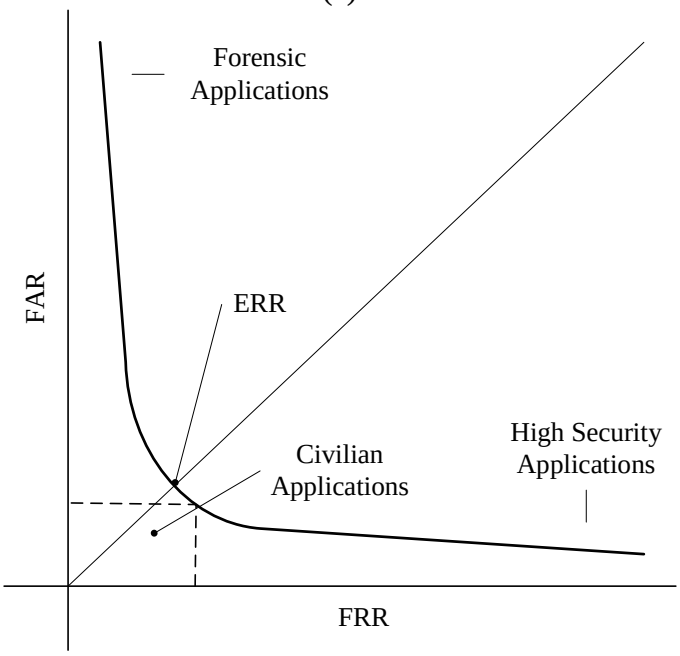
However, precision and recall neglect the ordering of the list of returned images. Variants of performance indexes based on precision and recall have been proposed (Ma and Zhang 1998; Heczko et al. 2004). Precision and recall (or other measures derived form them) are useful measurements only when the image database relies on the strong semantics provided by label or other textual description (Smith and Li 1998).

False Accept Rate (FAR), False Reject Rate (FRR) and Equal Error Rates (EER)

The distribution of scores generated from *mate pairs* is called *genuine* distribution. The distribution of scores generated from *nonmate pairs* is called *impostor* distribution. *False Accept(ance) Rate* (*FAR*) and *False Reject(ion) Rate* (*FRR*) are occurred in pairs as a function of a particular operating point, threshold *t*, that regulates the decision of a biometric system. (Prabhakar et al. 2003; Jain et al. 2004) Please see Figure 3.5 (a) for a graphical illustration.



(a)



(b)

Figure 3.5 (a) FAR and FRR for a given threshold display over the genuine and impostor distribution and (b) ROC curve with EER and typical operating points marked

FAR, which is also known as *False Match Rate* (*FMR*), refers the rate of mistaking biometric measurements from two different persons to be from the same person. FRR, which is also known as *False NonMatch Rate* (*FNMR*), refers the rate of mistaking biometric measurements from the same person to be from two different persons. (Prabhakar et al. 2003; Jain et al. 2004) *Genuine Accept(ance) Rate* (*GAR*)

is the complement of FRR. *Equal Error Rate* (*EER*) is the error rate (probability) where FAR and FRR curves cross (crossover rate) (Dawson 2003). Receiver Operating Characteristic (ROC) curve is a plot of GAR/FRR against FAR for all possible operating points. Please see Figure 3.5 (a) and (b) for an illustration.

3.4 Recapitulation

In this chapter, we have reviewed some tools for content-based storage and retrieval. They include some transformed based methods, learning methods and benchmarking methods.

For transformed based methods, we have described three popular transforms in the image processing community, namely, Fourier transform, Wavelet transform and Gabor filters.

We consider two unsupervised learning methods, Principle Component Analysis and Self Organizing Map, and one supervised learning method, Linear Discriminant Analysis.

We finally describe several benchmarking methods for performance evaluation of retrieval/recognition.

Chapter 4 A Study of Palmprint Images

Hand-based biometrics is always prevailing since they generally have relatively high user acceptance. Fingerprint and hand geometry are two well studied hand-based approaches (Baltscheffsky and Anderson 1986; Bhanu and Tan 2003; Clancy et al. 2003; Halici and Ongun 1996; Hong et al. 1998; Jain et al. 1997, 1999, 2000b; Lee and Wang 1999; Miller 1994; Ratha et al. 1996; Sanchez-Reillo et al. 2000). Current fingerprint systems cannot support real-time application in large databases in the absence of special computation assistance units (Ratha et al. 1996; Jain et al. 1997). They also cannot handle some fingerprints provided by, e.g. subjects with genetic problems (Jain et al. 1999). Hand geometry cannot provide high accuracy because geometric features, such as length and width of fingers, are of limited discriminating power.

Palmprint is emerging as an alternative hand-based biometrics (Zhang 2000, 2002; Zhang et al. 2003) with user friendliness, flexibility in adapting the environment and power of discrimination. Table 4.1 compares, in terms of public acceptance, accuracy and feature complexity, some biometrics that can be collected from hands, such as fingerprints, hand geometry and palmprints.

Table 4.1 Comparison of biometrics obtained from hands

	Public Acceptance	**Accuracy**	**Features**
Fingerprint	Medium	High	Minutiae Points
Hand Geometry	High	Low	3-D Geometry
Palmprint	High	High	Line & texture

4.1 Introduction to Palmprint Retrieval

A palmprint is made up of principal lines, wrinkles and ridges on a palm. Like fingerprints, palmprints can be used as a powerful means in law enforcement for criminal identification because of its stability and uniqueness (Zhang 2000, 2003).

Palmprint and fingerprint patterns appear to resemble each other in some ways. Both consist of a large amount of ridges. Although the minutiae based matching which utilizes terminations and bifurcations of the ridges is successful for fingerprint recognition, such an approach is not suitable for palmprint patterns because of the change in orientations.

Palmprint capture and retrieval can be distinguished as *on-line* and *off-line*. An on-line system (e.g. Zhang et al. 2003) equips a palmprint capture sensor connected to the system that takes the palmprint image in a real-time fashion. An off-line palmprint system (e.g. You et al. 2002) acquires palmprint images through digitizing inked palmprints images.

Several types of image-based techniques have been proposed for palmprint representation. They include point-based (Zhang and Shu 1999; Duta et al. 2001), line-based (Duta et al. 2001; Han et al. 2002) and texture-based (Zhang et al. 2003).

Principal lines, wrinkles, ridges, minutiae points, singular points and texture are regarded as useful features for palmprint representation (Shu and Zhang 1998). Various features can be extracted at different image resolutions. For features such as minutiae points, ridges and singular points, a high-resolution image, with at least 400 dpi (dots per inch), is required for feature extraction (Shi et al. 2001). However, features like principal lines and wrinkles, which are defined in Figure 4.1, can be obtained from a low-resolution palmprint image with less than 100 dpi (Zhang and Shu 1999; Shu and Zhang 1998).

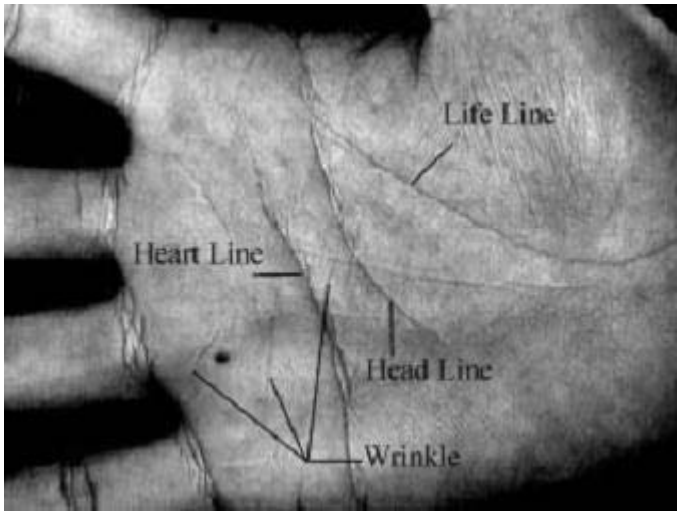


Figure 4.1 Wrinkles and principal lines in a palmprint

In general, high-resolution images are essential for some applications such as law enforcement, where ridges, singular points and minutiae points are extracted and matched in latent prints for identification and verification. Some companies, including NEC and PRINTRAK, have developed automatic palmprint identification and verification systems for law enforcement applications (NEC 2003; Printrak 2003). For civil and commercial applications, low-resolution palmprint images are more suitable than high-resolution images because of their smaller file sizes, which results in shorter computation times during preprocessing and feature extraction. Therefore, they are useful for many real-time palmprint applications.

4.2 Palmprint Database

We use the palmprint database constructed by Biometric Research Centre, The Hong Kong Polytechnic University. The database contains inkless digital palmprint images collected from 193 individuals using a CCD-based palmprint camera and an A/D converter. The resolution of images is 75 dpi.

The subjects are mainly volunteers from The Hong Kong Polytechnic University, i.e. students and staff. In the database, 131 of the subjects are male and 62 are female.

The age distribution of the subjects is as follows: about 86% of them are between 18 to 30 years old, 3% are older than 50 and the remainders are between 30 and 50.

The database contains palmprint images of two sessions collected at an interval of around two months. For each subject, about 10 images of each palm/hand have been collected in each of the two sessions. In total, the database contained 7,752 palmprint images from 386 palms of 193 individuals.

4.2.1 Palmprint preprocessing

Usually the central part of the palm is of interest and will be used for further processing. Images are preprocessed to crop out the area of interest. This is done through five steps with reference to those gaps between fingers depicted in Kong (2002; also in Zhang et al. 2003) and recapitulated as follows.

Step 1: Apply a lowpass filter, $L(u, v)$, such as Gaussian smoothing, to the original image, $O(x, y)$, see Figure 4.2 (a). A threshold, T_p, is used to convert the convolved image to a binary image, $B(x, y)$, see Figure 4.2(b).

Step 2: Obtain the boundaries of the gaps, $(F_i x_j, F_i y_j)$ (i=1, 2), between the fingers using a boundary tracking algorithm (see Figure 4.2(c)). The boundary of the gap between the ring and middle fingers is not extracted since it is not useful in the subsequent processing.

Step 3: Compute the tangent of the two gaps. Let (x_1, y_1) and (x_2, y_2) be any points on $(F_1 x_j, F_1 y_j)$ and $(F_2 x_j, F_2 y_j)$, respectively. If the line $(y = mx + c)$ passing though these two points satisfies the inequality, $F_i y_j \leq m F_i x_j + c$, for all i and j (see Figure 4.2(d)), then the line $(y = mx + c)$ is considered to be the tangent of the two gaps.

Step 4: Line up (x_1, y_1) and (x_2, y_2) to get the Y-axis of the palmprint coordinate system, and use a line passing through the midpoint of these two points, which

is perpendicular to the *Y*-axis, to determine the origin of the coordinate system (see Figure 4.2(d)).

Step 5: Extract fixed size area of interest based on the coordinate system. The area of interest is located within the palmprint image (see Figure. 4.2(e)). Then the palmprint subimage is cropped for feature extraction (see Figure. 4.2(f)).

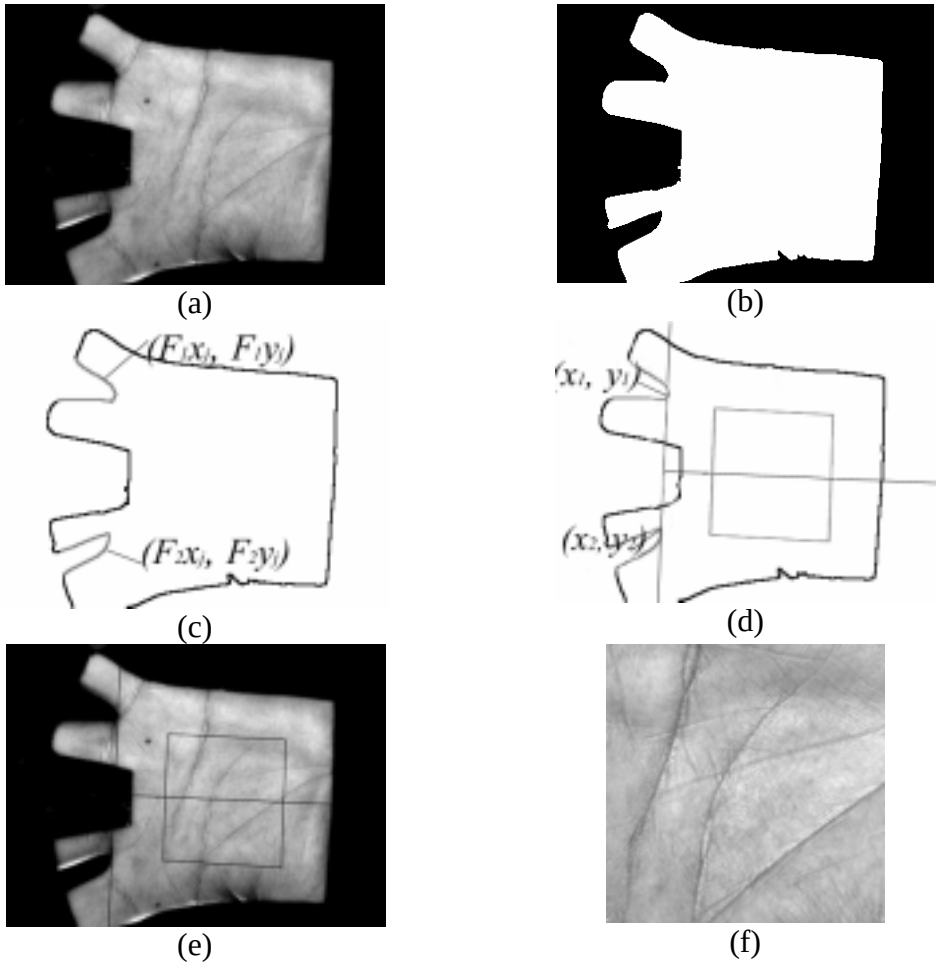


Figure 4.2 Preprocessing of palmprint images

4.3 A PCA + SOM Approach

We have preprocessed monochrome palmprint images of size *s*×*s* as input. In the training phase, as the training set first undergoes PCA (analysis), which is a noise-sensitive process, we have set a threshold to filter noisy images out of the candidate training samples to form the training set. The training set will form a

matrix T of dimension $s^2 \times M$ (M is the number of images in the training set). Each training image in the training set is deformed column-wisely into a column vector v of size $s^2 \times 1$ of T. T will then undergo PCA (see Chapter 3.2.2). We do this in our proposed method for two reasons. One is to generate feature values: coefficients of chosen Principal Components are used as global line and texture features to represent images; another is to perform dimensionality reduction, or more commonly referred to as feature selection. The system flow of our proposed PCA + SOM system is illustrated in Figure 4.3

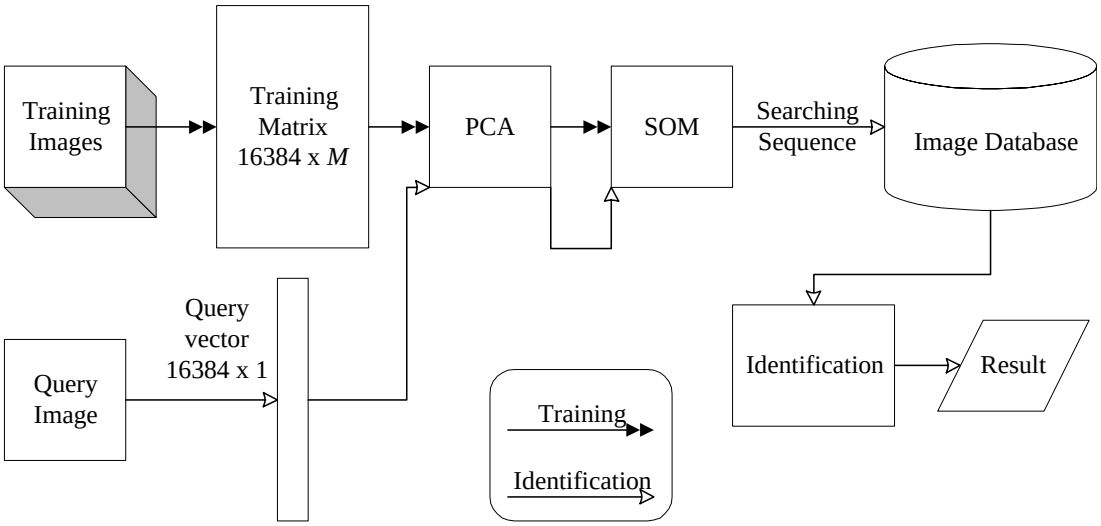


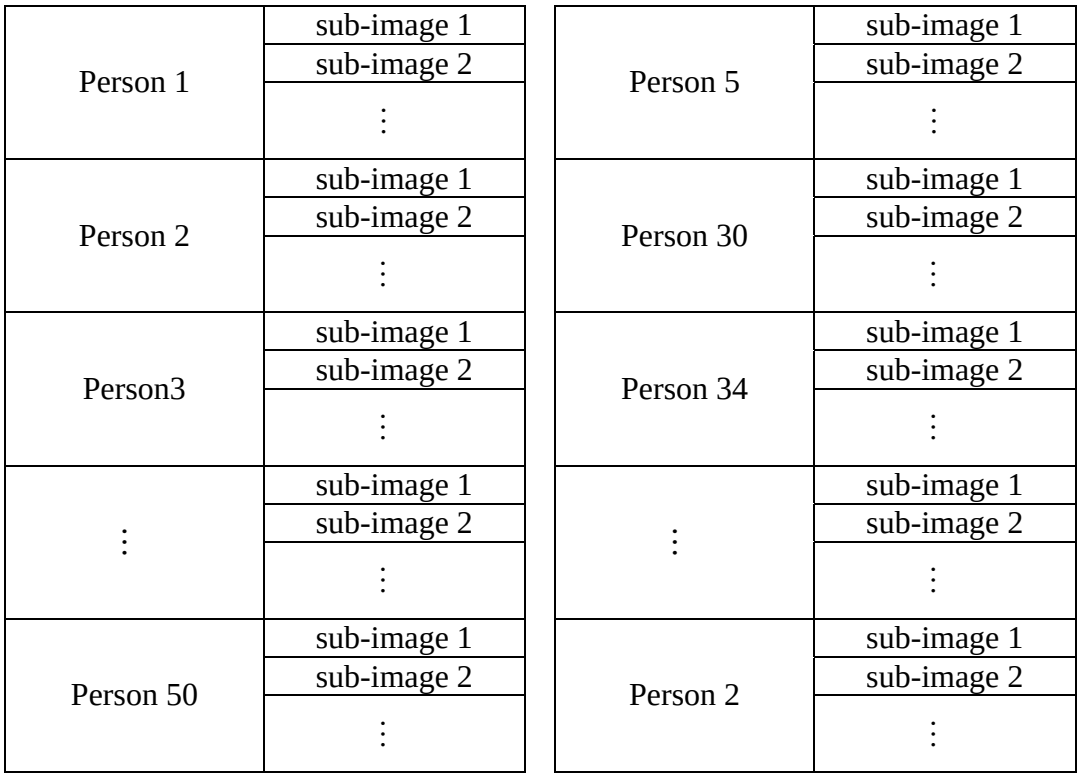
Figure 4.3 A system flowchart of PCA + SOM palmprint image retrieval system

Assume the first m Principal Components are chosen. By projecting each sample image in the training set onto the space spanned by the first m Principal Components, we obtained the m coefficients for each sample. They are then used as the training data to train the SOM. After training, SOM is calibrated and majority voting mechanism may be employed to resolve conflicts.

In the retrieval phase, query image is projected onto the principal subspace. The principal subspace coefficients obtained are passed into the trained SOM to generate a search sequence that guides the search of the Image Database. The trained SOM is

used as the engine to guide the searching by arranging, according to the query input for identification, the order of searching.

As an illustration, suppose the query input for identification is from Person 30 and the winning node of the SOM is the one containing Persons 5, 30, and 34. In Figure 4.4, the one on the left is the sequential searching sequence while the one on the right is generated by our PCA + SOM method that presents earlier to the identification engine the potential matching images in correspondence to the input.



Sequential Search PCA+SOM approach

Figure 4.4 Searching sequence generated by Sequential Search and PCA+SOM approach

4.3.1 Details of Experiment

Palmprint images of 50 different people are used in our experiment. Each subject has registered 10 palmprint images of his/her *left* hand by putting the hand on the palmprint capturing device. The image is then preprocessed to a size of 128×128 (see

Section 4.2.1 above for details); each subject has registered twice on two different dates (Zhang et al. 2003); therefore, there are 1,000 images in the database for this experiment. Three images of each set of images are selected as candidate training samples (300 images) and the remainders are used as the testing set.

Table 4.2 The accumulated variance explained by first *m* principal components

Feature Subspace Dimension (*m*)	**Accumulated Variance (%)**
5	99.42%
10	99.55%
20	99.68%
30	99.75%
40	99.79%
50	99.82%

According to Table 4.2, Principal Components of 5 dimensions can already preserve 99.42% of the original energy. However, the increase of the number of Principal Components used does not help the increase of Energy Preservation much, only 0.13%, 0.26%, 0.33%, 0.37% and 0.4% for the increase of 5, 15, 25, 35 and 45 dimensions used respectively. Thus, we choose to use 10 dimensions, which can preserve more than 99.5% of energy with a smaller number of dimensions.

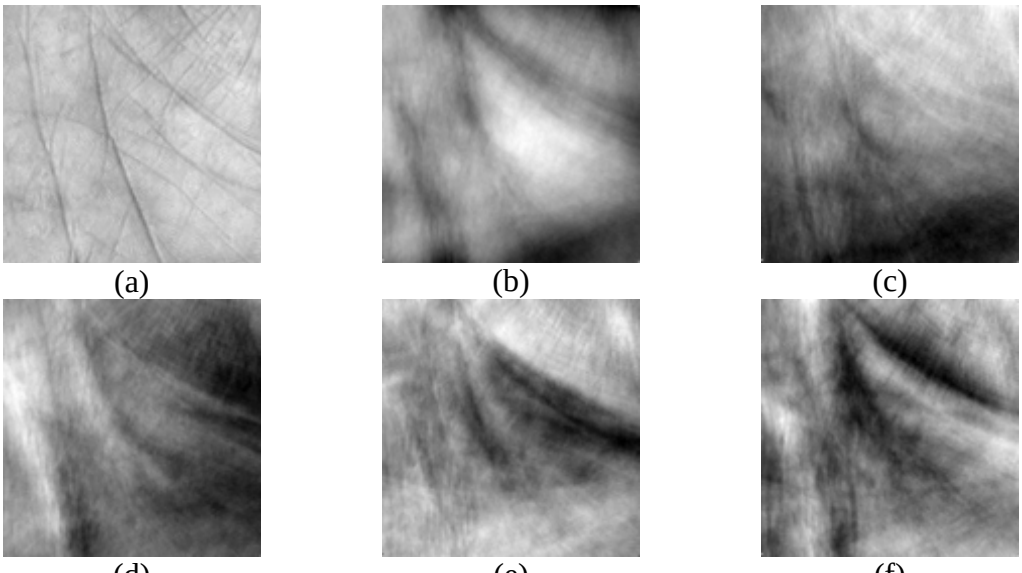


Figure 4.5 (a) a sample left hand sub-image in Palmprint Database, (b)–(f) first 5 Principal Components acquired after PCA

Figure 4.5 (a) shows a left hand palmprint sub-image from the Palmprint Database and Figure 4.5 (b)–(f) show respectively the first 5 Principal Components resulted from the PCA. It can be observed that the first Principal Component has captured the information of the three principal lines and the other Principal Components have captured texture information of various parts of the palm.

Table 4.3 SOM Training Parameters

	Ordering Phase	**Tuning Phase**
Learning Rate	0.1	0.01
Size of Neighborhood	ALL	1

Since there are at most 50 categories (50 different people), we choose SOMs of sizes 3×3 and 5×5 for experiments. SOMs of all sizes are trained for 3,000, 5,000 and 10,000 epochs respectively and the training parameters are shown in Table 4.3.

4.3.2 Experimental Results and Performance

Table 4.4 Average number of persons and images searched for PCA+SOM (3 Trails of 2 Sizes and 3 Training Times) and Sequential approaches

PCA+SOM approach

		Training Time (epochs)					
		3000		**5000**		**10000**	
Trails	**SOM size**	**Person**	**Image**	**Person**	**Image**	**Person**	**Image**
Trail 1	**3×3**	12.1167	67.2833	11.5233	63.97	11.7467	65.2467
	5×5	10.2133	56.7867	10.0033	55.3233	9.9267	54.84
Trail 2	**3×3**	11.9967	66.6633	11.6067	64.43	11.39	63.1933
	5×5	10.4	55.44	9.8233	54.72	10.0767	55.6067
Trail 3	**3×3**	11.4733	63.6733	11.29	63.04	11.3267	62.9867

Sequential approach
Total number of persons in Training Set (P) = 50
Average number of persons searched for Sequential Searching = $P/2$ = 25
Total number of images in Training Set (M) = 280
Average number of images searched for Sequential Searching = $M/2$ = 140

Total number of images in training set, i.e. discarding noisy ones, is 280. Therefore, the average number of persons searched for Sequential Searching is equal to half of the size of training set (50), i.e. 25 and the average number of images searched for

Sequential Searching is equal to half of the size of training set (280), i.e. 140.Our PCA + SOM method, under all experimental conditions: 3 trails of 2 SOM sizes and 3 different training times for each network size, performs much better than the sequential search by reducing the search space to 25%–30% of the original space (see Table 4.4).

Table 4.5 depicted distributions of persons and images over the 3 trails of SOM of sizes 3×3.

Table 4.5 Node distribution for 3×3 SOM: No. of Person and Image classified under each node

		Training Time (epochs)					
		3000		**5000**		**10000**	
Trails	**Node**	**Person**	**Image**	**Person**	**Image**	**Person**	**Image**
Trail 1	**1**	7	41	8	39	8	47
	2	7	41	5	27	6	33
	3	6	36	8	47	9	45
	4	5	27	4	24	5	30
	5	4	20	5	26	4	20
	6	3	16	4	24	5	30
	7	9	45	3	18	7	41
	8	5	30	5	28	3	16
	9	4	24	8	47	3	18
Trail 2	**1**	6	36	9	45	7	42
	2	8	47	4	24	6	34
	3	8	47	2	12	2	12
	4	5	27	4	21	6	35
	5	4	20	5	26	5	26
	6	4	22	7	40	5	30
	7	8	39	6	36	6	36
	8	4	24	7	41	4	21
	9	3	18	6	35	9	44
Trail 3	**1**	8	39	9	52	6	36
	2	5	27	6	36	5	28
	3	8	47	6	36	1	6
	4	4	24	4	21	5	30
	5	5	26	7	42	7	38
	6	4	24	5	28	6	36
	7	3	18	7	30	5	30
	8	5	28	5	29	6	32
	9	8	47	1	6	9	44

Table 4.6 depicted distributions of persons and images over the 3 trails of SOM of sizes 5×5.

Table 4.6 Node distribution for 5×5 SOM: No. of Person and Image classified under each node (a) Trail 1, (b) Trail 2 and (c) Trail 3

(a)

		Training Time (epochs)					
		3000		**5000**		**10000**	
Trails	**Node**	**Person**	**Image**	**Person**	**Image**	**Person**	**Image**
Trail 1	**1**	4	24	5	30	5	30
	2	2	12	3	17	0	0
	3	1	6	3	18	1	6
	4	3	17	1	2	3	18
	5	3	10	3	10	2	12
	6	0	0	0	0	3	18
	7	2	12	3	18	1	6
	8	0	0	2	12	3	17
	9	2	9	3	14	1	6
	10	4	23	1	6	2	11
	11	2	12	2	12	1	6
	12	3	17	1	6	1	6
	13	3	14	0	0	4	20
	14	1	6	2	12	3	18
	15	2	12	4	24	4	22
	16	4	24	3	18	3	14
	17	0	0	2	12	1	6
	18	3	18	1	6	1	6
	19	2	12	2	12	1	6
	20	0	0	0	0	1	6
	21	1	6	2	11	4	16
	22	2	12	1	6	2	12
	23	3	16	2	10	2	12
	24	0	0	1	6	0	0
	25	3	18	3	18	1	6

(b)

		Training Time (epochs)					
		3000		5000		10000	
Trails	**Node**	**Person**	**Image**	**Person**	**Image**	**Person**	**Image**
Trail 2	**1**	5	30	4	24	4	24
	2	1	6	0	0	1	6
	3	1	6	2	12	1	6
	4	3	18	3	18	2	12
	5	2	12	1	6	1	6
	6	3	17	2	12	2	12
	7	1	6	1	6	2	12
	8	2	12	3	17	3	17
	9	0	0	1	6	1	6
	10	2	12	4	23	2	12
	11	3	18	1	6	4	24
	12	0	0	2	12	0	0
	13	3	14	1	6	3	14
	14	3	18	0	0	3	18
	15	4	22	4	22	2	10
	16	2	8	4	15	2	12
	17	2	7	5	26	3	14
	18	2	12	1	6	1	6
	19	3	18	1	6	2	12
	20	0	0	1	6	0	0
	21	3	14	2	12	4	16
	22	2	12	2	9	3	17
	23	2	12	3	18	2	12
	24	0	0	0	0	0	0
	25	1	6	2	12	2	12

(c)

		Training Time (epochs)					
		3000		5000		10000	
Trails	**Node**	**Person**	**Image**	**Person**	**Image**	**Person**	**Image**
Trail 3	**1**	5	30	6	36	2	12
	2	0	0	0	0	1	6
	3	1	6	1	6	2	12
	4	4	24	4	24	4	24
	5	2	12	1	6	1	6
	6	1	6	2	12	1	6
	7	3	18	1	6	2	12
	8	3	17	3	17	3	17
	9	1	6	0	0	1	6
	10	1	6	3	17	1	6
	11	1	6	3	18	3	18
	12	2	9	0	0	1	6
	13	2	8	3	14	3	14
	14	3	18	1	6	2	12
	15	4	22	3	16	3	16
	16	2	11	4	20	3	15
	17	0	0	0	0	2	11
	18	2	12	2	12	1	6
	19	2	12	1	6	2	12
	20	0	0	1	6	0	0
	21	3	10	5	22	3	10
	22	2	11	2	12	4	23
	23	3	18	2	12	3	18
	24	0	0	0	0	0	0
	25	3	18	2	12	2	12

Table 4.7 recorded the minimum, maximum and mean search depth, which is in terms of node visited, of all experimental conditions.

Table 4.7 Minimum, Maximum and Mean Search Depth (in terms of Node) for 3 Trails of 2 Sizes and 3 Training Times

		Training Time (epochs)								
		3000			5000			10000		
Trails	**SOM size**	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
Trail 1	**3×3**	1	5	2.6033	1	5	2.4933	1	5	2.5233
	5×5	1	12.5	5.3167	1	13.6667	5.6	1	12.5	5.3167
Trail 2	**3×3**	1	5	2.53	1	5	2.48	1	5	2.4433
	5×5	1	12.5	5.3633	1	13	5.3	1	13.6667	5.3467
Trail	**3×3**	1	5	2.4833	1	5	2.4033	1	5	2.3733

It is observed that the more evenly the persons/classes are distributed over the nodes of SOM, the better the retrieval performance. Adequate nodes and training for SOM may make the SOM more evenly spread over the input space/training samples and thus improve the retrieval performance.

4.4 A Hierarchical Approach

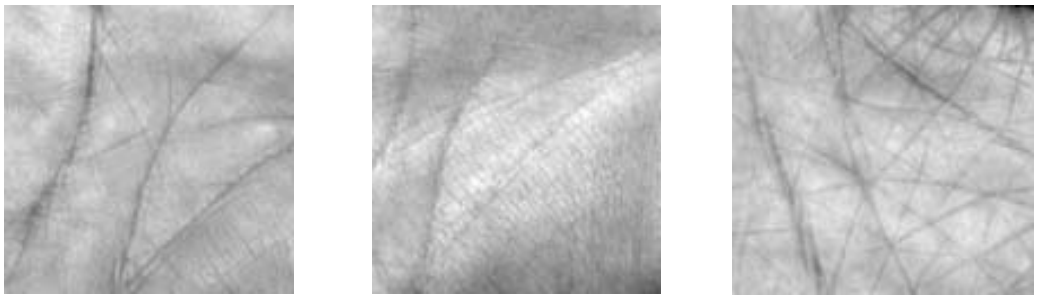
We propose a hierarchical palmprint coding scheme to facilitate coarse-to-fine matching for effectively and efficiently identifying a palmprint in a large database. More specifically, we extract different palmprint features at different levels.

Level-1. Global geometry based key point distance;

Level–2. Global texture energy;

Level–3. Fuzzy “interest’’ line;

Level–4. Local directional texture energy vector.



(a) (b) (c)

Figure 4.6 Palmprint samples of distinctive texture differences (a) strong principal lines (b) less wrinkle (c) strong wrinkle

We start with the global geometry feature to localize the region of interest of palmprint sample at coarse level and apply a distance measurement of the palm boundary to guide the dynamic selection of a small set of similar candidates from the database for further processing. We also use the global texture energy (GTE) for fast search for the best match. Such a mask-based texture feature representation is characterized with high convergence of inner-palm similarities and good dispersion of inter-palm discrimination. We then adopt fuzzy set theory to detect “interest”

feature lines to guide the search for the best match at fine level. Finally, we apply local texture measurement to establish a feature vector for palmprint matching. Figure 4.7 illustrates the general structure of our system.

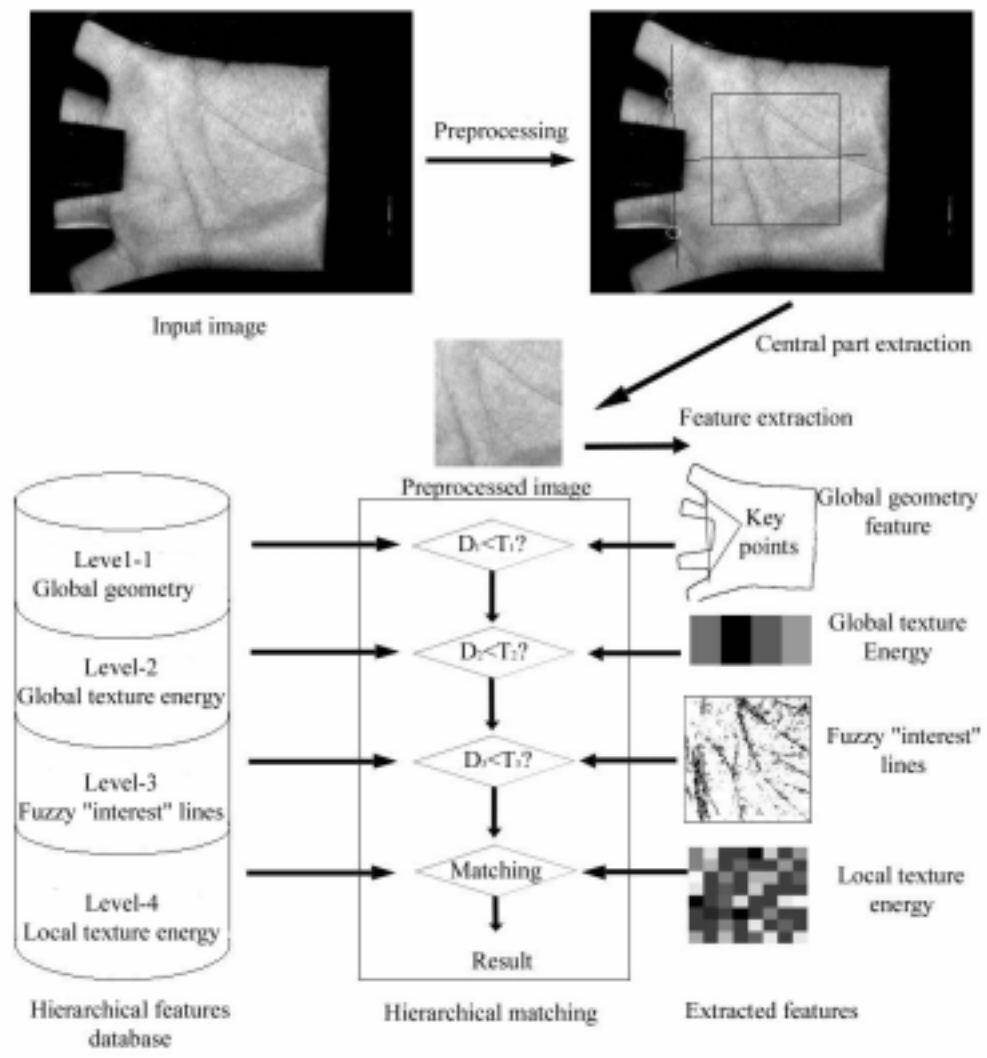


Figure 4.7 System diagram of hierarchical palmprint system

4.4.1 Level-1: Global geometry based key point distance

As detailed in 4.2.1, the points (x_1, y_1) and (x_2, y_2) defined in the preprocessing of palmprint images are regarded as key points and the distance (d) between the points is used as Level-1 feature. The similarity measure used at this level is as follows.

$$D_1 = \left| d_i - d_j \right| \tag{4.1}$$

4.4.2 Level–2: Global texture energy (*GTE*)

Global texture energy constitutes the averages of each of the four directional texture energies (*TE*). Directional *TE* at position (*i*, *j*) is obtained from summing, within a test window of size $W_x \times W_y$ centered at (*i*, *j*), the square of values after convolution of preprocessed palmprint images (*I*) against the corresponding directional "tuned" mask (A_k) (You and Cohen 1993; Laws 1980) and then normalizing with $P^2 = \sum (A_k)^2$.

$$TE(i, j) = \frac{\sum_{W_x} \sum_{W_y} (I * A_k)_{rs}^2}{P^2 W_x W_y} \tag{4.2}$$

Directional *TE* approximates the local variance after convolution. *GTE* is insensitive to noise and shifting, easy to compute and good for discrimination among different classes. The four directional "tuned" masks, horizontal, vertical, 45° and -45°, shown in Figure 4.8, are zero sum and of size 5×5.

-1	-2	-4	-2	-1
0	0	0	0	0
2	4	8	4	2
0	0	0	0	0
-1	-2	-4	-2	-1

(a)

-1	0	2	0	-1
-2	0	4	0	-2
-4	0	8	0	-4
-2	0	4	0	-2
-1	0	2	0	-1

(b)

0	-1	-4	0	2
-1	-6	0	8	0
-4	0	12	0	-4
0	8	0	-6	-1
2	0	-4	-1	0

(c)

2	0	-4	-1	0
0	8	0	-6	-1
-4	0	12	0	-4
-1	-6	0	8	0
0	-1	-4	0	2

(d)

Figure 4.8 Four directional "tuned" masks for global texture energy extraction. (a) Horizontal (b) Vertical, (c) 45° and (d) -45°

The similarity measure used at this level is as follows.

$$D_2 = \sum_{k=0}^{3} \left| v_{ki} - v_{kj} \right| \tag{4.3}$$

Figure 4.9 depicts the distribution of GTE of 10 palms, in total 80 images. It can be seen that GTE is an effective tool to distinguish similar palms from dissimilar ones. Two sets of images from similar palms are shown in Figure 4.10.

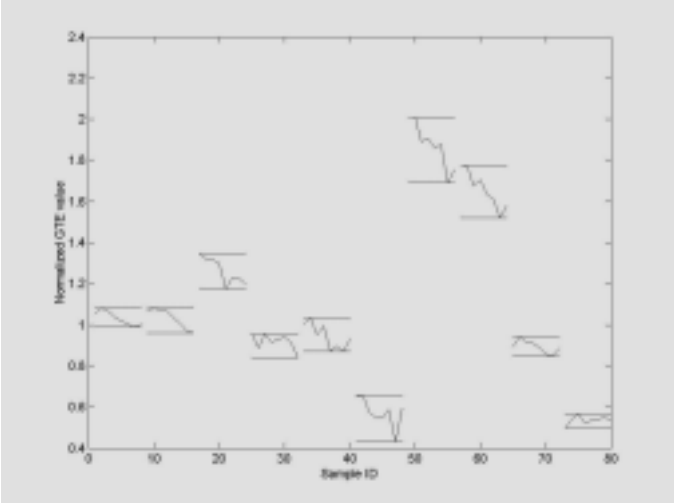
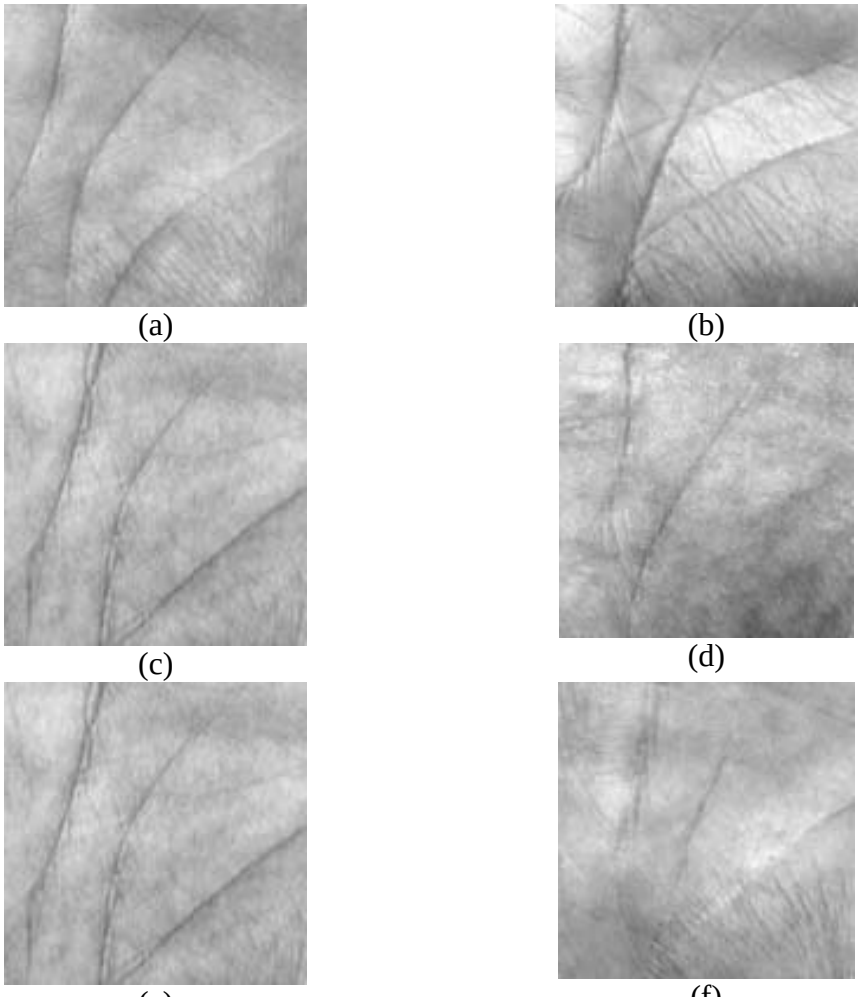


Figure 4.9 Comparison of palmprint GTE distribution (80 palmprint images of 10 palms): inter-palm dispersion vs. inner-palm convergence.



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

Figure 4.10 Two sets of similar palmprints. 1st Column (a) (c) and (e) Set one and 2nd Column (b), (d) and (f) Set two

4.4.3 Level–3: Fuzzy “interest” line

Dominant feature lines in palmprint, e.g. principal lines, wrinkles, are extracted as follows:

Step 1. Convert the pre-processed images into a feature image as described in You and Bhattacharya (2000).

Step 2. Apply the fuzzy rule to extract the “interest” lines. Let I_T be a feature image. The fuzzy output is given by the following piecewise linear membership function with $a = \mu$ and $b = \mu + \sigma$, where μ and σ are respectively the sample mean and standard deviation of all pixels in I_T.

$$S(l) = \begin{cases} 0 & if \quad l < a \\ (l-a)/(b-a) & if \quad a \le l \le b \\ 1 & if \quad b > 1 \end{cases} \tag{4.4}$$

A sample of fuzzy “interest” lines is shown in Figure 4.11.

(a) (b)

Figure 4.11 The detection of “interest” lines based on fuzzy theory.
(a) Original image and (b) its “interest” lines

Step 3. Take the mean of the fuzzy output in a window of size $M \times N$ to represent the local interest lines such as

$$v(I_T) = \frac{1}{MN} \sum_x \sum_y S(I_T(x, y)) \tag{4.5}$$

In our experiments, both M and N are set to 20. Thus, the “interest” lines can be represented by a 64 dimension feature vector obtained from 64 overlapped blocks. The similarity measure used at this level measuring the distance between the feature vectors X and Y is an angular distance as follows.

$$D_3 = \frac{X^T Y}{\|X\|\|Y\|} \tag{4.6}$$

4.4.4 Level–4: Local directional texture energy vector

In achieving high accuracy, a long feature vector formed from local directional texture energy is utilized. The local directional texture energy within a window of size $X \times Y$ is defined as follows.

$$u_i = \frac{1}{XY} \sum_x \sum_y \left| \sum_{W_X} \sum_{W_Y} (I * A_i)_{rs} \right| \tag{4.7}$$

The size of the window is set to 20×20 that strike the balance between the size of a feature vector and the accuracy of the system. Therefore, each block is overlapped with the adjacent blocks. The similarity measure used at this level measuring the distance between the feature vectors $\{y_{ki}\}$ and $\{y_{kj}\}$ is a local angular distance as follows.

$$D_4 = \frac{1}{64} \sum_{k=1}^{64} \frac{y_{ki} y_{kj}^T}{\|y_{ki}\|\|y_{kj}\|} \tag{4.8}$$

where $y_{ki} = \{u_c\}$ and u_c is defined in Eq. 4.7 with $k = \{1,\ldots, 64\}$ and $c = \{0,\ldots, 3\}$

4.4.5 Experimental Results and Performance

Due to some inappropriately placed palms at the stage of acquisition, only 5437 out of 7,752 palmprint images are used for the results reported below. In Figure 4.12, the effectiveness of the four level features is reported in the form of genuine and impostor distributions and, ROC curves. From Figure 4.12, we can observe that the discriminative power increases from Level-1 feature to Level-4 feature.

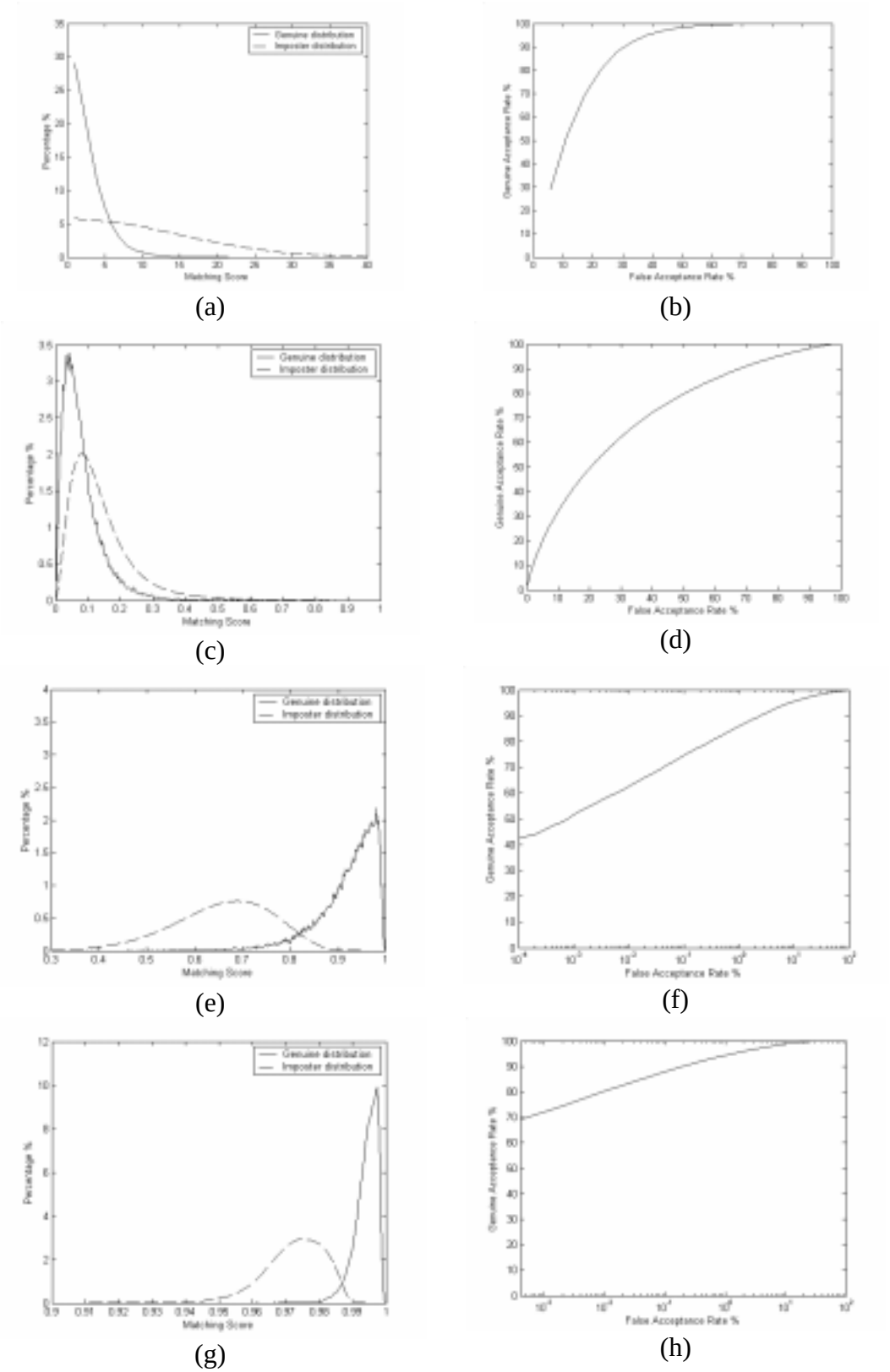


Figure 4.12 Performance of Level-1 to Level-4 features. (a) (c) (e) and (g) the genuine and imposter distributions and (b) (d) (f) and (h) ROC curves of the four level features respectively

We then have to determine the thresholds at each of the first three levels and evaluate the performance of the hierarchical approach. Three sets of tests, Hierarchical 1, 2 and 3, are formed with different thresholds selected for each of the first three levels, T_1 to T_3 respectively (see Table 4.8 for details).

Table 4.8 Thresholds for tests of hierarchical approach

	Hierarchical 1	**Hierarchical 2**	**Hierarchical 3**
T_1	20	23	25
T_2	0.0194	0.0216	0.0250
T_3	0.764	0.73	0.68

Retaining more candidates after the first two levels while restricting more at Level-3, the performance (of Hierarchical 3) is found to be the best (lower false and correct rejection rate) among tested (see Table 4.9 for details).

Table 4.9 Retrieval performance of tests of hierarchical approach

	Hierarchical 1	**Hierarchical 2**	**Hierarchical 3**
False Rejection Rate	6.13%	3.89%	1.98%
Correct Rejection Rate	88.23%	78.19%	60.51%

The performance of the three tests with comparison to using Level-4 feature only, i.e. without hierarchical approach, is plotted as ROC curves in Figure 4.13. We can observe the performance of using Level-4 feature only is significantly better than those with hierarchical approach only when the FAR is large, e.g. 1%. They perform more or less the same at the usual biometric system operating points, i.e. lower FAR.

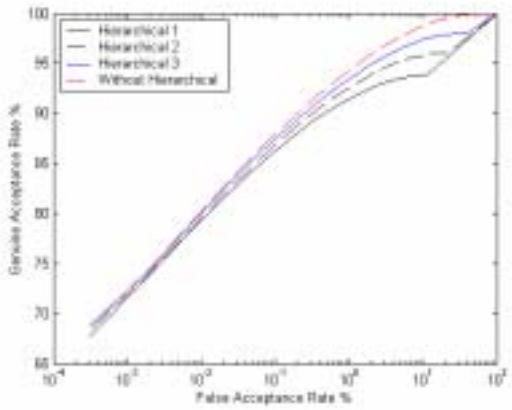


Figure 4.13 The ROC curves of hierarchical approach with different parameters

Our hierarchical method was tested on an embedded Intel Pentium III 500MHz processor PC and implemented with Visual C++ 6.0. The measured execution times of preprocessing, feature extraction, Level-1 to Level-4 matching for the database under test are presented in Table 4.10. The total execution time was around 0.75*s*.

Table 4.10 Execution time of each procedures considered

Operations	**Time used (ms)**
Preprocessing	538
Feature Extraction	215
Level-1 matching	4.7×10^{-5}
Level-2 matching	3.4×10^{-4}
Level-3 matching	0.009
Level-4 matching	0.059

Based on the execution times measured, we can model the execution time of a large database as follows. Let D be the size of the database and P_1, P_2 and P_3 be the percentages of the total number of discarded palmprint samples in the database respectively after Level-1, Level-2 and Level-3 matching, which are controlled by the thresholds T_1, T_2 and T_3 accordingly. Let S_p, S_f, and, S_1, S_2, S_3 and S_4 respectively be the execution time for preprocessing, feature extraction and, Level-1, Level-2, Level-3 and Level-4 matching. For a sequential search in a palmprint image database, the total execution time required (T_S) is

$$T_S = S_p + S_f + D \times S_4 \tag{4.9}$$

For searching the same palmprint image database with hierarchical approach, the total execution time required (T_H) is

$$\begin{aligned} T_H = \; & S_p + S_f + D \times S_1 + D \times (1\text{-}P_1) \times S_2 + D \times (1\text{-}P_1\text{-}P_2) \times S_3 \\ & + D \times (1\text{-}P_1\text{-}P_2\text{-}P_3) \times S_4 \end{aligned} \tag{4.10}$$

The total execution time required (T_S for sequential approach and T_H for hierarchical approach) is plotted against the size of the database D in Figure 4.14. The thresholds are set as listed in Table 4.8.

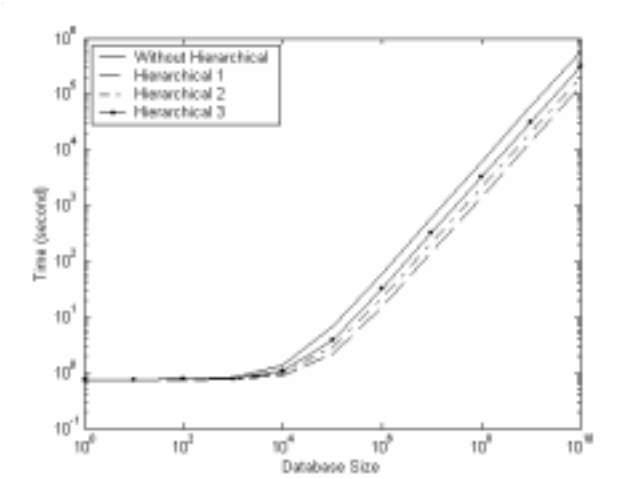


Figure 4.14 Computation time of hierarchical approach for large database

The modeled execution time of our hierarchical approach for a large database with 10^5 palmprint images is 2.8s while the traditional sequential approach requires 6.7s with EER of 4.5%. It is obvious that the hierarchical approach is much more effective than sequential method when the database is large.

A comparison of different palmprint retrieval systems is summarized in Table 4.11

Table 4.11 Comparison of different palmprint retrieval systems

	Feature Point Based Matching (Duta et al. 2001)	**Line based Matching (Zhang and Shu 1999)**	**Hierarchical Approach**
Database Size	30 samples	200 samples	7,752 samples
Feature Extraction	Feature points (single feature)	Lines (single feature)	Texture, geometry and lines (multiple features)
Matching Criteria	Distance measurement (fixed measurement)	Euclidian distance (fixed measurement)	Energy difference, angular difference (flexible measurement)
Search Method	1-to-1 comparison (sequential)	1-to-1 comparison (sequential)	Guided search (hierarchical)
Retrieval Accuracy	good	limited	good

4.5 Recapitulation

In this chapter, we have, first, given an introduction to palmprint and issues related to palmprint retrieval. A comprehensive description of one of the largest palmprint

databases available and its preprocessing operations are then given. We have presented two proposed palmprint image retrieval methods which have been tested on the described palmprint database. The PCA+SOM method addresses the inefficiency of sequential search which is a usual practice in biometric systems. The hierarchical method proposed is an initiative to deal with palmprint retrieval in large databases.

Chapter 5 A Study of Face Images

Many automated systems need to confirm the identity of an individual, whom is requesting services, in a reliable manner because the rendered service should only be accessed by a legitimate user. Biometric recognition, or simply Biometrics (Miller 1994), is concerned with the automatic recognition of individuals based on their biological/physiological or behavioral characteristics. Although no single biometric is expected to satisfy all identification requirements, the use of such unique, reliable and stable personal features has attracted considerable interest in the development of biometrics identification systems for civilian, military, and forensic applications. They have attained certain level of maturity. (Jain et al. 2004)

Among many body characteristics that have been used, face is one of the most commonly used characteristics (Jain et al. 1999, 2004; Zhang 2000, 2002) and has been studied (Heisele et al. 2003; Liu and Wechsler 2003; Wu et al. 2002; Zhang et al. 2002) across a number of research fields, e.g. computer vision and pattern recognition.

5.1 Introduction to Face Retrieval

Face recognition is a non-intrusive method that captures still images and/or video sequences, from controlled, static to uncontrolled, cluttered environment; recognition can be performed using 2D images and/or 3D models with Holistic/Global approaches, Feature-based/Structural approaches or Hybrid approaches. (Zhao et al. 2003; Chellappa et al. 1995; Jain et al. 2004)

Holistic/Global approaches generally include those methods that take the whole face region as the input to a face recognition system. (Zhao et al. 2003) One well

known holistic/global approach to face recognition is applying Principal Component Analysis (PCA) on preprocessed, or even transformed, face training images. A subset of principal components forms the global representation of face for recognition with a suitable (dis)similarity measure. (Turk and Pentland 1991; Draper et al. 2003; Martinez and Kak 2001; Liu and Wechsler 2003) Another common holistic/global approach is the introduction of artificial neural networks (Heisele et al. 2003) as the classifier for face recognition.

Feature-based/Structural approaches first locate feature points/regions of interest, such as eyes, noses and mouth. Then features are extracted with the use of various operators or transforms based on their geometric properties or appearance characteristics. (Zhao et al. 2003) Wiskott et al. (1997) is one of the designs using this approach. They locate a set of fiducial points as nodes of the elastic bunch graph. A small set of Gabor jets (i.e. filters responses) from an individual is stored at the relevant nodes of the elastic bunch graph for face recognition.

Hybrid approaches take the global (whole face) and local features into consideration, which is arguably to be potentially the best approach. Heisele et al. (2003) adopt this approach. They allocate each local feature region a Support Vector Machine (SVM, Vapnik 1995) detector to decompose the face into a set of facial components. There are 10 dominant components used. After normalization, they are formed into a single vector which is the input to SVMs for personal identification.

5.2 Face Databases

In this subsection, we will briefly introduce some commonly referred public face databases. We have made used some of those described below in our studies.

5.2.1 The AR database

The AR face database (Martinez and Benavente 1998) from Purdue University is used in our experiments. After careful examination of the database, we found that there are images from 135 subjects (76 men and 59 women) instead of the stated 126. We also found that only 120 subjects (65 men and 55 women) out of the 135 subjects have been taken images in the two sessions. Therefore, we used only those images from the 120 subjects who have been taken photographs in two sessions (with 14 days apart between the two sessions). As there are 13 images are taken in each session, the testing set we used contains 3,120 images (1,560 in each session); each of the 13 images is taken under various conditions: different facial expressions, illumination settings and occlusions (sunglasses and scarf). There is no limitation on the participant's clothes, make-up, ornaments, hair styles, etc. (Martinez and Benavente 1998; Liu and Wechsler 2003)

The AR face database provides still images of two different sessions. It has been used to study the effect of matching "duplicate" (Martinez and Kak 2001). It has also been used to test on the effect of facial expressions, illumination settings and occlusions (Martinez 2002).

5.2.2 The ORL database

The ORL face database (Samaria and Harter 1994), which contains 10 different images for each of 40 distinct subjects, is created by the Olivetti Research Laboratory, Cambridge, UK. For some of the subjects, the images were taken at different times, varying lighting slightly, facial expressions (open/closed eyes, smiling/non-smiling) and facial details (glasses/no-glasses). All the images are taken against a dark homogeneous background and the subjects are in up-right, frontal

position (with tolerance for some side movement). The size of each image is 92×112 of 8-bit grey levels.

The ORL database has been used for face detection and recognition studies. It has been used to test for the effect of pose variations. (Zhao et al. 2003)

5.3 An Aggregated 2D Gabor Features-based Approach

In many commercial/civilian systems, since the environment is static and under control, full elasticity of the automatic face recognition system may not be required, i.e. with cooperative subjects, proper 2D frontal image can be obtained. In such a case, the automatic face recognition problem can be simplified to the classical pattern recognition/image retrieval problem, which deals mainly with feature extraction and identification/verification.

We proposed a holistic approach with lower feature dimension to deal with the transformed face recognition problem; our aggregated Gabor approach uses the mean and standard deviation of Gabor filters responses, which is the result of the convolution of Gabor filters with grayscale face images, as features. Since Gabor filters of 3 scales and 6 orientations are adopted for feature extraction, for the mean and the standard deviation of filter responses, each is of 36 dimensions (18 each for real and imaginary part of filter's responses); that is there are totally 72 dimensions/features. Feature vector of a lower dimensional space can reduce the computation complexity, i.e. increasing speed, as well as the storage required for face recognition process.

Based on Eq. 3.20, Eq. 3.23 and Eq. 3.24, the mean (M) of the real part (Eq. 5.1) and the imaginary (Eq. 5.2) part of filter responses are

$$M_c(\omega,\theta) = \sum_x \sum_y k_c(x, y, \omega, \theta) \tag{5.1}$$

$$M_s(\omega,\theta) = \sum_x \sum_y k_s(x,y,\omega,\theta) \tag{5.2}$$

Correspondingly, the standard deviation (*SD*) of the real part (Eq. 5.3) and the imaginary part (Eq. 5.4) of filter responses are as follows.

$$SD_c(\omega,\theta) = \sum_x \sum_y k_c(x,y,\omega,\theta) \tag{5.3}$$

$$SD_s(\omega,\theta) = \sum_x \sum_y k_s(x,y,\omega,\theta) \tag{5.4}$$

The feature vector (*FV*) of a face image is of the following form

$$\begin{aligned} FV = [&M_c(\omega_1,\theta_1),\ldots, M_c(\omega_3,\theta_6), M_s(\omega_1,\theta_1),\ldots, M_s(\omega_3,\theta_6),\\ &SD_c(\omega_1,\theta_1), ,\ldots, SD_c(\omega_3,\theta_6), SD_s(\omega_1,\theta_1),\ldots, SD_s(\omega_3,\theta_6)] \end{aligned} \tag{5.5}$$

L_1- and L_2-norm are adopted as the (dis)similarity measures to define how similar is one image, in our testing databases, when it is compared to another. Based on the query face image, the N nearest neighbour rule with majority voting scheme helps to determine the best matched individual within the database.

We have converted the color images in The AR database into gray scale images of size 192×144 as images of our testing database. Since there is only one face per image and the face is roughly located at the center of the image, no face detection is performed.

5.3.1 Experimental results

Phase 1

Face subimages for processing are cropped from the center part of the image in our testing database using two strategies: 1) the testing set (64): images in our testing database are first resized from 192×144 to 128×128 using bicubic interpolation. Then, the central canvas, of size 64×64, of each interpolated image is used to extract features; 2) the testing set (96): the central canvas, of size 96×96, of each image in our testing database is cropped for feature extraction. A set of sample images of a subject in the testing database is shown in Figure 5.1, (a)–(z); images taken from 1st session are (a)–(m) and images taken from 1st session are (n)–(z). One sample from

each of the two testing sets is also shown along in Figure 5.1, marked as (64) and (96), for reference; the one marked with (64) is from the 1st session being resized and cropped using strategy one and image marked with (96) is from the 2nd session being cropped using strategy two.

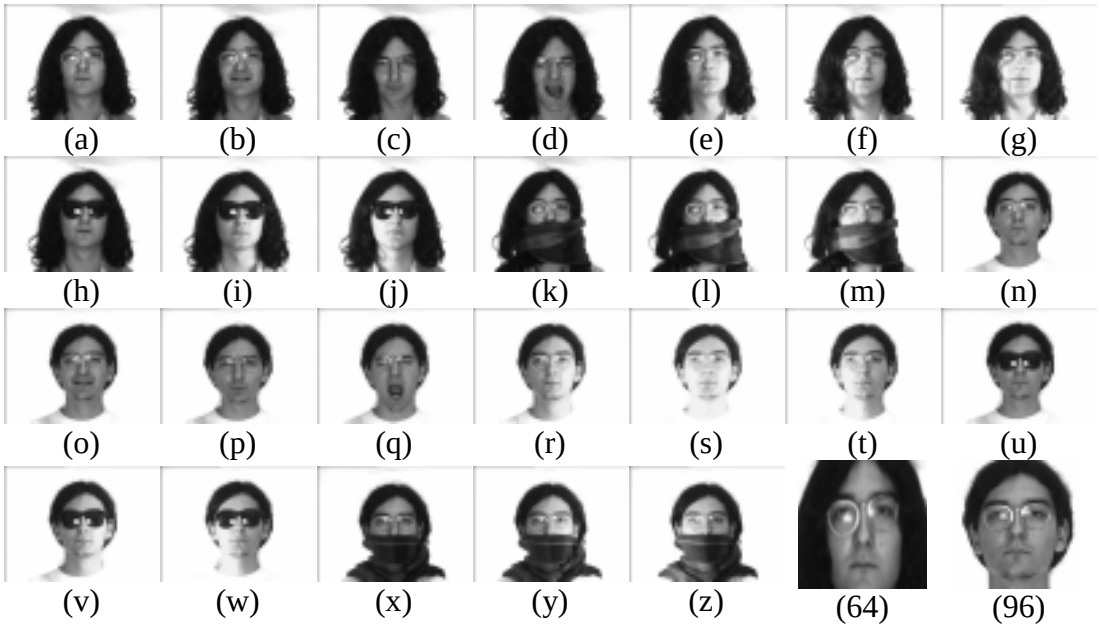


Figure 5.1 A set of images of one subject in the testing database: images taken from 1st session (a)-(m) and 2nd session (n)-(z); image marked (64) is from 1st session being preprocessed to be of size 64×64 as the testing set (64) and image marked (96) is from 2nd session being preprocessed to be of size 96×96 as the testing set (96).

Elliptical Gabor filters of 3 scales and 6 orientations, as derived in Chapter 3.1.3 Eq. 3.21, directly convolve with all images of the two testing sets; the mean and standard deviation of the real and imaginary parts of all filters responses are extracted as features, i.e. in total 72 dimensions (36 each for real and imaginary parts, 18 each for mean and standard deviation of each part of the filter response), to represent each image in our testing database for the first experiment. L_1- and L_2-norm are used to measure (dis)similarity across our experiments.

We experimented with our method under five situations: (1) the whole database, (2) 1st session match against 1st session (S1 vs. S1), (3) 1st session match against 2nd session (S1 vs. S2), (4) 2nd session match against 1st session (S2 vs. S1) and (5) 2nd

session match against 2nd session (S2 vs. S2). For case (1), images of the whole testing database are treated as registered images, i.e. no unseen image is expected. For cases (2) and (5), they are similar to case (1) except that only images of 1st session or 2nd session is used. For cases (3) and (4), however, either 1st or 2nd session is the registered set while the other one is the query set, i.e. matching "duplicates" (Martinez and Kak 2001).

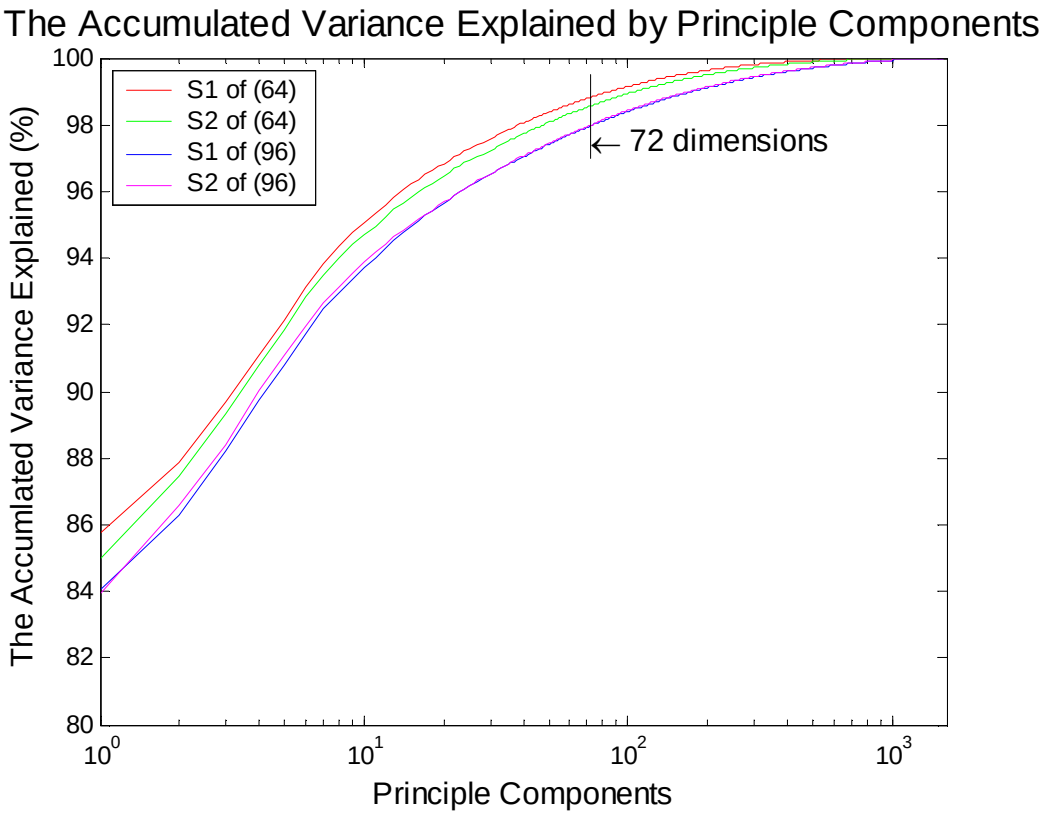


Figure 5.2 The Accumulated Variance Explained by Principle Components

Another test uses the first 72 principle components, which is associated with the largest variances, trained with all images obtained in one session, i.e. either S1 or S2, of the two testing sets (64) and (96). Since we use PCA as a reference for performance comparison, only case (2) to (5) described above are tested such that images in the training sets are from the same session. Moreover, we choose, for PCA, the same feature dimension as that of Aggregated 2D Gabor features because we can determine the effectiveness of aggregated Gabor method directly from the system

performance; we only use one feature dimensionality for experiment as it already captures 97.968%-98.823% of the total variability of the training sets (either session of the two testing sets). The accumulated variances explained by principle components of each training set are shown in Figure 5.2.

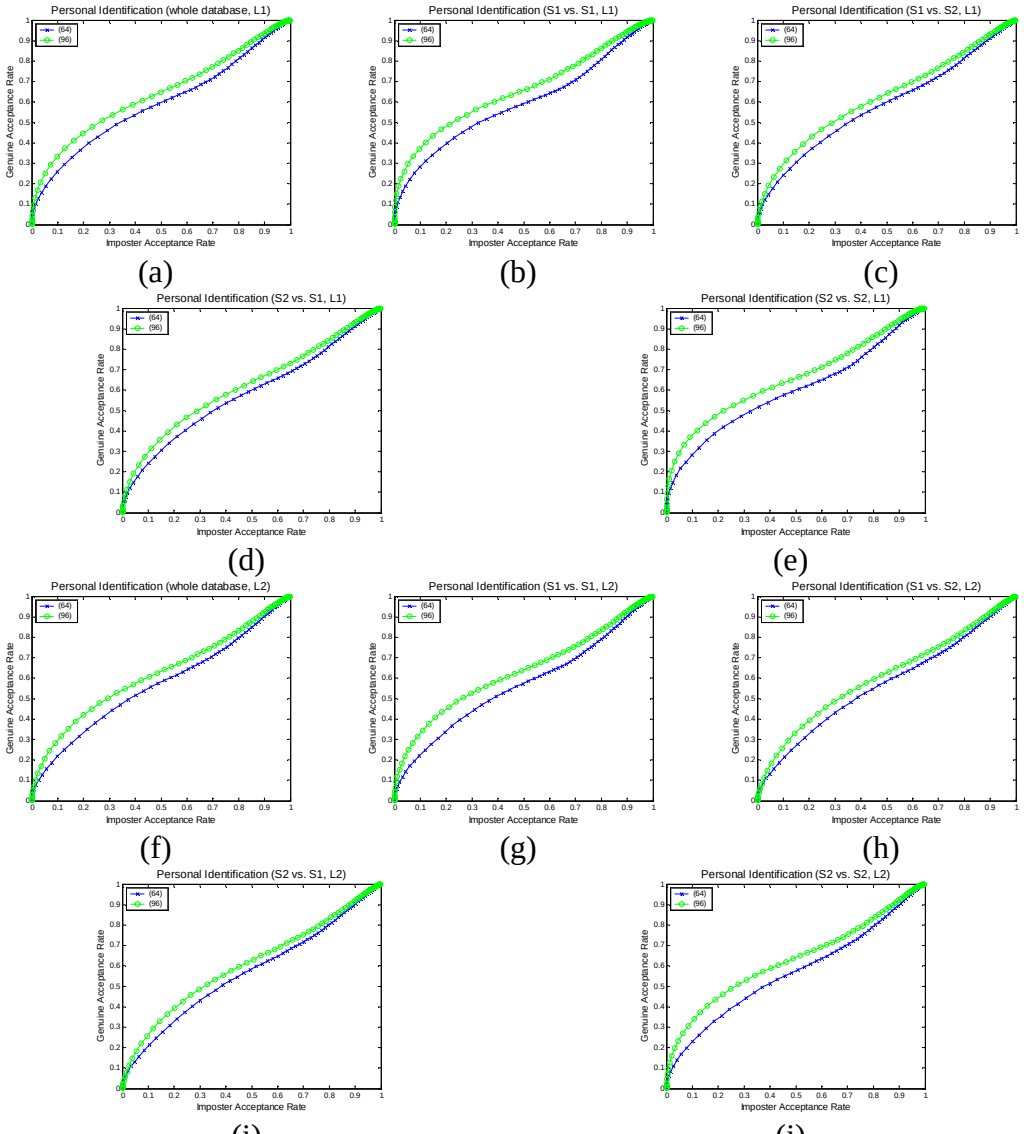


Figure 5.3 The effect of using L_1- and L_2-norm on the performance of aggregated Gabor method for testing sets (64) and (96)

The ROC curves depicting system performance of aggregated Gabor method under the five evaluation cases using L_1- and L_2-norm on the testing set (64) are shown in Figure 5.3 (a)-(e). The ROC curves depicting system performance of aggregated Gabor method under the five evaluation cases using L_1- and L_2-norm on

the testing sets (96) are shown in Figure 5.3 (f)-(j). It can be observed that the L_1-norm is performing slightly better than L_2-norm in all cases of experiments done on both testing sets.

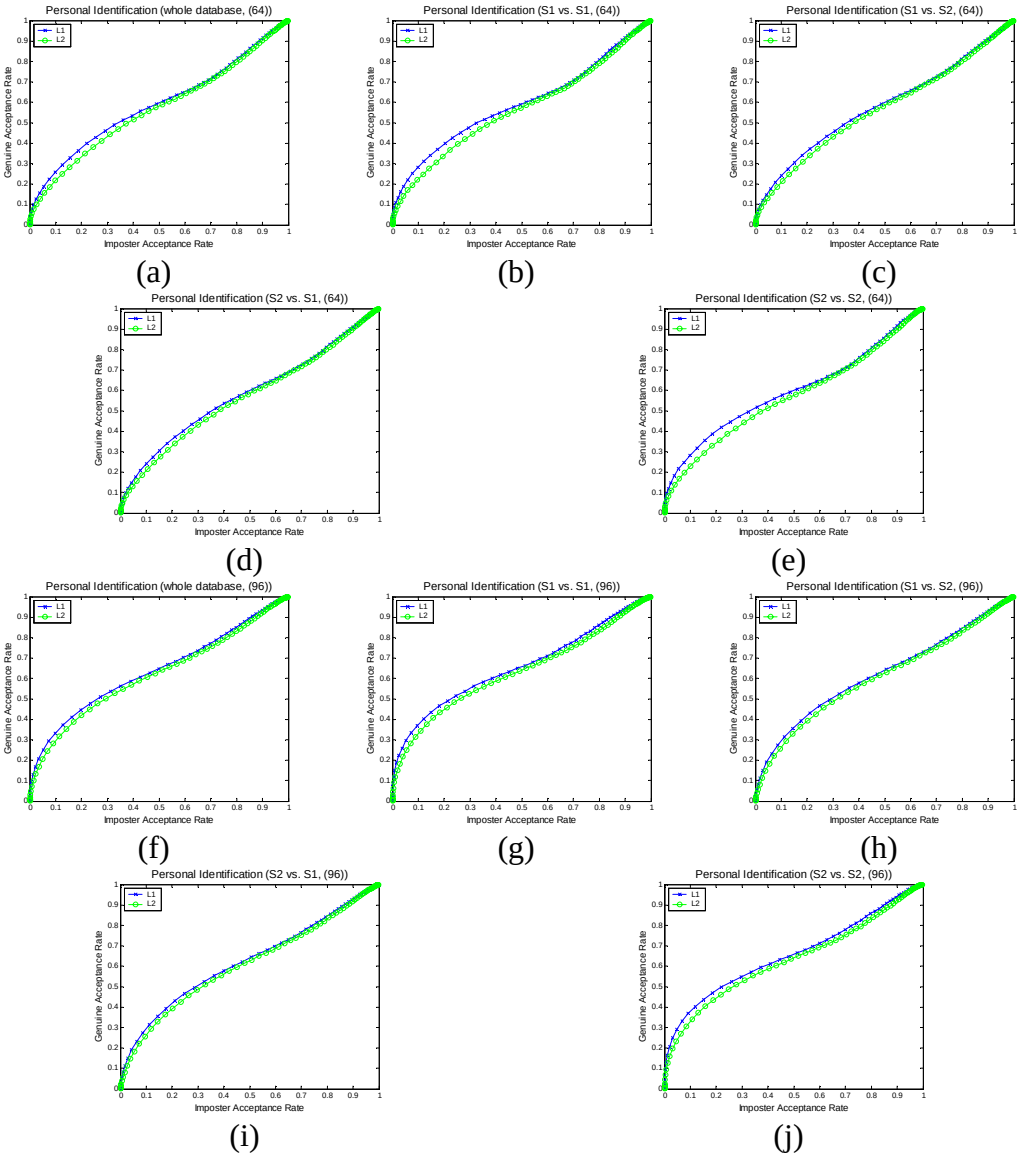


Figure 5.4 The effect of testing sets on the performance of aggregated Gabor method using L_1-norm and L_2-norm

The ROC curves depicting system performance of aggregated Gabor method under the five evaluation cases using L_1-norm on the testing sets (64) and (96) is shown in Figure 5.4 (a)-(e) while using L_2-norm is shown in Figure 5.4 (f)-(j). Based on the ROC curves in Figure 5.4, it is observed that the performance of our method

applied on the testing set (96) is better than the testing set (64) if a predefined threshold is used to decide whether it is a match or not.

Table 5.1 Recognition rates using aggregated 2D Gabor features and, L_1- and L_2-norm on the testing sets (64) and (96) based on the (a) 10, (b) 25 and (c) 50 nearest neighbours.

(a)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
Whole (64)	0.46154	0.74776	0.96154	0.26923	0.65705	0.88462
S1 vs. S1 (64)	0.30769	0.76026	1	0.30769	0.66346	1
S1 vs. S2 (64)	0	0.28397	0.84615	0	0.24679	0.61538
S2 vs. S1 (64)	0	0.32179	0.84615	0	0.27115	0.76923
S2 vs. S2 (64)	0.38462	0.72179	1	0.15385	0.66731	1
Whole (96)	0.46154	0.75673	1	0.30769	0.63558	0.96154
S1 vs. S1 (96)	0.30769	0.77564	1	0.15385	0.6609	1
S1 vs. S2 (96)	0	0.25064	0.76923	0	0.18462	0.69231
S2 vs. S1 (96)	0	0.26667	0.92308	0	0.19038	0.69231
S2 vs. S2 (96)	0.23077	0.76538	1	0.15385	0.64231	1

(b)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
Whole (64)	0.19231	0.60321	0.92308	0.076923	0.46763	0.76923
S1 vs. S1 (64)	0.15385	0.63462	1	0	0.47628	0.92308
S1 vs. S2 (64)	0	0.23846	0.69231	0	0.19808	0.69231
S2 vs. S1 (64)	0	0.24103	0.76923	0	0.20385	0.76923
S2 vs. S2 (64)	0.15385	0.5891	1	0	0.49872	1
Whole (96)	0.15385	0.60417	0.96154	0.11538	0.47628	0.92308
S1 vs. S1 (96)	0.15385	0.63782	1	0.076923	0.49936	0.92308
S1 vs. S2 (96)	0	0.21923	0.84615	0	0.15321	0.69231
S2 vs. S1 (96)	0	0.21474	0.92308	0	0.15897	0.84615
S2 vs. S2 (96)	0	0.6359	1	0	0.49808	1

(c)

	(Dis)Similarity Measure					
	L_1-norm			L_2-norm		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
Whole (64)	0.038462	0.48109	0.84615	0	0.35769	0.80769
S1 vs. S1 (64)	0	0.51987	1	0	0.36923	0.92308
S1 vs. S2 (64)	0	0.20192	0.76923	0	0.17115	0.69231
S2 vs. S1 (64)	0	0.20577	0.84615	0	0.15705	0.76923
S2 vs. S2 (64)	0.076923	0.48141	1	0	0.38654	1
Whole (96)	0.076923	0.48462	0.92308	0.038462	0.36635	0.84615
S1 vs. S1 (96)	0	0.49615	1	0	0.37436	0.92308
S1 vs. S2 (96)	0	0.18718	0.84615	0	0.12692	0.84615
S2 vs. S1 (96)	0	0.19487	0.92308	0	0.14359	0.92308
S2 vs. S2 (96)	0	0.51987	1	0	0.38462	1

Table 5.1 shows the recognition rates based on the 1st rank of (a) Top 10, (b) Top 25 and (c) Top 50 retrieved images of aggregated Gabor method under the five evaluation cases using L_1- and L_2-norm on the testing sets (64) and (96). Table 5.2 shows the recognition rates using the same determinant based on PCA for comparison.

Table 5.2 Recognition rates using PCA and, L_1- and L_2-norm on the testing sets (64) and (96) based on the (a) 10, (b) 25 and (c) 50 nearest neighbours.

(a)

	(Dis)Similarity Measure					
	L_1-norm			L_2-norm		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
S1 vs. S1 (64)	0.53846	0.92372	1	0.38462	0.88269	1
S1 vs. S2 (64)	0	0.3359	1	0	0.32244	1
S2 vs. S1 (64)	0	0.14872	1	0	0.15769	1
S2 vs. S2 (64)	0.69231	0.94487	1	0.53846	0.91474	1
S1 vs. S1 (96)	0.53846	0.95064	1	0.53846	0.93462	1
S1 vs. S2 (96)	0	0.35833	1	0	0.34423	1
S2 vs. S1 (96)	0	0.16987	0.92308	0	0.17372	1
S2 vs. S2 (96)	0.69231	0.96603	1	0.76923	0.94808	1

(b)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
S1 vs. S1 (64)	0.46154	0.85256	1	0.23077	0.79295	1
S1 vs. S2 (64)	0	0.30513	1	0	0.26474	1
S2 vs. S1 (64)	0	0.13974	1	0	0.13654	1
S2 vs. S2 (64)	0.23077	0.86154	1	0.076923	0.81282	1
S1 vs. S1 (96)	0.38462	0.86923	1	0.23077	0.8359	1
S1 vs. S2 (96)	0	0.34295	1	0	0.30385	1
S2 vs. S1 (96)	0	0.15064	0.92308	0	0.14615	0.92308
S2 vs. S2 (96)	0.38462	0.90385	1	0.15385	0.85641	1

(c)

	(Dis)Similarity Measure					
	L_1-norm			*L_2*-norm		
Image Sets	Worst	Average	Best	Worst	Average	Best
S1 vs. S1 (64)	0.30769	0.77564	1	0.076923	0.68141	1
S1 vs. S2 (64)	0	0.28846	1	0	0.25	1
S2 vs. S1 (64)	0	0.12564	1	0	0.12308	1
S2 vs. S2 (64)	0.30769	0.79038	1	0	0.72372	1
S1 vs. S1 (96)	0.15385	0.78269	1	0.15385	0.70577	1
S1 vs. S2 (96)	0	0.32821	1	0	0.27244	1
S2 vs. S1 (96)	0	0.13846	0.92308	0	0.11987	0.92308
S2 vs. S2 (96)	0.38462	0.83013	1	0.076923	0.76538	1

From Table 5.1 and 5.2, it is observed that under cases (2) and (5), the performance of PCA is better that aggregated Gabor method. Nonetheless, under cases (3) and (4), i.e. matching “duplicates”, the performance of aggregated Gabor method is better or comparable to that of PCA in terms of not only figures but also milder performance degradation caused by “duplicates”. In addition, the performance of case (3), S1 vs. S2, is always significantly better than that of case (4), S2 vs. S1; this reveals that the performance of PCA heavily depends on the training samples selected (registered images) while aggregated Gabor method is not affected by the selection of registered images.

Phase 2

In Phase 2; images of 50 randomly chosen subjects, which consisted of 25 males and 25 females, in our testing database are used. For the image sets used in Phase 2, the facial features (Martinez and Kak 2001), such as the center of eyes, in each of the image are first localized. Then each face is aligned upright based on the center of eyes (Martinez 2002) and the region containing the face is cropped to be of size 55×71; this is analogous to the planar transform suggested by Beymer (1995). They are finally warped to a "standard" face (Beymer 1995); the warping procedure, which is designed to normalize the face and to align the facial features to approximately the same pixels, has been shown to improve recognition results (Martinez 2002; Martinez and Kak 2001). Afterwards, an oval mask is applied to remove the highly probable background area so as to extract only responses (for Aggregated 2D Gabor Features) or pixels (for PCA) that are within the mask for feature extraction (Martinez and Kak 2001).

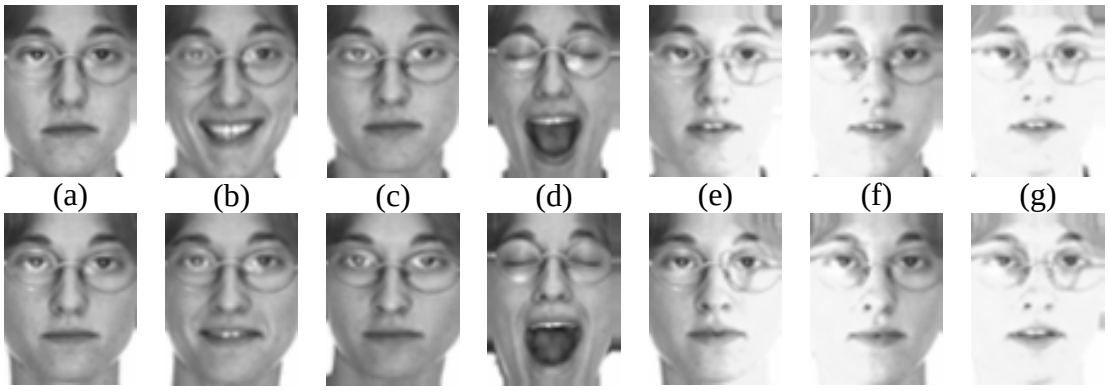


Figure 5.5 A set of warped images of one subject in our testing set for Phase 2. (a)-(g) are from 1st session while (h)-(n) are from 2nd session.

A sample set of the fourteen preprocessed (warped) images used in Phase 2 is shown in Figure 5.5. In Figure 5.5, the first row, i.e. (a)-(g), are images obtained in the 1st session (s1) and the second row, i.e. (h)-(n), are those obtained in the 2nd session (s2). In stage one of Phase 2, images of Figure 5.5 (a), (e)-(g) and (h), (l)-(n)

(corresponding to Figure 5.1 (a), (e)-(g) and (n), (r)-(t) and, 400 images in total) are used to test against aggregated Gabor method and PCA similarly as described in Phase 1. In stage two of Phase 2; images of Figure 5.5 (a)-(g) and (h)-(n) (corresponding to Figure 5.1 (a)-(g) and (n)-(t) and, 700 images in total) are used to test against the two methods.

We conduct the experiments of Phase 2 in two stages: First, only the four images of the neutral expression (under various illumination conditions) from both sessions (s1n and s2n) are used in stage one; Second, the seven images without occlusion (i.e. s1 and s2, the four images used in stage one plus the three with various facial expressions, s1f and s2f other than neutral) are used in stage two.

Since PCA are sensitive to pixel-wise variations, the warping (that aligns the features), the oval mask (that removes the noisy background) and the testing images (only the neutral expression is used, i.e. no variation introduced by facial expression) are all more favourable to PCA; it is expected that PCA will outperform aggregated Gabor method in stage one of the experiment. Both of the methods are expected to attain high accuracy rate, normally over 90 percent with a testing image set at this scale (Martinez 2002).

The results of stage one are summarized in Table 5.3 and Table 5.4 while the results of stage two are summed up in Table 5.5 and Table 5.6. Table 5.3 and Table 5.5 shows the recognition rates based on the (a) 5, (b) 10 and (c) 20 nearest neighbours of aggregated Gabor method under the four evaluation cases (i.e. cases (2) to (5) in Phase 1 described above) using L_1- and L_2-norm on the preprocessed (warped) image set. Table 5.4 and Table 5.6 show the recognition rates using the same determinant based on PCA for comparison.

Table 5.3 Recognition rates using aggregated 2D Gabor features and, L_1- and L_2-norm on warped images with neutral expressions, s1n and s2n, based on the (a) 5, (b) 10 and (c) 20 nearest neighbours

(a)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1n vs. s1n	0.5	0.97	1	0.5	.975	1
s1n vs. s2n	0.25	0.895	1	0	0.875	1
s2n vs. s1n	0.5	0.915	1	0.25	0.88	1
s2n vs. s2n	0.25	0.97	1	0.5	0.975	1

(b)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1n vs. s1n	0.5	0.945	1	0.25	0.96	1
s1n vs. s2n	0.25	0.89	1	0.25	0.85	1
s2n vs. s1n	0.5	0.915	1	0.25	0.885	1
s2n vs. s2n	0.5	0.945	1	0.25	0.955	1

(c)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1n vs. s1n	0.5	0.86	1	0.5	0.895	1
s1n vs. s2n	0.25	0.81	1	0	0.85	1
s2n vs. s1n	0.25	0.835	1	0.25	0.835	1
s2n vs. s2n	0.25	0.855	1	0.5	0.86	1

Table 5.4 Recognition rates using PCA and, L_1- and L_2-norm on warped images with neutral expressions, s1n and s2n, based on the (a) 5, (b) 10 and (c) 20 nearest neighbours

(a)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1n vs. s1n	1	1	1	1	1	1
s1n vs. s2n	0.75	0.985	1	0.25	0.955	1
s2n vs. s1n	0	0.96	1	0.25	0.94	1
s2n vs. s2n	1	1	1	1	1	1

(b)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1n vs. s1n	0.75	0.99	1	1	1	1
s1n vs. s2n	0.5	0.95	1	0.25	0.965	1
s2n vs. s1n	0.5	0.95	1	0.5	0.975	1
s2n vs. s2n	0.75	0.995	1	0.5	0.99	1

(c)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1n vs. s1n	0.75	0.985	1	0.25	0.945	1
s1n vs. s2n	0.25	0.93	1	0	0.9	1
s2n vs. s1n	0.25	0.915	1	0.5	0.935	1
s2n vs. s2n	0.75	0.995	1	0.5	0.95	1

Table 5.5 Recognition rates using aggregated 2D Gabor features and, L_1- and L_2-norm on warped images, s1 and s2, based on the (a) 5, (b) 10 and (c) 20 nearest neighbours

(a)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.42857	0.88571	1	0.42857	0.87714	1
s1 vs. s2	0.14286	0.78	1	0.14286	0.71143	1
s2 vs. s1	0.28571	0.79143	1	0.14286	0.585	1
s2 vs. s2	0.57143	0.92286	1	0.42857	0.88	1

(b)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.42857	0.82286	1	0.14286	0.77429	1
s1 vs. s2	0.14286	0.70857	1	0	0.61429	1
s2 vs. s1	0.28571	0.72286	1	0	0.51	1
s2 vs. s2	0.42857	0.82571	1	0	0.79714	1

(c)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.14286	0.7	1	0.14286	0.66857	1
s1 vs. s2	0	0.61714	1	0.14286	0.58571	1
s2 vs. s1	0	0.58857	1	0.14286	0.395	1
s2 vs. s2	0	0.70571	1	0	0.66286	1

Table 5.6 Recognition rates using PCA and, L_1- and L_2-norm on warped images, s1 and s2, based on the (a) 5, (b) 10 and (c) 20 nearest neighbours

(a)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.71429	0.96571	1	0.71429	0.95714	1
s1 vs. s2	0.42857	0.87143	1	0.28571	0.85429	1
s2 vs. s1	0.28571	0.82571	1	0.14286	0.79714	1
s2 vs. s2	0.85714	0.98571	1	0.71429	0.96857	1

(b)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.57143	0.92286	1	0.42857	0.92	1
s1 vs. s2	0.42857	0.82	1	0.14286	0.81714	1
s2 vs. s1	0.28571	0.74571	1	0.14286	0.76571	1
s2 vs. s2	0.71429	0.94286	1	0.57143	0.91714	1

(c)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.42857	0.88571	1	0.42857	0.87714	1
s1 vs. s2	0.14286	0.75143	1	0.42857	0.78	1
s2 vs. s1	0	0.68571	1	0.14286	0.71714	1
s2 vs. s2	0.42857	0.91143	1	0.42857	0.88286	1

From the tables, it can be observed that the performances of both methods are greatly improved, especially in matching the “duplicates”, after suitable processing is introduced. This suggested that proper preprocessing that minimizes the differences aroused, such as alignment and normalization, from capturing face images in different sessions is necessary. From Table 5.3 and Table 5.4, it is confirmed that our presumption about the performance of aggregated Gabor method and PCA is correct; PCA outperforms aggregated Gabor method in stage one of this experiment.

Nonetheless, with the introduction of variations in facial expression, a larger drop in performance of matching “duplicates” using PCA is observed by comparing Table

5.4 and Table 5.6 while a relatively mild degradation in performance is recorded for aggregated Gabor method by contrasting Table 5.3 and Table 5.5.

The overall performance of PCA is better, which is mainly attributed to the composition of the testing image set. In the testing image set, there are more than half (four out of seven) of the images are of neutral facial expression.

Robustness to Variations in Facial Expression

The high recognition rate of PCA for neutral faces (see Table 5.4 and Table 5.6) dominates the performance evaluation. This can be confirmed by a test using only faces with facial expressions other than neutral, i.e. s1f: Figure 5.5 (b)-(d) or s2f: Figure 5.5 (i)-(k), as query images to match against a database that contains the seven images from either s1 or s2.

The results of aggregated Gabor method and PCA are shown in Table 5.7 and Table 5.8 respectively. It can be noted that aggregated Gabor method is more robust than PCA against variations in facial expression.

Table 5.7 Recognition rates using aggregated 2D Gabor features and, L_1- and L_2-norm on warped images, s1f/s2f vs. s1/s2, based on the (a) 5, (b) 10 and (c) 20 nearest neighbours

(a)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1f vs. s1	0.33333	0.79333	1	0.33333	0.78	1
s1f vs. s2	0	0.68	1	0	0.6	1
s2f vs. s1	0	0.66	1	0	0.60667	1
s2f vs. s2	0.33333	0.88667	1	0.33333	0.81333	1

(b)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1f vs. s1	0	0.68667	1	0	0.60667	1
s1f vs. s2	0	0.55333	1	0	0.44	1
s2f vs. s1	0	0.54	1	0	0.47333	1
s2f vs. s2	0.33333	0.7	1	0	0.69333	1

(c)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1f vs. s1	0	0.54	1	0	0.48667	1
s1f vs. s2	0	0.44667	1	0	0.40667	1
s2f vs. s1	0	0.39333	1	0	0.42	1
s2f vs. s2	0	0.58	1	0	0.49333	1

Table 5.8 Recognition rates using PCA and, L_1- and L_2-norm on warped images, s1f/s2f vs. s1/s2, based on the (a) 5, (b) 10 and (c) 20 nearest neighbours

(a)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1f vs. s1	0	0.53333	1	0	0.55333	1
s1f vs. s2	0	0.45333	1	0	0.45333	1
s2f vs. s1	0	0.46667	1	0	0.46667	1
s2f vs. s2	0	0.53333	1	0	0.54667	1

(b)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1f vs. s1	0	0.51333	1	0	0.51333	1
s1f vs. s2	0	0.38667	1	0	0.43333	1
s2f vs. s1	0	0.37333	1	0	0.4	1
s2f vs. s2	0	0.52	1	0	0.53333	1

(c)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1f vs. s1	0	0.52667	1	0	0.49333	1
s1f vs. s2	0	0.39333	1	0	0.34667	1
s2f vs. s1	0	0.35333	1	0	0.40667	1
s2f vs. s2	0	0.52	1	0	0.46667	1

We can further shown that aggregated Gabor method is more robust to variations in facial expression by a test using faces with facial expressions other than neutral as query images to match against a database that contains only faces with neutral facial expression. Since the database does not contain any faces with facial expression other than neutral, they are unseen to PCA, i.e. not trained.

Table 5.9 Recognition rates using aggregated 2D Gabor features and, L_1- and L_2-norm on warped images, s1f/s2f vs. s1n/s2n, based on the (a) 5, (b) 10 and (c) 20 nearest neighbours

(a)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1f vs. s1n	0	0.59333	1	0	0.5	1
s1f vs. s2n	0	0.49333	1	0	0.46	1
s2f vs. s1n	0	0.5	1	0	0.41333	1
s2f vs. s2n	0	0.62667	1	0	0.56667	1

(b)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1f vs. s1n	0	0.61333	1	0	0.51333	1
s1f vs. s2n	0	0.54	1	0	0.52667	1
s2f vs. s1n	0	0.51333	1	0	0.47333	1
s2f vs. s2n	0	0.63333	1	0	0.6	1

(c)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1f vs. s1n	0	0.45333	1	0	0.41333	1
s1f vs. s2n	0	0.42	1	0	0.47333	1
s2f vs. s1n	0	0.41333	1	0	0.48	1
s2f vs. s2n	0	0.52	1	0	0.59333	1

Table 5.10 Recognition rates using PCA and, L_1- and L_2-norm on warped images, s1f/s2f vs. s1n/s2n, based on the (a) 5, (b) 10 and (c) 20 nearest neighbours

(a)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1f vs. s1n	0	0.26667	1	0	0.25333	1
s1f vs. s2n	0	0.23333	1	0	0.25333	1
s2f vs. s1n	0	0.22	1	0	0.2	1
s2f vs. s2n	0	0.26	1	0	0.26667	1

(b)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1f vs. s1n	0	0.23333	1	0	0.26	1
s1f vs. s2n	0	0.25333	1	0	0.25333	1
s2f vs. s1n	0	0.22667	1	0	0.20667	1
s2f vs. s2n	0	0.22667	1	0	0.24	1

(c)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1f vs. s1n	0	0.24	1	0	0.23333	1
s1f vs. s2n	0	0.26	1	0	0.22	1
s2f vs. s1n	0	0.23333	1	0	0.2	1
s2f vs. s2n	0	0.21333	1	0	0.2	1

Table 5.9 and Table 5.10 show the performance of aggregated Gabor method and PCA accordingly. PCA is ineffective to adapt variations in facial expression provided that no faces with similar facial expression are presented in training whereas aggregated Gabor method can still achieve a moderate accuracy.

Curse of dimensionality and Gender

In addition, we have considered the curse of dimensionality (Jain et al. 2000a). A test is conducted using only the images from male or female to test against the two methods but no significant difference in performance is spotted between using only male (Table 5.11 and Table 5.12) or female (Table 5.13 and Table 5.14) and using all 50 subjects (Table 5.5 and Table 5.6), i.e. the peaking phenomena does not emerge

with the present number of classes. It can also be noticed that the gender of subjects does not possess noticeable effect on the matching performance of both methods.

Table 5.11 Recognition rates using aggregated 2D Gabor features and, L_1- and L_2-norm on warped images of male subjects only, s1 and s2, based on the (a) 5, (b) 10 and (c) 20 nearest neighbours

(a)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.42857	0.85714	1	0.42857	0.85143	1
s1 vs. s2	0.42857	0.81714	1	0.42857	0.77143	1
s2 vs. s1	0.28571	0.76571	1	0.14286	0.70286	1
s2 vs. s2	0.57143	0.85714	1	0.42857	0.85714	1

(b)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.42857	0.77143	1	0.42857	0.75429	1
s1 vs. s2	0.14286	0.68571	1	0	0.60571	1
s2 vs. s1	0.28571	0.64	1	0.14286	0.62857	1
s2 vs. s2	0.42857	0.79429	1	0	0.74857	1

(c)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.14286	0.66286	1	0.28571	0.55429	0.85714
s1 vs. s2	0	0.45143	0.85714	0.14286	0.43429	0.85714
s2 vs. s1	0	0.46857	1	0.14286	0.42857	1
s2 vs. s2	0	0.6	1	0	0.53714	1

Table 5.12 Recognition rates using PCA and, L_1- and L_2-norm on warped images of male subjects only, s1 and s2, based on the (a) 5, (b) 10 and (c) 20 nearest neighbours

(a)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.71429	0.98286	1	0.57143	0.94857	1
s1 vs. s2	0.42857	0.83429	1	0.28571	0.78857	1
s2 vs. s1	0.42857	0.90286	1	0.28571	0.83429	1
s2 vs. s2	0.71429	0.92571	1	0.71429	0.93143	1

(b)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.71429	0.93714	1	0.57143	0.92571	1
s1 vs. s2	0.42857	0.71429	1	0.42857	0.75429	1
s2 vs. s1	0.42857	0.83429	1	0.14286	0.78857	1
s2 vs. s2	0.71429	0.87429	1	0.57143	0.89714	1

(c)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.42857	0.88	1	0.42857	0.74286	1
s1 vs. s2	0.14286	0.63429	1	0.14286	0.57143	1
s2 vs. s1	0.28571	0.76571	1	0.28571	0.62286	1
s2 vs. s2	0.14286	0.78857	1	0.28571	0.75429	1

Table 5.13 Recognition rates using aggregated 2D Gabor features and, L_1- and L_2-norm on warped images of female subjects only, s1 and s2, based on the (a) 5, (b) 10 and (c) 20 nearest neighbours

(a)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.57143	0.88571	1	0.57143	0.90286	1
s1 vs. s2	0.28571	0.74286	1	0.14286	0.69143	1
s2 vs. s1	0.28571	0.84	1	0.57143	0.81714	1
s2 vs. s2	0.57143	0.89143	1	0.57143	0.89143	1

(b)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.42857	0.77714	1	0.42857	0.75429	1
s1 vs. s2	0.42857	0.70286	1	0.14286	0.66286	1
s2 vs. s1	0.42857	0.77143	1	0.42857	0.79429	1
s2 vs. s2	0.57143	0.86286	1	0.42857	0.74857	1

(c)

	(Dis)Similarity Measure					
	***L_1*-norm**			***L_2*-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.28571	0.63429	1	0.28571	0.6	1
s1 vs. s2	0.14286	0.50857	0.85714	0.14286	0.54286	1
s2 vs. s1	0.28571	0.60571	0.85714	0.42857	0.59429	0.85714
s2 vs. s2	0.28571	0.67429	1	0.14286	0.6	1

Table 5.14 Recognition rates using PCA and, L_1- and L_2-norm on warped images of female subjects only, s1 and s2, based on the (a) 5, (b) 10 and (c) 20 nearest neighbours

(a)

	(Dis)Similarity Measure					
	***L*$_1$-norm**			***L*$_2$-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.71429	0.98286	1	0.71429	0.97714	1
s1 vs. s2	0.28571	0.83429	1	0.42857	0.81714	1
s2 vs. s1	0.42857	0.86857	1	0.28571	0.85143	1
s2 vs. s2	0.85714	0.97714	1	0.85714	0.97714	1

(b)

	(Dis)Similarity Measure					
	***L*$_1$-norm**			***L*$_2$-norm**		
Image Sets	**Worst**	**Average**	**Best**	**Worst**	**Average**	**Best**
s1 vs. s1	0.57143	0.94286	1	0.57143	0.92	1
s1 vs. s2	0.28571	0.78857	1	0.28571	0.81143	1
s2 vs. s1	0.14286	0.83429	1	0.42857	0.86857	1
s2 vs. s2	0.85714	0.95429	1	0.71429	0.95429	1

(c)

	(Dis)Similarity Measure					
	L_1-norm			L_2-norm		
Image Sets	Worst	Average	Best	Worst	Average	Best
s1 vs. s1	0.57143	0.92571	1	0.42857	0.72	1
s1 vs. s2	0	0.77143	1	0.14286	0.63429	1
s2 vs. s1	0.14286	0.72571	1	0.28571	0.73714	1
s2 vs. s2	0.85714	0.96	1	0.57143	0.78857	1

5.4 Recapitulation

In this chapter, we have briefly introduced the background of face image retrieval. Two publicly available face image databases are presented. A series of experiments on the proposed low dimensional, holistic appearance-based face image features, with reference to a benchmarking method, are reported and discussed.

Chapter 6 Some Design Issues in Storage Security of Biometric Databases

6.1 Overview of Cancelable Biometrics

Biometric authentications, nonetheless, is not without its challenges. Authentications are increasingly performed in unattended and/or over networked environments. Intruders have more time and opportunity to attack via the exposed communication channels. Ratha et al. (2001, 2003) pointed out eight potential points of threats in a generic biometric authentication system and one potential point of threat in the enrollment process as shown in Figure 6.1 numbered with 1 to 8 and 9 respectively. Schneier (1999); Tuyls and Goseling (2004) and Uludag et al. (2004) also address the importance of the threats.

Vulnerable points in a generic biometrics-based system

The presentation of a fake biometric at the Sensor (A) is Threat 1. Channel attack between the Sensor (A) and the biometric system (B) is Threat 2. Being taken over the process of the Template Extractor (B) is Threat 3. Channel attack between the Template Extractor (B) and the Matcher (C) is Threat 4. Being taken over the process of the Matcher (C) is Threat 5. Unauthorized modification made to the Template Databases (F) is Threat 6. Channel attack between the Template Database (F) and the Matcher (C) is Threat 7. Channel attack between the Matcher (C) and the Application (D) is Threat 8. Gaining unauthorized access to the enrollment office is Threat 9.

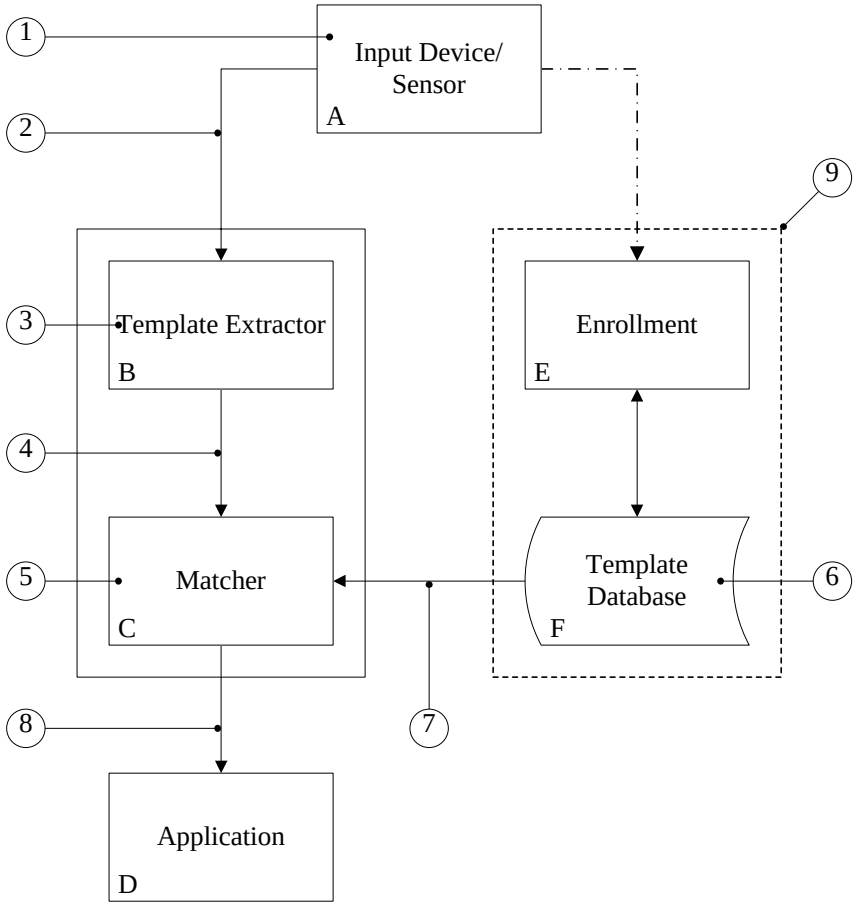


Figure 6.1 A biometric system and its potential points of threat

There is a growing concern from public (Ratha et al., 2001; Braithwaite et al., 2002) about the sharing of biometric templates obtained from different organizations and different applications. Since biometrics are not secrets and are limited, once they are compromised, they are compromised forever and perhaps everywhere they are adopted. They, moreover, cannot be changed, canceled/destroyed and reissued/updated. Sensitive personal information associated with biometrics may therefore be uncovered. The strengths of biometrics become their weaknesses as well. (Note: Cancelable biometric cannot help safeguarding the sharing of biometric databases of raw templates, e.g. raw image.)

Ratha et al. (2001) introduce the concept of cancelable biometrics (referred as private template by Uludag et al., 2004). This provides a mean to protect privacy in biometric authentication systems. Cancelable biometrics is achieved through

applying intentional and repeatable distortions (or transformations) on biometrics in either the signal domain (the raw/preprocessed information obtained from capture devices and before feature extractor is applied) or the feature domain (the information obtained after feature extraction is performed). The transformations for cancelable biometrics are ideally noninvertible. In practice, nevertheless, the transformation can be invertible. Alternatively, templates protection can be preformed by other methods, such as encryption (Uludag et al., 2004) that has to be implemented by invertible transforms. Cancelable biometrics may be easily confused with encryption. The major difference is that cancelable biometrics performs matching in the transformed domain while encryption does not. Ratha et al. (2001) have given some example transforms for cancelable biometrics. As invertible examples, grid morphing and block permutation are offered and as a noninvertible example, high order polynomial is offered.

Cancelable biometric templates are essential for biometric authentication systems, especially for those operated under unattended and/or over networked environments. (Schneier, 1999; Ratha et al., 2001, 2003; Tuyls and Goseling, 2004; Uludag et al., 2004) In addition to the previous concerns, cancelable biometrics can be extended for biometric cryptosystems. (Clancy et al., 2003; Juels and Sudan, 2002; Juels and Wattenberg, 1999; Soutar et al., 1999; Tuyls and Goseling, 2004).

6.2 Issues to be considered in Cancelable Biometrics

Cancelable biometrics can provide the means to enhance biometric authentication systems but involve tradeoffs between the System Security, System Requirements and System Performance. Figure 6.2 shows the relations among them. In the following subsections, we are going to take a deeper look at the three aspects,

System Security, System Requirements and System Performance and, the tradeoffs between them.

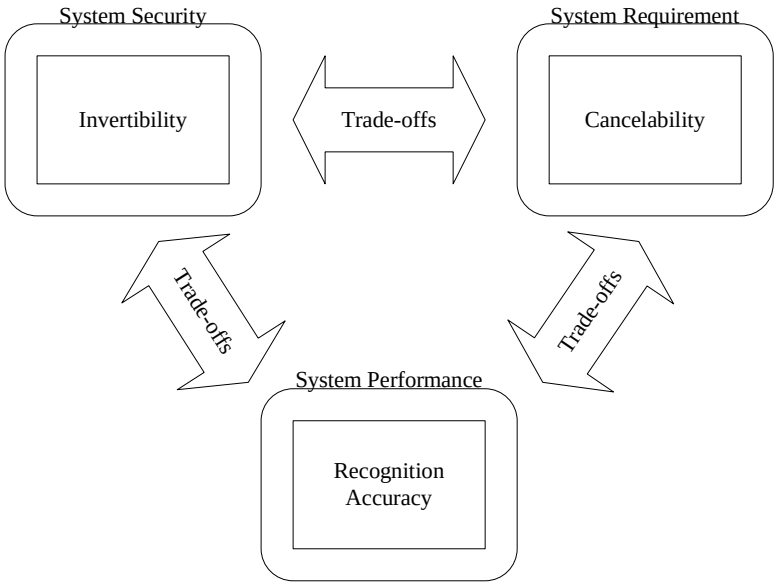


Figure 6.2 The mutual relationship among accuracy, cancelability and invertibility in a biometric system

6.2.1 System performance — Recognition Accuracy

A biometric system is useful only if the recognition accuracy is high enough for certain application. In cancelable biometrics, however, the transforms are preferably noninvertible implying possible loss of information. Thus, noninvertible transforms can lead to loss of discriminating ability resulting in deterioration of accuracy. This argument applies to the noninvertible transforms for both signal and feature domains.

To preserve information, invertible transforms may be used to replace noninvertible ones. We, nevertheless, have to choose transforms that preserve some meaningful relationships for feature extraction (for signal domain transforms) and similarity measure (for both domain) afterwards. Alternatively, we may have to define a suitable feature extraction as well as similarity measure in the transformed domain. Otherwise, the accuracy may not be guaranteed. For example, if a monotonically increasing function is the transform and L_2-norm is the similarity

measure, it will be inaccurate to measure the similarity in the transformed domain using L_2-norm. To sustain cancelability, certain transforms are still adopted even though we are not able to define a suitable feature extraction and/or similarity measure in the transformed domains to maintain the original recognition accuracy.

6.2.2 System requirement — Cancelability

Cancelable biometrics have the ability to produce cancelable templates, i.e. cancelability. They are designed for preventing templates sharing and for templates re-issuances. Any templates from two different transforms should be regarded as two different templates as from two different persons. In the sense of recognition, they will be treated as unmatched, and therefore, cancelable. Thus, if someone presents a compromised template to the biometric system, it will be classified as an impostor.

It should be noted that the possible number of reissued templates is limited. Ideally, transforms are issued randomly by the system and we do not memorize the previously issued transforms. We may use the following probability model to describe the situation of cancelability.

Suppose that X is a biometric template obtained from a registered person Y. The probability of two different cancelable templates of a person Y being unmatched is $\Pr(D(T_i(X),T_j(X)) > \boldsymbol{t}) = \alpha$ where $D(\cdot)$ is a distance/dissimilarity measure, T_i and T_j are two different transforms for the cancelable templates and $\boldsymbol{t}$ is the threshold that determines a match. The false acceptance rate, therefore, is $\Pr(D(T_i(X),T_j(X)) \leq \boldsymbol{t}) = 1-\alpha$. Assume an attacker has n compromised templates of the registered person Y in hand to attack the system. The probability of breaking in the system is $1-\alpha^n$.

Using real numbers to see the effect, suppose $\alpha = 0.99$, the probability of breaking in for n = 10, 50, 100 are 0.09562 (9.562%), 0.3950 (39.50%), 0.6340 (63.40%) respectively. Thus, the system is more vulnerable to attack as the number of

compromised templates increases. This demonstrates that we may not have unlimited, number of templates for re-issuance, which depends on α.

6.2.3 System security — Invertibility

For any invertible transform $y = f(x)$, we can derive $x = f^{-1}(y)$ where f^{-1} is the inverse of f. From the security point of view, invertible transforms will collapse easily if attackers know all the details about the transform.

The exact inverse g^{-1} of any noninvertible transform $y = g(x)$ does not exist. Noninvertible transforms, therefore, are more secure than invertible transforms but they do not completely free from attacks. We can, still, approximate x from y by the approximated inverse g^* of g, i.e. $x' = g^*(y)$. If x' is close enough to x, it can be used to fake the biometric system to break in because (dis)similarity measure, i.e. inexact match, is adopted. This is a potential risk that we should pay attention.

6.3 Discussion on Existing Cancelable Biometrics

We have discussed the three issues in cancelable biometrics, System Security, System Requirements and System Performance and, the tradeoffs between them. We are going to review some proposed solutions to cancelable biometrics to illustrate further the issues discussed above.

Braithwaite et al. (2002) have outlined the concept of application (/transaction)-specific transforms to protect biometric templates from unauthorized exchange. An authorized central management of application-specific transforms is suggested to establish. Setting up such management center is expensive and it requires connecting all the (iris) devices via internets/ land but this can facilitate the reuse of templates by authorized conversion of application-specific biometric templates for one application to others without re-enrollment, Their proposed

transform is invertible and operates in the feature domain. Since the transform is invertible, there is potential risk that the raw templates can be recovered.

Savvides et al. (2004) have integrated minimum average correlation energy (MACE) filter and random kernel to provide cancelable biometrics. It has been proven to be of no degradation in accuracy of matching the transformed features because the MACE filter possesses a nice property that eliminates the effect of the random kernel. The transform (random kernel) used is invertible and operates in feature domain.

Connie et al. (2004b), Pang et al. (2004) and, Teoh and Ngo (2005) presented prototypes of cancelable biometrics for palmprint and face based on BioHashing (Teoh et al. 2004c). BioHashing consists of two major steps, feature domain random transformation, and discretization (a two level quantization). Their transform is noninvertible and operates in feature domain. As the transform is noninvertible, the raw template can be better protected. Random transform can provide excellent cancelabiltiy (see Chapter 6.4.3). Nonetheless, we believe that noninvertible random transformations will destroy the optimality of most feature representations and thus the recognition accuracy deteriorates (see Chapter 6.4.2). There is a tradeoff between optimality and cancelability in representation and similarity matching in the feature domain of cancelable biometrics, i.e. increase in error rate of the authentication (Uludag et al., 2004).

6.4 A Case Study

We use BioHashing (Teoh et al. 2004c) and its variants (Connie et al. 2004; Pang et al. 2004; Teoh and Ngo 2005), which is a random transformed feature-based cancelable biometrics, as a case study to highlight some possible situations. BioHashing and its variants are chosen because they are, in existing approaches, the

closest, i.e. using noninvertible transform, to cancelable biometrics established by Ratha et al. (2001). They, also, have been tested on several biometrics, including face, fingerprint and palmprint.

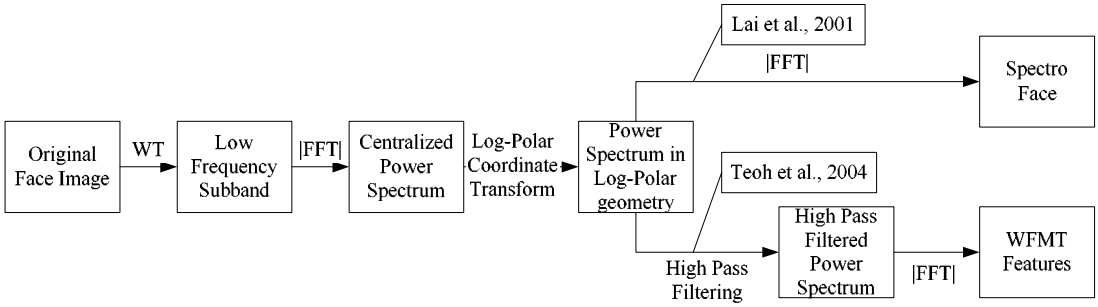


Figure 6.3 Comparison of process flow of Holistic Fourier Invariant Features (HFIF) extraction (Lai et al. 2001) and Integrated Wavelet and Fourier-Mellin Transform (Teoh et al. 2004)

We have re-implemented the experiments in Teoh and Ngo (2005) for our case study because face recognition is a well-known problem and public databases are available and the experimental details are given as follows.

Transform integrating wavelet transform (WT) and Fourier-Mellin Transform (FMT) has been exploited by Lai et al. (2001, see also Chapter 3.1.1 for more) and Teoh and Ngo (2005) for face recognition and, Teoh et al. (2004c) for fingerprint recognition. The flowchart in Figure 6.3 shows the flow of HFIF and WFMT. Teoh and Ngo (2005) and Teoh et al. (2004c) differ from Lai et al. (2001) by undergoing a highpass filter (Srinivasa Reddy and Chatterji, 1996) between the first and the second Fourier transforms.

We note down the details of the WFMT that we implemented based on Teoh and Ngo (2005) and Lai et al. (2001) in Table 6.1 and the details of BioHashing in Table 2. We would like, also, to make a note of how Teoh and Ngo (2004c) determine the genuine and impostor distribution. They match all images among the same person to determine the genuine distribution; while, they match only the n^{th} image of one person to the n^{th} image of other persons to determine the impostor distribution where

n = 1…10. We will present the performance based on both criteria, Teoh and Ngo (2005) and the usual practice, which matches one image against all images of the same person and of others persons in the database to determine the genuine and impostor distributions respectively.

Table 6.1 The details of WFMT implementation

Processes/Variables/Parameters	Values/Descriptions
Raw image sizes	92×112, no preprocessing
Wavelet	db7
Level of Wavelet Decomposition	1
Wavelet transformed image sizes (LL band)	52×62
Log-polar transformation	Largest inscribed circle, bicubic interpolation, 62 logarithmic levels
Highpass Filter	same as Teoh and Ngo (2004) $H(x,y) = (1-\cos(\pi x)\cos(\pi y)) \times (2-\cos(\pi x)\cos(\pi y))$

Table 6.2 Thresholds used for various dimensions of BioCode

Dimension	Threshold(τ)
20	0
40	0
60	0
80	0

6.4.1 Overview of BioHashing and its variants

In Figure 6.4, the two major processes, feature domain random transformation and discretization, of BioHashing are shown. Different biometric signals exploit different signal acquisition, preprocessing and feature extraction techniques. The feature domain random transformation process includes the generation of random matrix, orthonormalization and feature transformation. Discretization is performed afterwards based on a threshold, τ. Feature domain random transformation and discretization is conducted as detailed below.

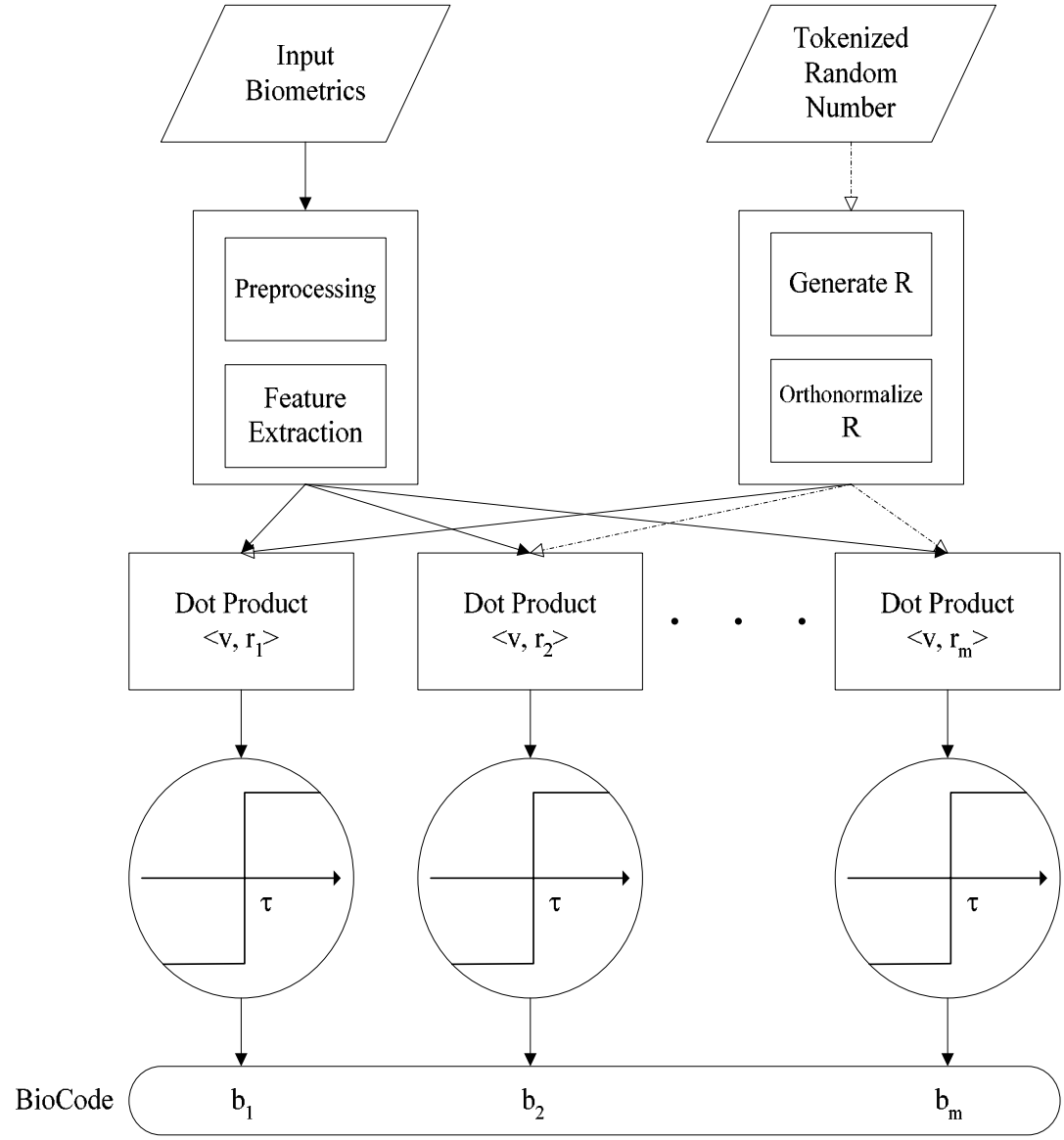


Figure 6.4 A schematic diagram of BioHashing

Step 1. Employ the input token to generate a set of pseudo-random vectors, $\{r_i \in \Re^M \mid i = 1,....., m\}$ based on a seed.

Step 2. Apply the Gram-Schmidt process to $\{r_i \in \Re^M \mid i = 1,....., m\}$ to obtain, a set of orthonormal vectors $\{p_i \in \Re^M \mid i = 1,....., m\}$.

Step 3. Calculate the dot product of ***v,*** the feature vector obtained from Step 1 and each orthonormal vector, $\boldsymbol{p_i}$, such that $\langle v,\ p_i \rangle$.

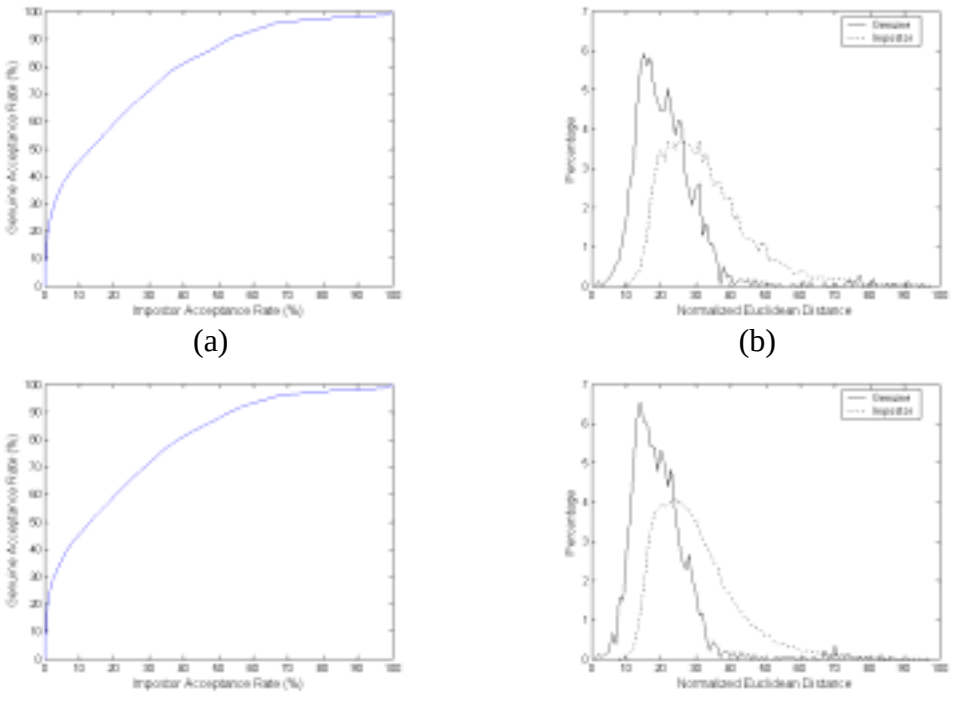
Step 4. Use a threshold τ to obtain BioCode, ***b*** whose elements are defined as

$$b_i = \begin{cases} 0 & if \quad \langle v, p_i \rangle \leq \tau \\ 1 & if \quad \langle v, p_i \rangle > \tau \end{cases},$$

where i is between 1 and m, the dimensionality of $\boldsymbol{b}$. Two BioCodes (or BioHash in Teoh et al. 2004c) are compared by hamming distance.

6.4.2 Recognition Accuracy

In order to ensure cancelable biometrics is practical, we have to look at the system performance. We understand there is tradeoff between cancelability and recognition accuracy. So, we now look at the amount of recognition accuracy traded for cancelability.



(a) (b)

(c) (d)

Figure 6.5 Test on accuracy of BioCodes.
ROC curves of Wavelet Fourier Mellin Transform using L_2-norm (a) Teoh and Ngo's matching scheme and (c) whole database and, their corresponding genuine and impostor distributions (b) and (d)

Using WFMT and Euclidean distance (L_2-norm), the ROC curves, which is a plot of the genuine acceptance rates (GAR) against the impostor/false acceptance rates (FAR), and their corresponding genuine and impostor distributions based on Teoh

and Ngo's matching scheme and the usual practice are plotted in Figure 6.5. This will be used as the reference for the performance comparisons to cancelable biometrics thereafter. The number of matching for estimating the genuine and impostor distributions based on the usual practice are 1,800 and 78,000 respectively and the number of matching for estimating the genuine and impostor distribution based on the Teoh and Ngo's matching scheme are 1800 and 7800, respectively. We, however, cannot reproduce their result (case *wfm*) reported in Teoh and Ngo (2005). We, therefore, have simulated the experiments in Lai et al. (2001) and confirmed the correctness of our implementation of WFMT by producing similar performance.

The genuine and impostor distributions of various dimensions of BioCode, which are based on Teoh and Ngo's matching scheme and the usual practice, are shown in Figure 6.6. The number of matching producing genuine and impostor distributions based on Teoh and Ngo (2005) are 1,800 and 7,800 respectively. The number of matching producing genuine and impostor distributions based on the usual practice are 1,800 and 78,000 respectively. Their corresponding ROC curves with comparison to WFMT and Euclidean distance (L_2-norm) are shown in Figure 6.7. It is obvious that our hypothesis is correct. The optimality of feature representation is destroyed by the noninvertible random transform and quantization and thus the recognition accuracy deteriorates. The performance of BioHashing is even worse than that of using WFMT and Euclidean distance (L_2-norm).

It may be noticed that the performance of BioHashing is different from those given in (Teoh et al. 2004c; Connie et al. 2004; Pang et al. 2004; Teoh and Ngo 2005). Their performances reported here are their true performances. The detail explanation to this issue is given as an analysis of BioHashing and its variants in Chapter 7.

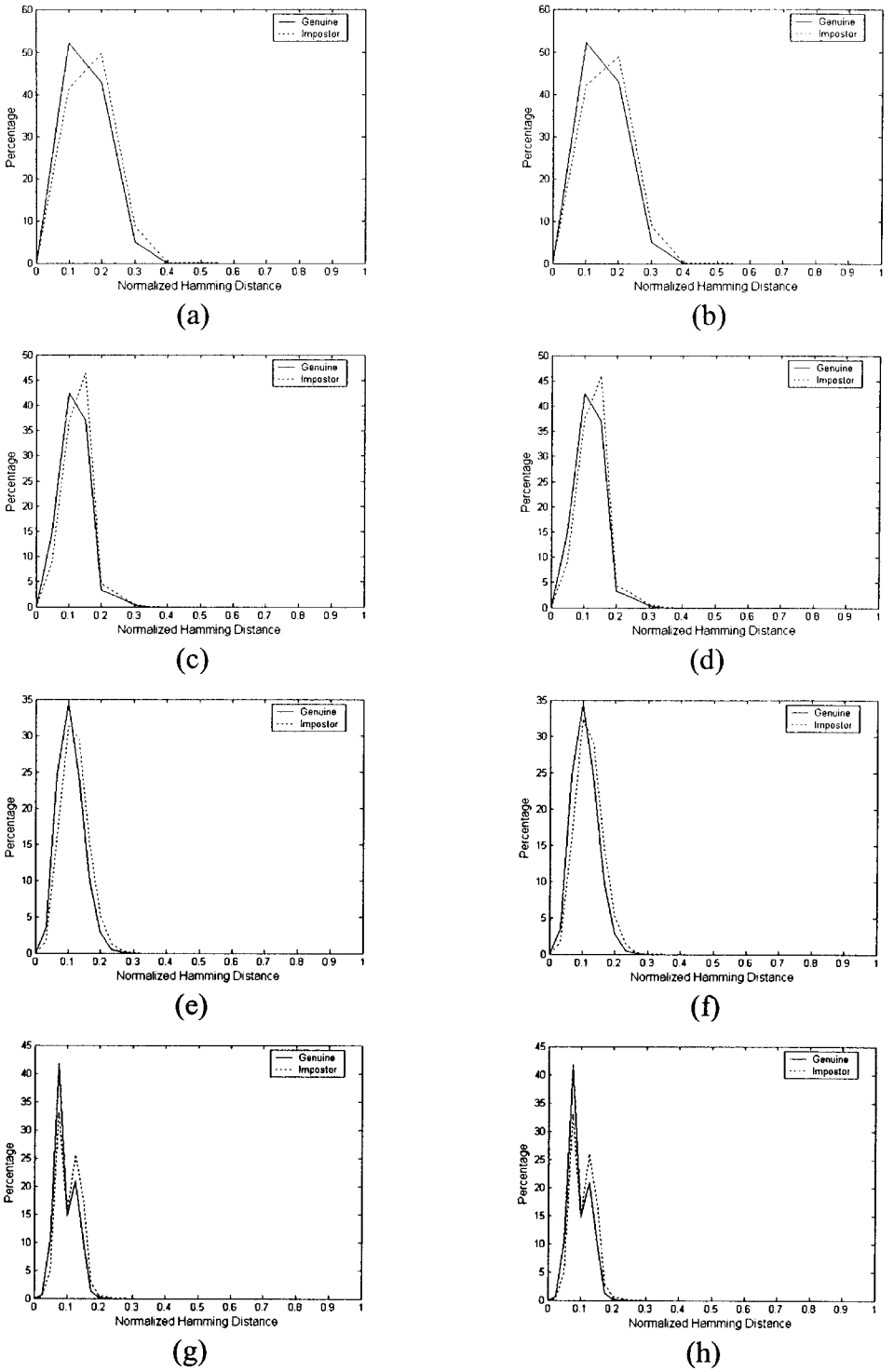
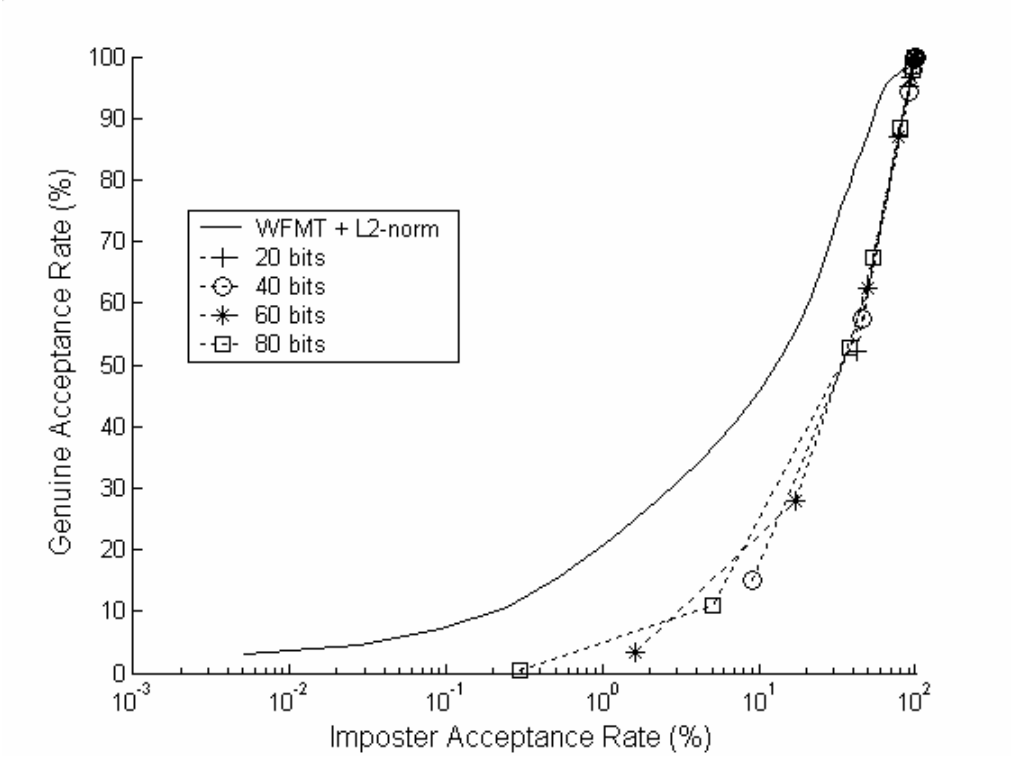
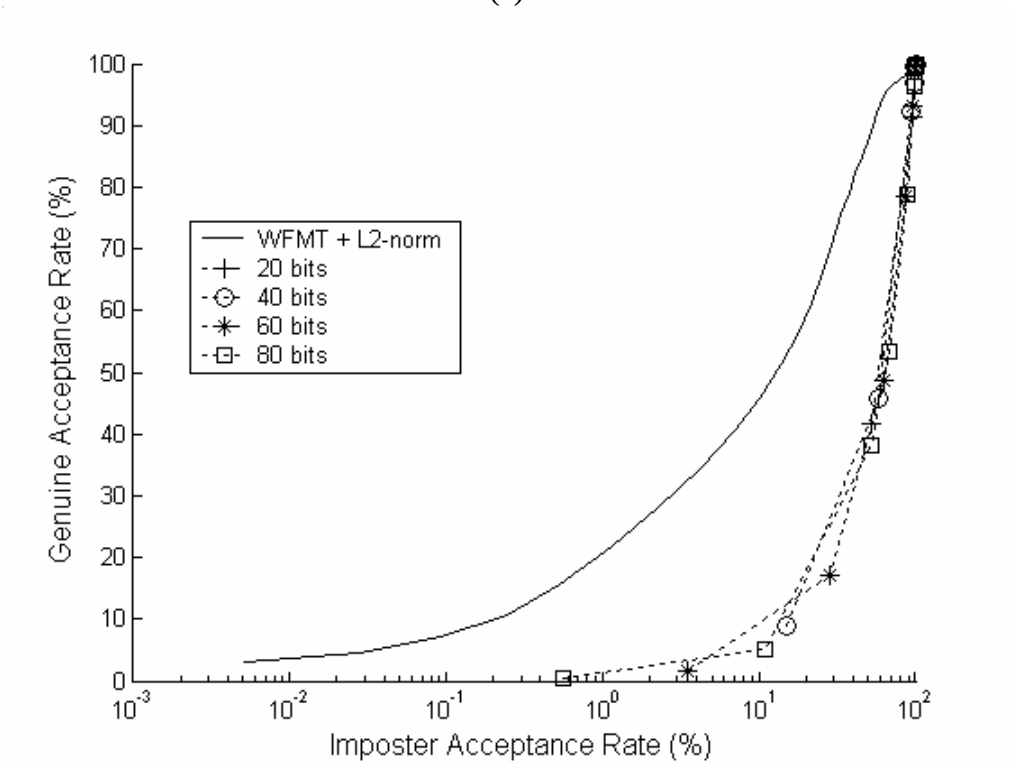


Figure 6.6 Test on accuracy of BioCodes.
The dimensions are (a) and (b) 20, (c) and (d) 40, (e) and (f) 60, (g) and (h) 80
Genuine and impostor distributions of various dimensions of BioCode (a), (c), (e), (g) based on matching criteria in Teoh and Ngo 2004 and (b), (d), (f), (h) based on the usual practice.



(a)



(b)

Figure 6.7 Test on accuracy of BioCodes.
(a) ROC curves of distributions in Fig. 5 (b) and Fig 6 (a), (c), (e), (g).
(b) ROC curves of distributions in Fig. 5 (d) and Fig 6 (b), (d), (f), (h).

6.4.3 Cancelability

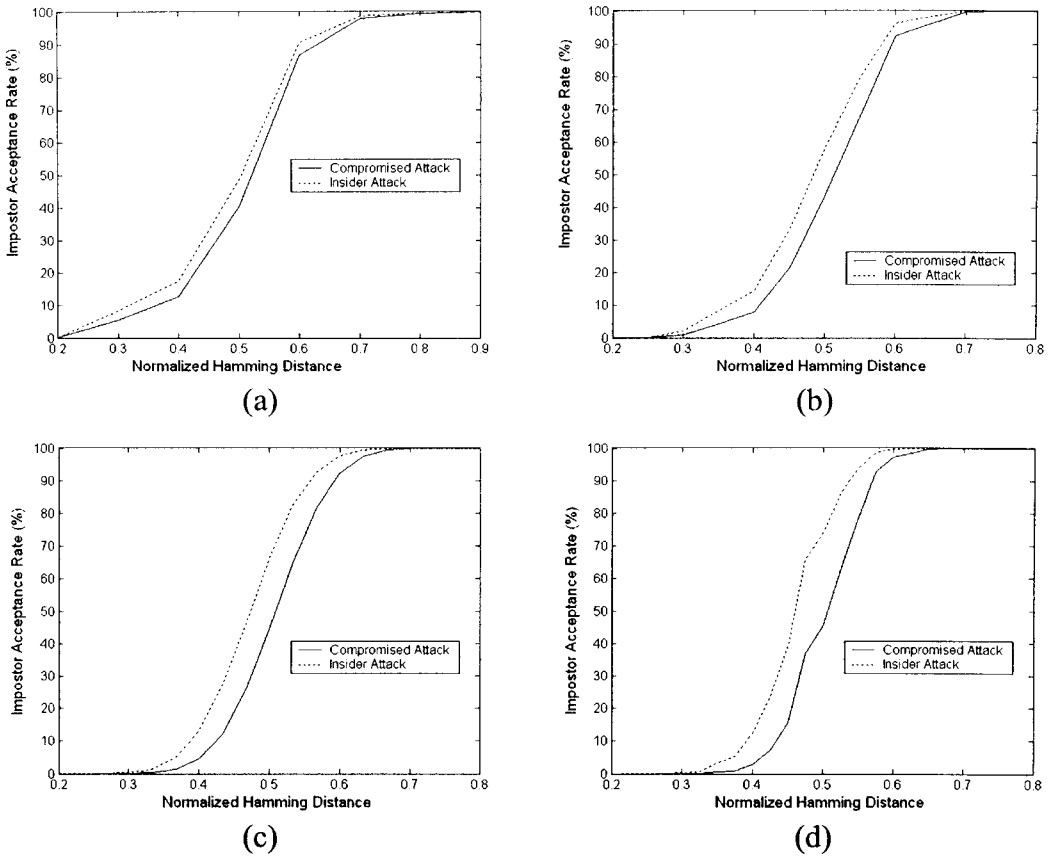


Figure 6.8 Test on cancelability and invertibility of BioCodes.
Impostor distributions of compromised and insider attack with feature dimensions (a) 20 (b) 40 (c) 60 (d) 80

We assume that the biometric template of each individual for generating cancelable templates is identical (see above for more implementation details). One hundred cancelable templates are generated for each individual in our demonstration and ninety-nine templates are compromised at a time. This simulates the situation that compromised cancelable templates of an individual are being used to attack, i.e. to match against the current cancelable template (the remaining one). The number of matching producing impostor distributions due to *Compromised Attack* is 198,000 ($=40\times(100^2 - 100)/2$). The resulting FAR and impostor distributions are plotted as solid lines in Figure 6.8 and Figure 6.9 respectively for BioCode of dimension 20, 40, 60 and 80. From those solid lines in Figure 6.8, those left tails vanishes well beyond

zero indicating a clear distinction between current and compromised templates. Therefore, the cancelability of BioHashing is ensured.

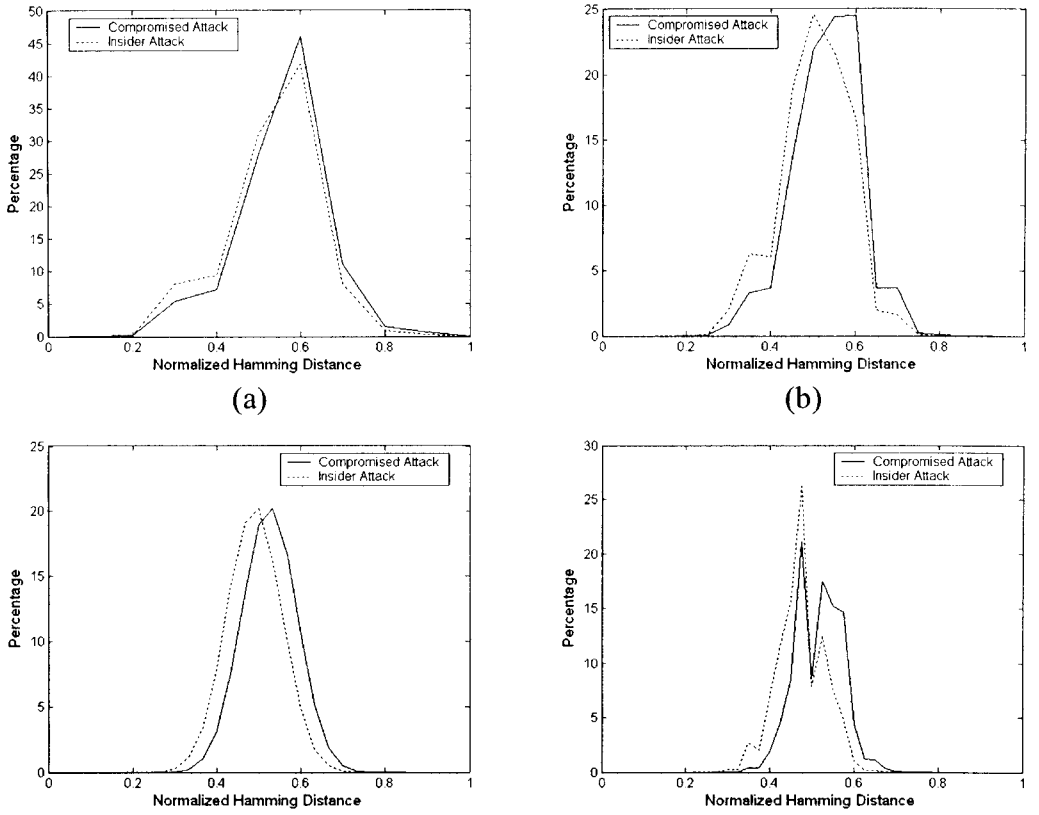


Figure 6.9 Test on cancelability and invertibility of BioCodes.
Impostor acceptance rates of compromised and insider attack with feature dimensions (a) 20 (b) 40 (c) 60 (d) 80

6.4.4 Invertibility

We now examine whether noninvertible random transform in BioHashing can be approximated. It is assumed that the random transform T_i and the feature vector for matching B_i (BioCode for BioHashing) are compromised at one time. Further, the newly issued random transform T_j, is also compromised but it is unknown to be compromised already. So, the attacker has two random transforms and one feature in hand to perform the attack. We have to point out that, in order to make such close estimation, one must know the system well. So we called this type of break-in an *Insider Attack*.

Since the

$$B_i = Q(T_i \cdot F) \text{ and,} \quad (6.1)$$

$$B_j = Q(T_j \cdot F) \quad (6.2)$$

where, F is the biometric feature and $Q(\cdot)$ is the quantization function depicted in Chapter 6.4.1 Step 4. The estimation is done as follows.

$$\hat{B}_j = Q(T_j \cdot (T_i^{-1} B_i')) \quad (6.3)$$

where $\hat{B}_j$ is the estimated feature vector for matching, T_i^{-1} is the pseudo inverse of T_i and B_i' is the resultant vector of the mapping $\{B_i \rightarrow B'_i$: [1] → [1] and [0] → [-1]}

We have used the first image of each subject in ORL database as the biometric feature. For each of them, one hundred random transform is applied. The threshold, τ is set to 0. The acceptance rate of impostors and distribution of impostors due to Insider Attack are plotted with dotted lines in Figure 6.8 and Figure 6.9 respectively for BioCode of dimension 20, 40, 60 and 80. It can be observed that the acceptance rate of impostors and distribution of impostors both shift to the left if Insider attack is adopted. This indicates the biometric system collapse more easily when the details of transformation are known by attackers even if noninvertible transforms are used.

6.4.5 Discussion on the case study

Cancelable biometrics via feature domain transformation, e.g. BioHashing, uses a noninvertible random projection, which does not serve for any objective function. Thus, there is certainly loss of information after the random transform is applied. Although cancelability is achieved, recognition accuracy is traded. If information has to be preserved, invertible transformation, e.g. square random (invertible) transformation matrix, is required. The raw biometric template, however, may be recovered if any of the cancelable templates is compromised.

Noninvertible transformation can provide better security. Yet, we still need to keep the details of the transformation as secrets. Otherwise, the system may be breaking in more easily. In our case study, it is noted that the FAR due to impostors is much larger than that due to insider attack. That is to say the accuracy is sacrifices too much in achieving cancelability. The problem of insider attack exists but it is less critical, compared to accuracy sacrificed in BioHashing.

6.5 Recapitulation

In this chapter, a security concept and its realization evolved initially from the research of biometrics systems known as cancelable biometrics is first introduced. Three issues concerning cancelable biometrics: accuracy, cancelability and invertibility, which have not been addressed before, are then discussed and a case study for better illustrating the issues is given.

Chapter 7 An Analysis of BioHashing and Its Variants

This analysis is motivated by identifying an implicit false assumption that is used for designing a biometric verification system. It has recently been reported that, along with its variants, BioHashing, a new technique that combines biometric features and a tokenized (pseudo-) random number (TRN), has achieved perfect accuracy, having zero equal error rates (EER) for faces, fingerprints and palmprints. There are, however, anomalies in this approach.

There are two classical personal authentication approaches (Jain et al. 1999; Jain et al. 2004): token based approach, which relies on physical items such as smart cards, physical keys and passports and knowledge based approach which replies on private knowledge such as passwords. Both of these approaches have their limitations: it is possible for both “tokens” and “knowledge” to be forgotten, lost, stolen, or duplicated. Further, authorized users may share their “knowledge” and “tokens” with unauthorized users. These limitations do not apply to biometric means of authentication, which identify a person based on physiological characteristics, such as the iris, fingerprint, face or palmprint, and/or by behavioral characteristics, such as a person’s signature or gait (Jain et al. 1999).

Although biometric authentication has several inherent advantages over the classical approaches, all of biometric verification systems make two types of errors (Jain et al. 2004): 1) misrecognizing measurements from two different persons to be from the same person, called false acceptance and 2) misrecognizing measurements from the same person to be from two different persons, called false rejection. The performance of a biometric system is commonly described by its false acceptance

rate (FAR) and false rejection rate (FRR). These two measurements can be controlled by adjusting a threshold, but it is not possible to exploit this threshold by simultaneously reducing FAR and FRR. FAR and FRR must be traded-off, as reducing FAR increases FRR and vice versa. Another important performance index of a biometric system is its equal error rate (EER) defined as at the point where FAR and FRR are equal. A perfect system in terms of accuracy would have EER of zero. Unfortunately, however, over thirty years' investigation, a perfect biometric verification system has not been developed. Numerous biometric researchers thus continue to work in this area, looking for new biometrics, developing new feature representations and matching methods and combining the existing techniques (Chellappa et al. 1995; Bhanu and Tan 2003; Jain et al. 1997; Zhang et al. 2003; Kim et al. 2004; Ross and Jain 2003; Israel et al. 2005).

7.1 An Overview of Biometric Verification System

We focus on biometric systems for verification tasks in which BioHashing and its variants are operated. Figure 7.1 illustrates the operation flow of such biometric systems. To validate the user's claimed identity, verification systems conduct one-to-one comparison using two pieces of information: a claimed identity and biometric data. The input biometric data is compared with biometric templates associated with the claimed identity in a given database. Generally speaking, user identities can be inputted via various devices such as keypads and smart card readers and should be unique to each person, like a primary key in a database. It should be noted that user identities in such forms can be shared, lost, forgotten and duplicated, just like the keys/cards of token-based and the PINs/passwords of knowledge-based authentication systems. Nonetheless, because biometric authentication is also required, to pass through the verification system requires more than the mere

possession of a valid user identity i.e. an impostor cannot gain access unless the input biometric data matches the template of the claimed identity. We have to emphasize here that the performance of a biometric verification system should not depend solely on user identity or its equivalent and therefore, many biometric systems accept obvious user identities such as personal names.

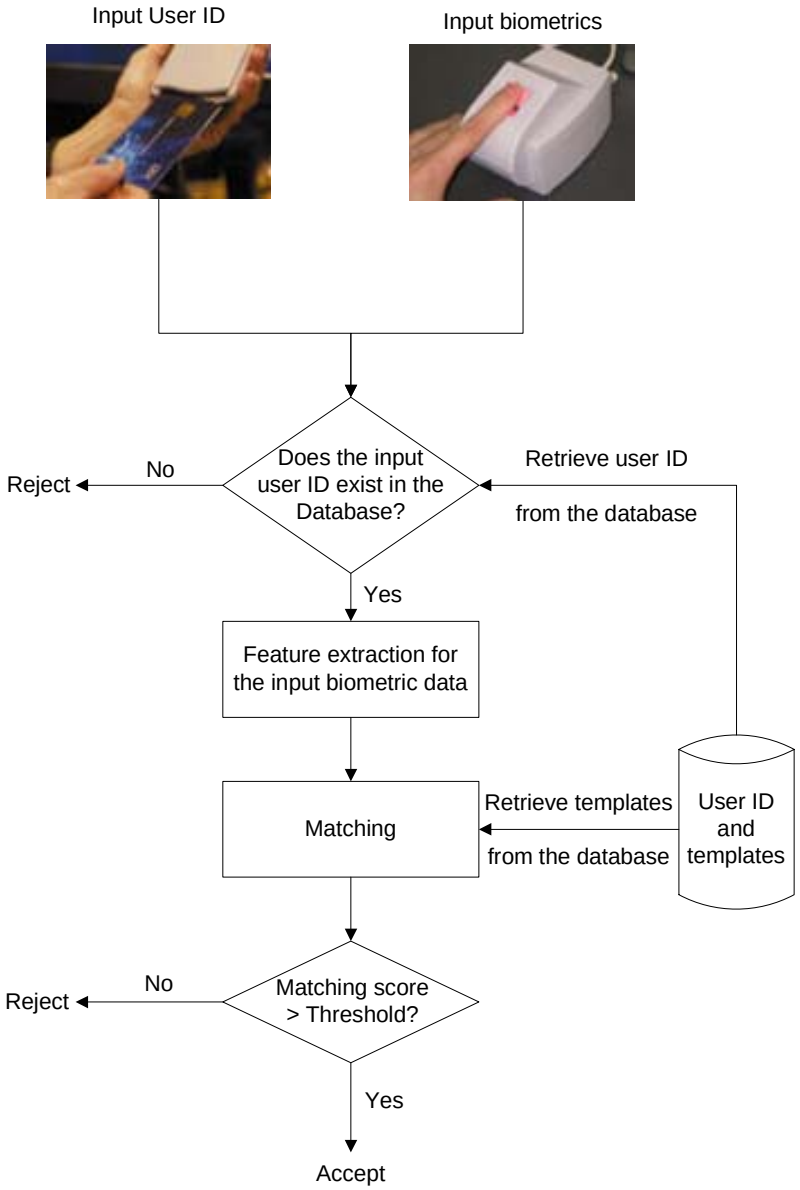


Figure 7.1 Operation flow of a biometric verification system

A biometric verification system has five possible input combinations. For the sake of convenience, we use the notation {*I*, *B*} to represent the pair-wise information, a claimed user identity *I* and its associated biometric data *B*. A registered user *X* stores

his/her user identity and biometric template, $\{I_X, B_{XD}\}$, in a database after enrollment. Suppose user X provides his/her user identity and biometric data, $\{I_X, B_{XV}\}$ at the time of verification. Even though B_{XD} and B_{XV} are from the same person, because of various noises, they are not identical. We also assume that an impostor Y, an unregistered person has an invalid user identity I_Y (which may have been obtained by, for example guessing) and biometric data B_{YV}.

Case 1 $\{I_X, B_{XV}\}$ vs. $\{I_X, B_{XD}\}$. User X inputs his/her user identity and biometric data $\{I_X, B_{XV}\}$ to the system to compare $\{I_X, B_{XD}\}$ in the database. The system would compare B_{XD} and B_{XV} and give a matching score. From this matching score, there are two possible responses: either "correct acceptance" or "false rejection".

Case 2 $\{I_X, B_{YV}\}$ vs. $\{I_X, B_{XD}\}$. Assume an impostor Y has the user identity of X and inputs X's identity together with his/her biometric $\{I_X, B_{YV}\}$ to the system. The input will then be compared with $\{I_X, B_{XD}\}$ in the database. As Case 1, we have two possible responses, "false acceptance" and "correct rejection".

Case 3$\{I_Y, B_{YV}\}$. An impostor Y provides his/her user (invalid) identity and biometric data to the system. Since the user identity, I_Y, does not match any identity in the system, the system simply rejects the user without error and will not attempt to match any biometric information.

Case 4 $\{I\sim_X, B_{XV}\}$. A registered user X inputs a wrong user identity $I\sim_X$, i.e. not his/her valid user identity, I_X. If $I\sim_X$. is a valid user identity, the system would output its decision based on matching the biometric information and the threshold, $\{I\sim_X, B_{XV}\}$ vs. $\{I\sim_X, B\sim_{XD}\}$, as in Case 2. If $I\sim_X$ is an invalid user identity, the system would simply reject the user, as in Case 3.

Case 5 {$NULL$, B_{XV}} OR {$NULL$, B_{YV}}. No matter who operates the verification system, a registered user or an unregistered user, a user identity is required. A biometric verification system cannot operate without user identities, i.e. NULL. It would also be unreasonable to assign a temporary user identity to any user who did not provide a user identity at the time of verification.

From these analyses, we conclude the following: Case 5 is invalid; and testing a verification system on Case 3 produces trivial rejection; and Case 4 can be resolved to be either Case 2 or Case 3 depending on the user identity provided. As it happens, to evaluate the performance of a verification system, all biometric researchers assume a situation in which impostors have obtained a valid user identity. Performance evaluations for genuine distributions are estimated using the matching scores from Case 1, for impostor distributions, those from Case 2.

If “knowledge” or “token” representing the user identity in verification would not be forgotten, lost or stolen, it made the introduction of biometric system less meaningful except for guarding against multiple users using the same identity through sharing or duplicating “knowledge” or “token”. If, further, “knowledge” or “token” would not be shared or duplicated, introducing biometrics became meaningless.

7.2 Details of Experiment

We have re-implemented FaceHashing, one of the variants of BioHashing, to show the anomaly. Publicly available face database, the AR face database (Martinez and Benavente 1998) and The ORL database (Samaria and Harter, 1994), and a well known feature extraction technique, Principal Component Analysis (PCA), also called Eigenface for face recognition (Turk and Pentland 1991; Martinez and Kak 2001; see also Chapter 3.2.1) are chosen for this demonstration so that all the results

reported in this paper are reproducible. We do not employ other effective face recognition algorithms and other accurate biometrics such as fingerprint in order that the performance differences due to the unrealistic assumption can be clearly observed. Although we demonstrate here only FaceHashing, it does not lose any generality to analyze BioHashing and its variants. For a summary of BioHashing, please refer to Chapter 6.4.1. For details of face databases, the AR database and the ORL database, please refer to Chapter 5.2.1 and Chapter 5.2.2 respectively.

The face subimages of the AR database for feature extraction are 96×96 in size and were cropped from the central canvas of a gray scale face image. Since we are only doing a demonstration to facilitate our analysis and discussion, 50 subjects were randomly selected from the face database. Eight images of each subject were selected, four from each session. The facial expression in the four images from each session is neutral but the illuminations vary (see Figure 5.1 (a), (e)-(g), (n), (r)-(t) for sample images and Chapter 5.3.1 Phase 1 for image preparation details).The face images of the ORL database are used directly for feature extraction, i.e. no preparation or preprocessing is done. Sample face images from the ORL database is shown in Figure 7.2.

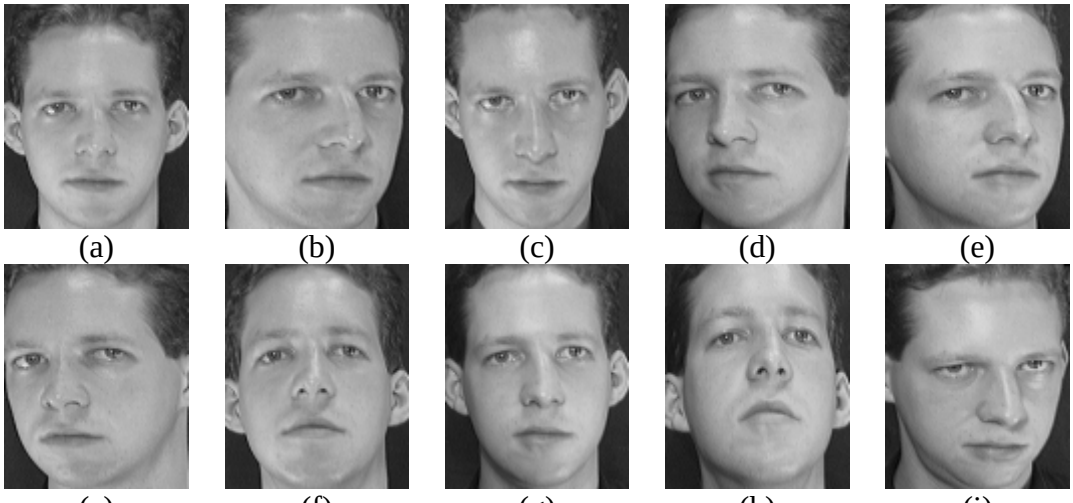


Figure 7.2 Sample face images in the ORL database.

All ten images of all subjects within the ORL database are used to determine the principal components, i.e. treating as same session matching. While for the AR database, the four images from the first session are used to determine the principal components. The four images from second session are matched against those from first session, i.e. matching "duplicate" (Martinez and Kak 2001).

Table 7.1 Thresholds used for various dimensions of BioCode tested on the AR database

BioCode dimension	Threshold for BioCode (τ)
10	0
50	0
100	0
150	0
200	0

Table 7.1 and Table 7.2 list the dimensions of the BioCode and the corresponding thresholds (τ) according to the deduction of made in (Connie et al. 2004a; Teoh et al. 2004b) tested on the AR database and the ORL database respectively.

Table 7.2 Thresholds used for various dimensions of BioCode tested on the ORL database

BioCode dimension	Threshold for BioCode (τ)
10	0
25	0
50	0
75	0
100	0

7.3 Analysis of BioHashing and Its Variants

In our demonstration below, we are able to obtain a zero EER by applying only a simple feature extraction approach, PCA, but in general, even with advanced classifiers, such as support vector machines, PCA is impossible to yield 100% accuracy along with zero EER. Obviously, the high performance of BioHashing is not from the biometric features.

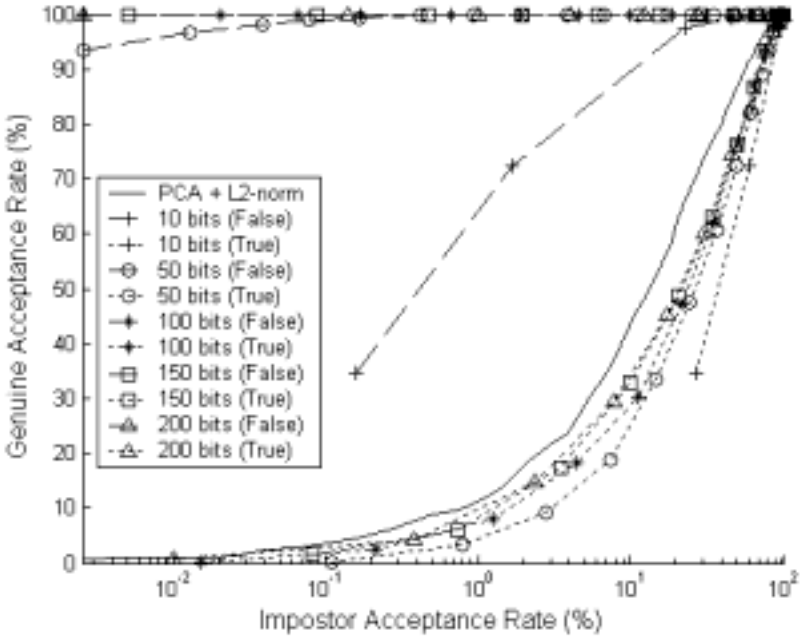


Figure 7.3 Comparison of ROC curves of various dimensions of BioCode under different assumptions tested on the AR database

We simulated FaceHashing (Ngo et al. 2004; Teoh and Ngo 2005; Teoh et al. 2004a, 2004b) with different dimensions of BioCode and their performances are reported in the form of ROC curves in Figure 7.3 and Figure 7.4 respectively for tests on the AR database and the ORL database.

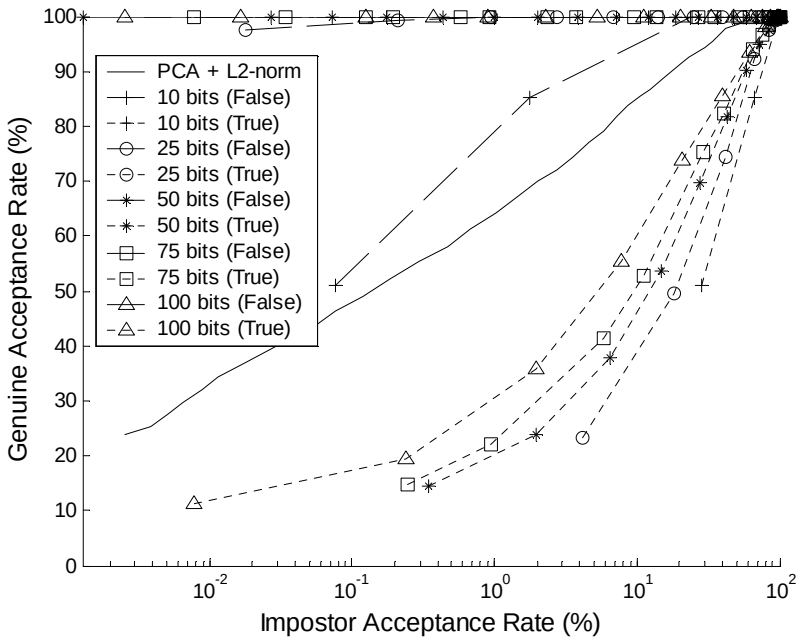


Figure 7.4 Comparison of ROC curves of various dimensions of BioCode under different assumptions tested on the ORL database

7.3.1 The secret of BioHashing and its variants

The authors (Connie et al. 2004a; Ngo et al. 2004; Teoh et al. 2004a, 2004b, 2004c) mentioned that a unique seed among different persons and applications is used to generate a set of pseudo random numbers, called Tokenized Random Number (TRN). It is followed that 1) the seed and thus the TRN for each user used in enrollment and verification is the same; 2) different users (and applications) have different seeds and thus different TRNs. In other words, the seed and TRN are unique across users as well as applications. They also pointed out that the seed for generating the TRN can be stored in a USB token or smart card. Comparing the properties of the seed in BioHashing (and its variants) and the user identity of a biometric verification system as described in Chapter 7.1 above, it is obvious that the seed, and thus the TRN can serve as a user identity. As the seed is stored in a physical media, TRN also suffers from the problems of "token" in traditional authentication methods (Jain et al. 2004), e.g. they can be lost, stolen, shared and duplicated.

The TRN has a central role in BioHashing and its variants and is requisite for achieving zero EER. The authors assume that no impostor has the valid seed/TRN. That is, they assume that the "token" will not be lost, stolen, shared and duplicated. If their assumption is true, introducing any biometric becomes meaningless since the system can rely solely on the "tokens" without any risk. Undoubtedly, their assumption does not hold in general. In their experiments (Connie et al. 2004a, 2004b; Ngo et al. 2004; Pang et al. 2004; Teoh and Ngo 2005; Teoh et al. 2004a, 2004b, 2004c), they determine the genuine distribution using, in our notation, Case 1. However, they determine the impostor distribution using Case 3 in which no biometric should be involved because of the mismatch of with the "pseudo user identity", the seed/TRN.

7.3.2 Performance analysis of BioHashing and its variants

Based on their invalid assumption, it is possible to achieve zero EER provided that the BioCode is long enough, i.e. 75 bits or above for test on the ORL database (same session matching) and 100 bits or above for the test on the AR database ("duplicate" matching). This is shown in Figure 7.3 and Figure 7.4, for all possible operating points, as dashed lines with markers.

In Figure 7.3 and Figure 7.4, the dotted lines with markers are the ROC curves using the general assumption for evaluating a biometric verification system. The solid line without any marker is the ROC curve using PCA and Euclidean distance. It can be observed that the true performance of BioCode can even be worse than that of using PCA and Euclidean distance. This is possibly because BioHashing uses a random projection, which does not serve for any objective function.

7.4 Recapitulation

We have revealed that the outstanding achievements of BioHashing and its variants, zero EER are under a hidden and unpractical assumption — that the TRN would never be lost, stolen, shared or duplicated that does not hold generally. We also point out that if this assumption held, there would be no need for biometrics to combine the TRN since the TRN could serve as a perfect token. To further support our argument, we used a public face database and PCA to simulate their experiments. It is possible to achieve a zero EER by using the combination of a TRN and a biometric under their assumption. Adopting an assumption generally used in the biometric community, our experimental results show that the true performance of BioHashing is far from perfect.

Chapter 8 Conclusion and Future Works

8.1 Conclusion

We have investigated the storage and retrieval of two image-based biometric systems, i.e. the image databases in a narrow domain. Two image domains that have been investigated are palmprint and face.

Regarding palmprint images, we have integrated Principle Component Analysis and Self Organizing Map to generate a better searching sequence. The technique has been tested on the palmprint image database. It is found to be capable of reducing the searching scope and thus saving searching time. We have formulated, moreover, a hierarchical searching scheme based on four different image features and similarity measures that is tested on the palmprint database and found to be effective and yet efficient.

Concerning face images, we have developed a holistic appearance-based feature representation of low dimensionality based on aggregated information extracted by Gabor filters. It has been tested on a publicly available database and demonstrated to be more robust to facial expression variations in comparison to the benchmarking technique, Principle Component Analysis. Of its low feature dimensionality, it is efficient to be computed.

Information security of some systems like biometric systems is vital. Cancelable Biometrics is a newly proposed approach to fight against potential information security threats in the biometric system. Following the idea of Cancelable Biometrics, we raise three issues, that worth discussion and have not been addressed before. The three issues: accuracy, cancelability and invertibility have been illustrated further through a case study of an existing approach to cancelable biometrics.

8.1.1 Major Contributions of This Thesis

The major contributions of this thesis are summarized as follows.

1. We have considered the storage and retrieval of biometric templates and incorporated some primitive visual-based image features, texture, lines and points, as content descriptors of biometric images. Various intelligent system techniques, including learning and fuzzy methods, have been used to extract and represent the visual-based image features of palmprint and face images;

2. We are among the first to consider the retrieval in large palmprint databases and have developed methods that can compactly represent and effectively retrieve palmprint image templates;

3. We originally identify and study three design issues of cancelable biometrics and we are among the first to propose the consideration of the three issues integrally when designing and evaluating cancelable biometrics;

4. We have identified, in the biometrics research community, that a group of researchers have proposed the use of “safe” token. We have given a thorough analysis of their method, BioHashing and its variants, revealing that their assumption is wrong. We have raised the issue so as to avoid the use “safe” token for biometric systems.

8.2 Future Works

Multimodal biometric systems, which incorporate more than one biometric, with appropriate security measures are underway. Storage and retrieval of data in those systems poses challenging requirements on the system design, the choice of features and representations, and the choice of similarity measure and retrieval/recognition method.

Incorporating protective measures on user template for privacy and security in all future biometrics systems is a must. Public are increasingly concern the use of biometrics in daily activities. Protection of such personal information poses new requirements that should be pursued seriously before the systems can be put into practical use.

Appendix 1 List of Publications

Referred Journal Article

1. Kong, A.; Cheung, K.-H.; Zhang, D.; Kamel, M.; You, J., 2005. An Analysis of BioHashing and Its Variants. Pattern Recognition **39**(7), pp. 1359-1368.
2. You, J.; Kong, W.-K.; Zhang, D.; Cheung, K.-H., 2004. On hierarchical palmprint coding with multi-features for personal identification in large databases. IEEE Transactions on Circuit Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics **14**(2), pp. 234-243.

Working Journal Article

3. Cheung, K.-H.; Kong A.; Zhang, D.; Kamel, M.; You J., 2005. On Cancelable Biometrics for Personal Authentication. To be submitted.
4. Cheung, K.-H.; Kong A.; Zhang, D.; Kamel, M.; You J., 2005. A comparative study of Gabor features for biometric recognition. To be submitted.
5. Cheung, K.-H.; Kong A.; You J.; Zhang, D.; Kamel, M., 2006. Evaluating Eigenpalm from an Application Perspective. Submitted to Pattern Recognition.

Book Chapter

6. Cheung, K.-H.; Kong, A.; Zhang, D.; Kamel, M.; You, J., 2006. Revealing the secret of FaceHashing. Lecture Notes in Computer Science **3832**, Springer-Verlag GmbH, pp.106-112.
7. You, J.; Li, Q.; Cheung, K.H.; Bhattacharya, P., 2005. An Integration of Biometrics and Mobile Computing for Personal Identification. Lecture Notes in Computer Science **3687**, Springer-Verlag GmbH, pp. 226-235.

8. Cheung, K.-H.; Kong, A.; Zhang, D.; Kamel, M.; You, J.; Lam, H.-W.T., 2005. An Analysis on Accuracy of Cancelable Biometrics based on BioHashing. Lecture Notes in Computer Science **3683**, Springer-Verlag GmbH, pp.1168-1172.
9. Cheung, K.-H.; You, J.; Liu, J.; Ao Ieong, W.H.T., 2004. Appearance-based Face Recognition using Aggregated 2D Gabor Features. Lecture Notes in Computer Science **3214**, Springer-Verlag GmbH, pp. 572-579.
10. You, J.; Kong, W.K.; Zhang, D.; Cheung, K.-H., 2004. A new approach to personal identification in large database by hierarchical palmprint coding with multi-features. Lecture Notes in Computer Science **3072**, Springer-Verlag GmbH, pp. 739-745.

Referred Conference Article

11. Cheung, K.H.; You, J.; Li, Q.; Bhattacharya, P., 2005. A new approach to appearance-based face recognition. Proc. of the IEEE International Conference on Systems, Man and Cybernetics (IEEE SMC'2005) **2**, pp. 1686-1691.
12. Cheung, K.H.; You, J.; Li, Q.; Bhattacharya, P., 2005. On Aggregated 2D Gabor Features for Appearance-based Face Recognition. Proc. of The 2005 International Conference on Imaging Science, Systems, and Technology: Computer Graphics (CISST'2005), pp. 33-39.
13. Cheung, K.-H.; Kong, A.; You, J.; Zhang, D., 2005. An Analysis on Invertibility of Cancelable Biometrics based on BioHashing. Proc. of The 2005 International Conference on Imaging Science, Systems, and Technology: Computer Graphics (CISST'2005), pp. 40-45.
14. You, J.; Cheung, K.H.; Kong, W.K.; Li, Q.; Bhattacharya, P., 2005. Dynamic feature selection and coarse-to-fine search for content-based image retrieval.

Proc. of the Workshop on Pattern Recognition in Information Systems (PRIS'2005), pp. 3-11.

15. Cheung, K.H.; You, J.; Li, Q.; Bhattacharya, P., 2005. Appearance-based face recognition using aggregated 2D Gabor features. Proc. of the Workshop on Pattern Recognition in Information Systems (PRIS'2005), pp. 81-93.
16. Cheung, K.H.; You, J.; Kong, W.K.; Zhang, D., 2004. A Study of Aggregated 2D Gabor Features on Appearance-based Face Recognition. Proc. of the Third International Conference on Image and Graphics (ICIG 2004), pp. 310-313.
17. You, J.; Cheung, K.H., 2004. WebGuard: A new approach to biometrics-based personal identification using mobile agents. Proc. of 2004 International Conference on Image Science, Systems and Technology (CISST'2004), pp. 223-229.
18. Cheung, K.H.; Kong, W.K.; You, J.; Zhang, D., 2003. An integration of principal component analysis and self-organizing map for effective palmprint retrieval. Proc. of the 16th International Conference on Computer Applications in Industry and Engineering (CAINE-2003), pp. 101-104.
19. You, J.; Kong, W.K.; Zhang, D.; Cheung, K.H., 2003. On hierarchical palmprint coding with multi-features for personal identification in large databases. Proc. of the 16th International Conference on Computer Applications in Industry and Engineering (CAINE-2003), pp. 105-107.
20. Cheung, K.H.; Kong, W.K.; You, J.; Zhang, D., 2003. On Effective Palmprint Retrieval for Personal Identification. Proc. of 2003 International Conference on Imaging Science, Systems, and Technology (CISST'2003), pp. 111-117.

21. You, J.; Cheung, K.H.; Liu, J., 2003. A hierarchical approach to content-based image retrieval. Proc. of 2003 International Conference on Imaging Science, Systems, and Technology (CISST'2003), pp. 127-133.
22. You, J.; Zhang, D.; Cheung, K.H.; Kong, W.K., 2003. A new approach to personal identification via hierarchical palmprint coding. Proc. of 2003 International Conference on Imaging Science, Systems, and Technology (CISST'2003), pp. 531-537.
23. You, J.; Cheung, K.H.; Li, L.; Liu, J., 2002. A new approach to content-based image retrieval. Proc. of the 15th International Conference on Computer Applications in Industry and Engineering (CAINE-2002), pp. 53-56.
24. You, J.; Liu, J.; Li, L.; Cheung, K.H., 2002. Smart eShopping using mobile agent and web-mining technology. Proc. of the 15th International Conference on Computer Applications in Industry and Engineering (CAINE-2002), pp. 150-153.
25. You, J.; Liu, J.; Li, L.; Cheung, K.H., 2002. On data mining and data warehousing for multimedia information retrieval. Proc. of IASTED International Conference on Artificial and Computational Intelligence (ACI'02).
26. You, J.; Cheung, K.H.; Liu, J., 2002. On similar shape retrieval by coarse-to-fine curve matching. Proc. of 2002 International Conference on Image Science, Systems and Technology (CISST'2002) **II**, pp. 451-457.
27. Cheung, K.H.; Dillon, T.; You, J.; Liu, J., 2001. ColorGuide: A wavelet-based data warehousing approach to hierarchical image retrieval. Proc. of International Conference on Image Science, Systems and Technology (CISST'2001), pp. 386-392.

References

Ahuja, N.; Schachter, B.J., 1981. Image Models. ACM Computing Surveys **13**(4), pp.373-397.

Albanesi, M.G.; Bertoluzza, S., 1995. Human Vision Model and Wavelets for High-quality Image Compression. Proc. of Eighth International Conference on Image Processing And Its Applications, pp. 311-315.

Albanesi, M. G.; Ferretti, M.; Giancane, A., 1999. A Compact Wavelet Index for Retrieval in Image Database. Proc. of Tenth International Conference on Image Analysis and Processing, pp. 927-932.

Albuz, E.; Kocalar, E.; Khokhar, A.A., 1999. Vector-Wavelet Based Scalable Indexing and Retrieval System for Large Color Image Archives. Proc. of 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing **6**, pp. 3021-3024.

Baltscheffsky, P.; Anderson, P., 1986. The palmprint project: Automatic identity verification by hand geometry. Proc. 1986 International Carnahan Conference on Security Technology, Gothenburg, Sweden, pp. 228-235.

Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D.J., 1997. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Machine Intell. **19**(7), pp. 711-720.

Benke, K.K.; Skinner, D.R.; Woodruff, C.J., 1988. Convolution operators as a basis for objective correlates for texture perception. IEEE Trans. Syst. Man, and Cybern. **18**(1), pp. 158-163.

Beymer, D., 1995. Vectorizing Face Images by Interleaving Shape and Texture Computations. AI Memo No. 1537, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Bhanu, B.; Tan, X., 2003. Fingerprint indexing based on novel features of minutiae triplets. IEEE Trans. Pattern Anal. Machine Intell. **25**(5), pp. 616-622.

Böhm, C.; Berchtold, S.; Keim, D.A., 2001. Searching in High-Dimensional Spaces — Index Structures for Improving the Performance of Multimedia Databases. ACM Computing Surveys **33**(3), pp.322-373.

Borowski, M.; Bröcker, L.; Heisterkamp, S.; Löffler, J., 2000. Structuring the Visual Content of Digital Libraries Using CBIR Systems. Proc. of IEEE International Conference on Information Visualization, pp.288-293.

Bovic, A. C.; Clark, M.; Geisler, W. S., 1990. Multichannel texture analysis using localized spatial filters. IEEE Trans. Pattern Anal. Machine Intell. **12**(1), pp. 55-73.

Braithwaite, M.; von Seelen, U.C.; Cambier, J.; Daugman, J.; Class, R.; Moore, R.; Scott, I., 2002. Applications-Specific Biometric Template. Proc. of IEEE Workshop on Automatic Identification Advanced Technologies, Tarrytown, pp. 167-171.

Burns, J.B.; Hanson, A.R.; Riseman, E.M., 1986. Extracting straight lines. IEEE Trans. Pattern Anal. Machine Intell. **8**(4), pp. 425-455.

Carson, C.; Belongie, S.; Greenspan, H.; Malik, J., 1997. Region-Based Image Querying. Proc. of IEEE Workshop on Content-Based Access of Image and Video Libraries, pp.42-49.

Casasent, D.; Psaltis, D., 1976. Position, rotation, and scale invariant optical correlation. Applied Optics **15**(7), pp. 1795-1799.

Chang, S.-F.; Smith, J.R.; Beigi, M.; Benitez, A., 1997. Visual Information Retrieval from Large Distributed Online Repositories. Communications of the ACM **40**(12), pp. 63-71.

Chávez, E.; Navarro, G.; Baeza-Yates, R.; Marroquín, J.L., 2001. Searching in Metric Spaces. ACM Computing Surveys **33**(3), pp.237-321.

Chellappa, R.; Wilson, C.L.; Sirohey, A., 1995. Human and machine recognition of faces: A survey. Proc. IEEE **83**(5), pp. 705-740.

Chen, T., 1998. The Past, Present, and Future of Image and Multidimensional Signal Processing. IEEE Signal Processing Magazine **15**(2), pp. 21-58.

Chua, T.-S.; Pung, H.-K.; Lu G.-J.; Jong, H.-S., 1994. A Concept-Based Image Retrieval System. Proc. of the Twenty-Seventh Annual Hawaii International Conference on System Sciences, pp. 590-598.

Clancy, T.C.; Kiyavash, N.; Lin D.J., 2003. Secure Smartcard-Based Fingerprint Authentication. Proc. of ACM SIGMM 2003 Multimedia, Biometrics Methods and Applications Workshop, pp. 45-52.

Connie, T.; Teoh, A.; Goh, M.; Ngo, D., 2004a. PalmHashing: A Novel Approach for Dual-Factor Authentication. Pattern Analysis and Application **7**(3), pp. 255-268.

Connie, T.; Teoh, A.; Goh, M.; Ngo, D., 2004b. PalmHashing: a novel approach to cancelable biometrics. Information Processing Letter **93**(1), pp. 1-5.

Cox, I.J.; Miller, M.L.; Minka, T.P.; Papathomas, T.V.; Yianilos, P.N., 2000. The Bayesian Image Retrieval System, PicHunter: Theory, Implementation, and Psychophysical Experiments. IEEE Trans. Image Processing, **9**(1), pp.20-37.

Daugman, J.G.., 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. Journal of the Optical Society of America A. **2**(7), pp. 1160-1169.

Dawson, B., 2003. Biometrics and Machine Vision. Vision Systems Designs. (available online at http://downloads.pennnet.com/pnet/surveys/vsd/vsd03_dawson_biometrics.pdf)

Draper, B.A.; Baek, K.; Bartlett, M. S. and Beveridge, J. R., 2003. Recognizing faces with PCA and ICA. Computer Vision and Image Understanding 91(1-2), pp. 115-137.

Duchene, J.; Leclercq, S., 1988. An Optimal Transformation for Discriminant and Principal Component Analysis. IEEE Trans. Pattern Anal. Machine Intell. **10**(6), pp. 978-983.

Duta, N.; Jain, A.K.; Mardia, K.V., 2001. Matching of palmprint. Pattern Recognition Letters **23**(4), pp. 477-485.

Enser, P.G.B., 1995. Pictorial Information Retrieval. Journal of Documentation **51**(2), pp. 126-170.

Flickner, M.; Sawhney, H.; Niblack, W.; Ashley, J., 1995. Query by image and video content: The QBIC system. IEEE Computer **28**(9), p. 23-32.

Foley, D.H.; Sammon Jr., J.W., 1975. An Optimal set of discriminant vectors. IEEE Trans. Comput. **24**(3), pp. 281-289.

Forsyth, D.A.; Fleck, M.M., 1999. Automatic Detection of Human Nudes. International Journal of Computer Vision **32**(1), pp. 63-77.

Frederix, G.; Caenen, G.; Pauwels, E.J., 2000. PARISS: Panoramic Adaptive and Reconfigurable Interface for Similarity Search. Proc. of International Conference on Image Processing **3**, pp. 222-225.

Gabor, D., 1946. Theory of Communications. Journal of Institution of Electrical Engineers 93, pp. 429-457.

Garcia, C.; Tziritas, G., 1999. Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis. IEEE Trans. Multimedia **1**(3), pp. 264-277.

Gevers, T.; Smueulders, A.W.M., 2000. Pictoseek: Combining Color and Shape Invariant Features for Image Retrieval. IEEE Trans. Image Processing **9**(1), pp.102-119.

Gong, Yihong, 1998. Chapter 2: Survey of Contemporary Content-Based Image Retrieval Systems. Intelligent Image Database: Toward Advanced Image Retrieval, Boston: Kluwer Academic Publishers.

Gonzalez, R.C.; Woods, R.E., 1992. Digital Image Processing, Addison Wesley.

Gorsky, W. I.; Mehrotra, R. (Guest Editors), 1989. Image Database Management. IEEE Computer **22**(12), pp. 7-8.

Gudivada, V. N.; Raghavan, V. V. (Guest Editors), 1995. Content-Based Image Retrieval Systems. IEEE Computer **28**(9), pp. 18-22.

Gupta, A.; Jain, R., 1997. Visual Information Retrieval. Communications of the ACM **40**(5), pp. 71-79.

Halici, U.; Ongun, G., 1996. Fingerprint Classification Through Self-Organizing Feature Maps Modified to Treat Uncertainties. Proc. IEEE **84**(10), pp. 1497-1512.

Han, C.C.; Cheng, H.L.; Fan, K.C.; Lin, C.L., 2002. Personal authentication using palmprint features. Pattern Recognition **36**(2), pp. 371-382.

Haralick, R.M.; Shanmugam, K.; Dinstein, I., 1973. Texture features for image classification. IEEE Trans. Syst. Man, and Cybern. 3(6), pp. 610-621.

Haralick, R.M., 1979. Statistical and structural approaches to texture. Proc. IEEE **67**, pp. 786-804.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation, 2nd Edition, Upper Saddle River, N.J.: Prentice Hall.

Heczko, M.; Hinneburg, A.; Keim, D.; Wawryniuk, M., 2004. Multiresolution similarity search in image database. Multimedia Systems **10**(1), pp. 28-40.

Heisele, B.; Ho, P.; Wu, J.; Poggio, T., 2003. Face Recognition: component-based versus global approaches. Computer Vision and Image Understanding **91**(1-2), pp. 6-21.

Hiroike, A.; Musha, Y.; Sugimoto, A.; Mori, Y., 1999. Visualization of Information Spaces to Retrieve and Browse Image Data. Proc. of the Third International Conference on Visual Information and Information Systems, pp. 155-162.

Hong, L.; Wan, Y.; Jain A., 1998. Fingerprint Image Enhancement: Algorithm and Performance Evaluation. IEEE Trans. Pattern Anal. Machine Intell. **20**(8), pp. 777-789.

Huet, B.; Hancock, E.R., 1999. Line Pattern Retrieval Using Relational Histograms. IEEE Trans. Pattern Anal. Machine Intell. **21**(12), pp.1363-1370.

Israel, S.A.; Irvine, J.M.; Cheng, A.; Wiederhold, M.D.; Wiederhold, B.K., 2005. ECG to identify individuals. Pattern Recognition **38**(1), 133-142.

Jain, A.; Hong, L.; R. Bolle, 1997. On-line fingerprint verification. IEEE Trans. Pattern Anal. Machine Intell. **19**(4), pp. 302-314.

Jain, A.K.; Vailaya, A., 1998. Shape-Based Retrieval: A Case Study with Trademark Image Databases. Pattern Recognition **31**(9), pp. 1369-1390.

Jain, A.; Bolle, R.; Pankanti, S., 1999. Biometrics: Personal Identification in Networked Society, Kluwer Academic Publishers.

Jain, A.K.; Duin, R.P.W.; Mao, J., 2000a. Statistical Pattern Recognition: A Review. IEEE Trans. Pattern Anal. Machine Intell. **22**(1), pp. 4-37.

Jain, A.K.; Prabhakar S.; Hong, L.; Pankanti, S., 2000b. Filterbank-Based Fingerprint Matching. IEEE Trans. Image Processing **9**(5), pp.846-859.

Jain, A.K.; Ross, A.; Prabhakar, S., 2004. An Introduction to Biometric Recognition. IEEE Trans. Circuits Syst. Video Technol. **14**(1), pp. 4-20.

Jin, Z.; Yang, J.-Y.; Hu, Z.-S.; Lou, Z., 2001a. Face recognition based on the uncorrelated discriminant transformation. Pattern Recognition **34**(7), pp. 1405-1416.

Jin, Z.; Yang, J.-Y.; Tang, Z.-M.; Hu, Z.-S., 2001b. A theorem on the uncorrelated optimal discriminant vectors. Pattern Recognition **34**(10), pp. 2041-2047.

Juang, B.H.; Chen, T., 1998. The Past, Present, and Future of Speech Processing. IEEE Signal Processing Magazine **15**(3), pp. 24-48.

Juels, A.; Wattenberg, M., 1999. A Fuzzy Commitment Scheme. Proc. of 6th ACM Conference of Computer and Communications Security, pp. 28-36.

Juels, A.; Sudan, M., 2002. A Fuzzy Vault Scheme. Proc. of IEEE International Symposium on Information Theory, pp. 408.

Jung, Keechul; Kim, Kwang In; Jain, Anil K., 2004. Text Information Extraction in Images and Video: A survey. Pattern Recognition, **37**(5), pp. 977-997.

Karlekar, J.; Desai, U.B., 2000. Finding Faces in Wavelet Domain for Content-based Coding of Color Images. Proc. of 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing **6**, pp. 2338-2341.

Kato, T.; Kurita, T.; Otsu, N.; Hirata, K., 1992. A Sketch Retrieval Method for Full Color Image Database — Query by Visual Example. Proc. of International Conference on Pattern Recognition **1**, pp. 530-533.

Kim, H.-K., 1996. Efficient Automatic Text Location Method and Content-Based Indexing and Structuring of Video Database. Journal of Visual Communication and Image Representation, **7**(4), pp. 336-344.

Kim, J.O.; Lee, W.; Hwang, J.; Baik, K.S.; Chung, C.H., 2004. Lip print recognition for security systems by multi-resolution architecture, Future Generation Computer Systems **20**(2), pp. 295-301.

Kohonen, T.; Oja, E.; Simula O.; Visa, A.; Kangas, J., 1996. Engineering Applications of Self-Organizing Map. Proc. IEEE **84**(10), pp.1358-1384.

Kohonen, T., 1997. Self-Organizing Maps, 2nd Edition, Berlin: Springer-Verlag.

Kong, A.W.-K., 2002. Using Texture Analysis on Biometric Technology for Personal Identification, MPhil. Thesis. The Hong Kong Polytechnic University.

Lai, J.H.; Yuen, P.C.; Feng, D.C., 2001. Face recognition using holistic Fourier invariant features. Pattern Recognition **34**(1), pp. 95-109.

Laws, K.I., 1980. Textured Image Segmentation, PhD. Thesis. University of Southern California.

Lee, T.S., 1996. Image representation using 2D Gabor wavelet. IEEE Trans. Pattern Anal. Machine Intell. **18**(10), pp. 959-971.

Lee, C.-J.; Wang, S.-D., 1999. Fingerprint feature extraction using Gabor filters. Electronics Letters **35**(4), pp. 288-290.

Liang, K.C.; Kuo C.C.J., 1997. Progressive Image Indexing and Retrieval Based on Embedded Wavelet Coding. Proc. of International Conference of Image Processing, pp. 572-575.

Liang, K.C.; Kuo C.C.J., 1999. WaveGuide: A Joint Wavelet–Based Image Representation and Description System. IEEE Trans. Image Processing **8**(11), pp.1619-1629.

Liu, C.; Wechsler, H., 2002. Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition. IEEE Trans. Image Processing **11**(4), pp. 467-476.

Liu, C.; Wechsler, H., 2003. Independent Component Analysis of Gabor Features for Face Recognition. IEEE Trans. Neural Networks, **14**(4), pp. 919-928.

Lu, G., 2001. Indexing and Retrieval of Audio: A Survey. Multimedia Tools and Applications 15(3), pp. 269-290.

Lu, J.; Plataniotis, K.N.; Venetsanopoulos, A.N., 2003. Face Recognition Using LDA-Based Algorithms. IEEE Trans. Neural Networks, **14**(1), pp. 195-200.

Ma, W.-Y.; Zhang, H.J., 1998. Benchmarking of Image Features for Content-based Retrieval. Conference Record of the Thirty-Second Asilomar Conference on Signals, Systems & Computers **1**, pp.253-257.

Ma, W.-Y.; Manjunath, B.S., 2000. EdgeFlow: A Technique for Boundary Detection and Image Segmentation. IEEE Trans. Image Processing **9**(8), pp. 1375-1388.

Mandal M.K.; Aboulnasr T.; Panchanathan S., 1996. Image Indexing Using Moments and Wavelets. IEEE Trans. Consumer Electron. **42**(3), pp. 557-565.

Martinez, A.M.; Benavente, R., 1998. The AR face database. CVC Tech. Report #24 see also: http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html

Martinez, A.M.; Kak, A.C., 2001. PCA versus LDA. IEEE Trans. Pattern Anal. Machine Intell. **23**(2), pp. 228-233.

Martinez, A.M., 2002. Recognizing Imprecisely Localized, Partially Occluded, and Expression Variant Faces from a Single Sample per Class. IEEE Trans. Pattern Anal. Machine Intell. **24**(6), pp. 748-763.

McCabe, A., February 1997. A Theory of Spatio-Chromatic Image Encoding, PhD. Thesis, School of computing Curtin University of Technology.

Melzer, T., 2004. Singular Value Decomposition—
http://www.prip.tuwien.ac.at/~melzer/lehre/statpr/apponly.pdf

Miller, B., 1994. Vital signs of identity. IEEE Spectr. **31**(2), pp. 22-30.

Mirmehdi, M.; Petrou, M., 2000. Segmentation of Color Texture. IEEE Trans. Pattern Anal. Machine Intell. **22**(2), pp.142-159.

Misiti, M.; Misiti, Y.; Oppenheim, G.; Poggi, J.-M., 1996. Wavelet Toolbox For Use with MATLAB®.

Mitra, S.; Pal, S.K., 1994. Self-Organizing Neural Network As A Fuzzy Classifier. IEEE Trans. Syst., Man and Cybern. **24**(3), pp. 385-399.

Mojsilovic, A.; Kovacevic, J.; Hu, J.; Safranek, R.J.; Ganapathy, S.K., 2000. Matching and Retrieval Based on the Vocabulary and Grammar of Color Patterns. IEEE Trans. Image Processing **9**(1), pp.38-54.

Narasimhalu, A.D.; Kankanhalli, M.S.; Wu, J., 1997. Benchmarking Multimedia Databases. Multimedia Tools and Applications **4**(3), pp. 333-356.

NEC Automatic Palmprint Identification System—
http://www.nectech.com/afis/download/PalmprintDtsht.q.pdf, 2003.

Ngo, D.C.L.; Teoh, A.B.J.; Goh, A., 2004. Eigenspace-based face hashing, Proc. of the First International Conference on Biometric Authentication (ICBA 2004), Hong Kong, pp. 195-199.

Pang, Y.H.; Teoh, A.B.J.; Ngo, D.C.L., 2004. Palmprint based cancelable biometric authentication system. International Journal of Signal Processing, **1**(2), pp. 98-104.

Pauwels, E.J.; Frederix, G., 2000. Nonparametric Clustering for Image Segmentation and Grouping. Image Understanding **75**(1), pp. 73-85.

Pentland, A.; Picard, R.W.; Sclaroff, S., 1996. Photobook: tools for content-based manipulation of image databases. International Journal of Computer Vision **18**(3), pp. 233-254.

Phillips, P.J.; Martin, A.; Wilson, C.L.; Przybocki, M., 2000a. An Introduction to Evaluating Biometric Systems. IEEE Computer **33**(2), pp. 56-63

Phillips, P.J.; Moon, H.; Rauss, P.J.; Rizvi, S., 2000b. The FERET Evaluation Methodology for Face Recognition Algorithms. IEEE Trans. Pattern Anal. Machine Intell. **22**(10), pp. 1090-1104.

Picard, R.W.; Minka, T.P., 1995. Vision Texture for Annotation. Multimedia Systems **3**(1), pp. 3-14.

Prabhakar, S.; Pankanti, S.; Jain, A.K., 2003. Biometric Recognition: Security and Privacy Concerns. IEEE Security Privacy **1**(2), pp. 33-42.

Printrak Automatic Palmprint Identification System— http://www.printrakinternational.com/omnitrak.htm, 2003.

Ratha, N.K.; Karu, K.; Chen, S.; Jain, A.K., 1996. A Real-Time Matching System for Large Fingerprint Databases. IEEE Trans. Pattern Anal. Machine Intell. **18**(8), pp.799-813.

Ratha, N.K.; Connell, J.H.; Bolle, R.M., 2001. Enhancing security and privacy in biometrics-based authentication systems. IBM Systems Journal **40**(3), pp. 614-634.

Ratha, N.K.; Connell, J.H.; Bolle, R.M., 2003. Biometrics break-ins and band-aids. Pattern Recognition Letters **24**(13), 2105-2113.

Ross A.; Jain, A.K., 2003. Information Fusion in Biometrics, Pattern Recognition Letters **24**(13), pp. 2115-2125.

Rubner, Y., 1999. Perceptual Metrics For Image Database Navigation, PhD. Thesis, Stanford University

Rudnicky, A.I.; Hauptmann, A.G.; Lee, K.-F., 1997. Survey of Speech Technology. Communications of the ACM **37**(3), pp. 52-73.

Rui, Y.; Huang, T.S.; Mehrotra S., 1997a. Content-based Image Retrieval with Relevance Feedback in MARS. Proc. of IEEE International Conference on Image Processing **2**, pp. 815-818.

Rui, Y.; Huang, T.S.; Mehrotra S.; Ortega, M., 1997b. A Relevance Feedback Architecture for Content-based Multimedia Information Retrieval Systems, Proc. of IEEE Workshops on Content-based Access of Image and Video Libraries, pp. 82-89.

Rui, Y.; Huang, T.S.; Chang, S.F., 1999. Image Retrieval: Current Techniques, Promising Directions, and Open Issues. Journal of Visual Communication and Image Representation **10**(1), pp. 39-62.

Saha, S.; Vemuri, R., 1999. Adaptive Wavelet Coding of Multimedia Images. Proc. of the Seventh ACM International Conference (Part 2) on Multimedia 1999, pp. 71-74.

Samaria, F.; Harter, A., 1994. Parameterisation of a stochastic model for human face identification. Proc. of the 2nd IEEE Workshop on Applications of Computer Vision, Sarasota (Florida), pp. 138-142. (paper and ORL face database both available online at http://www.uk.research.att.com/facedatabase.html)

Sanchez-Reillo R.; Sanchez-Avilla C.; Gonzalez-Marcos A., 2000. Biometric identification through hand geometry measurements, IEEE Trans. Pattern Anal. Machine Intell. **22**(10), pp. 1168-1171.

Savvides, M.; Vijaya Kumar, B.V.K.; Khosla, P.K., 2004. Cancelable biometric filters for face recognition. Proc. of the 17th International Conference on Pattern Recognition (ICPR'2004) **3**, pp. 922-925.

Schmid, C.; Mohr, R., 1997. Local Grayvalue Invariants for Image Retrieval. IEEE Trans. Pattern Anal. Machine Intell. **19**(5), pp.530-535.

Schneier, B., 1999. The Uses and Abuses of Biometrics. Communications of the ACM **42**(8), 136.

Sebe, N.; Tian, Q.; Loupias, E.; Lew, M.S.; Huang, T.S., 2000. Color Indexing Using Wavelet-based Salient Points. Proc. of IEEE Workshop on Content-based Access of Image and Video Libraries, pp. 15-19.

Servetto, S.; Ramchandran, K.; Huang, T. S., 1997. A Successively Refinable Wavelet-based Representation for Content-based Image Retrieval. Proc. of IEEE First Workshop on Multimedia Signal Processing, pp.325-330.

Sheng, Y.; Arsenault, H.H., 1986. Experiments on pattern recognition using invariant Fourier-Mellin descriptors. Journal of Optical Society of America A. **3**(6), pp. 771-776.

Shi, W.; Rong, G.; Bain Z.; Zhang, D., 2001. Automatic palmprint verification. International Journal of Image and Graphics **1**(1), pp. 135-152.

Shu, W.; Zhang, D., 1998. Automated personal identification by palmprint. Optical Engineering **37**(8), pp. 2659-2362.

Smeulders, A.W.M.; Kersten, M.L.; Gevers, T., 1998. Crossing the Divide between Computer Vision and Data Bases in Search of Image Databases. Proc. of the IFIP TC2/WG 2.6 Fourth Working Conference on Visual Database Systems, pp.223-239.

Smeulders, A.W.M.; Worring, M.; Santini, S.; Gupta, A.; Jain, R., 2000. Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Machine Intell. **22**(12), pp. 1349 -1380.

Smith, S.M.; Li, C.-S., 1998. Image Retrieval Evaluation. Proc. of the IEEE Workshop on Content-Based Access of Image and Video Libraries, pp. 112-113.

Smith, J.R.; Chang, S.-F., 1996. Automated binary texture feature sets for image retrieval. Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing **4**, pp. 2239-2242.

Soutar, C.; Roberge, D.; Stoianov, A.; Gilroy, R.; Vijaya Kumar, B.V.K., 1999. Biometric Encryption. In: Nichols, R.K. (ed.), ICSA Guide to Cryptography. McGraw-Hill, New York, Chapter 22, 649-675.

(available online at http://www.bioscrypt.com)

Srinivasa Reddy, B.; Chatterji, B.N., 1996. An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration. IEEE Trans. Image Processing **5**(8), pp. 1266-1271.

Strang, G.; Nguyen, T., 1996. Wavelets and Filter Banks, Wellesley-Cambridge Press.

Stricker M.; Orengo, M., 1995. Similarity of Color. Images Proc. of SPIE – storage and Retrieval for Image and Video Databases III **2420**, San Jose, pp.381-392.

Swain, M.J.; Ballard, D.H., 1991. Color Indexing. International Journal of Computer Vision **7**(1), pp.11-32.

Swets, D.; Weng, J., 1996. Using discriminant eigenfeatures for image retrieval. IEEE Trans. Pattern Anal. Machine Intell. **18**(8), pp. 831 -836.

Teoh, A.B.J.; Ngo, D.C.L., 2005. Cancellable biometrics featuring with tokenized random number. Pattern Recognition Letters **26**(10), pp 1454-1460.

Teoh, A.B.J.; Ngo, D.C.L.; Goh, A., 2004a. An integrated dual factor authenticator based on the face data and tokenized random number. Proc. of the First International Conference on Biometric Authentication (ICBA 2004), Hong Kong, pp. 117-123.

Teoh, A.B.J.; Ngo, D.C.L.; Goh, A., 2004b. Personalised cryptographic key generation based on FaceHashing. Computers and Security Journal **23**(7), pp. 606-614.

Teoh, A.B.J.; Ngo, D.C.L.; Goh, A., 2004c. BioHashing: two factor authentication featuring fingerprint data and tokenized random number. Pattern Recognition **37**(11), pp. 2245-2255.

Trémeau, A.; Colantoni, P., 2000. Regions adjacency Graph Applied to Color Image Segmentation. IEEE Trans. Image Processing **9**(4), pp. 735-744.

Tuceryan, M.; Jain, A.K., 1998. Texture Analysis. In Chen, C.H.; Pau, P.S.; Wang, P.S.P. (eds.). The Handbook of Pattern Recognition and Computer Vision (2nd ed.), World Scientific Publishing Co., pp. 207-248.

Turk, M.; Pentland, A., 1991. Eigenfaces for recognition. Journal of Cognitive Neuroscience **3**, pp. 71-86.

Tuyls, P.; Goseling, J., 2004. Capacity and Examples of Template Protecting Biometric Authentication Systems. Cryptology ePrint Archive, Report 2004/106. (available online at http://eprint.iacr.org/2004/106/)

Uludag, U.; Pankanti, S.; Prabhakar, S.; Jain, A.K., 2004. Biometric Cryptosystems: Issues and Challenges. Proc. IEEE **92**(6), pp. 948-960.

Vapnik, V., 1995. The Nature of Statistical Learning Theory, Springer, N.Y.

Vazirgiannis, M.; Theodoridis, Y.; Sellis, T., 1998. Spatio-temporal composition and indexing for large multimedia applications. Multimedia Systems **6**(4), pp. 284-298.

Wallace, I.; Mitchell, O., 1981. Three-dimensional shape analysis using local shape descriptors. IEEE Trans. Pattern Anal. Machine Intell. **3**(3), pp.310-323.

Webb, Andrew, 1999. Statistical Pattern Recognition, Arnold.

Wiskott, L.; Fellous, J.M.; Kruger, N.; von der Malsburg, C., 1997. Face Recognition by Elastic Bunch Graph Matching. IEEE Trans. Pattern Anal. Machine Intell. **19**(7), pp. 775-779.

Wu, H.; Yoshida, Y.; Shioyama, T., 2002. Optimal Gabor Filters for High Speed Face Identification. Proc. Of 16th International Conference on Pattern Recognition **1**, pp. 107-110.

Wu, P.S.; Li, M., 1997. Pyramid edge detection based on stack filter. Pattern Recognition Letters **18**(3), pp. 239-248.

Yang, J.; Yang, J.-Y.; Zhang, D., 2002. What's wrong with Fisher criterion? Pattern Recognition **35**(11), pp.2665-2668.

Yang, J.; Zhang, D.; Frangi, A.F.; Yang, J.-Y., 2004. Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. IEEE Trans. Pattern Anal. Machine Intell. **26**(1), pp. 131-137.

Yang, J.; Frangi, A.F.; Yang, J.-Y.; Zhang, D.; Zhong J., 2005. KPCA plus LDA: a Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition. IEEE Trans. Pattern Anal. Machine Intell. **27**(2), pp. 131-137.

Yang, J.; Yang, J.-Y., 2003 Why can LDA be performed in PCA transformed space? Pattern Recognition **36**(2), pp. 563-566.

Yang, J.; Yang, J.-Y.; Zhang, D.; Lu, J.-F., 2003. Feature fusion: parallel strategy vs. serial strategy. Pattern Recognition **36**(6), pp.1369-1381.

Yang, S.-Y.; Mitra, S., 1998. Wavelet Based Adaptive Vector Quantization for Encoding Large Color Images Proc. of 11th IEEE Symposium on Computer-Based Medical Systems, pp. 208-213.

Yoshitaka, A.; Kishida, S.; Hirakawa, M.; Ichikawa, T., 1994. Knowledge-Assisted Content-Based Retrieval for Multimedia Databases. IEEE Multimedia, pp. 12-21.

You, J.; Cohen, H.A., 1993. Classification and segmentation of rotated and scaled textured images using texture 'tuned' masks, Pattern Recognition **26**(2), pp. 245-258.

You, J.; Bhattacharya, P., 2000. A Wavelet-based coarse-to-fine image matching scheme in a parallel virtual machine environment. IEEE Trans. Image Processing **9**(9), pp. 1547-1559.

You, J.; Li, W.; Zhang, D., 2002. Hierarchical palmprint identification via multiple feature extraction. Pattern Recognition **35**(4), pp. 847-859.

Zamora, G.; Yang, S.; Wilson, M.; Mitra, S., 2000. Segmentation by Color Space Transformation Prior to Lifting and Integer Wavelet Transformation for Efficient Lossless Coding and Transmission. Proc.. 4th IEEE Southwest Symposium on Image Analysis and Interpretation, pp. 136-140.

Zhang, D.; Shu, W., 1999. Two novel characteristics in palmprint verification: datum point invariance and line feature matching. Pattern Recognition **32**(4), pp. 691-702.

Zhang, D., 2000. Automated Biometrics – Technologies and Systems, Kluwer Academic Publishers.

Zhang, D., 2002. Biometrics Resolutions for Authentication in An e-World, Kluwer Academic Publishers.

Zhang, D.; Peng, H.; Zhou, J.; Pal, S.K., 2002. A Novel Face Recognition System Using Hybrid Neural and Dual Eigenspaces Methods. IEEE Trans. Syst. Man, Cybern. A **32**(6), pp. 787-793.

Zhang, D.; Kong, W.K.; You, J.; Wong, M., 2003 On-line palmprint identification. IEEE Trans. Pattern Anal. Machine Intell. **25**(9), pp.1041-1050.

Zhao, W.; Chellappa, R.; Phillips, P.J.; Rosenfeld, A., 2003. Face Recognition: A Literature Survey. ACM Computing Surveys **35**(4), pp. 339-458.

Zucker, S.W.; Hummel, R.A., 1981. A Three-Dimensional Edge Operator. IEEE Trans. Pattern Anal. Machine Intell. **3**(3), pp. 324-331.