AUTOMATED BIOMETRICS

Technologies and Systems

David D. Zhang

KLUWER ACADEMIC PUBLISHERS

AUTOMATED BIOMETRICS
Technologies and Systems

The Kluwer International Series on
ASIAN STUDIES IN COMPUTER AND INFORMATION SCIENCE

Series Editor

Kai-Yuan Cai

Beijing University of Aeronautics and Astronautics
Beijing, CHINA

Also in the Series:

FUZZY LOGIC AND SOFT COMPUTING, *by Guoqing Chen, Mingsheng Ying, Kai-Yuan Cai*; ISBN: 0-7923-8650-7

INTELLIGENT BUILDING SYSTEMS, *by Albert Ting-pat So, Wai Lok Chan*; ISBN: 0-7923-8491-1

PERFORMANCE EVALUATION, PREDICTION AND VISUALIZATION OF PARALLEL SYSTEMS *by Xingfu Wu*; ISBN: 0-7923-8462-8

ROBUST MODEL-BASED FAULT DIAGNOSIS FOR DYNAMIC SYSTEMS *by Jie Chen and Ron J. Patton;* ISBN: 0-7923-8411-3

FUZZY LOGIC IN DATA MODELING: ***Semantics, Constraints, and Database Design*** *by Guoqing Chen;* ISBN: 0-7923-8253-6

SOFTWARE DEFECT AND OPERATIONAL PROFILE MODELING *by Kai-Yuan Cai;* ISBN: 0-7923-8259-5

AUTOMATED BIOMETRICS
Technologies and Systems

by

David D. Zhang
Hong Kong Polytechnic University

SPRINGER SCIENCE+BUSINESS MEDIA, LLC

Library of Congress Cataloging-in-Publication Data

Zhang, David, 1949-
Automated biometrics : technologies and systems / by David D. Zhang.
p. cm. -- (The Kluwer international series on Asian studies in computer and information science ; 4)
Includes bibliographical references and index.
ISBN 978-1-4613-7038-3 ISBN 978-1-4615-4519-4 (eBook)
DOI 10.1007/978-1-4615-4519-4
1. Pattern recognition systems. 2. Identification--Automation. 3. Biometry. I. Title. II. Series.

TK7882.P3 Z43 2000
006.4—dc21 00-038637

Originally published by Kluwer Academic Publishers, New York in 2000
Softcover reprint of the hardcover 1st edition 2000

Printed on acid-free paper.

CONTENTS

PREFACE

In the modern, automated world, there is an ever-growing need to authenticate and identify individuals. The current technology of using a personal identification numbers (PIN) or password for these purposes hardly meets the requirements of an identification system because one has to remember too many passwords and the password or PIN is very insecure. As a result, in today's complex, geographically mobile, increasingly electronically wired information society, the problem of identifying a person continues to pose a great challenge.

Biometrics-based authentication and identification are emerging as the most reliable method. Biometrics requires that the person to be identified be physically present at the point-of-identification and relies on "something which you are or you do" to provide better security, increased efficiency, and improved accuracy. It overcomes some of the limitations of the traditional automatic personal identification technologies such as ID cards and PIN: ID cards may be lost, stolen, forgotten, or misplaced whereas PIN may be forgotten or guessed by the impostors. In addition, the traditional identification methods are unable to differentiate between an authorized person and an imposter who fraudulently acquires the "knowledge" or "token" of the authorized person. Automated biometrics deals with physiological or behavioral characteristics such as fingerprints, signature, palmprint, iris, hand, voice and face that can be used to authenticate a person's identity or establish an identity from a database. With rapid progress in electronic and Internet commerce, there is also a growing need to authenticate identity of a person for secure transaction processing.

Designing an automated biometrics system to handle large population identification, accuracy and reliability of authentication are challenging tasks. Currently, there are over ten different biometrics systems that are either widely used or under development. Some automated biometrics, such as fingerprint identification and speaker verification, have received considerable attention over the past 25 years, and some issues like face recognition and iris-based authentication have been studied extensively resulting in successful development of biometrics systems in commercial applications. However, according to the author's best knowledge, so far, very few books have been found exclusively devoted to such issues of automated biometrics.

As such a book about automated biometrics, it will systematically introduce the relative technologies and systems, and explore how to design the corresponding systems with in-depth discussion. But, this is not meant to suggest a low-relevance of the book to biometrics in general. Rather, the issues of the book addresses are highly relevant to many fundamental concerns of both researchers and practitioners of automated biometrics in security/detective. The materials in the book are the outgrowth of research the author has conducted for many years, and present the author's recent academic achievements made in the field, although, for the sake of completeness, related work of other authors will also be addressed.

This book is organized into four parts. As an introduction, Chapter 1 first describes the basic concepts necessary for a good understanding of biometrics and answers some questions about biometrics like why, how, what and where. Then, Part I focuses on fundamental biometrics technologies. In Chapter 2, some basic definitions and notations of biometrics from human body are given. Image processing and pattern recognition technologies related to biometrics are discussed in Chapters 3 and 4, respectively.

Part II explores physical biometrics recognition systems. Chapter 5 describes how fingerprint system can be built. In Chapter 6, we develop a biometrics system by palmprint. Chapter 7 defines face system with facial feature extraction, face database and matching techniques and reviews various face recognition approaches. Chapter 8 introduces iris recognition based on coordination system to solve head tilting problem and texture energy.

Part III describes how some behavioral biometrics systems are designed. Chapter 9 presents voice system that deals with speech acquisition, feature extraction, and speaker verification. As another main behavioral biometrics, signature system is discussed in Chapter 10. In addition, Chapter 11 summarizes the other three behavioral biometrics systems, including keystroke biometrics, gesture recognition, and gait biometrics.

Part IV deals with biometrics applications. As an important stage, biometrics user interface is first discussed in Chapter 12. Chapter 13 describes personal authentication application. In Chapter 14, we apply biometrics to Traditional Chinese Medicine (TCM) and such a biometrics Chinese medicine system is then developed. Finally, some main research directions in biometrics are discussed in Chapter 15.

This book is not a primer in biometrics, in that a certain amount of prior knowledge, such as pattern recognition and image processing, is assumed. It is my hope that this book will contribute to our understanding of this new and exciting discipline: Automated Biometrics: Technologies and Systems.

David D. Zhang
Hong Kong Polytechnic University

ACKNOWLEDGEMENTS

My sincere thank goes to Professor Zhaoqi Bian at Tsinghua University, Beijing, China, for his advices throughout this research. I would like to thank Professors Kuanquan Wang, Jane You and Jeffery Liu for their useful comments and supports on special biometrics applications. I would also like to express my gratitude to my research associates and assistants, Dr. Wei Shu, Dr. Changshui Zhang, Dr. Yiying Zhang, Dr. Jie Zhou, Dr. Yong Fang and Mr. Lei Huang for their useful helping and supporting. My graduate students are excellent companions on a difficult road towards understanding Biometrics. I am very grateful to Wenxin Li, Adams Kong, Pang Bo, Xiangqian Wu, Yanlai Li, Shen Xu, George Bao and Cheng Cao, who have contributed to this book. Also thank is due to Mr. Carlos Chan for his help in the preparation of this book. The financial support from UGC/CRC Grants in Hong Kong is appreciated.

SERIES EDITOR'S ACKNOWLEDGMENTS

I am pleased to acknowledge the assistance to the editorial work by Beijing University of Aeronautics and Astronautics and the National Natural Science Foundation of China

Kai-Yuan Cai
Series Editor
Department of Automatic Control
Beijing University of Aeronautics and Astronautics
Beijing 100083
China

1 INTRODUCTION TO BIOMETRICS

In this chapter, we give an overall perception to biometrics technologies and systems. First, the background of biometrics is introduced in Section 1.1. After some basic concepts, taxonomy and history of biometrics are described in Section 1.2. Section 1.3 discusses the principle and performance evaluation of biometrics systems. Next, biometrics applications are briefly reviewed. A whole book perspective is finally given in Section 1.5.

1.1 Why Biometrics?

Along with the automation of our modern life, the need for security becomes more and more important. Every day questions such as "Should this person be given access to a secure system?", "Does this person have authorization to perform a given transaction?", "Is this person a citizen of our country?" are asked millions of times. All these questions are dealing with the same security issue – how to correctly identify human beings.

Currently there are two popular ways for solving such security problems. One is related to "something that you have", such as credit cards, physical keys, etc., and the other depends on "something that you know", such as passwords, Personal Identification Number (PIN), etc. Both methods give the authority to some media, such as password or keys, other than end users. If a user gets the password or other media, he will get the authority; otherwise, he will lose the authority (see Figure 1-1). Under such security schema, people have to keep various cards and remember tens of passwords. Losing card or forgetting password may bring users into great trouble. In the meanwhile, banks, telecommunication companies and governments are suffering from losing hundreds of millions dollars per year due to the breaches of current card or password based security systems [1].

In order to solve this problem, researchers are trying various ways and biometrics approach is most promising. Biometrics should be a technology that uses human being's unique physical or behavioral features to identify or verify persons. It relies on "something that you are" to make personal identification and therefore can inherently differentiate between an authorized person and a fraudulent impostor [2,3]. Because one's unique characteristics can not be stolen, forgotten, duplicated, shared or observed, biometrics based security system is nearly impossible to fraud. The following is a typical verification process using an ID3D Handkey biometrics system: A user enters an identification code on a keypad, then positions the right hand on a plate between a set of pins. A CCD camera records the hand from above and, with the help of a mirror, from the side. Software analyzes the resulting black-and-white image to compare features of the hand's geometry to a record stored in the system's database [4].

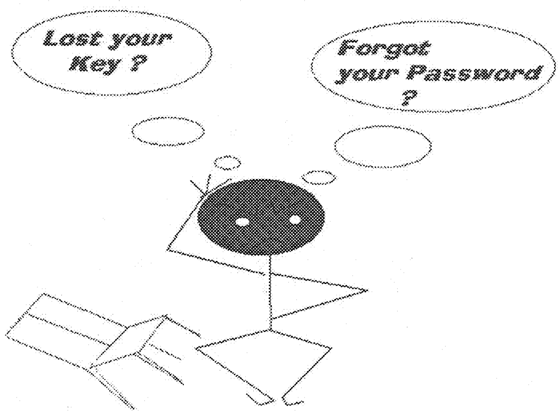


Figure 1-1. Why biometrics.

According to the Gartner Group's Business Technology Journal, biometrics technology is one of the top ten technologies to watch in 1998 [10]. In the past several years, the transactions on biometrics technologies have increased greatly, as shown in Figure 1-2.

1.2 What is Biometrics?

DEFINITIONS

The term "biometrics", strictly speaking, refers to a science involving the statistical analysis of biological characteristics [1]. However, we are usually concerned with technologies that analyze human characteristics for automatically recognizing or verifying identity, where a biometrics way is a measurable physical or behavioral characteristic of an individual. The physical and behavioral characteristics chosen to identify identity should basically satisfy the following conditions [3,5]:

- *Universality*, which indicates that every person should have the characteristic;
- *Uniqueness*, which means that any two persons should be different enough to distinguish each other based on this characteristic;
- *Permanence*, which indicates that the characteristic should be stable enough and not change significantly with environment or time;
- *Collectability*, which means that the characteristic can be measured quantitatively;
- *Acceptability*, which indicates to what extent people are willing to accept the biometrics system;

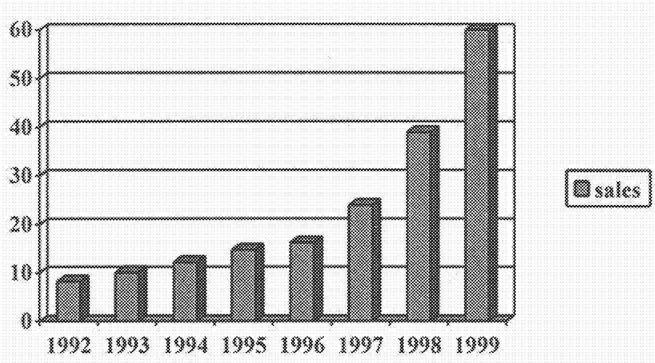


Figure 1-2. Biometrics technology sales (millions of dollars).

- *Performance*, which refers to the achievable identification accuracy, the resource requirements to achieve an acceptable identification accuracy, and the working or environment factors that affect the identification accuracy;
- *Circumvention*, which refers to how easy it is to fool the system by fraudulent techniques.

Whether a human characteristic is suitable for biometrics system can only be told after test by large quantity samples.

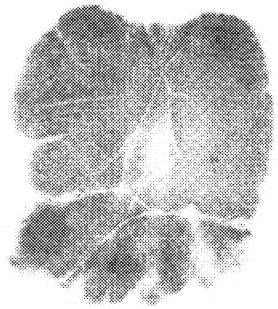
Generally, physical and behavioral characteristics used by biometrics include the following [6-11]:

1) Physical characteristics:
   — Chemical composition of body odor;
   — Facial features and thermal emissions;
   — Features of the eye, i.e., retina and iris;
   — Fingerprints;
   — Palm-prints;
   — Hand geometry;
   — Skin pores;
   — Wrist/hand veins.

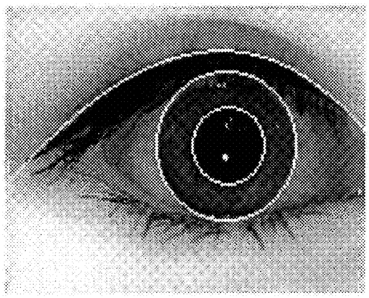
2) Behavioral characteristics:
   — Handwritten signature;
   — Keystrokes or typing;
   — Voiceprint;
   — Gait;
   — Gesture.

The typical physical and behavioral characteristics which biometrics employed are shown in Figure 1-3.

(a) Fingerprint

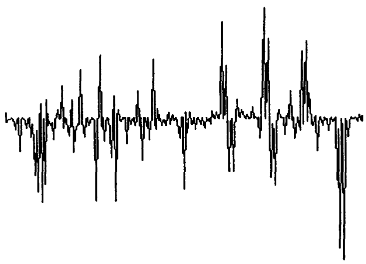


(b) Palmprint

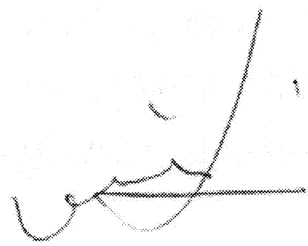


(c) Iris

(d) Face



(e) Voiceprint



(f) Handwring signature

Figure 1-3.

Typical physical characteristics: (a) Fingerprint, (b) Palmprint (c) Iris and (d) Face, and behavioral characteristics: (e) Voiceprint and (f) Handwriting signature.

TAXONOMY OF BIOMETRICS

According to biometrics applications and technologies, in this book, we can mainly classify biometrics systems into the following two types:

Application Type

As the general knowledge, biometrics technology is basically applied to security services. In fact, a few other applications using biometrics are also very effective. Based on our summary of different applications, biometrics systems can be divided into four categories (see Figure 1-4):

- *Personal Authentication*: We may use biometrics technology to identify individuals. Since this is a popular biometrics application, in this book, we will focus on this category.
- *Medical Diagnosis*: Tongue, color of face, beat of heart and other aspects of our body can be also used as biometrics features for medical diagnosis. Traditional Chinese Medicine (TCM) particularly needs to measure such kind of body characteristics.
- *Future Expectation*: Do you have the experience of looking at a handsome face at the first glance and thinking of 'Oh, he must be a nice guy'? Yes, our outside does can represent our inside in some degree. Egypt and China have some specialists on this biometrics application. By looking at one's palm, the specialists can tell the personality as well as the future direction of a person.
- *Ethnology exploration*: Measuring body characteristics can be also used to decide one's ethnic. We may use this biometrics technology to monitor the population shifting among different areas.

Technology Type

In this classification type, biometrics systems are categorized in terms of the employed physical or behavioral characteristics (see Figure 1-5). Since we will emphasize on the application of personal authentication, there are two main parts in our body dealing with physical characteristics, including head (face, iris, and others) and hand (fingerprint, palmprint and others). Behavioral characteristics cover signature, voice and others (gesture, gait, keystroke, etc.). In this book, we will use this technology type given to organize our chapters.

In fact, we also can define other classifications according to signal type (one-dimension or two-dimension) or sensor types (touching or non-touching). They will be discussed in Chapter 12.

IDENTIFICATION, RECOGNITION AND VERIFICATION

A distinction must be made between the terms identification, recognition and verification that constantly circulate throughout the biometrics community [1,12]. Both identification and recognition can be grouped together. An identification system answers the question "Who am I?" and a verification system answers the question like "Am I the person I claimed to be?" [6-13].

In an identification system, usually a huge database is used to store hundreds of thousands persons' digital biometrics features. The identification process is as follows:

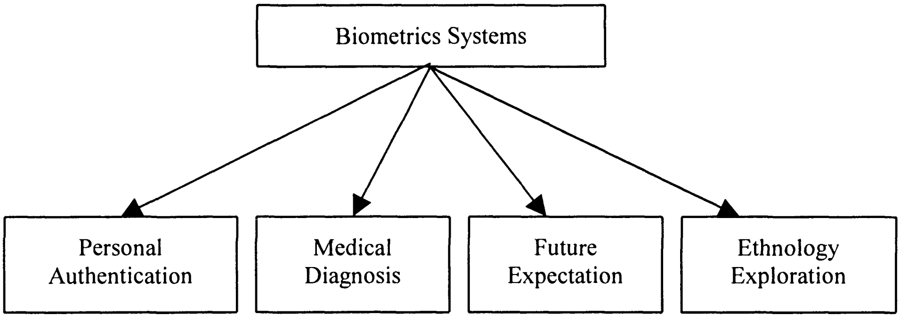


Figure 1-4. Taxonomy by application type.

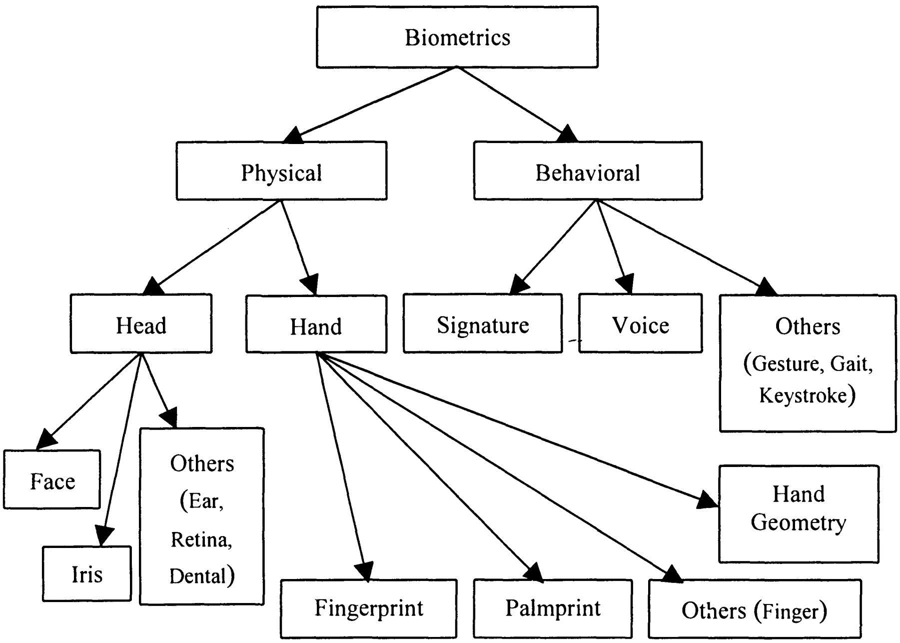


Figure 1-5. Taxonomy by technology type.

- The person who is being checked should have his/her biometrics features captured by the system.
- The system processes the captured data to obtain a pattern and then use it to compare with patterns existed in the database.

- The system gives the result whether the being checked person is a registered one or not. The answer could be in the style of 'Yes, he is.' or ' No, he is not.' or 'Yes. He is ---.'

In a verification system, it is not necessary to have a database since the objective of such kind of system is to verify a person which means the registered pattern is given when the comparison is conducted. The verification process is like this:

- The person being checked enters a password or slides his/her ID card.
- The system captures his/her biometrics feature(s).
- Captured data is processed and compared with the one decided by the password or ID card.
- The system gives verification result. The answer could be as 'Yes, he is.' Or 'No, he isn't.'

HISTORY OF BIOMETRICS

In a non-sophisticated way, biometrics has existed for centuries. Parts of our bodies and aspects of our behavior have historically been used as a means of identification. The study of finger images dates back to ancient China; we often remember and identify a person by his/her face or by the sound of his/her voice; and signature is the established method of authentication in banking, for legal contracts and many other walks of life.

However, automated biometrics has only 40 years' history. As everyone knows, matching finger images against criminal records is always an important way for law enforcers to find the criminal. But the manual process of matching is laborious and uses too much manpower. In late 1960s, the FBI (Federal Bureau of Investigation) began to automatically check finger images and by middle 1970s a number of automatic finger scanning systems had been installed. Among these systems, Identimat is the first commercial one. This system measured the shape of hand and looked particularly at finger length. Though the production of Identimat ceased in late 1980s, its use pioneered the application of hand geometry and set a path for biometrics technologies as a whole. Besides finger and hand, some other biometrics techniques are also developed. Recognition based on retina, iris, voice evolved during 1970s, while signature and facial verification are relatively newcomers to this field [1,2].

The role of biometrics in law enforcement has mushroomed since 1960s and Automated Fingerprint Identification Systems (AFIS) are used by a significant number of police forces throughout the globe now. Building on the success in law enforcement, biometrics is now exploring a range of civilian markets. Network security, physical sites access control and various security applications in banking are all potential markets for biometrics.

From 1996, especially in 1998, more funds had been given to biometrics technology research and development. Therefore research on biometrics becomes more active and exceeds the stage of separate research dispersed in pattern recognition, signal processing, image processing, computer vision, computer security and other subjects. By its distinguished features such as live scan, identical person maximum likelihood and different person minimum likelihood, biometrics grew up into an independent research field. The current biometrics researches concentrate on the iris [14-16,37-39], fingerprints [17-20,40-64], palmprints [21-23,65-67], face [24-26,68-80], voice [27-30,81-84] and handwritten signature [31-33,85-93]. A series of prominent events also show that biometrics is garnering much more attention in both academia and industry. For example, in September 1997, Proceedings of IEEE published a special issue on automated biometrics [34]; April 1998, the BioAPI Consortium was formed to develop a widely available and accepted API that will serve for various biometrics technologies.

A very remarkable phenomenon in the biometrics research is that companies, governments and universities are all involved in the development of biometrics technology. The number of companies who provide biometrics based security products or integrity security solutions is growing steadily recent years. The products vary from fingerprint capture devices, special camera for digital retina image capture, to software development toolkits and totally solution software packages. Governments play an important role in applying biometrics. On one hand, some government departments are important clients of biometrics, especially for large-scale applications; on the other hand, some government branches are the organizations that develop biometrics techniques. Regulations and law pieces relative to biometrics also have great impact with governments. Universities are cradles of biometrics. Like any other techniques, new biometrics techniques are emerging in universities and campuses are good places for educating both developers and end users of biometrics.

1.3 How does a Biometrics System Work?

BIOMETRICS SYSTEM STRUCTURE

Generally, all biometrics systems contain two parts, enrollment part and identification part. The enrollment part functions to have a user's characteristic registered so that it can be used as criterion when identification is performed; the identification part provides the user interface to have the end user's characteristic captured and verified. Though may be based on different characteristic or trait, all the systems follow the same function procedure. For the enrollment part, the procedure consists of sample capturing, feature extraction and storage; For the identification part, the procedure is formed by four stages - capture, feature extraction, comparison and decision [1], as shown in Figure 1-6.

Because the capture and feature extraction stages in the enrollment part are the same as the first two stages in the identification part, usually biometrics systems' general function procedure is considered to cover four processing parts (see Figure 1-7). As the useful biometrics technologies, their related topics are introduced in the following chapters – biometrics sensor (Chapter 12), signal and image processing (Chapter 3), pattern recognition and decision (Chapter 4).

Capture Stage

This is the process that a physical or behavioral sample is input into the system. Different systems use different devices to get the samples. The original device signals are then translated into digital codes with or without some degree's preprocessing.

Generally, physical biometrics data are captured by some type of cameras and the sample is stored as a digital image for processing. Video cameras are used in capturing iris and face while thermal cameras are trying to get face and hand back images. Because automatic fingerprint recognition has a relatively long history, some special fingerprint scanners are manufactured and sold on the market. One type of fingerprint scanner has a small camera in it and another type is using thermal sensor technique. Retina capture device is a special kind of camera with sharp light. Behavioral trait capture devices are different from each other. Voice-based system uses only a PC speaker and signature-based system uses an on-line written board.

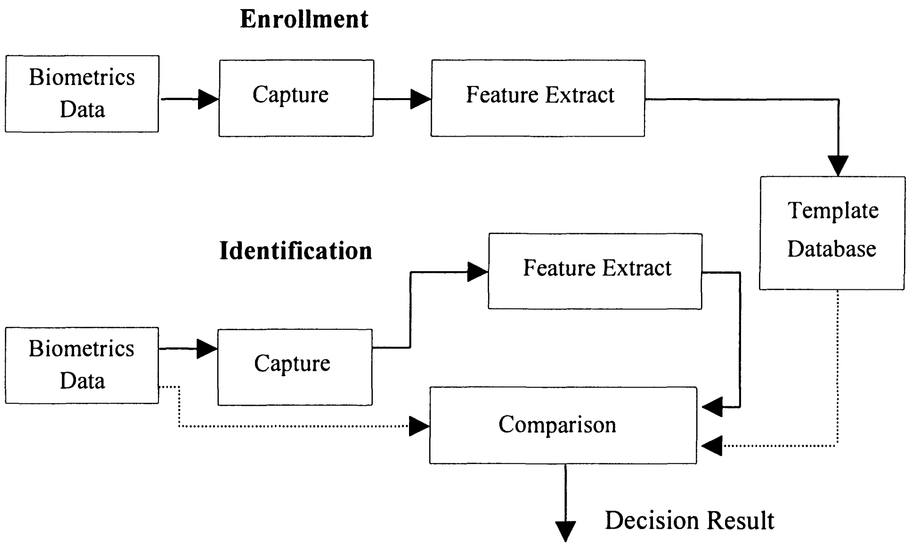


Figure 1-6. General procedure of biometrics systems.

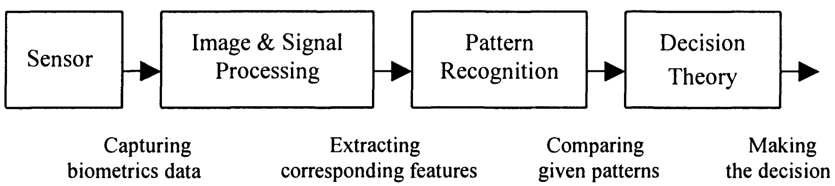


Figure 1-7. Biometrics systems' general procedure with four processing stages.

The capture is the first stage of automatic person identification, so the quality of the captured data is very important and then the less noisy, the better.

Feature Extraction Stage

Feature extraction is the process that unique data are extracted from the sample and a template is created. The templates for any two persons should be different and for different samples from the same person should be similar enough.

Two kinds of approaches are used to extract templates. First is that some meaningful sole features are predefined and proved. For example, minutiae and pores on fingerprint and eyes, mouth, nose and ears on face. Then in this stage, these features are extracted and changed into mathematical code. Most of these approaches use many image-processing algorithms to extract feature from the sample.

The second is that no meaningful feature is found and the sample is transformed or projected into other dimensions so that the noise is levered and the overall data quantity is decreased. Let's call this transformed sample 'refined data'. Then many times' tests are performed and the refined data should be proved to vary very slightly for samples from the same person but strongly differ from samples of any two persons. After this proof, the refined data becomes template. Voice, iris, retina and some of the face identification use this approach. Fourier Transformation and Wavelet Transformation are the most usually used algorithms in this approach. This approach is like the statistic method in traditional pattern recognition in some degree.

Comparison Stage

It is the process that the newly extracted template from a sample is compared with a registered template in the system. Because even the samples from the same person may vary from time to time, the comparison algorithm should tolerate the tiny change from the same person yet distinguish different persons. For example, the finger may contact with the live scanner at different place, direction and pressure, so in practice, no two samples are exactly the same.

Biometrics systems are divided into two categories, i.e., identification and verification. Identification means the new template should be compared with all registered templates in the system while verification means that the new template need only to compare with a particular registered template.

As to the identification systems, the registered templates should be organized into several levels so that the whole comparison process can be shorter.

Decision Stage

This stage is the process that the system decides whether the template extracted from the new sample matches the registered one.

In this stage, a matching score is gained to show the similar degree between the new template and the data stored in the system. In order to give a definite answer of 'yes' or 'no', a threshold is set. When the matching score is greater than the threshold, an answer 'yes' is given, otherwise 'no' is as an output.

PERFORMANCE EVALUATION

As described above, when a biometrics system is put to use, it will either match or not match the extracted biometrics data. A score is given for the comparison between the template and new sample. If the score exceeds the stated threshold for an application, a match is returned. This technique gives biometrics far more flexibility than the "yes or no" approach used by the PIN or password-based technologies.

The biometrics industry has used two performance measurements to rank the level of matching accuracy for many years [1,13,35]. These are known as the false rejection rate (FRR) and false acceptance rate (FAR). The FRR is sometimes referred to as the Type I error rate while the FAR is the Type II error rate. The FRR is concerned with the number of instances an authorized individual is falsely rejected by the system. The FAR refers to the number of instances a non-authorized individual is falsely accepted by the system. Both rates are expressed as a percentage using the following simple calculations:

$$\mathrm{FRR} = \frac{\mathrm{NFR}}{\mathrm{NAA}} \times 100\%, \tag{1-1}$$

$$\mathrm{FAR} = \frac{\mathrm{NFA}}{\mathrm{NIA}} \times 100\%. \tag{1-2}$$

NFR and NFA are the numbers of false rejections and false acceptances, respectively. NAA is the number of authorized identification or verification attempts and NIA is the number of impostor identification attempts.

Less frequently used is a rate known as the Equal Error Rate (EER) or Crossover Rate which is the point at which the FRR and FAR meet or crossover, as shown in Figure 1-8. For example, a system with an FRR and FAR of 1% will also have an EER of 1%. The use of FRR and FAR has led to a significant amount of debate within the biometrics industry. This particularly concerns the statistical significance of such a simplified calculation. The above formulas are staggeringly simple given the large number of variables that may be presented to the biometrics system. For example, the performance of a biometrics system, and therefore the FRR and FAR, may be affected by:

- Environmental conditions (For example, extreme temperature and humidity can affect a system's performance.)
- The age, gender, ethnic background and occupation of the user
- The beliefs, desires and intentions of the user (If a user does not wish to interact with the system, then performance will be affected.)
- The physical make-up of the user (A user with no limbs cannot use signature biometrics.)

Each of the above factors can influence the FRR and FAR. Both rates, when published by biometrics vendors to illustrate performance, can be the result of a laboratory test under controlled conditions. Performance claims must therefore be treated with care. As soon as the above variables are introduced, the rates will be adversely affected. FRR and FAR must also be treated with care because they can be calculated using a 'one-try' or 'three-try' method, depending on the nature of the biometrics application and the type of biometrics system being put to use. This means that a user may either be given a single attempt, or three attempts, at comparing the new sample with the template. Common sense dictates that if three attempts are used, this will improve the FRR. The other inherent problem in the objectivity of the FRR and FAR is that the rates can be manually configured. For example, a bank will require a very high level of security at the door of its main vault. Unauthorized individuals cannot be granted access. The bank may therefore decide that the FAR must be less than 0.1%. In other words, that the system might only grant unauthorized access for one in a thousand instances. The bank may be even more insistent and require a FAR of 0.001%. A biometrics vendor can alter the system's FAR so that these rates can be achieved. However, to do this, the FRR will suffer as a consequence.

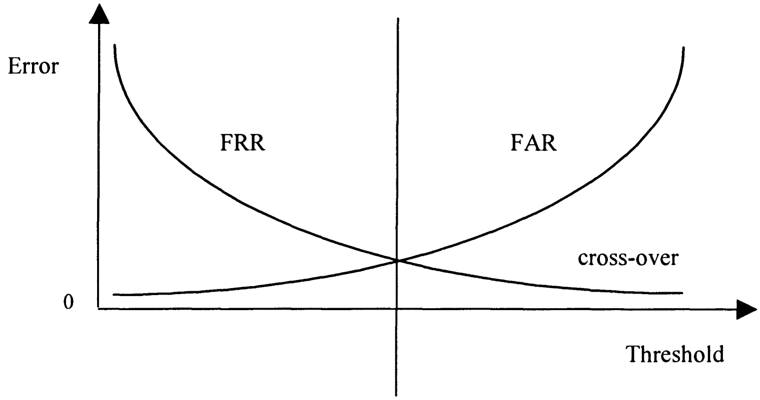


Figure 1-8. FRR, FAR and ERR.

The FRR and FAR are a simplistic means of evaluating performance. There are both an advantage and a disadvantage. It is helpful to understand the basic accuracy of a biometrics system, but it must be remembered that the circumstances of an application can affect the performance rates. Looking at the EER and considering the FRR and FAR in unison is a handy way of determining basic performance. Noting the circumstances of an application must not be forgotten, especially if a biometrics system is expected to operate within the confines of that application.

To provide a more reliable assessment of a biometrics system, some more descriptive performance measures are necessary. Receiver Operating Curve (ROC) and d' are the two other commonly used measures [17,36]. The ROC is a measure of correct acceptance rate against FAR, which gives a good representation of the trade-off between false acceptance and false rejection errors and can be used to select an appropriate operating point for a particular application. The statistical metric d' gives an indication of the separation between the genuine distribution and impostor distribution. It is defined as the difference between the means of the genuine distribution and imposter distribution divided by a conjoint measure of their derivation

$$d' = \frac{\left\| M_{\text{impostor}} - M_{\text{genuine}} \right\|}{\sqrt{(\text{SD}_{\text{impostor}}^2 + \text{SD}_{\text{genuine}}^2)/2}}, \tag{1-3}$$

where $M_{\text{genuine}}, M_{\text{impostor}}, SD_{\text{genuine}}, SD_{\text{impostor}}$ are the means and standard derivations of the genuine distribution and impostor distribution, respectively. Like FAR, FRR and ERR, both ROC and d' also depend heavily on test data and test environment [17]. For such performance metrics to be able to generalize precisely to the entire population of interest, enough samples should be available, and the samples should be representative of the population and adequately represent all the categories (impostor and genuine).

1.4 Where can Biometrics Systems be Applied?

Now worldwide, there are many applications of biometrics being used or considered. Four kinds of basic biometrics applications have been defined in Section 1.2 (see Figure 1-4). Most of the applications are still in the stage of testing and are optional for end users. The accuracy and effectiveness of these systems need to be verified in the real time operation environment. As an example, in this section, we will discuss various applications of personal authentication based on biometrics (see Table 1-1).

Any situation that allows an interaction between man and machine is capable of incorporating biometrics. Such situations may fall into a range of application areas. Biometrics is currently being used in various areas such as computer desktops, networks, banking, immigration, law enforcement, telecommunication networks and monitoring the time and attendance of staff. Each of these applications, and the ever-increasing range of new application areas, has one thing in common: people. Governments across the globe are tremendously involved in using and developing biometrics. National identity schemes, voting registration and benefit entitlement programs involve the management of millions of people and are rapidly incorporating biometrics solutions. Fraud is an ever-increasing problem and security is becoming a necessity in many walks of life. Biometrics applications of personal authentication can be simply categorized as follows [1]:

Law Enforcement
The law enforcement community is perhaps the largest biometrics user group. Police forces throughout the world use AFIS technology to process criminal suspect, match finger images and bring guilty criminals to justice. A number of biometrics vendors are earning significant revenues in this area. It is primarily geared towards AFIS and palm-based technologies.

Banking
Banks have been evaluating a range of biometrics technologies for many years. Automated Teller Machines (ATMs) and transactions at the point of sale are particularly vulnerable to fraud and can be secured by biometrics. Other emerging markets such as telephone banking and Internet banking must also be totally secure for bank customers and bankers alike. A variety of biometrics technologies are now striving to prove themselves throughout this range of diverse market opportunities.

Computer Systems (also known as Logical Access Control)
Biometrics technologies are proving to be more than capable of securing computer networks. This market area has phenomenal potential, especially if the biometrics industry can migrate to large-scale Internet applications. As banking data, business intelligence, credit card number, medical information and other personal data become the target of attack, the opportunities for biometrics vendors are rapidly escalating.

Table 1-1. Overview of various biometrics applications.

Biometrics	*Applications*	*Summary of Applications*
Dental	Recognition	Forensics
DNA	Recognition	Forensics, medicine, genetics
Fingerprint	Recognition / Verification	Immigration & naturalization, welfare distribution, military identification, forensics, access control
Face	Verification / Recognition	Suspect description & identification, missing persons, licenses, credit card, welfare distribution
Hand	Verification	Access control, immigration & naturalization, services distribution
Signature	Verification / Recognition	Signature verification, identification from handwriting
Iris	Verification	Access control
Voice	Verification/ Recognition	Speaker verification, phone service, speaker identification, access control

Physical Access
Schools, nuclear power stations, military facilities, theme parks, hospitals, offices and supermarkets, across the globe, employ biometrics to minimize security threats. As security becomes more and more important for parents, employers, governments and other groups - biometrics will be seen as a more acceptable and therefore essential tool. The potential applications are infinite. Cars and houses, for example, the sanctuary of the ordinary citizen, are under constant threat of theft and biometrics - if appropriately priced and marketed - could offer the perfect security solution.

Benefit Systems
Benefit systems like welfare especially need biometrics to struggle with fraud. Biometrics is well placed to capitalize on this phenomenal market opportunity and vendors are building on the strong relationship currently enjoyed with the benefits community.

Immigration
Terrorism, drug-running, illegal immigration and an increasing throughput of legitimate travelers is putting a strain on immigration authorities throughout the

world. It is essential that these authorities can quickly and automatically process law-abiding travelers and identify the law-breakers. Biometrics are being employed in a number of diverse applications to make this possible. The US Immigration and Naturalization Service is a major user and evaluator of a number of biometrics. Systems are currently in place throughout the US to automate the flow of legitimate travelers and deter illegal immigrants. Elsewhere biometrics are capturing the imagination of countries such as Australia, Bermuda, Germany, Malaysia and Taiwan.

National Identity
Biometrics are beginning to assist governments as they record population growth, identify citizens and prevent fraud from occurring during local and national elections. Often this involves storing a biometrics template on a card that in turn acts as a national identity document. Finger scanning is particularly strong in this area and schemes are already under way in Jamaica, Lebanon, The Philippines and South Africa.

Telephone Systems
Global communication has truly opened up over the past decade. While telephone companies are under attack from fraud. Once again, biometrics are being called upon to defend this onslaught. Speaker ID is obviously well suited to the telephone environment and is making in-roads into these opportune markets.

Time, Attendance and Monitoring
Recording and monitoring the movement of employees as they arrive at work, have breaks and leave for the day were traditionally performed by 'clocking-in' machines. Replacing the manual process with biometrics prevents any abuses of the system and can be incorporated with time management software to produce management accounting and personnel reports.

Researchers, developers and end users have posed diverse questions to biometrics technology. As the application of biometrics, more and more people involved in the discussion on biometrics. People from different groups concern this new technology at different angles. Researchers are interested in finding new ways to bypass the problems encountered in current technologies. Under such consideration, body odor, keystroke, gait and other biometrics features are used to identify people recently. But these new methods are still on the drawing board. While developers focus on exploiting market. Industry standard, real time performance, large-scale application performance and people resistance are all big issues for developers. End users who use biometrics system want to know whether the system is harmful to their bodies or intrusive to their privacy or whether the system interface is user friendly. Despite different groups have different emphasis, one thing is in common – biometrics has gained more and more attentions.

Magazines, web pages and conference tutorials are good resources to seek biometrics information. Each biometrics association has its own periodical on biometrics on which reports the news and trends of biometrics. But deep discussion to the algorithms and foundational theories still appears on some conventional magazines such as IEEE Transactions. Web pages are the most active places for propagating new advance and up coming events. Companies, associations, universities as well as privacy organizations and persons are all giving out their opinions through WWW.

1.5 Book Perspective

In this book, there are four parts to deal with automated biometrics technologies and systems. These parts are arranged in the following:

1) *Biometrics Technologies*: As a background, some basic technologies related to biometrics, including biology, signal & image processing and pattern recognition, are given in Part I.
2) *Physical Biometrics*: Four typical biometrics systems using physical characteristics, such as fingerprint, palmprint, face and iris system, are discussed in this part.
3) *Behavioral Biometrics*: Part III deals with some typical biometrics systems using behavioral characteristics, which involve voice, signature, gesture, keystroke and gait.
4) *Biometrics Applications*: Biometrics user interface is first introduced in Part IV. Among four kinds of biometrics applications (see Figure 1-4), two basic examples, personal authentication and biometrics Chinese medicine, are then discussed.

In Chapter 2, we begin with introducing the background of human body which gives the basis of biometrics technology. Then, we discuss the basic mechanism of human face, eye, skin and hand and finger. After that, behavioral characteristics are briefly reviewed. We also describe the principle of voice produced by human. Finally, we introduce the human ways of behaving.

In Chapter 3, some main technologies of image processing are summarized as biometrics background. Image transformation methods are first introduced. Then the concept and methods of image enhancement related to biometrics are given. Image restoration and image compression technologies are finally reviewed, respectively.

Chapter 4 deals with pattern and pattern recognition, as well as an illustrated configuration of general pattern recognition system. We review the techniques of image segmentation, including pixel classification, gradient-based segmentation, and edge detection. A brief introduction to feature selection and pattern classification is discussed. The theory of neural networks is also involved.

Fingerprint identification is the most widespread application of biometrics technology. In Chapter 5, we firstly review the current fingerprint systems and define some different types of feature points. Then the rules of how to distinguish the true minutiae from the false ones and the corresponding minutiae determination algorithms are systematically described, respectively. Fingerprint matching and the corresponding experiment results are finally given.

Chapter 6 presents automatic palmprint verification which is an important complement of biometrics authentication. As the first attempt of personal identification by palmprint, this chapter explores some new methods for three main processing stages in palmprint verification, including datum point registration, line feature extraction and palmprint classification. Various palmprint images have been tested to illustrate the effectiveness of the methods proposed.

In Chapter 7, we first overview the face recognition system so as to give an outline of what face recognition system does and its structure. Then, two key steps in face recognition system are reviewed in detail. A novel and effective method for face recognition is given. At the end of this chapter, we discuss what we need to do in the near future.

Iris is an information density object, which is suitable for personal identification. Chapter 8 describes some definitions and notations for iris recognition. Some current iris systems, including Daugman's approaches and others, have been reviewed. Then, two novel methods, coordination system to solve head tilting problem and texture energy, are developed, respectively. Their experimental results are finally shown.

In Chapter 9, the background of speaker recognition and some useful concepts associated with it are introduced. The principles of speaker recognition, including production of speech, basic structure of speaker recognition systems, feature parameters, speaker modeling, likelihood normalization and speaker recognition methods, are briefly reviewed. After that, we present a novel and effective speaker verification method.

Chapter 10 deals with handwritten signature, which is one of the most popular ways to verify one's identity. Based on an introduction of basic principles and methods of signature verification, two kinds of signature systems, off-line and on-line, are discussed. Then, an on-line signature verification system based on Dynamic Time Warping (DTW) is developed. Also, we present an on-line signature verification application in the Internet/Intranet.

In Chapter 11, we introduce the background of behavioral biometrics and review briefly behavioral biometrics characteristics. Then, other three behavioral biometrics, including keystroke, gesture and gait biometrics, are described respectively.

Chapter 12 discusses biometrics user interfaces, which provide the digital form of the end user's biometrics characteristics to the computer. It serves two entities of a biometrics system: 1) the human end user, and 2) the host computer. In this chapter, we will define the general structure of biometrics human-machine interface, and explain related elements in the structure in detail. As a result, a typical biometrics user interface in palmprint verification system is described.

In Chapter 13, we focus on a main biometrics application: personal authentication. Some examples of current applications are given first. Potential application areas and how to select a biometrics system are then discussed, respectively. Also, the application programming interface standards are defined in this chapter. Some information resources about personal authentication are listed finally.

As a first attempt to medical diagnosis, in Chapter 14, we apply biometrics technology to Chinese medicine. Firstly, we introduce Traditional Chinese Medicine (TCM). Four kinds of Chinese diagnosis methods and the various observation types in TCM are then discussed. Based on our previous works, a tongue diagnosis system is finally presented.

Automated biometrics will be promising and can play an important role in the modern society because it offers a new approach for identity identification based on the unique, reliable and stable personal characteristics. But biometrics technology is still in its developing stage and there are many works to do. In Chapter 15, we try to put forward three topics, including integration of biometrics, VLSI biometrics and Internet-Ecommerce biometrics, as our main directions in future.

References

[1] G. Roethenbaugh, "Biometrics Explained," Available: *http://www.icsa.net/services/ consortia/cbd-c/sec4.html.*

[2] E. Newham, "The Biometrics Report," New York: SJB Services, 1995. Available: *http:// www.sjb.co.co.uk.*

[3] R. Clarke, "Human Identification In Information Systems: Management Challenges And Public Policy Issues," *Info. Technol. People*, vol. 7, no. 4, pp. 6-37, 1994.

[4] D. Sims, "Biometrics Recognition: Our Hands, Eyes and Faces Give Us Away," *IEEE Computer Graphics and Applications,* 0272-17-16/94, 1994.

[5] J.D. Woodward, "Biometrics: Privacy's Foe or Privacy's Friend?," *Proc. IEEE Special Issue on Automated Biometrics,* vol. 85, no. 9, pp. 1480-1492, 1997.

[6] B. Carter, "Biometrics Technologies, What They Are and How They Work," in *Proc. CTST'97*, Orlando, EL, 1997.

[7] R. Chandrasekaran, "Brave New Whorl: ID System Using The Human Body Are Here, But Privacy Issues Persist," *Washington Post*, Mar. 30, 1997.

[8] A. Davis, "The Body as Password," *Wired*, July 1997.

[9] D.R. Richards, "Rules of Thumb For Biometrics Systems," *Security Manage*, Oct. 1, 1995.

[10] G. Lawton, "Biometrics: A New Era in Security," *IEEE Computer*, pp. 16-18, Aug. 1998.

[11] R. Mandelbaum, "Vital Signs of Identity," *IEEE Spectrum,* pp. 22-30, Feb. 1994.

[12] Association for Biometrics (AFB) and International Computer Security Association (ICSA), 1998 Glossary of Biometrics Terms, *http://www.afb.org.uk/glossuk1.html*, 1998.

[13] M. Golfarelli, D. Maio and D. Maltoni, "On the Error-Reject Trade-Off in Biometrics Verification Systems," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 786-796, 1997.

[14] R.P. Wildes, "Iris Recognition: An Emerging Biometrics Technology," *Proc. IEEE Special Issue on Automated Biometrics,* vol. 85, no. 9, pp. 1348-1363, 1997.

[15] D. McMordie, "Texture Analysis of the Human Iris for High Security Authentication," Available: *http://www.ee.mcgill.ca/~mcmordie/iris/iris.ps*, 1997.

[16] C. Seal, D. McCartney and M. Gifford, "Iris Recognition for User Validation," *British Telecommunications Engineering,* vol. 16, July 1997.

[17] A.K. Jain, H. Lin, P.Harath and R. Bolle, "An Identity-Authentication System Using Fingerprints," *Proc. IEEE Special Issue on Automated Biometrics,* vol. 85, no. 9, pp. 1365-1388, 1997.

[18] A. Jain, H. Lin and R. Bolle, "On-line Fingerprint Verification," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 4, pp. 302-313, 1997.

[19] A.R. Roddy and J.D. Stosz, "Fingerprint Features: Statistical Analysis and System Performance Estimates," *Proc. IEEE Special issue on automated biometrics,* vol. 85, no. 9, pp. 1390-1421, 1997.

[20] M. Kawagoe and A. Tojo, "Fingerprint Pattern Classification," *Pattern Recognition,* vol. 17, no. 3, pp. 295-303, 1984.

[21] D. Zhang and W. Shu, "Two Novel Characteristics in Palmprint Verification: Datum Point Invariance and Line Feature Matching," *Pattern Recognition,* vol. 32, pp. 691-702, 1999.

[22] W. Shu and D. Zhang, "Automated Personal Identification by Palmprint," *Optical Engineering,* vol. 37, no. 8, pp. 2359-2362, 1998.

[23] W. Shu and D. Zhang, "Palmprint Verification: An Implementation of Biometrics Technology," *Proc. of ICPR'98*, vol. I, pp. 219-221, Brisbane AUSTRALIA, 1998.

[24] K.C. Yow and R. Clipolla, "Feature-based Human Face Detection," *Image and Vision Computing,* vol. 15, pp. 713-735, 1997.

[25] J. Zhang, Y. Yan and M. Lades, "Face Recognition: Eigenface, Elastic Matching and Neural Nets," *Proc. IEEE Special Issue on Automated Biometrics,* vol. 85, no. 9, pp. 1423-1435, 1997.

[26] R. Chellappa, C.L. Wilson and A. Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proceedings of the IEEE,* vol. 83, no. 5, pp. 705-740, 1995.

[27] N.Z. Tishby, "On the Application of Mixture Hidden Markov Models to Text Independent Speaker Recognition," *IEEE Trans. Signal Processing,* vol. 39, no. 3, March 1991.

[28] R.P. Ramachandran, M.S. Zilovic and R.J. Mammone, "A Comparative Study of Robust Linear Predictive Analysis Methods with Applications to Speaker Identification," *IEEE Transactions Speech and Audio Processing,* vol. 3, no. 2, March 1995.

[29] J.P. Campbell, "Speaker Recognition: A Tutorial," *Proc. IEEE Special Issue on Automated Biometrics,* vol. 85, no. 9, pp. 1437-1462, 1997.

[30] D.A. Reynolds and R.C. Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models," *IEEE Trans. Speech and Audio Processing,* vol. 3, no. 1, January 1995.

[31] K. Huang and H. Yan, "Off-Line Signature Verification Based on Geometric Feature Extraction and Neural Network Classification," *Pattern Recognition,* vol. 30, no. 1, pp. 9-17, 1997.

[32] R. Bajaj and S. Chaudhury, "Signature Verification Using Multiple Neural Classifiers," *Pattern Recognition,* vol. 30, no. 1, pp. 1-7, 1997.

[33] Q.Z. Wu, S.Y. Lee and I.C. Jou, "On-Line Signature Verification Based on Split-and-Merge Matching Mechanism," *Pattern Recognition Letters,* vol. 18, pp. 665-673, 1997.

[34] Proc. of IEEE, Special Issue on Automated Biometrics, vol. 85, no. 9, Sept. 1997.

[35] W. Shen, M. Surette and R. Khanna, "Evaluation of Automated Biometrics-Based Identification And Verification Systems," *Proc. IEEE Special Issue on Automated Biometrics,* vol. 85, no. 9, pp. 1464-1478, 1997.

[36] J.G. Daugman and G.O. Williams, "A Proposed Standard for Biometrics Decidability," *in Proc. CardTech/SecureTech Conf.*, Atlanta, GA, pp. 223-234, 1996.

[37] J.G. Daugman, "High Confidence Visual Recognition of Persons By a Test of Statistical Independence," *IEEE Trans. on PAMI,* vol. 15, 1993.

[38] W.W. Boles, "A Security System Based on Human Iris Identification Using Wavelet Transform," *First International Conference on Knowledge-Based Intelligent Electronic Systems,* Adelaide, 21-23 May 1997.

[39] G.O. Williams, "Iris Recognition Technology," *IEEE, 0-7803-3537-6-9/96,* 1996.

[40] K. Morita, K. Asai, "Automatic Fingerprint Identification Terminal for Personal Verification," *Hybrid Image Processing,* vol. 638, 1986.

[41] X. Shen, M. Cheng, Q. Shi, and G. Qiu, "A New Automated Fingerprint Identification System," *Journals of Computer Science & Technology,* vol. 4, no. 4, 1989.

[42] L. O'Gorman, J. V. Nickerson, "An Approach to Fingerprint Filter Design," *Pattern Recognition,* vol. 22, no. 1, pp. 29-38, 1989.

[43] Q. Xiao, H. Raafat, "Fingerprint Image Postprocessing: A Combined Statistical and Structural Approach," *Pattern Recognition,* vol. 24, no. 10, pp. 985-992, 1991.

[44] Q. Xiao, H. Raafat, "Combining Statistical and Structural Information for Fingerprint Image Processing, Classification and Identification," *Pattern Recognition: Architectures, Algorithms and Applications,* 1991.

[45] V.S. Sprinivasan, N.N.Murthy, "Detection of Singular Points in Fingerprint Images," *Pattern Recognition,* vol. 25, no. 2, pp. 139-153, 1992.

[46] N. K.Ratha, S. Chen, and A. K. Jain, "Adaptive Flow Orientation-based Feature Extraction in Fingerprint Images," *Pattern Recognition,* vol. 28, no. 11, pp.1657-1672, 1995.

[47] S. Igaki, S. Eguchi, F. Yamagishi, H. Ikeda, and T. Inagaki, "Real-time Fingerprint Sensor Using a Hologram," *Applied Optics,* vol. 31, no. 11, April 1992.

[48] M.R. Verma, A.K.Majumdar, and B. Chatterjee, "Edge Detection in Fingerprints," *Pattern Recognition,* vol. 20, no. 5, pp. 513-523, 1987.

[49] L. Coetzee, E. C. Botha, "Fingerprint Recognition in Low Quality Images," *Pattern Recognition,* vol. 26, no. 10, pp. 1441-1460, 1993.

[50] B.M. Mehtre, "Fingerprint Image Analysis for Automatic Identification," *Machine Vision and Application,* 1993.

[51] D. Maio, D. Maltoni, "A Structural Approach to Fingerprint Classification," *IEEE, Proceedings of ICPR'96,* 1996.

[52] M.T. Leung, W.E.Engeler, and Frank, "Fingerprint Image Processing Using Neural Network," *IEEE Region 10 Conference on Computer and Commnicatin Systems, Hong Kong,* Sept. 1990.

[53] Y. Shan, P. Shi, and J. Li, "Fingerprint Preclassification Using Key-Points," *International Symposium on Speech, Image Processing and Neural Networks,* Hong Kong, 13-16 April 1994.

[54] Shmurun, V. Bjorn, S. Tam, and M. Holler, "Extraction of Fingerprint Orientation Maps Using A Radial Basis Function Recognition Accelerator," *IEEE, 0-7803-1901-X/94,* 1994.

[55] W. Chang, H. S. Soliman, and A. H. Sung, "Fingerprint Image Compression by Natural Clustering Neural Network," *IEEE, 0-8186-6950-0/94,* 1994.

[56] M. Kamijo, "Classifying Fingerprint Images using Neural Network: Deriving the Classification State," *IEEE, 0-7803-0999-5/93,* 1993.

[57] J. W. Lee, I. S. Kweon, "Extraction of Line Features in A Noisy Image," *Pattern Recognition,* vol. 30, no. 10, pp. 1651-1660, 1997.

[58] B.M.Mehtre, N.N. Urthy, S. Kapoor, and B. Chatterjee, "Segmentation of Fingerprint Images Using the Directional Image," *Pattern Recognition,* vol. 20, no. 4, pp. 429-435, 1987.

[59] A.K. Hrechak, J.A. Mchugh, "Automated Fingerprint Recognition Using Structural Matching," *Pattern Recognition,* vol. 23, no. 8, pp. 893-904, 1990.

[60] L. Wu, Z. Xie, "On Fingerprint Theorems," *IEEE, CH2614-6/88/0000/1216,* 1988.

[61] D.K. Isenor, S.G. Zaky, "Fingerprint Identification Using Graph Matching," *Pattern Recognition,* vol. 19, no. 2, pp. 113-122, 1986.

[62] L. Coetzee, E.C. Botha, "Preprocessing of Two-dimentional Fingerprint Images for Fingerprint Recognition," *IEEE, TH0370-7/91/0000-0069,* 1991.

[63] N.K. Ratha, K. Karu, S. Chen, and A.K. Jain, "A Real Time Matching System for Large Fingerprint Databases," *IEEE Trans. on PAMI,* vol. 18, no. 8, Aug. 1996.

[64] J.F. Keegan, "How Can You Tell if Two Line Drawings Are the Same," *Computer Graphics Image Processing,* vol. 6, no. 1, pp. 90-92, 1997.

[65] R.A. McLaughlin, M.D. Alder, C.J.S. DeSilva, "Inference of Structure: Hands," *Pattern Recognition Letters,* Oct. 1994.

[66] P. Baltscheffsky, P. Anderson, "The Palmprint Project: Automatic Identity Verification by Hand Geometry," *International Carnahan Conference on Security Technology,* Gothenburg, Sweden, Aug. 12-14, 1986.

[67] E. Mandler, "Advanced Preprocessing Technique for On-Line Recognition of Handprinted Symbols," *Computer Recognition and Human Production of Handwriting,* pp. 19-36, 1989.

[68] I.A. Essa, A.P. Pentland, "Coding, Analysis, Interpretation, and Recognition of Facial Expressions," *IEEE Transactions on PAMI,* vol. 19, no. 7, July 1997.

[69] S.H. Jeng, H.Y. Liao, C.C. Han, M. Y. Chern, and Y. T. Liu, "Facial Feature Detection Using Geometrical Face Model: An Efficient Approach," *Pattern Recognition,* vol. 31, no. 3, pp. 273-282, 1998.

[70] Q. Chen, H. Wu, and M. Yachida, "Face Detection by Fuzzy Pattern Matching," *IEEE, 0-8186-704208/95,* 1995.

[71] A. Tankus, Y. Yeshurun, and N. Intrator, "Face Detection by Direct Convexity Estimation," *Pattern Recognition Letters,* vol. 18, pp. 913-922, 1997.

[72] N. Foeder, X. Li, "Accuracy Analysis for Facial Feature Detection," *Pattern Recognition,* vol. 29, no. 1, pp. 143-157, 1996.

[73] J. Huang, S. Gutta, and H. Wechsler, "Detection of Human Faces Using Decision Trees," *IEEE, 0-8186-7713-9/96,* 1996.

[74] C.L.Huang, C.W.Chen, "Human Facial Feature Extraction for Face Interpretation and Recognition," *Pattern Recognition* , vol. 25 no. 12, pp. 1435-1444, 1992.

[75] R. Herpers, M. Michaelis, K.H. Lichtenauer, and G. Sommer, "Edge and Keypoint Detection in Facial Regions," *IEEE, 0-8186-7713-9/96,* 1996.

[76] T.F. Cootes and C.J.Taylor, "Locating Faces Using Statistical Feature Detectors," *IEEE, 0-8186-7713-9/96,* 1996.

[77] K.C. Yow, R. Cipolla, "Detection of Human Faces under Scale, Orientation and Viewpoint Variations," *IEEE, 0-8186-7713-9/96,* 1996.

[78] A.J. Colmenarez, T.S. Huang, "Maximum Likelihood Face Detection," *IEEE, 0-8186-7713-9/96,* 1996.

[79] B. Takacs, H. Wechsler, "Visual Filters for Face Recognition," *IEEE, 0-8186-7713-9/96,* 1996.

[80] M.S. Lew, "Information Theoretic View-Based and Modular Face Detection," *IEEE, 0-8186-7713-9/96,* 1996.

[81] R.C. Rose, E.M. Hofstetter, and D.A. Reynolds, "Integrated Models of Signal and Background with Application to Speaker Identification in Noise," *IEEE Trans. on Speech and Audio Processing,* vol. 2, no. 2, April 1994.

[82] Q. Lin, E. Jan, C. Che, and J. Flanagan, "Speaker Identification in Teleconferencing Environments Using Microphone Arrays and Neural Networks," *Esca Workshop on Automatic Speaker Recognition, Identification and Verification.*

[83] S. Furui and M.M. Sondhi (eds.), *Advances in Speech Signal Processing,* Marcel Dekker, NY, 1992.

[84] S. Dutta, "Algorithimic Approach to Automatic Speech Recognition Technique," *IEEE, CH2998-3/91/0000-1233,* 1991.

[85] R. Sabourin, G. Genest, and F.J. Preteux, "Off-line Signature Verification by Local Granulometric Size Distributions," *IEEE Trans. on PAMI,* vol. 19, no. 9, Sept. 1997.

[86] R. Sabourin, R. Plamondon, L. Beaumier, "Structural Interpretation of Handwritten Signature Images," *International Journal of Pattern Recognition and Artificial Intelligence,* vol. 8, no. 3, 1994.

[87] G. Dimauro, S. impedovo, G. Pirlo, "Component-Oriented Algorithms for Signature Verification," *International Journal of Pattern Recognition and Artificial Intelligence,* vol. 8, no. 3, pp. 771-793, 1994.

[88] K. Huang, "On-line Signature Verification Based on Dynamic Segmentation and Global and Local Matching," *Optical Engineering,* vol. 34, no. 12, Dec. 1995.

[89] R. Martens, L. Claesen, "On-line Signature Verification by Dynamic Time-Warping," *IEEE, Proceedings of ICPR '96,1015-4651/96,* 1996.

[90] M. Parizeau, R. Plamondon, "What Types of Scripts Can Be Used for Personal Identity Verification?, " *Computer Recognition and Human Production of Handwriting,* pp. 77-90, 1989.

[91] M. Ammar, Y. Yoshida, and T. Fukumura, "Feature Extraction and Selection for Simulated Signature Verification," *Computer Recognition and Human Production of Handwriting,* pp. 61-76, 1989.

[92] N.A. Murshed, R. Sabourin, F. Bortolozzi, "A Cognition Approach to Off-line Signature Verification," *International Journal of Pattern Recognition and Artificial Intelligence,* vol. 11, no. 5, pp. 801-825, 1997.

[93] G. Dimauro, S. Impedovo, G. Pirlo, and A. Salzo, "A Multi-Expert Signature Verification System for Bankcheck Processing," *International Journal of Pattern Recognition and Artificial Intelligence,* vol.11, no. 5, pp. 827-844, 1997.

2 HUMAN BODY AND BIOMETRICS

In this chapter, we will firstly introduce our human body that gives the basis of biometrics technology. Then, some physical characteristics of human body, including face, eye, skin, hand and finger, are described in Section 2.2. After that, behavioral characteristics dealing with the body are briefly reviewed in Section 2.3. Section 2.4 discusses some human ways of behaving that may use as biometrics.

2.1 Our Body: Overview

The term "biometrics" is composed of three parts: a prefix "bio-" which means "life or living", a root "metr(o)" having the meaning of "measure" and a suffix "-ics" which means "the scientific study of". Thus, "biometrics" can be literally interpreted as "the scientific study of the life / living measure". Since we use biometrics technology as an important research direction, it is necessary for us to understand our human body.

LEVELS OF ORGANIZATION

Biology is the study of living things, and human biology is the study of humans. According to human biology [1], human body is a precisely structured container of chemical reactions and is organized in structural and functional levels of increasing complexity. Each higher level incorporates the structures and functions of the previous level. The simplest level is the chemical level, and others, in complexity increasing order, are cells, tissues, organs, and organ systems.

Chemicals

The chemicals that make up the body may be divided into two major categories: inorganic and organic. *Inorganic chemicals* are usually simple molecules made of one or two elements other than carbon. Examples of inorganic chemicals are water (H_2O), oxygen (O_2), and minerals such as iron (Fe), calcium (Ca), and sodium (Na). Despite their simplicity, however, some inorganic compounds are essential to normal structure and functioning of the body. *Organic chemicals* are often very complex and always contain the elements carbon and hydrogen. In this category, organic chemicals are carbohydrates, fats, proteins, and nucleic acids.

Cells

Cells are the smallest living subunits of a multicellular organism such as a human being. A cell is a complex arrangement of the chemicals, which is living, and carries out specific activities. Unlike microorganisms that are single cells, which function independently, human cells, however, must work together and function interdependently.

The human body consists of innumerable (about one hundred million million) cells. The cell is the building block of the human body. The cell theory states that [2]: Cells are not only the basic structural units of all plants and animals, but also the smallest functioning units of life. They are produced only by the division of preexisting cells, and each maintains homeostasis. Cells do not normally exist on their own. Usually large numbers of them are massed together into a tissue.

Tissues

Tissues are collections of specialized cells and cell products that perform a limited range of functions. There are four major tissue types in the body (Table 2-1): epithelial tissue, connective tissue, muscle tissue, and neural tissue.

Table 2-1. Tissues found in the human.

Name of tissue	*Main functions*	*Examples*
Epithelial tissue	Cover or line body surfaces; produce secretions.	Outer layer of the skin, sweat glands.
Connective tissue	Connect and support parts of the body; transport or store materials.	Blood, bone, adipose tissue
Muscle tissue	Bring about movement	Skeletal muscles, heart
Nerve tissue	Generate and transmit electro- chemical impulses	Brain, optic nerves

Organs

An *organ* is a group of tissues precisely arranged so as to accomplish specific functions. Examples of organs are the kidneys, liver, lungs, and stomach. The stomach is lined with epithelial tissue that secretes gastric juice for digestion. Muscle tissue in the wall of the stomach contracts to mix food with gastric juice and propel it to the small intestine. Nerve tissue carries impulses that increase or decrease the contractions of the stomach.

Organ systems

An *organ system* is a group of organs that all contribute to a particular function. The human body is considered to have 11 major organ systems [3]: integumentary, skeletal, muscular, nervous, endocrine, cardiovascular, lymphatic, respiratory, digestive, urinary, and reproductive. These systems work together to support you, hold you up, make you move, let you eat, make you breath, among many other functions (see Figure 2-1).

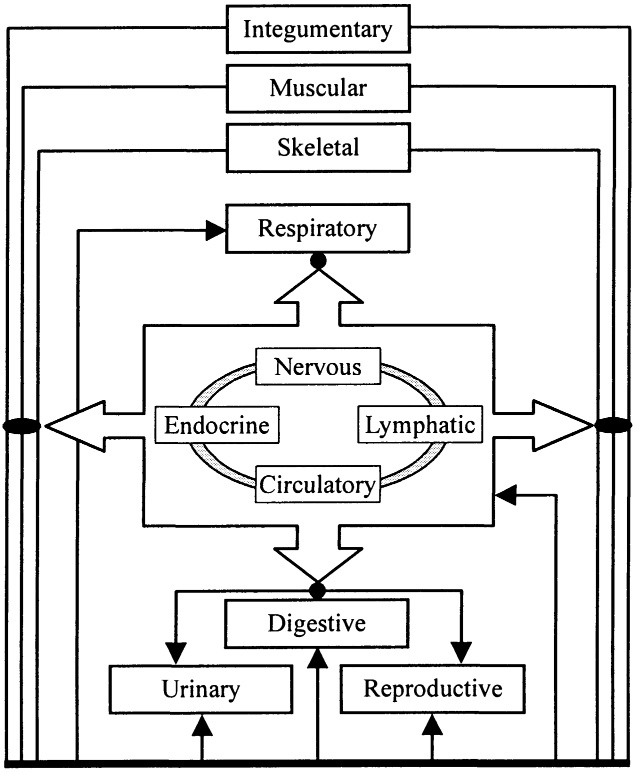


Figure 2-1. Organ systems in the body.

ORGAN SYSTEMS OF THE BODY

Integumentary System

The *integumentary system* consists of the skin and accessory structures such as hair, nails, and glands. Because the integumentary system is located on the outside of the body, it is an organ system with which people are already familiar. The integumentary system has many functions, including the protection of internal structures, prevention of infectious agent entry, regulation of temperature, production of vitamin D, prevention of water loss, and the detection of stimuli such as touch, pain, and temperature.

Skeletal System

Skeletal system is made up of bones, ligaments, and tendons. It determines the shape and symmetry of the body. In addition, it acts as both a protective device for the organs and a firm base for the attachments of muscles (without bones, your muscles would not function properly). The marrow tissues in the cavity of the bones produce red cells and some white cells, required in the blood.

Muscular System
The human body contains more than 650 individual muscles attached to the skeleton, which provides the pulling power for us to move around. The main job of the *muscular system* is to provide movement for the body. The muscular system consist of three different types of muscle tissues: skeletal, cardiac, and smooth. Each of these different tissues has the ability to contract, which then allows body movements and functions. There are two types of muscles in the system, the involuntary muscles and the voluntary muscles. The muscles in which we can control by ourselves are called the voluntary muscles and the ones we can not control are the involuntary muscles. The heart, or the cardiac muscle, is an example of involuntary muscle.

Nervous System
The *nervous system* of the human being is responsible for sending, receiving, and processing nerve impulses throughout the body. All the organs and muscles inside your body rely upon these nerve impulses to function. It could be considered as the master control unit inside your body. Sense organs provide the nervous system with information about the environment by means of such senses as sight, hearing, smell, taste, tough, pressure, and pain. Nerves are connected throughout the whole body to the brain. They carry the information throughout the body in the form of electrochemical signals called *impulses*. These impulses travel from the brain and spinal cord to the nerves located throughout the body. It is largely made up of specialized cells called *neurons*. Each of these neuron has a cell body, or *cyton*, which contains the nucleus and organelles. It takes the corporation of three systems to carry out the mission of the nervous system. They are the central, the peripheral, and the autonomic nervous system.

Endocrine System
The endocrine system is a collection of special organs in the body that produce hormones. These organs are usually called the *glands*, organs of the body that manufacture needed substances, which are necessary for normal bodily functions. They are located in different parts of the body.

Cardiovascular System
The *cardiovascular system,* also called *circulatory system,* is composed of vessels and muscles that control the flow of blood around the body. This process of blood flowing around the body is called *circulation.* The main components of the circulatory system are the heart, arteries, capillaries and veins.

Lymphatic System
The lymphatic system and the cardiovascular system are closely related structures that are joined by a capillary system, which is important to the body's defense mechanisms. It filters out organisms that cause disease, produces certain white blood cells and generates antibodies. It is also important for the distribution of fluids and nutrients in the body, because it drains excess fluids and protein so that tissues do not swell up. The lymphatic system consists of *lymph nodes* (or *lymph glands*) and *lymphatic* (the small vessels that link the lymph nodes).

Respiratory System
The *respiratory system* provides the energy needed by cells of the body. The respiratory system allows oxygen from the air to enter the blood and carbon dioxide to leave the blood and enter the air. The cardiovascular system transports oxygen from the lungs to the cells of the body and carbon dioxide from the cells of the body to the lungs. Thus, the respiratory and cardiovascular systems work together to supply oxygen to all cells, and to remove carbon dioxide. The respiratory system consists of the *nose*, the *nasal cavity*, the *pharynx*, the *larynx*, the *trachea*, the *bronchi*, and the *lungs*.

Digestive System
The digestive system is a group of organs that work like wrecking equipment to break down the chemical components of food, through the use of *digestive juices*, into tiny nutrients which can be absorbed to generate energy for the body. The main components of the digestive system are *oral cavity*, *pharynx*, *esophagus*, *stomach*, *small intestine*, *liver*, *pancreas*, *large intestine*, and *peritoneum*.

Urinary System
The Urinary System consists of the *ureters* (tubes leading from the kidneys to the bladder), *kidneys*, *urethra* (a tube leading from the bladder to the exterior of the body), and *bladder*. The Urinary System is the system that releases waste fluids out of the body in the form of urine. These wastes are formed from use of foods in the cells and dissolved dead cells.

Reproductive System
There are two kinds of reproductive system: female reproductive system and male reproductive system. The major function of the male and female reproductive systems is to produce offspring. Although a functional reproductive system is not required for the survival of the individual, it is essential for the survival of the species.

2.2 Physical Structures

MECHANISM OF HUMAN FACE

Faces are rich in information about individual identity, mood and mental state. There are two eyes, two bows, a nose, a mouth and a chin on every face (see Figure 2-2). Distance between two eyes, shape of face, eyes, nose and mouth are different one from another. As sizes, angles or expression of faces vary, the face image will change, too.

Position relationship between face parts, such as eyes, nose, mouth and chin, as well as their shapes and sizes have strong contribution to classify faces (See Chapter 7). Extracting geometry features like shapes and contribution of face parts to recognize faces are very different from subspace method [10-12].

Among all the biometrics identification methods, face recognition has attracted much attention in recent years because it has potential to be most non-instructive and user-friendly. Face recognition from images has lower uniqueness than fingerprint recognition and iris recognition, while it provides a more direct, friendly and convenient identification way and is more acceptable compared with individual identification ways by other biometrics features. Thus the research on face recognition has become one of the most important parts in biometrics.

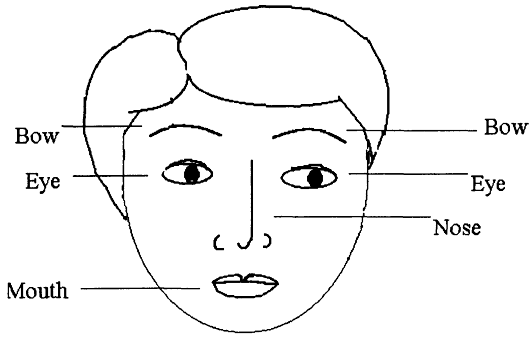


Figure 2-2. Overview of human face.

What features should be derived? How to get these features and how about their classification capacity to faces are problems encountered in the research of this kind of methods. In early research period, researchers mark face feature points such as eyes, nose and mouth manually, calculate all kinds of geometry features, such as distances, separation angles and areas of feature points, then study and observe the recognition rate according to these deduced features. Recently, more research works focused on automatically deriving geometry features [27-28]. These works include automatically deriving some important features like eyes, mouth and nose, and deriving geometry features of eyes, width and length of nose, position of mouth and shape of chin etc.

Facial verification is being touted as the answer to security that concerns electronic commerce. Smart cards can be stolen or copied, but faces stay with their owners and are virtually impossible to duplicate [19].

HUMAN EYE STRUCTURES

Eyes are one of our five sensory organs which include ear, nose, taste receptors, and sensory receptors in the skin, mucous membranes, and other tissues. The eye is the best defined of these structures (see Figure 2-3) [20].

Accessory Structures

Accessory structures protect, lubricate, and move the eye. They include eyebrows, eyelids, conjunctiva, lacrimal apparatus, and extrinsic eye muscles.

Eyebrows

The *eyebrows* protect the eyes by preventing perspiration, which can irritate the eyes, from running down the forehead and into them. They also help shade the eyes from direct sunlight.

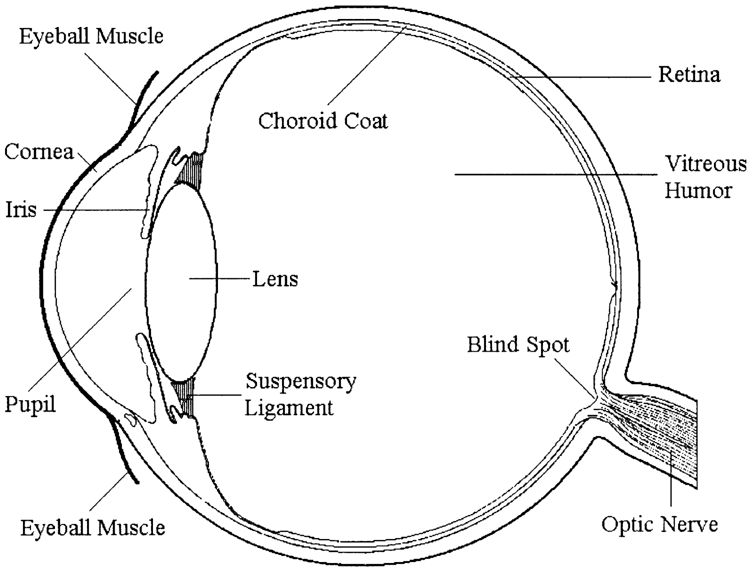


Figure 2-3. The structure of human eye.

Eyelids

The *eyelids*, with their associated lashes, protect the eyes from foreign objects. If an object suddenly approaches the eye, the eyelids protect the eye by closing and then opening quite rapidly (blink reflex). Blinking, which normally occurs about 20 times per minute, also helps to keep the eyes lubricated by spreading tears over the surface of the eye.

Conjunctiva

The *conjunctiva* is a thin, transparent mucous membrane that covers the inner surface of the eyelids and the anterior surface of the eye.

Lacrimal Apparatus

Tears are produced by the *lacrimal glands*, located at the upper, outer corner of the eyeball, within the orbit. Small ducts take tears to the anterior of the eyeball, and blinking spreads the tears and washes the surface of the eye. Tears serve to lubricate the eye and cleanse it. In addition, tears contain an enzyme that helps combat eye infections.

Extrinsic Eye Muscles
Movement of each eyeball is accomplished by six skeletal muscles called the *extrinsic eye muscles.* Four of these muscles run more or less straightly from their origins in the orbit posterior to the eye, to attach to the four quadrants of the eyeball. They are the superior, inferior, medial, and lateral *rectus muscles.* Two muscles, the superior and inferior *oblique muscles*, are placed at an angle to the long axis of the eyeball.

Eyeball
The *eyeball* is a hollow, fluid-filled sphere. The sphere has a larger, posterior part, which makes up about five sixths of the eye, and a much smaller anterior part, which makes up about one sixth of the eye.

Layers of the Eyeball
In its wall, the eyeball has three layers: outer sclera, middle choroid layer, and inner retina. The *sclera* is the thickest layer and is made of fibrous connective tissue which is visible as the white of the eye. The most anterior portion is the *cornea*, which differs from the rest of the sclera in that it is transparent and has no capillaries. The cornea is the first part of the eye that *refracts*, or bends, light rays.

The *choroid layer*, contains blood vessels and a dark blue pigment that absorbs light within the eyeball and thereby prevents glare. The anterior portion of the choroid is modified into more specialized structures: the ciliary body (muscle) is a circular muscle that surrounds the edge of the lens and is connected to the lens by *suspensory ligaments*. The *lens* is made of a transparent, elastic protein, and like the cornea, has no capillaries. The shape of the lens is changed by the ciliary muscle, which enables the eye to focus light from objects at varying distances from the eye.

Just in front of the lens is the circular *iris*, the colored part of the eye. Two sets of smooth muscle fibers change the diameter of the *pupil*, the central opening. Contraction of the radial fibers dilates the pupils; this is a sympathetic response. Contraction of the circular fibers constricts the pupils; this is a parasympathetic response (oculomotor nerves). Pupillary construction is a reflex that protects the retina from intense light or that permits more acute near vision, as when reading. Based on the mechanism of iris, we can extract useful features for biometrics applications (see Chapter 8).

The *retina* lines the posterior two thirds of the eyeball and contains the visual receptors, the rods and cones. *Rods* detect only the presence of light, whereas *cones* detect colors which ,as you may know from physics, are the different wavelengths of visible light. Cones are most abundant in the center of the retina. The *fovea*, which contains only cones, is a small depression directly behind the center of the lens and is the area for best color vision. Rods are proportionally more abundant toward the periphery, or edge, of the retina. Our best vision in dim light or at night, for which we depend on the rods, is at the sides of our visual fields. Research has shown that the pattern formed by veins beneath the retinal surface in an eye is stable and unique [13], based on which retinal pattern based identification is an important biometrics applications and is also very accurate.

Neurons called *ganglion neurons* carry the impulses generated by the rods and cones. These neurons all converge at the *optic disc* and pass through the wall of the eyeball as the *optic nerve*. There are no rods or cones in the optic disc, so this part of the retina is sometimes called the "blind spot." We are not aware of a blind spot in our field of vision, however, in part because the eyes are constantly moving, and in part because the brain "fills in" the blank spot to create a "complete" picture.

Cavities of the Eyeball
There are two cavities within the eye: the posterior cavity and the anterior cavity. The larger, *posterior cavity* is found between the lens and retina and contains *vitreous humor*. This semisolid substance keeps the retina in place.

The *anterior cavity* is found between the front of the lens and the cornea and contains *aqueous humor*, the tissue fluid of the eyeball. Aqueous humor is formed by capillaries in the ciliary body, flows anteriorly through the pupil, and is reabsorbed by the *canal of Schlemm* (small veins also called the scleral venous sinus) at the junction of the iris and cornea.

SKIN COVERING THE BODY

The skin is the outer covering of the body. It is made of several different tissue types and is considered as an organ. Since the skin covers the surface of the body, one of its functions is readily apparent: it separates the body from the external environment and prevents the entry of many harmful substances.

The skin is made up of two major tissue layers. The *dermis* is a layer of dense connective tissue, and the *epidermis* (upon the dermis) is a layer of epithelial tissue that rests on the dermis. The structure of skin is shown in Figure 2-4.

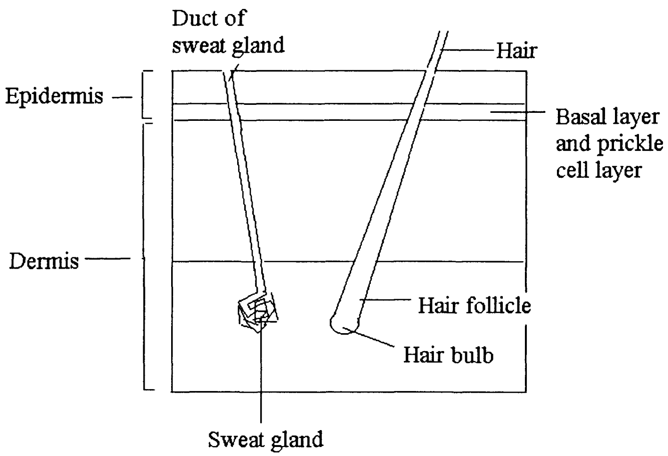


Figure 2-4. The structure of skin.

Epidermis

The epidermis consists of stratified squamous epithelium. It has several layers but its thickness varies over different parts of the body. It is thin, for example, in the eyelids and thick on the soles of the feet. This variation is usually due to a variation in the most superficial layers of non-cellular keratin, which forms the outermost layer of the epidermis. It is especially thick over some areas of palms of the hands and soles of the feet.

In the deepest layer of the epidermis, cells are produced by *mitosis*. As new cells are formed , they push older cells to the surface. The outermost cells protect the cells underneath, and the deeper replicating cells replace cells lost from the surface. During their movement the cells change shape and chemical composition. This process is called *keratinization* because the cells become filled with the protein *keratin*. During keratinization these cells eventually die and produce an outer layer of cells that resists abrasion and forms a permeability barrier.

Although keratinization is a continuous process, distinct layers called *strata* are recognized. The deepest stratum, the *stratum basale*, consists of cuboidal or columnar cells. One daughter cell becomes a new stratum basale cell and can divide again. The other cell is pushed toward the surface. As the cell moves to the surface, it undergoes keratinization.

The *stratum corneum* is the outer layer of the epidermis. It consists of cornified cells, which are dead, squamous cells filled with the hard protein keratin. Keratin gives the stratum corneum its structural strength. The cornified cells are also coated and surrounded by lipids, which help prevent fluid loss through the skin.

Dermis

The dermis consists of vascular connective tissue with fibroblasts, a new fat cells and macrophages. Collagen and elastic fibers are responsible for most of the structural strength of the skin. The collagen fibers are oriented in many different directions and can resist stretch. This produces cleavage or tension lines in the skin, and the skin is most resistant to stretch along the cleavage lines.

The upper part of the dermis has projections called *dermal papillae*, which extend into the epidermis. The dermal papillae contain many blood vessels that supply the overlying avascular epidermis with nutrients, remove waste products, and aid in regulating body temperature. The dermal papillae in the palms of the hands, the soles of the feet, and the tips of the digits are in parallel, curving ridges that shape the overlying epidermis into fingerprints and footprints. The ridges increase friction and improve the grip of the hands and feet.

Palmprints and Fingerprints

The surface of the skin may be marked by deep lines, for example, the palm of the hand (flexure lines) or finer lines, known as *palmprints* which are used mainly to foretell the future (palmistry [30]) as well as diagnose diseases [14,15]. A fingerprint is made of a series of ridges and furrows on the surface of the finger, and everyone is known to have unique, immutable fingerprints. The uniqueness of a fingerprint can be determined by the pattern of ridges and furrows as well as the minutiae points [29]. Minutiae points are local ridge characteristics that occur at either a ridge bifurcation or a ridge ending. Among all the biometrics techniques, fingerprint-based identification is the oldest method which has been successfully used in numerous applications [16] (see chapter 5).

Accessory Skin Structures

Hair

The presence of *hair* is one of the characteristics common to all mammals. In humans, hair is found everywhere in the skin except the palms, soles, lips, nipples, parts of the genitalia, and the distal segments of the fingers and toes.

The *shaft* of the hair protrudes above the surface of the skin, whereas the *root* and *hair bulb* are below the surface (Figure 2-4). A hair has a hard *cortex*, which surrounds a softer center, the *medulla*. The cortex is covered by the *cuticle*, a single layer of overlapping cells that holds the hair in the hair follicle. The *hair follicle* is an extension of the epidermis deep into the dermis, and it can play an important role in tissue repair.

Hair is produced in the hair bulb, which rests on a dermal papilla (Figure 2-4). Blood vessels within the papilla supply the hair bulb with the nourishment needed to produce the hair. Hair is produced in cycles: when the next growth stage begins, a new hair is formed ,and the old hair falls out.

Hair color is determined by varying amounts and types of *melanin*. The production and distribution of melanin by *melanocytes* occurs in the hair bulb by the same method as in the skin. With age the amount of melanin in hair can decrease, causing the hair to become faded in color, or the hair can have no melanin and be white.

Muscles

Associated with each hair follicle are smooth muscle cells, the *arrector pili*. Contraction of the arrector pili causes the hair to "stand on end" and also produces a raised area of skin called "goose flesh."

Glands

The major glands of the skin are the *sebaceous glands* and the *sweat glands* (Figure 2-4). Most sebaceous glands are connected by a duct to the upper part of a hair follicle. They produce *sebum*, an oily, white substance rich in lipids. The sebum lubricates the hair and the surface of the skin, which prevents drying and protects against some bacteria.

There are two kinds of sweat glands. *Merocrine sweat glands* are located in almost every part of the skin and are most numerous in the palms and soles. They produce a secretion that is mostly water with a few salts. Merocrine sweat glands have a duct that opens onto the surface of the skin through sweat pores. When the body temperature starts to rise above normal levels, the sweat glands produce sweat, which evaporates and cools the body.

Sweat pores [4] have been used historically to assist in forensic matching. Although most matching methods have emphasized minutia comparisons and used pores as ancillary comparison features, the ability to match prints based on pore information alone has been documented [5-7]. Considerable research has shown that pores do not disappear, move or spontaneously generate over time [6], which is the basis of its as a biometrics feature.

Apocrine sweat glands produce a thick secretion rich in organic substances. They open into hair follicles, but only in the axillae and genitalia. Apocrine sweat glands become active at puberty because of the influence of sex hormones. Their secretion, which is essentially odorless when released, is quickly broken down by bacteria to cause what is commonly known as body odor, which has potential application in identity recognition and Traditional Chinese Medicine (TCM) oriented diagnosis, which will be introduced in Chapter 14.

Nails

The distal ends of the digits of humans and other primates have nails. The *nail* is a thin plate, consisting of layers of dead stratum corneum cells that contain a very hard type of keratin. The visible part of the nail is the *nail body*, and the part of the nail covered by skin is the *nail root*. The nail grows from the *nail matrix*, located under the proximal end of the nail. Unlike hair, nails grow continuously and do not have a resting stage.

According to the theory of TCM [17-18], the configuration of nails is an important feature that can be used in the diagnoses of many diseases.

HAND

Authentication of a user's identity using the geometric shape of the hand is an interesting problem. Individual hand features are not descriptive enough for identification. However, it is possible to devise a method by combining various individual features to attain robust verification.

Unlike fingerprints, the human hand isn't unique. One can use finger length, thickness, and curvature for the purposes of verification but not for identification. For some kinds of access control like immigration and border control, invasive biometrics (e.g., fingerprints) may not be desirable as they infringe on privacy. In such situations, it is desirable to have a biometrics system that is sufficient for verification. As hand geometry is not distinctive, it is the ideal choice. Furthermore, hand geometry data is easier to collect. With fingerprint collection, good frictional skin is required by imaging systems, and with retina-based recognition systems, special lighting is necessary. Additionally, hand geometry can be easily combined with other biometrics, namely fingerprint. One can envision a system where fingerprints are used for (infrequent) identification and hand geometry is used for (frequent) verification. The biometrics technologies based on fingerprint and palmprint are discussed in Chapters 5 and 6, respectively.

PULSE

A *pulse* is the heartbeat that is felt at an arterial site [8]. What is felt is not actually the force exerted by the blood, but the force of ventricular contraction transmitted through the walls of the arteries. This is why pulses can only be felt at locations where large arteries are close to the surface of the body, but can not be felt in veins: they are too far from the heart for the force to be detectable. From the definition of the pulse, it can be seen that the contraction of heart is the essential origin of pulse.

Cardiac Cycle

The *cardiac cycle* is the sequence of events in one heartbeat. In its simplest form, the cardiac cycle is the simultaneous contraction of the two *atria*, followed a fraction of a second later by the simultaneous contraction of the two *ventricles*. *Systole* is another term for contraction. The term for relaxation is *diastole*. There is a significant difference between the movement of blood from the atria to the ventricles and the movement of blood from the ventricles to the arteries.

Blood is constantly flowing from the veins into both atria. As more blood accumulates, its pressure forces open the right and left AV (*atrioventricular*) valves. Two thirds of the atrial blood flows passively into the ventricles; the atria then contract to pump the remaining blood into the ventricles.

Following their contraction, the atria relax and the ventricles begin to contract. Ventricular contraction forces blood against the flaps of the right and left AV valves and closes them; the force of blood also opens the aortic and pulmonary semilunar valves. As the ventricles continue to contract, they pump blood into the arteries. Notice that blood that enters the arteries must all be pumped. The ventricles then relax, and at the same time blood continues to flow into the atria, and the cycle will begin again. The cardiac cycle is this precise sequence of events that keeps blood moving from the veins, through the heart, and into the arteries.

Pulse Sites

It is helpful to know the major locations where the pulse can be detected because monitoring the pulse is important clinically. The most commonly used pulse sites are:

- Radial artery — the radial artery on the thumb side of the wrist;
- Carotid — the carotid artery lateral to the larynx in the neck;
- Temporal — the temporal artery just in front of the ear;
- Femoral — the femoral artery at the top of the thigh;
- Popliteal — the popliteal artery at the back of the knee;
- Dorsalis pedis — the dorsalis pedis artery on the top of the foot (commonly called the pedal pulse).

Other locations include facial artery, axillary artery, brachial artery, and posterior tibial artery.

Pulse and Disease Diagnosis

The heart rate, heart rhythm, and other many characteristics can be determined by feeling the pulse. When taking a pulse, the doctor also notes the rhythm and force of the pulse. Abnormal rhythms may reflect cardiac arrhythmias, and the force of the pulse (strong or weak) is helpful in assessing the general condition of the heart and arteries. For example, weak pulses usually indicate a decreased stroke volume or increased constriction of the arteries.

Moreover, "feeling the pulse", as one of the four methods of diagnosis used in TCM, is a highly important diagnostic method, allowing the doctor to draw valuable conclusions as to energetic changes in the body. The method is much more sophisticated than that used in Western medicine. A Chinese doctor differentiates not only between the pulse on the wrists of the right and left hand but also between three positions about a finger's breadth wide on each wrist. These positions relate to various functional systems in the body. The doctor also feels the quality of the pulse, differentiating, for example, between a superficial and a deep pulse; an exhausted pulse which can only be tested by feeling it very carefully; a rough pulse, which feels as though a knife were being scraped over bamboo; a long pulse, which is wider than a finger's breadth, or a very tense pulse, which feels like a taut lute string, etc.

2.3 Behavioral Characteristics

PERSONAL VOICE AND SOUND

Speech — Distinguishing Attribute of Human

Verbal thinking and the articulation of speech, although closely connected, involve quite different neurological activities. Human speech differs qualitatively from the communication sounds made by other animals, and involves considerable activity in the cortex of the brain. In the other primates, vocalization appears to be in limbic system rather than by the cortex. So it is probably correct to suggest that speech is one of the distinguishing attributes of the human being. Based on the attribute, human voice can be used as the biometrics feature in personal authentication applications (see chapter 9).

Brain & Speech Production System

There are five readily distinguishable brain areas concerned with language. The area in the frontal lobe, near the motor cortex, is concerned with the control of the many muscles in the face, tongue, throat and jaw involved in the articulation of speech. The other areas, further back along the temporal lobe of the brain (conveniently lumped together as Wernicke's area [31]), are concerned with all the sensory aspects of speech. Wernicke's area is connected to Broca's area [32] by a thick tract of nerve fibres. Speech is believed to be formulated in Wernicke's area and the information passed to Broca's area where the programme for the necessary complex, and muscular activity are put together. This is then passed to the motor cortex so that the appropriate muscles can be activated. Wernicke's area receives input from the visual cortex for written or printed language and from the auditory cortex for heard speech.

The speaker-specific characteristics of speech are due to differences in physiological and behavioral aspects of the speech production system in humans [9]. The main physiological aspect of the human speech production system is the vocal tract shape. The vocal tract is generally considered as the speech production organ above the vocal folds, which consists of the following:

1. Laryngeal pharynx (beneath the epiglottis);
2. Oral pharynx (behind the tongue, between the epiglottis and velum);
3. Oral cavity (forward of the velum and bounded by the lips, tongue, and palate);
4. Nasal pharynx (above the velum, rear end of nasal cavity);
5. Nasal cavity (above the palate and extending from the pharynx to the nostrils).

It is the vocal tract that modifies the spectral content of an acoustic wave as it passes through it, thereby producing speech. Hence, it is common in speaker verification systems to make use of features derived only from the vocal tract.

Production of Voice

The vocal cords are folds of mucous membrane inside the voice box (larynx) (see Figure 2-5). They have associated muscles that change their tension and bring them together, edge to edge. During normal breathing the cords are held apart; the production of voice sounds begins with the moving together with the cords during expiration. As they come together the pressure in the lungs rises, the degree of the rise varies with the loudness of the sounds we wish to make. Air passing between the tightly pressed cords forces them suddenly apart. This leads to a sudden drop of pressure so that the cords are able to come together again and the pressure under them rises once more. The effect is rapidly repeated series of separations and closures so that a succession of compressions and easings is imposed on the column of air in the throat, mouth and nose. The vibrations thus set up in the column of air are known as sound waves.

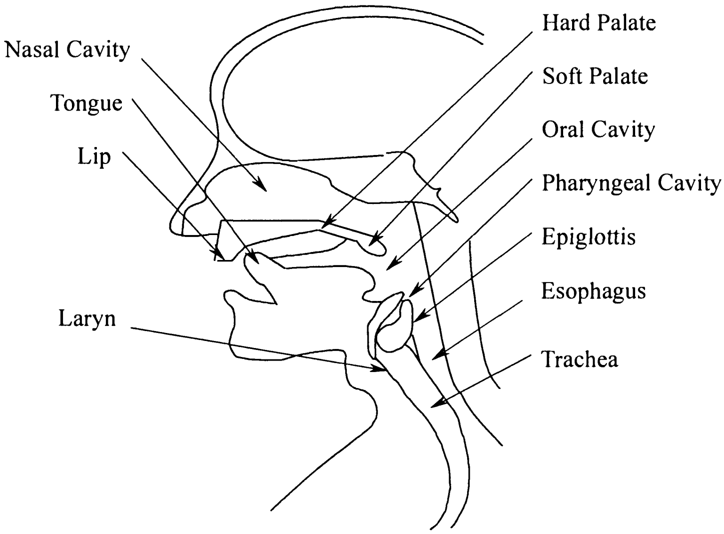


Figure 2-5. A cross-section of head and neck showing different parts of pharynx.

The acoustic wave is produced when the airflow from the lungs is carried by trachea through the vocal folds. This source of excitation can be characterized as phonation, whispering, frication, compression, vibration, or a combination of these. Phonated excitation occurs when the airflow is modulated by the vocal folds. Whispered excitation is produced by airflow rushing through a small triangular opening between the arytenoid cartilage at the rear of the nearly closed vocal folds. Frication excitation is produced by constrictions in the vocal tract. Compression excitation results from releasing a completely closed and pressurized vocal tract. Vibration excitation is caused by air being forced through a closure other than the vocal folds, especially at the tongue. Speech produced by phonated excitation is called voiced, that produced by phonated excitation plus frication is called mixed voiced, and that produced by other types of excitation is called unvoiced.

Sound Pitch

The pitch of the sounds produced by the vocal cords depends on the frequency of vibration: the higher the frequency, the higher the pitch. This, in turn, depends on the tension on the vocal cords and on the length of cord allowed to vibrate. A short length produces a high pitch and the full length and lowest tension produce the deepest pitch. However, the operation of the cords is not a simple matter as they vibrate in a complex manner and different parts can vibrate simultaneously at different frequencies. In general, women and children have shorter vocal cords than men, so their voices are pitch higher. The range of the lowest pitch of the voice extends from about 80 Hz in men to about 400 Hz in women.

Speech Quality

The basic sound produced by the vocal cords has a waveform described as a sawtooth. This is the kind of wave produced by any double–reed musical instrument such as an oboe or a bassoon. The sawtooth waveform is quite different from the sine (bell-shaped) waveform produced by a turning fork or a flute played softly. Pure sine waves have a fundamental tone of a particular frequency (the recognizable pitch of the tone) and nothing else. While sawtooth waves have a fundamental tone they also have a rich collection of harmonics. These are additional tones, or overtones, having frequencies that are simple multiples of the fundamental frequency. The harmonics are usually at a lower volume (or amplitude) than the fundamental tone and so are not heard separately but blend with the tone to alter its quality (or timbre). Different musical instruments produce different combinations of overtones, which is why they have their own distinctive characters.

However, the quality of the voice has little to do with richness and relative amplitude of the various harmonics produced by vibration of the vocal cords. Its final quality depends instead on the principle of sympathetic resonance. Like musical instruments many objects can vibrate at a characteristic frequency within the audible range so as to produce a particular note, the pitch of the note being determined by the size of the object. Such objects will sound if exposed to a range of pitches that include those of the natural resonant frequency of the object. The effect of this 'sympathetic vibration' is to amplify that particular tone in the same way as small, well-timed (in-phase) pushes on a swing quickly build up the amplitude of movement. Hollow cavities also have a natural frequency of vibration, as can be shown by blowing across the top of a bottle. There are many such cavities in the human head and neck — the throat cavity, nasal cavity, mouth cavity and the bony sinuses. Each of these reinforces, to a greater or lesser degree, the particular harmonics in the basic tone produced by the larynx that correspond to their nature resonant frequency.

The effect of these cavities on voice quality varies with their volume, with the force of the expiration of the air, and with the degree of communication between them. Speech quality is, for instance, greatly affected by alteration in the degree of swelling of the lining of the nose, as during a cold. Laryngeal tone, by itself, is thin, weak and lacking in character.

2.4 Ways of Behaving

Behavior is an action that alters the relationship between the organism and its environment. It is externally directed activity and does not include the many internal changes that are constantly taking place in living things.

Behavior may occur as a result of an external stimulus. Receptors are necessary to detect the stimulus, the nerves are needed to coordinate the response, and the effectors actually carry out the action. Behavior may also occur as a result of an internal stimulus. More often than not, the behavior of an organism results from the combination of both external and internal stimuli. Internal stimuli, such as hunger, provide the motivation for the action taken when food is actually seen or smelled.

In studying behavior, it is worthwhile to try to distinguish between innate and learned forms of behavior. The formers are responses whose nature is, in large measure, determined by inherited pathways of nervous coordination. They are usually quite inflexible, a given stimulus giving rise to a given response. Learned behavior, on the other hand, is behavior that has become more or less permanently altered as a result of the individual organism. Only by diligent practice can one learn to play basketball well.

REFLEXES

Reflexes are the simplest innate responses found in animals having a nervous system. A reflex is an automatic response of part of the body to a stimulus. The response is inborn, that is, its nature is determined by an inherited pattern of receptors, nerves, and effectors.

The familiar knee jerk reflex is a stretch reflex. You have undoubtedly had your doctor tap you just below the knee cap with a rubber-headed hammer. Your response should have been a sudden kick with the lower leg. The response is quite automatic. It requires a properly functioning spinal cord, but the brain need play no role (although you can use your brain to try to override the reflex).

LEARNED BEHAVIOR

Learned behavior is behavior that is more or less permanently acquired or modified as a result of the experiences of the individual.

HABITUATION & IMPRINT

Almost all animals are able to learn not to respond to repeated stimuli which has been proved to be harmless. This phenomenon is known as habituation and is an example of true learning. One of the most narrowly specialized, clear-cut examples of learning is imprinting.

THE CONDITIONED REFLEX

Perhaps the simplest form of learned behavior is the conditioned response. Basically, this is a response, comes to be caused by a stimulus different from that which originally triggered it.

We assume that the physiological basis of conditioned response is the transfer, by appropriate neurons, of nervous activity in the auditory area of the brain to the motor neurons controlling salivation in Pavlov‘s experiment. In fact, some psychologists feel that all learned behavior arises from the development of conditioned responses. They feel that the conditioned response is the fundamental unit of even the most complex forms of human behavior. While our knowledge of our own higher mental processing is still far too poor to accept or reject this theory outright, there is no question that humans can be conditioned to some degree.

THE VISUAL ANALYSIS OF HUMAN MOVEMENT

The ability to recognize humans and their activities by vision is key for a machine to interact intelligently and effortlessly with a human-inhabited environment. Because of many potentially important applications, "looking at people" is currently one of the most active domains in computer vision. This part identifies a number of promising applications and provides an overview of recent development in this domain. The emphasis is on discussing the various methodologies that are grouped in 2-D approaches with or without explicit shape model and 3-D approaches, where appropriate systems are reviewed. We also conclude with some thoughts about future directions.

A new application domain of computer vision has emerged over the past few years dealing with the analysis of images involving humans. This domain (sometime called "looking at people") covers, among others, face recognition, hand gesture recognition, and whole-body tracking. The strong interest in this domain has been motivated by the desire for improved man-machine interaction for which there are many promising applications.

One of the general goals of artificial intelligence has been to design machines which act more intelligently or human-like. Natural language understanding, large knowledge bases, and sophisticated reasoning have all made contributions toward reaching this goal, as embodied by the Turing test. Yet, they provide only a partial solution, for a machine, to be truly intelligent and useful, requires the ability to perceive the environment in which it is embedded. It needs to be able to extract information from its environment independently, rather than rely on information supplied to it externally by keyboard input (as in the original conception of the Turing test). Perhaps the most relevant information to be retrieved for interaction is where and who are the humans in the environment and what are their activities. Here, computer vision can play an important role. An added benefit of such a capability is that it makes communication with machines easier for humans, allowing input modalities such as gestures.

Traditionally, there has been keen interest in human movement from a wide variety of disciplines. In psychology, there have been the classic studies on human perception by Johansson. His experiments with moving light displays (MLD) attached to body parts showed that human observers can almost instantly recognize biological motion patterns even when presented with only few of these moving dots. This raised the question whether recognition of moving patterns could be achieved directly from motion, without structure recovery. In the hand gesture area, there have been many studies on how humans use and interpret gestures.

In kinesiology (i.e., biomechanics) the goal has been to develop models of the human body that explain how it functions mechanically and how one might increase its movement efficiency. A typical procedure involves obtaining 3-D joint data, performing kinematic analysis, and computing the corresponding forces and torques for a movement of interest. 3-D data is typically obtained in an intrusive manner, e.g., by placing markers on the human body [21].

In choreography, there has been long-term interest in devising high-level descriptions of human movement for the notation of dance, ballet, and theatre. Some of the more popular notations have been the labanotation, the Ekshol-Wachmann, and the effort-shape notation. Across the variety of notation systems there has been little consensus, though, what these general purpose descriptions should be [22].

Computer graphics has dealt with the synthesis of human movement. This has involved devising realistic models of human bodies for applications in crash simulations, workplace assessment, and entertainment. Some of the issues have been how to specify spatial interactions and high-level tasks for the human models [23].

The recent interest in the "looking at people" domain in vision is hardly surprising. From a technical point of view, this domain is rich and challenging because of the need to segment rapidly changing scenes in natural environments involving non-rigid motion and (self) conclusion. A number of potentially important applications exist. Additional momentum has been provided by recent technological advances, chief among them is the introduction of real-time capture, transfer, and processing of images on standard hardware systems. The extensive coverage in the vision literature is apparent form the many special workshops devoted to this topic: the Looking at People workshop in Chambery, the Motion of Non-Rigid and Articulated Objects workshop in Austin, and the two Automatic face and Gesture Recognition workshops in Zurich and Killington. Some of the materials have now also reached the popular scientific press [24-25].

This development has been driven by the many interesting applications that lie ahead in this area and the recent technological advances involving the real-time capture, transfer, and processing of images on widely available low-cost hardware platforms (e.g., PCs).

A number of promising application scenarios were discussed: virtual reality, surveillance systems, advanced user interfaces, and motion analysis. Three main approaches were discussed: 2-p approaches without explicit shape models, 2-D approaches with explicit shape models, and 3-D approaches. It was argued that which of the above approaches to pursue depends on the application, and at the same time some general guidelines were given. Action recognition was considered in the context of matching time-varying feature data.

Although one appreciates from this survey the large amount of work that already has been done in this area, many issues are still open, e.g., regarding image segmentation use of models, tracking versus initialization, multiple persons, occlusion, and computational cost. One of the challenges for 2-D systems is to show that the approaches scale up to allow pose recovery for a large set of movements from different viewpoints. 3-D systems still have to resolve issues dealing with model acquisition, detail of modeling, and obtaining ground truth. An interesting question is whether a set of generic human actions can be defined which is useful across applications and if so, what the features of interest would be. Added functionality and performance is likely to be gained by adding a symbolic component. Work on different sensor modalities (range, infrared, and sound) will furthermore lead to systems with combined strengths [26].

References

[1] Martini and Bartholomew, *Essentials of Anatomy and Physiology*, Prentice-Hall, 1999.

[2] The Cell Theory, http://pc65.frontier.osrhe.edu/hs/science/bcell1.htm.

[3] R. Seeley, D. Stephens, and P. Tate, *Essentials of Anatomy & Physiology*, 2nd ed., Mosby, St. Louis, 1996.

[4] J.G. Webster, *Medical Instrumentation*, 2nd ed., Boston: Houghton Mifflin Company, 1992, Chapter 5, pp.249-250.

[5] D.R. Ashbaugh, "Poroscopy," *RCMP Gazette*, Vol. 45, pp. 12-17, 1983.

[6] E. Locard, "Les Pores et L'Identification des Criminels," *Biologica, revue scientifique de medecin*, Vol. 22, pp. 357-362, 1912.

[7] J.D. Stosz and L.A. Alyea, "Automated system for fingerprint authentication using pores and ridge structure," *Proceedings of the SPIE*, Automatic Systems for the Identification and Inspection of Humans, Vol. 2277, pp. 210-223, July, 1994.

[8] C.S. Valerie and S. Tina, *Essentials of Anatomy and Physiology*, 2nd ed., F. A. Davis Company, Philadelphia, 1995.

[9] S. Furui, "Recent Advances in Speaker Recognition," *Pattern Recognition Letters*, 18:859-872, 1997.

[10] Y. Noguchi, "Subspace Method and Projection Operators," *Proc. 4th Int. Conf. on Pattern Recognition*, Kyoto, pp.448-451, 1978.

[11] E. Oja and J. Karhunen, "An Analysis of Convergence for a Learning Version of the SubSpace Method," *J. Math. Anal.* Appl., pp.102-111, 1983.

[12] E. Oja and M. Kuusela, "The ALSM Algorithm—An Improved Subspace Method of Classification," *Pattern Recognition*, 1983.

[13] R.B. Hill, "Apparatus and method for identifying individuals through their retinal vasculature patterns," US Patent No. 4109237, 1978.

[14] Z.X. Long, *Diagnostics of Traditional Chinese Medicine*, Academic Press, Beijing, 1998.

[15] X. Yang and S.G. Cheng, *Hand Diagnosis with Shape and Color*, Tian Jin Science and Technology Press, Tian Jin, 1998.

[16] A.K. Jain, L. Hong, S. Pankanti, and R. Bolle, "An identity-authentication system using fingerprints," *Proceedings IEEE*, 85(9):1365-1388, 1997.

[17] T.T. Den and Z.Q. Guo, *Diagnostics of Traditional Chinese Medicine*, Shang Hai Science and Technology Press, Shang Hai, 1984.

[18] H.H. Yin and B.N. Zhang, *Elementary Theory of Traditional Chinese Medicine*, Shang Hai Science and Technology Press, Shang Hai, 1984.

[19] G.M. David, *Exploring Psychology*, 2nd ed., Worth Publishers, 1993.

[20] N. Badler, C. Phillops, and B. Weber, *Simulating Humans*, Oxford Univ. Press, Oxford, 1993.

[21] W.K. John, *Biology*, 4th ed., Addison, Wesley Publishing Company, 1977.

[22] D.M. Gavrila, "The Visual Analysis of Human Movement: A Survey," *Computer Vision and Image Understanding*, Vol.73, No.1, January, pp.82-98, 1999.

[23] G. Johanson, "Visual perdeption of biological motion and a model for its analysis,"*Perception Psychophys*, pp.201-211, 1973.

[24] T. Calvert and A. Chipamn, "Analysis and synthesis of human movement, in Handbook of Pattern Recognition and Image Processing,"*Computer Vision (T. Young, Ed.)*, pp.432-474, Academic Press, San Diego, 1994.

[25] N. Badler and S. Smoliar, "Digital representations of human movement," *ACM Comput.* Surveys 11(1), pp.19-38, 1979.

[26] J. Joseph, *Essential Anatomy*, 2nd ed., *Published by MTP Press Limited*, Lancaster, England, 1979.

[27] A. Yuille, P. Hallinan, and D. Cohen, "Feature Extraction from Faces using Deformable Templates," *International Journal of Computer Vision*, 8(2): 99-111, 1992.

[28] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10): 1042-1052, 1993.

[29] About Fingerprints, http://www.digitalpersona.com/html/technology/finger2.htm.

[30] The Palmistry Center, http://www.palmistry.com/.

[31] Neurology and Pathology of Language, http://www.ldc.upenn.edu/myl/lx1/neurology.html.

[32] Broca's Area, http://asa.calvin.edu/archive/evolution/199511-12/0584.html.

3 SIGNAL AND IMAGE PROCESSING

In this chapter, we will summarize some technologies about image processing related to biometrics. Digital image processing will be simply introduced in Section 3.1. Section 3.2 discusses some kinds of image transformation. Section 3.3 contains the concept and methods of image enhancement. Image restoration and data compression are introduced in Section 3.4 and Section 3.5, respectively.

3.1 Digital Image Processing

Digital image processing is one of the most important techniques in Biometrics. It can be defined as subjecting a numerical representation of an object to a series of operations in order to obtain a desired result [1]. Digital image processing deals with problems in which both input and output are pictures. The major topics within the field of digital image processing include image enhancement, image restoration, and image compression [2], which are often used in Biometrics. Image enhancement involves taking an image and improving it visually, typically by taking advantage of the human visual system's response. Image restoration is the process of taking an image with some known, or estimated, degradation, and restoring it to its original appearance. Image compression involves reducing the typically massive amount of data needed to represent an image. In many cases, digital images have to be transformed from spatial domain to other domains, for example frequency domain, before they are enhanced, restored or compressed. So image transformation is also an important part used in Biometrics. Digital image processing has many applications, such as enhancement of medical images, restoration of remotely sensed satellite images of outer space, improvement of definition in television receivers without altering the transmission standards, and image compression for transmission and storage purposes [3]. Fingerprint system, where is concerned with all technologies about digital image processing, is such an example [4-10]. Overexposed or underexposed or blurred pictures can be improved with contrast enhancement techniques. Sometimes it is desirable to apply more drastic transformations. An image with a wide range of illumination may be reduced into an image where one sees only two levels of illumination. The resulting silhouettes may be reduced further into stick type figures. Other times we may even create a new image from a set of others, such as constructing images of cross-sections of the human body from lateral x-ray picture [11]. Interest in digital image processing methods stems from two principal application areas: improvement of pictorial information for human interpretation, and processing of scene data for autonomous machine perception [12]. Biometrics belongs to the second area.

3.2 Transformation Technology

In biometrics, images sometimes can not be processed effectively in the spatial domain. In this case, the images should be transformed from the spatial domain to other domains. Transform theory plays a key role in image processing. The major applications of transforms are in the design of filters for image enhancement and also in certain data compaction techniques. They also provide the foundations for image reconstruction from projections. There are many transforms in image processing [1, 11-12]: the Fourier Transform (FT), the Hadamard Transform (Walsh Transform), Discrete Cosine Transform (DCT), etc.

THE FOURIER TRANSFORM

The Fourier transform [13-16] is a powerful tool in linear system analysis. It allows us to quantify the effects of digitizing systems, sampling spots, electronic amplifiers, convolution filters, noise, and display spots [1]. The Fourier transform is among the most widely used tools for transforming data sequences and functions from what is referred to as the time domain to the frequency domain [13]. Applications of the transform range from designing filters for noise reduction in audio-signals, such as music or speech, to fast multiplication of polynomials.

The discrete two-dimensional Fourier transform is defined as [1]

$$F(u,v)=\frac{1}{N}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}f(m,n)e^{-j2\pi(u\frac{m}{N}+v\frac{n}{N})}, \tag{3-1}$$

for $u=0,1,2,\cdots,N-1$ and $v=0,1,2,\cdots,N-1$. The inverse transform is

$$f(m,n)=\frac{1}{N}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}F(u,v)e^{j2\pi(m\frac{u}{N}+n\frac{v}{N})}, \tag{3-2}$$

for $x=0,1,2,\cdots,N-1$ and $y=0,1,2,\cdots,N-1$.

The significance of the discrete Fourier transform is enhanced not only by its use to approximate continuous transform, but also because it may be efficiently computed via the Fast Fourier Transform (FFT) algorithm [17-19]. An important mathematical fact is that the discrete Fourier transform may be used to determine the eigenvalues of a circulant matrix [20].

There are many useful properties in the discrete two-dimensional Fourier transform [1,12,20]. A few of them are listed in Table 3-1 and summarized below.

Linearity

Linearity and superposition apply in both domains. The spectrum of a linear sum of images is the linear sum of their spectra. Further, any function may be regarded as a sum of component parts and the spectrum is the sum of the component spectra.

Table 3-1. Properties of the discrete two-dimensional Fourier transform.

Property	*Spatial domain*	*Frequency domain*
Linearity	$af_1(m,n)+bf_2(m,n)$	$aF_1(u,v)+bF_2(u,v)$
Scale Change	$f(am,bn), \quad a,\ b \neq 0$	$\frac{1}{\lvert ab \rvert} F(\frac{u}{a},\frac{v}{b})$
Convolution Theorem	$f(m,n) * g(u,v)$	$F(u,v)G(u,v)$
Multiplication	$f(m,n)g(m,n)$	$F(u,v) * G(u,v)$
Correlation Theorem	$f(m,n) \circ g(m,n)$ $f(m,n)g^*(m,n)$	$F(u,v)G^*(u,v)$ $F(u,v) \circ G(u,v)$
Shift of Position	$f(m-a,n-b)$	$F(u,v)\exp[-j(ua+vb)]$
Modulation	$\exp[j(u_0 m+v_0 n)/N]f(m,n)$	$F(u-u_0,v-v_0)$
Rotation	$f(m\cos\theta+n\sin\theta,$ $-m\sin\theta+n\sin\theta)$	$F(u\cos\theta+v\sin\theta,$ $-u\sin\theta+v\cos\theta)$
Differentiation	$f(m,n)-f(m-1,n)$	$F(u,v)[1-\exp(-jk/N)]$
Integration	$f(m,n)+f(m-1,n)$	$F(u,v)[1+\exp(-jk/N)]$

Scale Change

Space-bandwidth invariance. Compressing a spatial function expands its spectrum in frequency and reduces its amplitude by the same factor. The amplitude reduces because the same energy is spread over a greater bandwidth. For $a=b=-1$, the spatial function is reversed. The frequency axes are also reversed, which, for real images, changes only the phase spectra.

Convolution Theorem

The two-dimensional convolution of $f(m,n)$ and $g(m,n)$ is given by the relation

$$f(m,n) * g(m,n) = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} f(i,j)g(m-i,n-j) \tag{3-3}$$

for $x=0,1,2,\cdots,M-1$ and $y=0,1,2,\cdots,N-1$.

The convolution of two discrete spatial functions corresponds to the product of the individual spectra. In the periodic convolution described, care must be taken to prevent undesirable wrap around. A periodic convolution may also be computed with an extended periodic convolution.

Multiplication

The product of two discrete spatial functions corresponds to the convolution of their discrete spectra.

Correlation Theorem

The discrete two-dimensional correlation can be defined as

$$f(m,n) \circ g(m,n) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(i,j) g(m+i, n+j), \tag{3-4}$$

for $x = 0,1,2,\cdots,\ M-1$ and $y = 0,1,2,\ \cdots,\ N-1$.

In Correlation Theorem (see Table 3-1), "*" represents the complex conjugate. The correlation of two spatial functions corresponds to the product of one spectrum and the conjugate of the other. If the two functions are identical, the spectrum is the magnitude squared or power spectral density.

Shift of Position

Shifting or translating the discrete spatial function in integer amounts adds a linear phase spectrum. Conversely, a linear phase addition to the spectrum produces a translation of the image. The magnitude spectrum is invariant to translation.

Modulation

Multiplying a spatial function by a discrete complex sinusoid translates its spectrum to center about the frequencies of the sinusoid.

Rotation

Rotation of the discrete spectrum by multiples of 90° corresponds to rotation of the spectrum by exactly the same angle. Other rotations involve interpolation.

Differentiation

The derivative must be approximated by an appropriate difference formula but will generally produce a form of high-pass filter.

Integration

The integral must be approximated by an appropriate numerical integration formula but the result will generally be a form of low-pass filter.

THE HADAMARD TRANSFORM

The Hadamard transform [21] is a symmetric, separable unitary transformation that has only +1 and −1 as elements in its kernel matrix [1]. It exits for $N = 2^n$, where n is an integer.

For the two-by-two case, the kernel matrix is

$$\frac{1}{\sqrt{2}} H_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{3-5}$$

and for successively larger N, these can be generated from the block matrix form

$$\frac{1}{\sqrt{N}} H_N = \frac{1}{\sqrt{N}} \begin{bmatrix} H_{N/2} & H_{N/2} \\ H_{N/2} & -H_{N/2} \end{bmatrix}. \tag{3-6}$$

For any size $N = 2^n$, the matrix contains only elements that are ± 1, provided that the $N^{-1/2}$ factor is kept out in front. This makes the transform less expensive to compute.

The Hadamard transform corresponds to expansions along square waves rather than the sine and cosine waves of the Fourier transform [11].

The basis functions of the Hadamard transform actually are Walsh functions. Thus, the Hadamard transform is also referred to as the Walsh transform [1].

THE DISCRETE COSINE TRANSFORM

The discrete cosine transform (DCT) [22] consists of a set of basis vectors that are sampled cosine functions. The transform pair is

$$C(u,v) = \begin{cases} \dfrac{1}{N} \displaystyle\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y), & u = v = 0 \\ \dfrac{1}{2N^3} \displaystyle\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y)[\cos(2x+1)u\pi][\cos(2y+1)v\pi], & otherwise \end{cases} \tag{3-7}$$

and

$$f(x,y) = \frac{1}{N} C(0,0) + \frac{1}{2N^3} \sum_{u=1}^{N-1} \sum_{v=1}^{N-1} C(u,v)[\cos(2x+1)u\pi][\cos(2y+1)v\pi]. \tag{3-8}$$

The basis vectors of the DCT are related to a class of discrete Chebyshev polynomials, which are orthogonal polynomials. Thus the DCT is that its basis vectors are directly related to the eigenvectors of a corresponding Toeplitz matrix.

Like the DFT, the DCT can be computed by a fast algorithm. Unlike the DFT, the DCT is real valued. It has found wide usage in image compression [1].

3.3 Image Enhancement

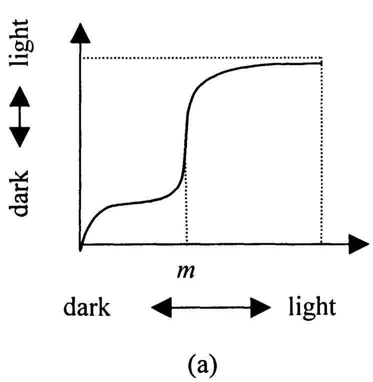
Just as other images, the Biometrics images may be corrupted by various kinds of noise. For example, the fingerprint images can be corrupted by these noises: ridges of different widthes, ridge breaks, bridges between ridges, and inner holes [23]. So we have to enhance the images. Image enhancement techniques are designed to improve image quality for human viewing [20]. The principal objective of enhancement techniques is to process a given image so that the result is more suitable than the original image [12]. The image enhancement approaches can be divided into two broad categories: frequency-domain methods and spatial-domain methods [12]. Processing techniques in the first category are based on modifying the Fourier transform of an image. The spatial-domain, on the other hand, refers to the image plane itself, and approaches in this category are based on manipulation of the pixels in an image.

THE SPATIAL-DOMAIN METHOD

Spatial domain techniques are based on gray-level mappings, where the type of mapping used depends on the criterion chosen for enhancement [24-25].

Contrast Enhancement

Let r and s denote any gray level in the original and enhanced image, respectively, and suppose that for every pixel with level r in the original image, we create a pixel in the enhanced image with level $s = T(r)$. If $T(r)$ has the form shown in Figure 3-1(a), the effect of this transformation will be to produce an image of higher contrast than the original by darkening the levels below a value m and brightening the levels above m in the original pixel spectrum. In this technique, the levels of r below m are compressed by the transformation function into a narrow range of s toward the dark end of the spectrum; the opposite effect takes place for values of r above m. In the limiting case shown in Figure 3-1(b), $T(r)$ produces a two-levels (i.e., binary) image.



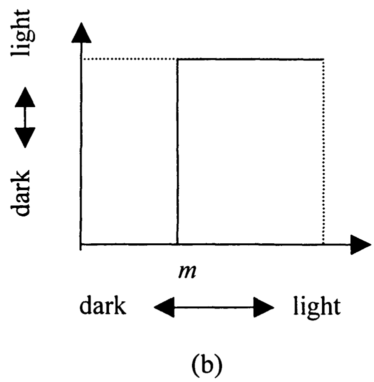


Figure 3-1. Gray-level transformation functions for contrast enhancement.

Histogram Modification

The gray-level histogram is a function showing, for each gray level, the number of pixels in the image that have that gray level [1]. It is one of the simplest and most useful tools in digital image processing. A histogram of gray level content provides a global description of the appearance of an image. We can achieve enhancement by modifying the histogram of a given image in a specified manner. The type and degree of enhancement obtained depends on the nature of the specified histogram.

Histogram Equalization
Suppose we want to take a given input image into an output image with equally many pixels at every gray level (a flat histogram). This can be useful for putting images into a consistent format prior to comparison or segment.

Suppose n_k is the number of times this level appears in the image, and n is the total number of pixels in the image, the probability of the kth gray level $p_r(r_k)$ can be defined as

$$p_r(r_k) = \frac{n_k}{n}, \quad 0 \le r_k \le 1, \ k = 0, 1, 2, \cdots, \ L-1, \tag{3-9}$$

where L is the number of levels of an image. A plot of $p_r(r_k)$ vs. r_k is usually called a histogram, and the technique used for obtaining a uniform histogram is known as histogram equalization or histogram linearization.

The histogram equalization is given by the relation

$$s_k = T(r_k) = \sum_{j=0}^{k} \frac{n_j}{n} = \sum_{j=0}^{k} p_r(r_j), \qquad 0 \le r_k \le 1, \ k = 0, 1, 2, \cdots, L-1 \tag{3-10}$$

After a histogram equalization point operation, the actual histogram will usually take on a rather ragged appearance due to the finite number of available gray levels. Some gray levels will be unoccupied and others highly populated. On the average, however, the histogram will be approximately flat.

Histogram Matching
Sometimes it is desirable to transform an image so that its histogram matches that of another image or a specified functional form. This could be used, for example, before comparing two images of the same scene when they have been digitized under different lighting conditions.

The histogram matching procedure can be described as follows: [20]

(1) Histogram-equalize the original image, $s = T(x)$.

(2) Specify the desired density function and obtain the equalizing transformation $U(y)$.

(3) Apply the inverse transformation $y = U^{-1}(s)$ to the previously equalized image s.

This procedure may be compressed into a single transformation of the original image by combining the transformations

$$y = U^{-1}[T(x)]. \tag{3-11}$$

Noise Smoothing
Noise arising from an electronic sensor generally appears as random, uncorrelated, additive errors or "snow". This type of noise may give rise to extreme pixel-to-pixel changes rather than the small changes normally observed from natural scene [20]. In these cases, the replacement of each pixel of the noisy image by a weighted sum of its

neighbors will reduce the variability among adjacent pixels, and we will obtain a picture that is closer to the original. In this way, we are led to an equation that describes the relation between the original image $f(m,n)$ and the filtered image $g(m,n)$:

$$g(m,n) = \sum_i \sum_j f(i,j)H(m-i+1, n-j+1), \tag{3-12}$$

where H is a convolution matrix.

In order to smooth the noise, H can be the low-pass form as below:

$$H = \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad H = \frac{1}{10}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad H = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \tag{3-13}$$

Edge Enhancement

Edge enhancement, sharpening, or crispening techniques are designed to increase the visibility of general low-contrast edges, and often lead to increased perception of detail.

In order to enhance the edges, H in Equation (12) can be the high-pass form as below:

$$H = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad H = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad H = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 5 & -2 \\ 1 & -2 & 1 \end{bmatrix} \tag{3-14}$$

Edge enhancement may also be approached by considering the derivative of the image functions [20]. The digital gradient at each point is approximated by differences. Two typical approximation are given by the relations

$$G[f(x,y)] \approx |f(x,y) - f(x+1,y)| + |f(x,y) - f(x,y+1)| \tag{3-15}$$

and

$$G[f(x,y)] \approx |f(x,y) - f(x+1,y+1)| + |f(x+1,y) - f(x,y+1)|. \tag{3-16}$$

The relationship between pixels in Equations (15)-(16) is shown in Figures 3-2(a) and (b), respectively. Equation (16) sometimes is called the Roberts gradient.

THE FREQUENCY-DOMAIN METHOD

The foundation of frequency-domain techniques is the Fourier transform and the convolution theorem. Let $g(x,y)$ be an image formed by the convolution of an image $f(x,y)$ and a position-invariant operator $h(x,y)$, that is,

$$g(x,y) = h(x,y) * f(x,y). \tag{3-17}$$

Then, from the convolution theorem, we have that the following frequency-domain relation holds:

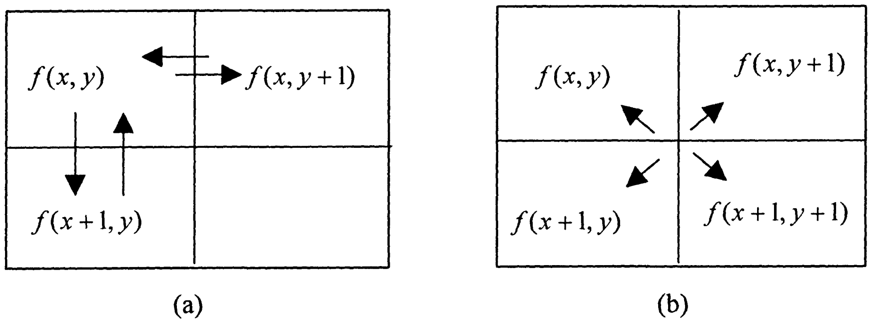


Figure 3-2. Two procedures for computing a two-dimensional, discrete gradient.

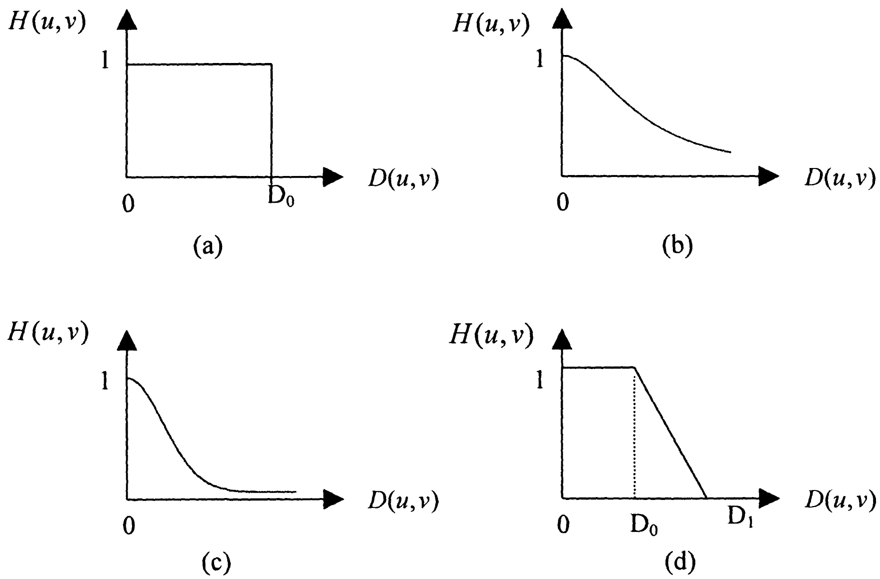


Figure 3-3. (a) Ideal lowpass filter cross section. (b) Butterworth lowpass filter cross section. (c) Exponential lowpass filter cross section. (d) Trapezoidal lowpass filter cross section.

$$G(u,v) = H(u,v)F(u,v)\,, \tag{3-18}$$

where $G, H,$ and F are the Fourier transforms of $g, h,$ and f, respectively. The transform $H(u,v)$ is sometimes called the transfer function of the process. All the image enhancement techniques in frequency domain are based on the Equation (18).

Image Smoothing

Edges and other sharp transitions (such as noise) in the gray levels of an image contribute heavily to the high-frequency content of its Fourier transform. It follows, therefore, that blurring can be achieved via the frequency domain by attenuating a specified range of high-frequency components in the transform of a given image. In Equation (18), the problem is to select a function $H(u,v)$, which yields $G(u,v)$ by attenuating the high-frequency components of $F(u,v)$. The inverse transform of $G(u,v)$ will then yield the desired smoothed image, $g(x,y)$. Since high-frequency components are "filtered out", and information in the low-frequency range is "passed" without attenuation, this method is commonly referred to as low-pass filtering [12]. The function $H(u,v)$ is referred to in this context as a filter transfer function. Several useful lowpass filters [12, 20] are shown in Figure 3-3. In all cases, the filters are functions that affect corresponding real and imaginary components of the Fourier transform in exactly the same manner. Such filters are referred to as Zero-phase-shift-filters because they do not alter the phase of the transform. In Figure 3-4, D_0 is a specified nonnegative quantity, and $D(u,v)$ is the distance from point (u,v) to the origin of the frequency plane, that is

$$D(u,v) = (u^2 + v^2)^{\frac{1}{2}} . \tag{3-19}$$

Image Sharpening

It was shown in last section that an image could be blurred by attenuating the high-frequency components of its Fourier transform. Since edges and other abrupt changes in gray levels are associated with high-frequency components, image sharpening can be achieved in the frequency domain by a highpass filtering process which attenuates the low-frequency components without disturbing high-frequency information in the Fourier transform [12]. The term high-frequency enhancement filter, or highpass filter, is generally taken to describe a transfer function $H(u,v)$ in Equation (18) that is unity at zero frequency and increases with increasing frequency [1]. Such a transfer function may either level off at some value greater than unity or, more commonly, fall back toward zero at higher frequencies. In the latter case, the high-frequency enhancement filter is actually a type of bandpass filter [1] with the restriction of unity gain at zero frequency.

Several useful highpass filters [12, 20] are shown in Figure 3-4, where D_0 is a specified nonnegative quantity, and $D(u,v)$ which is defined by the Equation (19) is the distance from point (u,v) to the origin of the frequency plane.

3.4 Image Restoration

Image restoration deals with images that have been recorded in the presence of one or more sources of degradation [26-30]. There are many degradations in Biometrics images, including the blurring that can be introduced by optical systems, image motion, and the like, as well as noise from electronic and photometric sources. Some types of degradation affect only the gray levels of the individual picture points without introducing spatial blur, these are sometimes called point degradations. Other types which do involve blur are called spatial degradations. Still other types involve chromatic or temporal effect [26]. We have to restore these images. In this section, we will only be concerned with point and spatial degradations.

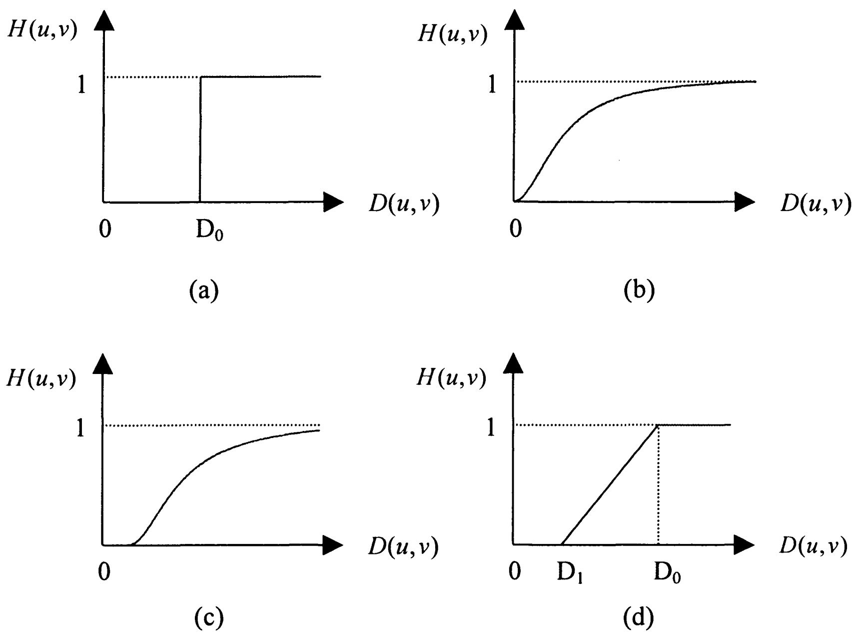


Figure 3-4. (a) Ideal high pass filter cross section. (b) Butterworth highpass filter cross section. (c) Exponential highpass filter cross section. (d) Trapezoidal highpass filter cross section.

The Image enhancement techniques are often used for improving the quality of images for human viewing or as preprocessing techniques such as normalization of a set of images for automatic pattern recognition. On the other hand, Image restoration may be defined as the improvement of image quality under an objective evaluation criterion [20]. The aim of image restoration is to bring the image toward what it would have been if it had been recorded without degradation [1].

DEGRADATION MODEL

As shown in Figure 3-5, the degradation process will be modeled as an operator (or system), H, which together with an additive noise term, n, operates on an input image, f, to produce a degraded image, g. The digital image restoration problem may be viewed as that of obtaining an approximation to f, given g and knowledge of the degradation in the form of the operator H. It is assumed that our knowledge about n is limited to information of a statistical nature [12].

The input-output relationship in Figure 3-4 is given by the expression

$$g = Hf + n. \tag{3-20}$$

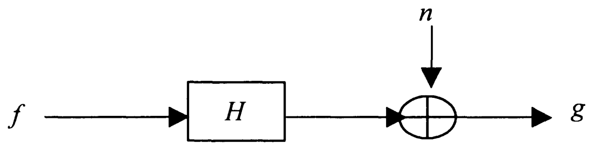


Figure 3-5. A model of the image degradation process.

DECONVOLUTION

In the mid-1960s, deconvolution (inverse filtering) began to be applied broadly to digital image restoration. In most images, adjacent pixels are highly correlated, while the gray levels of widely separated pixels are only loosely correlated [1]. From this, we can argue that the auto-correlation function of typical images generally decreases away from the origin. Since the power spectrum of an image is the (real and even) Fourier transform of its auto-correlation function, we can argue that the power spectrum of an image generally decreases with frequency.

Typical noise sources have either a flat power spectrum or one that decreases with frequency more slowly than typical image power spectra [1]. Thus, the expected situation is for the signal to dominate the spectrum at low frequencies while the noise dominates at high frequencies. Since the magnitude of the deconvolution filter generally increases with frequency, the filter enhances high-frequency noise.

Helstrom adopted the minimum mean-square error estimation (MSE) procedure and presented the Wiener deconvolution filter, which has the two-dimensional transfer function

$$G(u,v) = \frac{H^*(u,v)}{|H(u,v)|^2 + P_n(u,v)/P_f(u,v)}, \tag{3-21}$$

where P_f and P_n are both the power spectra of the signal and noise, respectively.

Wiener deconvolution affords an optimal method for rolling off the deconvolution transfer function in the presence of noise, but it is plagued with three problems that limit its effectiveness [1]. First, the mean square error (MSE) criterion of optimality is not particularly good if the image is being restored for the human eye. The problem is that the MSE criterion weights all errors equally, regardless of their location in the image, while the eye is considerably more tolerant of errors in dark areas and high-gradient area than elsewhere. In minimizing the mean square error, the Wiener filter also tends to smooth the image more than the eye would prefer.

Second, classical Wiener deconvolution can not handle a spatially variant blurring PSF. This occurs with coma, astigmatism, curvature of field, and with motion blur that involves rotation.

Finally, the technique can not handle the common case of non-stationary signals and noise. Most images are highly non-stationary, having large flat areas separated by sharp transitions (edges). Furthermore, several important noise sources are highly dependent on local gray level (signal-dependent noise).

To solve these problems, several alternatives to and improvements upon Wiener deconvolution [1] have been developed.

ALGEBRAIC APPROACHES TO RESTORATION

As the degradation model shown in Figure 3-4 and the Equation (20) indicated, the objective of image restoration is to estimate an original image f, given a degraded image g and some knowledge or assumption about H and n [12]. From Equation (20), it is possible to formulate a class of image restoration problems in unified linear algebraic framework.

Central to the algebraic approach is the concept of seeking an estimate of f, denoted by $\bar{f}$, which minimizes a predefined criterion of goodness.

Unconstrained Restoration

In the absence of any knowledge about n, a meaningful criterion function is to seek a $\bar{f}(x,y)$ such that $H\bar{f}$ approximates g in a least-squares sense by assuming that the norm of the noise term is as possible [12]. In other words, we wish to find an $\bar{f}$ such that

$$\| n \|^2 = \| g - H\hat{f} \|^2 \tag{3-22}$$

is minimum, where, by definition, $\| n \|^2 = n'n$, and $\| g - H\hat{f} \|^2 = (g - H\hat{f})'(g - H\hat{f})$ are the squared norms of n and $(g - H\hat{f})$, respectively.

By minimizing Equation (22), we can obtain the relation

$$\bar{f} = H^{-1}g\,. \tag{3-23}$$

Constrained Restoration

The least squares restoration approach is characterized by the possibility of singular solutions. A general approach to problems of this nature that permit a variety of solutions is to consider natural constraints that may reduce this variety, perhaps to a unique solution. The constrained restoration solution adds two important steps to the unconstrained restoration problem [20]. First, natural constraints in mathematical form must be developed. Then, solutions to the restoration problem subject to these constraints must be developed. Also, these two steps must be compatible. In a practical sense, it is undesirable to have either a natural constraint with no solution or a constrained solution for an unnatural constraint.

The constrained least squares approach to restoration performs very well for equality constraints. However, a closed form solution is not easily obtained if the constraint is in the form of an inequality [20].

A summary of several important constrained least square filters is shown in Table 3-2 [20].

Table 3-2. Constrained least-squares restoration summary.

Filter	*Criteria*	*Algebraic solution*
Inverse filter	$\parallel g-\hat{g}\parallel^2$	$\bar{f}=([H^*]'[H])^{-1}[H^*]'g$
Energy constrained	$\parallel g-Hf\parallel^2+\lambda(C-\parallel f\parallel^2)$	$\bar{f}=\{[H^*]'[H]+\lambda I\}^{-1}[H^*]'g$
Smoothness constrained	$\parallel g-[H]f\parallel^2+\lambda(C-\parallel [Q]f\parallel^2)$	$\bar{f}=\{[H^*]'[H]+\alpha[Q^*]'[Q]\}^{-1}[H^*]'g$
Maximum entropy	$\parallel g-[H]f\parallel^2+\lambda(f\ln f+C)$	$\bar{f}=\exp\{-1+2\lambda[H^*]'(g-[H]\bar{f})\}$

STOCHASTIC RESTORATION

The class of stochastic restoration techniques is based upon statistical considerations of images and noise as stochastic or random process [20]. The linear image restoration model, Equation (20), can be considered in several ways [20]. If the only random process involved is the additive noise n, then solving for the minimum mean squared estimate of f will be called the regression problem. If the image f is also considered as a random image with known first- and second-order moments, then the problem will be referred to as Wiener estimation. If the probability density function of g given f is known completely or parametrically, then maximum likelihood estimation may be used. Finally, if the probability density of f given g is known, the Bayes or maximum a posteriori estimation techniques may be used. This overview of the stochastic restoration techniques is shown in a flowchart form in Figure 3-6 [20]. In considering random processes in restoration one may proceed down the flowchart, which is ordered from difficult to very difficult in terms of the problem solutions.

3.5 Data Compression

Biometrics image processing generally creates significant numbers of large files containing digital image data. This call for efficient methods for the storage and transfer of Biometrics image data files. The possibility of an image data reduction is indicated by two observations. First, there is a large amount of statistical redundancy or correlation in normal images. Second, there is a large amount of psychovisual redundancy in most images [12]. The main idea of digital data compression [31-33] is to represent data bits with a lesser amount of storage while containing the essence of data. This means that useful information in data is represented in more compact form, while less important information may be represented with less precision or even be discarded. Figure 3-7 illustrates the procedure of data compressing and image reconstruction.

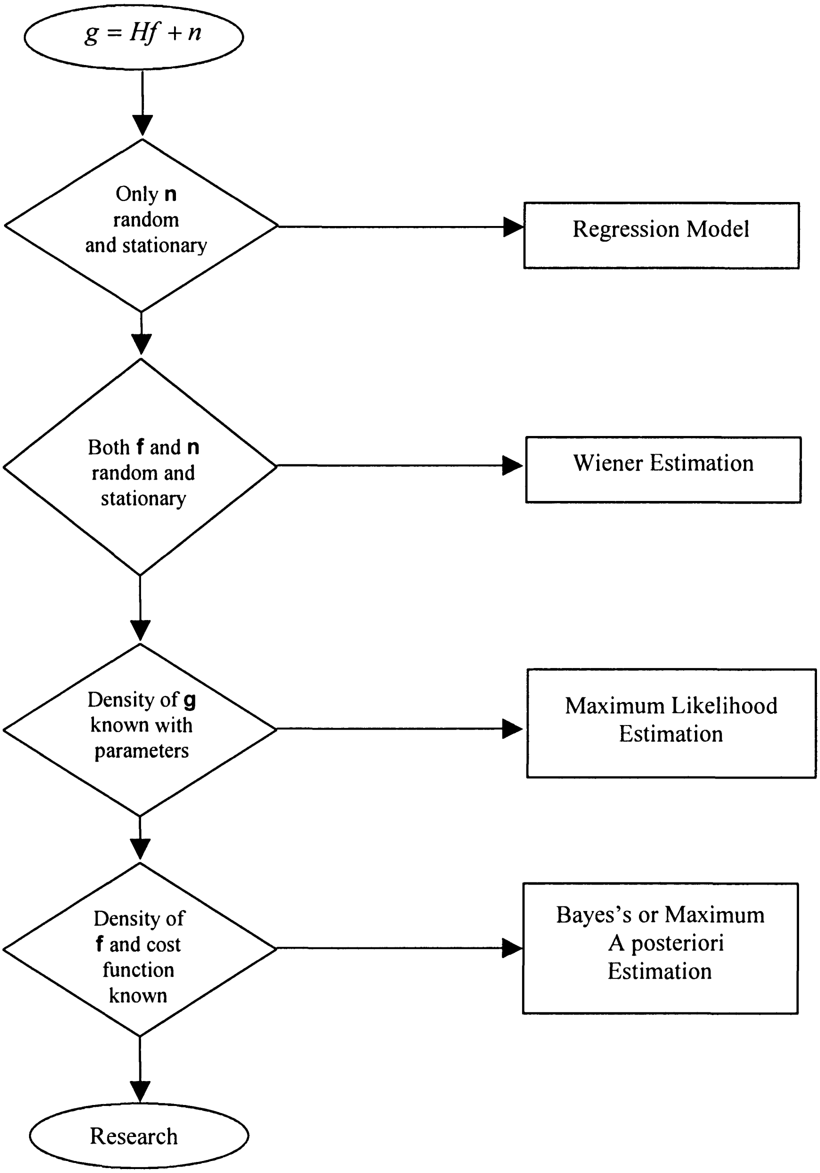


Figure 3-6. Stochastic restoration model decision flowchart.

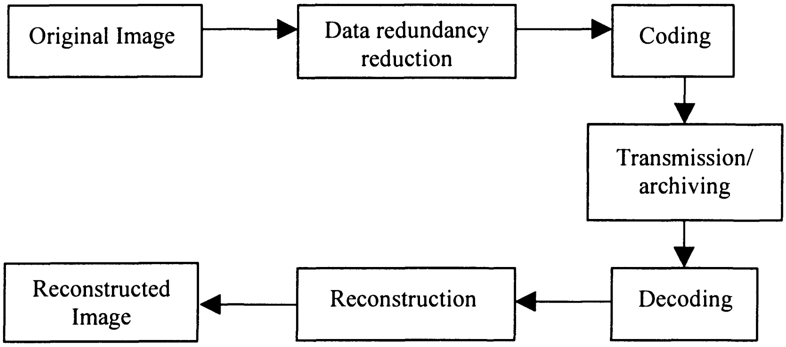


Figure 3-7. Data compression and image reconstruction.

By considering the characteristics of reconstructed data after the process of compression and decompression, data compression can be divided generally into two major types: lossless and lossy.

LOSSLESS COMPRESSION TECHNIQUES

A lossless algorithm eliminates only redundant information, so that one can recover the image exactly upon decompression of the file. Lossless data compression algorithms fall into broad categories: dictionary-based techniques and statistical methods [1]. Dictionary-based techniques generate a compressed file containing fixed-length codes (usually 12 to 16 bits), each represents a particular sequence of values in the original file. Statistical methods implement data compression by representing frequently occurring characters in the file with fewer bits than they do less commonly occurring ones.

Dictionary-based Techniques

Run-Length Encoding

The simplest dictionary-based data compression technique is run length encoding (RLE). In an image being stored line by line, a series of pixels having the same gray-level value is called a *run* [1]. One can store a code specifying that value, followed by the length of the run, rather than simply storing the same value many times over. This is run-length encoding. An RLE image is thus encoded as a sequence of (value, length) pair. It achieves considerable compaction, for example, with graphics and with images of objects residing upon a constant background. Other types of images compress poorly. Under worst case conditions (for example, where every pixel differs from its neighbors) RLE can actually double the size of the file.

LZW Encoding

LZ coding is a lossless technique first described by Lemple and Ziv [34]. It was extended by Welch [35] to form the widely used, proprietary LZW algorithm. Like

RLE, it effects compression by encoding strings of characters. However, unlike RLE, it builds up a table of strings (particular sequences of bytes) and their corresponding codes as it encodes the file [1]. A file of 8-bit bytes can be encoded, for example, into 12-bit codes. Of the 4,096 possible codes, 256 of them represent all possible single bytes. The remaining 3,840 are assigned to strings as they are encountered in the data during compression.

The first time a string not already in the table occurs, it is stored in full, along with the code that is assigned to it. Thereafter, when that string occurs again, only its code is stored. This squeezes redundancy out of the file. Not only is the string table built dynamically during compression, but it need not be stored with the compressed file: The decompression algorithm can reconstruct it from the information in the compressed file [1].

Statistical Encoding Methods

Huffman Encoding

Huffman coding [36] is a classical data compression technique. It has been used in various compression applications, including image compression. It is a simple but yet elegant compression technique that can supplement other compression algorithms. The main idea of Huffman coding is to represent data by codes of variable length, with more frequent data being represented by shorter codes. This simple idea causes a reduction in the average code length, and thus the overall size of compressed data is smaller than the original. The elegantly simple procedure of Huffman guarantees that one may obtain a uniquely decodable code with the minimum average number of bits per message R with [20]

$$H \leq R \leq H+1, \text{ where } H = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{3-24}$$

where $p_1, p_2, \cdots, p_n$ are the probability functions of these messages over the appropriate ensemble of digitized picture.

A Huffman code can be constructed [12] by first ordering the input probabilities according to their magnitudes. The two smallest probabilities are combined by addition to form a new set of probabilities. The new set of probabilities, which has one fewer probability than the original set, is again ordered according to magnitude. Equal probabilities can be ordered in any way. When we get down to two probabilities we stop. Code words are generated by starting at the last step and working backwards. We start by assigning a 0 to one of the last two combined probabilities, and a 1 to the other. We now process backwards, decomposing probabilities and generating code words as we go. The same procedure is repeated until it gets the input probabilities, at which point we have a code word assigned to each input symbol.

Arithmetic Encoding

Arithmetic coding [37] is also a kind of statistical coding algorithm similar to Huffman coding. However, it uses a different approach to utilize symbol probabilities, and is performing better than Huffman coding. In Huffman coding, optimal codeword length is obtained when the symbol probabilities are of the form, where x is an integer because Huffman coding assigns a code with integral number of bits. Obviously, this form of symbol probabilities is rare in practice. Arithmetic coding is a statistical coding method that solves this problem. The code from it is not restricted to integral number of bit, it can assign a code as a fraction of a bit. Therefore, when the symbol probabilities are more arbitrary, arithmetic coding has better compression ratio than Huffman coding. In brief, this is can be considered as grouping input symbols and coding them into one long code. Therefore different symbols can share a bit from the long code. The main procedure of arithmetic coding is to try to find an interval of number between 0 and 1 to represent an alphabet in the source. According to the actual order of these alphabets in the stream, these intervals are then combined until the last alphabet is accounted for, then a final interval of numbers is obtained. For example, when 'B' has the interval (0.2,0.3), and 'I' has the interval (0.5,0.6), then 'BI' should have the interval (0.25,0.26). In other words, as the stream of alphabets becomes longer, the interval representing it becomes finer and finer in precision. Then a number representing the whole stream can be found by choosing a number from this final interval. For decoding operation, encoded symbols can be decoded back to their original order by using the number we got from the encoder. The first symbol is decoded according to the interval of the number. Then the probability fraction for the first symbol is removed from the number by subtract the lower bound of the first symbol and then divided by the interval width of the first symbol. The resulting number is then used to find the second symbol. This process goes on until the last symbol is decoded. Although it is more powerful than Huffman coding in compression ratio, arithmetic coding is more number crunching that requires more computational power and memory. Huffman coding is more attractive comparing with arithmetic coding when simplicity is the major concern.

LOSSY IMAGE CODING

A lossy compression algorithm eliminates irrelevant information as well, and thus permit only an approximate reconstruction of the original, rather than an exact duplicate [1]. As one might expect, lossy compression algorithms achieve higher compression ratios. For images, a slight loss of fidelity is often an acceptable trade-off for a much higher degree of compaction.

Vector Quantization

Vector Quantization (VQ) [38] is a kind of lossy compression method that can be applied in coding digital images. It uses a codebook containing pixel patterns with corresponding index on each of them. The main idea of VQ in image coding is then to represent arrays of pixels by an index in the codebook. In this way, compression is achieved since the size of the index is usually a small fraction of that of the block of pixels.

Transform Image Coding

The basic idea of the transform image coding is that image data is represented by coefficients of discrete image transforms. The transform coefficients are ordered according to their importance and the least important (low contribution) coefficients can be omitted. Suppose we have an ensemble of images to be encode into a compact data representation. We can transform the images, discard those coefficients that are near zero, and coarsely quantize those that are small, thereby concentrating our data transmission and storage resources upon the coefficients that contain the most information about the image [1]. When the image is reconstructed later, little important content will have been lost. To remove correlated image data, the Karhunen-Loeve transform is the most important method. This transform builds a set of non-correlated variables with decreasing variance. Cosine, Fourier, Hadamard transforms are all suitable for image data compression. If an image is compressed using discrete transforms, it is usually divided into subimages of 8×8 or 16×16 pixels to speed up calculations, and then each subimage is transformed and processed separately. This simplifies the transformation process, particularly if eigenvectors must be computed. The subsequent elimination of some transform coefficients and the coarse quantization of others, however, can cause noticeable changes in gray level at edges of the blocks. This is called blocking artifact, and it can make the boundaries of the block obtrusive.

References

[1] K.R. Castleman, *Digital Image Processing*, Prentice Hall, 1996.

[2] S.E. Umbaugh, Ph.D., *Computer Vision and Image Processing: A Practical Approach Using CVIPtools*, Prentice Hall PTR, 1998.

[3] M.A. Sid-Ahmed, *Image Processing – Theory, Algorithms, and Architectures*, McGraw-Hill, Inc., 1994.

[4] L. O'Gorman and J.V. Nickerson, "An Approach to Fingerprint Filter Design," *Pattern Recognition*, vol. 22, no. 1, pp. 29-38, 1989.

[5] P.E. Danielsson and Q.Z.Ye, "Rotation-Invariant Operators Applied To Enhancement of Fingerprints," *Proc. Eighth ICPR*, pp. 329-333, Rome, 1988.

[6] D.C.D. Hung, "Enhancement and Feature Purification of Fingerprint Images," *Pattern Recognition*, vol. 26, no. 11, pp. 1661-1671, 1993.

[7] L.Coetzee and E.C. Botha, "Fingerprint Recognition in Low Quality Images, " *Pattern Recognition*, vol. 26, no. 10, pp. 1441-1460, 1993.

[8] A.Sherstinsky and R.W. Picard, "Restoration and Enhancement of Fingerprint Images Using M-Lattice-A Novel Nonlinear Dynamical System," *Proc. 12th ICPR-B*, pp. 195-200, Jerusalem, 1994.

[9] D.B.G. Sherlock, D.M. Monro, and K. Millard, "Fingerprint Enhancement by Directional Fourier Filtering," *IEEE Proc. Vis. Image Signal Processing*, vol. 141, no. 2, pp. 87-94, 1994.

[10] J.N. Bradley, C. M. Brislawn, and T. Hopper, "The FBI wavelet/scalar quantization standard for gray-scale fingerprint image compression," *SPIE Proceedings*, vol. 1961, Visual Information Processing II, Orlando, Florida, Apr. 1993.

[11] T. Pavlidis, *Algorithms for Graphics and Image Processing*, Computer Science Press, 1982.

[12] R.C. Gonzalez, P. Wintz, *Digital Image Processing*, Addison-Wesley Publishing Company, Inc., 1977.

[13] H. Shatkay, "The Fourier Transform - A Primer," *ftp://ftp.cs.brown.edu/pub/techreports/95c9-5-37.ps.Z*, 1995.

[14] R.N. Bracewell. *The Fourier Transform and Its Applications (2d revised ed.)*, McGraw-Hill, New York, 1986.

[15] R.Tolimieri, *Algorithms for Discrete Fourier Transform and Convolution*, Springer-Verlag, New York, 1989.

[16] J.S. Walker, *Fourier Analysis*, Oxford University Press, Oxford, U.K., 1988.

[17] E.O. Brigham and R.E. Morrow, "The Fast Fourier Transform," *IEEE Spectrum*, vol. 4, pp. 64-70, December 1967.
[18] D.F. Elliott and K.R. Rao, *Fast Transforms: Algorithms and Applications*, Academic Press, New York, 1983.
[19] R.W. Ramirez, *The FFT, Fundamentals and Concepts*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
[20] E. L. Hall, *Computer Image Processing and Recognition*, Academic Press, 1979.
[21] W.K. Pratt, H.C. Andrews, and J. Kane, "Hadamard Transform Image Coding," *Proc. IEEE*, vol. 57, no. 1, pp. 58-68, 1969.
[22] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transforms," *IEEE Trans. Comp.* vol. C-23, pp. 90-93, 1974.
[23] D.C.Douglas Hung, "Enhancement and Feature Purification of Fingerprint Images," *Pattern Recognition*, vol. 26, no. 11, pp. 1661-1671, 1993.
[24] J.J. Soltis, M.A. Sid-Ahmed, and M. Srdanovic, "Edge Enhancement in Digital Images Using Phase Contrast Filtering," *Can. J. Elect. & Comp. Eng.*, vol. 15, no. 1, pp. 22-26, 1990.
[25] M.A. Sid-Ahmed and J.J. Soltis, "Processing of Digital Images Using Phase contrast Filters Realized as 2-D IIR Filters," *Signal Processing*, vol. 18, pp. 231-237, 1989.
[26] A. Rosenfeld, A.C. Kak, *Digital Picture Processing*, Academic Press, 1982.
[27] H.C. Andrews and B.R. Hunt, *Digital Image Restoration*, Prentice-Hall, 1977.
[28] H.C. Andrews, "Digital Image Restoration: A Survey," *IEEE Computer*, vol. 7, no. 5, pp. 36-45, May 1974.
[29] A.K. Katsaggelos, ed., *Digital Image Restoration*, Springer-Verlag, Berlin, New York, 1991.
[30] I. Sezan and A.M. Tekalp, *Image Restoration*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1992.
[31] J.A. Storer, *Data Compression*, Computer Science Press, Rockville, MD, 1998.
[32] M. Rabbani and P.W. Jones, "Digital Image Compression Techniques," *SPIE-International Society for Optical Engineering*, 1991.
[33] R.K. Miller and T.C. Walker, *Image Compression*, SEAI Technical Publications, 1991.
[34] J. Ziv and J. Lempel, "A Universal Algorithm for Sequential Data Compression," *IEEE Trans. Info. Theory*, vol. IT-23, no. 3, pp. 337-343, 1977.
[35] T.A. Welch, "A Technique for High-Performance Data Compression," *IEEE Computer*, vol. 17, no. 6, pp. 8-19, 1984.
[36] D.A. Huffman, "A Method for the Construction of Minimum Redundancy Codes," *Proc. IRE*, vol. 40, no. 10, pp. 1098-1101, 1952.
[37] I.H. Witten, R.M. Neal, and J.G. Cleary, "Arithmetic Coding for Data Compression.," *Communications of the ACM*, vol. 30, no. 6, pp. 520-540, 1987.
[38] R.M. Gray, "Vector Quantization," *IEEE ASSP Magazine*, pp. 4-29, April 1984.

4 PATTERN RECOGNITION

In this chapter, we will discuss some basic concepts about pattern and pattern recognition dealing with biometrics, as well as an illustrated general configuration of pattern recognition system. Then, the techniques for image segmentation, including pixel classification, gradient-based segmentation, and edge detection, are reviewed in Section 4.2. Section 4.3 presents a brief introduction to feature selection. In Section 4.4, the principles of pattern classification are described. Finally, Section 4.5 is devoted to the theory of neural networks.

4.1 Pattern, Pattern Recognition and System

PATTERN AND PATTERN RECOGNITION

The patterns we encounter can fall into two categories: abstract and concrete. Examples of abstract items include ideas and arguments. Recognition of such patterns, termed *conceptual recognition* [1], is beyond the scope of this book. Examples of concrete items include characters, symbols, pictures, biomedical images, three-dimensional physical objects, target signatures, speech waveforms and seismic waves. Some of these items are spatial, whereas others are temporal.

A problem of pattern recognition usually denotes classification and / or description of a set of processes or events. The set of processes or events to be classified could be a set of physical objects or a set of more abstract ones, such as mental states [2]. The processes or events with some similar properties are grouped into a class. A total number of pattern classes in a particular problem are often determined by the particular application in mind. For example, considering the problem of English character recognition: we should have a problem of 26 classes. On the other hand, if we are interested in discriminating English characters from Russian characters, we have only a two-class problem. In some problems, the exact number of classes may not be known initially, and it may have to be determined from the observation of many representative patterns. In this case, we would like to detect the possibility of having new classes of patterns as we observe more and more patterns.

In the last couple of decades, the researchers' interests have focused on two types of pattern recognition problems:

(1) The mechanism of the pattern recognition system possessed by living organisms.
(2) The development of theories and techniques of computer implementation for a given recognition task.

CONFIGURATION OF PATTERN RECOGNITION SYSTEM

In pattern recognition related to biometrics, we can divide an entire task into four phases: data acquisition, data preprocessing, feature extraction, and decision classification, as shown in Figure 4-1. In the data acquisition phase, analog data from the physical world are gathered through a transducer and converted to digital format suitable for computer processing. In this stage, the physical variables are converted into a set of measured data. The measured data are then used as the input to the second phase (data preprocessing). In the next step (feature extraction), the preprocessed data are grouped into a set of characteristic features. The last phase is actually a classifier that is in the form of a set of decision functions.

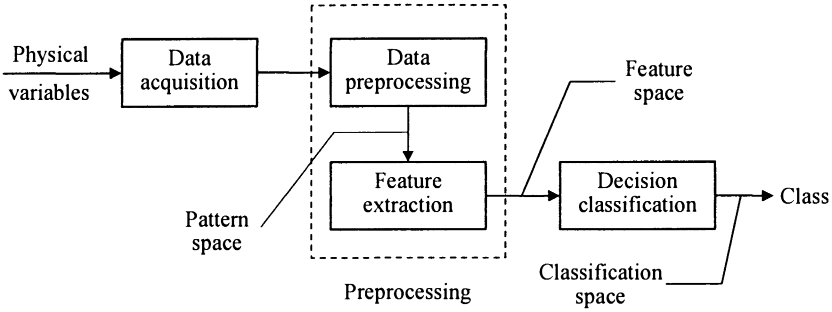


Figure 4-1. Configuration of a pattern recognition system.

4.2 Image Segmentation

INTRODUCTION

In image compression or enhancement, the desired output is a picture — an approximation to, or an improved version of, the input picture. Another major branch of picture processing deals with *image analysis* or *scene analysis*; here the input is still pictorial, but the desired output is a *description* of the given picture or scene. The following are examples of image analysis problems which have been extensively studied:

1) The input is text (machine printed or handwritten), and it is desired to read the text; here the desired description of the output consists of a sequence of names of characters.
2) The input is a nuclear bubble chamber picture, and it is desired to detect and locate certain types of "events"; here the description consists of a set of coordinates and names of event types.
3) The input is an aerial photograph of terrain; the desired output is a map showing specific types of terrain features (forests, urban areas, bodies of water, roads, etc.). Here again the output is pictorial, and is even in registration with the input; but construction of the output requires location and identification of the desired terrain types.
4) The input is a medicinal image; the desired output is a special interested part which can be used to make diagnoses.

In all of these examples, the description refers to specific parts (regions or objects) in the picture or scene; to generate the description, it is necessary to segment the picture into these parts. Thus to identify the individual characters in text, they must first be singled out; to locate bubble chamber events, the bubble tracks and their ends or branches must be found; and so on.

We can define the image segmentation process as one that partitions a digital image into disjoint (nonoverlapping) regions [3]. For general purposes, a *region* is a connected set of pixels — that is, a set in which all the pixels are adjacent or touching [4]. The formal definition of *connectedness* is as follows: Between any two pixels in a connected set, there exists a connected path wholly within the set, where a *connected path* is a path that always moves between neighboring pixels. Thus, in a connected set, you can trace a connected path between any two pixels without ever leaving the set.

There are two rules of *connectivity*. If only laterally adjacent pixels (up, down, right, left) are considered to be connected, this is *four-connectivity*, and the objects are *four-connected*. Thus, each pixel has only four neighbors to which it can be connected. If, in addition, diagonally adjacent (45-degree neighbor) pixels are also considered to be connected, we have *eight-connectivity*, and the objects are *eight-connected*. Each pixel would then have eight neighbors to which it could be connected. Often, eight-connectivity yields results that lie close to one's intuition [5].

Image segmentation can be approached from three different philosophical perspectives. In the case we call the *region approach*, one assigns each pixel to a particular object or region on the basis of their gray levels or their spectral or spatial signatures. In the *boundary approach*, one attempts only to locate the boundaries that exist between the regions. In the *edge approach*, one seeks to identify edge pixels and then link them together to form the required boundaries.

PIXEL CLASSIFICATION BY THRESHOLDING

Segmentation is basically a process of *pixel classification*; the image is segmented into subsets by assigning the individual pixels to classes. In addition, thresholding is a particularly useful region-approach technique for scenes containing solid objects resting upon a contrasting background. It is computationally simple and never fails to define disjoint regions with closed, connected boundaries.

When using a threshold rule for image segmentation, one assigns all pixels at or above the threshold gray level to the object. All pixels with gray level below the threshold fall outside the object. The boundary is then a set of interior points, each of which has at lease one neighbor outside the object. It is obvious that threshold selection is essential to this technique.

If the objects differ from their background by some property other than gray level. For example, given a color picture or a picture obtained by a multispectral scanner, we can first use an operation that converts that property to gray level. Then gray-level thresholding can segment the processed image.

Global Thresholding

In the simplest implementation of boundary location by thresholding, the value of the threshold gray level is held constant throughout the image. If the background gray level is reasonably constant throughout, and if the objects all have approximately equal contrast above the background, then a fixed global threshold will usually work well, provided that the threshold gray level is properly selected.

Adaptive / Multilevel Thresholding

In many cases, the background gray level is not constant, and the contrast of objects varies within the image. In such case, a threshold that works well in one area of the image might work poorly in other areas. In these cases, it is convenient to use a threshold gray level that is a slowly varying function of position in the image.

Figure 4-2 [5] shows a microscope image of the chromosomes from a single human blood cell, of which the background gray level varies due to non-uniform illumination and contrast varies from one chromosome to the next. In Figure 4-2(a), a constant threshold (global thresholding) gray level has been used throughout the image to isolate the chromosome. In Figure 4-2(b), the threshold was varied from one chromosome to the next commensurately with local background and the contrast of the chromosomes [6,7]. This produced fewer segmentation errors — cases where multiple chromosomes were stuck together or individual chromosomes were broken up. In Figure 4-2(b), the threshold for each chromosome was set approximately midway between its mean interior gray level and the local background gray level [6,8].

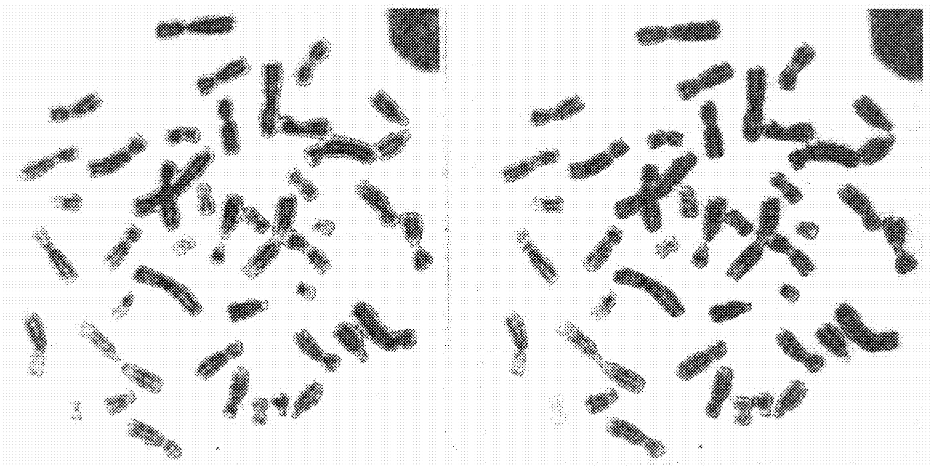


Figure 4-2. Global and adaptive thresholding.

Histogram Techniques

An image containing an object on a contrasting background has a bimodal gray-level histogram (Figure 4-3). The two peaks correspond to the relatively large numbers of points inside and outside the object. The dip between the peaks corresponds to the relatively few points around the edge of the object. In cases like this, the histogram is commonly used to establish the threshold gray level [9-12].

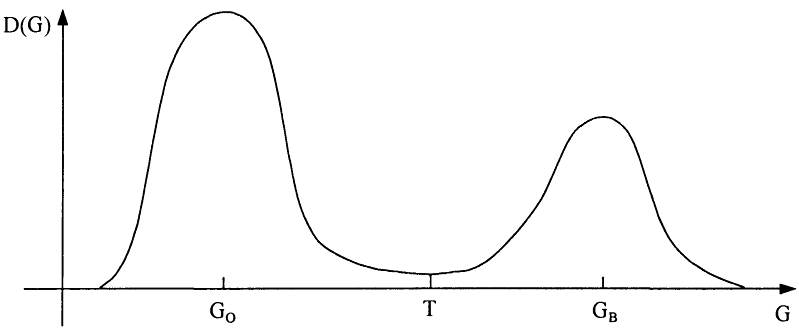
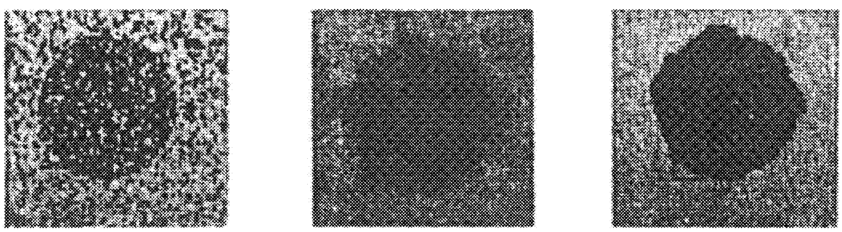


Figure 4-3. The bimodal gray-level histogram.

Also, one can form a histogram of only those pixels having a relatively high gradient magnitude [12] — for example, the highest 10 percent [13]. This eliminates the large number of interior and exterior pixels from consideration and may make the dip in the histogram more accessible, which is essential to pixel classification by thresholding. One can also divide the histogram by the average gradient of pixels at each gray level to further enhance the dip [12], or average the gray level of high-gradient pixels to determine a threshold [13,14].

Smoothing and Thresholding

If the image or the region of the image containing the object is noisy and not large, the histogram itself will be noisy. Unless the dip is uncommonly sharp, the noise will make its location obscure or at least unreliable from one image to the next. This can be overcome to some extent by smoothing the image before thresholding. For example, we can simply locally average the picture, i.e., replace the gray level at each point by an average of neighboring gray levels. Of course, averaging also blurs the borders of the regions, but thresholding will still extract the regions more or less correctly, though it will smooth out irregularities in their borders. An extreme example of histogram improvement by local averaging is shown in Figure 4-4 [15].



(a) (b) (c)

Figure 4-4. Extreme example of region extraction by averaging and thresholding. (a) Original picture; (b) Results of smoothing (a); (c) Results of thresholding (b).

GRADIENT-BASED SEGMENTATION

The preceding pixel approaches accomplish segmentation by partitioning the image into sets of interior and exterior points. By contrast, boundary approaches attempt to find the edges directly by their distinct character — high gradient magnitudes.

Gradient Image

The gradient is defined as follows: Given a scalar function $f(x,y)$ and a coordinate system with unit vectors **i** in the x-direction and **j** in the y-direction, the gradient is the vector function,

$$\nabla f(x,y) = \mathbf{i}\frac{\partial f(x,y)}{\partial x} + \mathbf{j}\frac{\partial f(x,y)}{\partial y}\ , \tag{4-1}$$

where ∇ indicates the vector gradient operator. An important scalar function is the gradient magnitude, given by

$$|\nabla f(x,y)| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}\ . \tag{4-2}$$

Also, since the square root operation is computationally expensive, Equation (2) is often approximated by the form

$$|\nabla f(x,y)| = \max\left[|f(x,y) - f(x+1,y)|,|f(x,y) - f(x,y+1)|\right] \tag{4-3}$$

The gradient magnitude takes on large values in areas of steep slope, such as at the edges of objects.

Boundary Tracking

When dealing with the gradient magnitude image computed from an image containing a single object on a contrasting background, we can implement the boundary-tracking process [5] by the following steps:

1. Identify the pixel of the highest gray level (i.e., gradient magnitude) as the first boundary point, since it must be on the boundary. If several points have the maximum gray level, then we choose arbitrarily.
2. Search the three-by-three neighborhood centered on the first boundary point and take the neighbor with the maximum gray level as the second boundary point. If two neighbors have the same maximum gray level, we choose arbitrarily.
3. At this point, we begin the iterative process of finding the next boundary point, given the current and previous boundary points. Working in the three-by-three neighborhood centered on the current boundary point, we examine the neighbor diametrically opposite the previous boundary point and the neighbors on each side of it (Figure 4-5). The next boundary point is one of those three that has the highest gray level. If all three or two adjacent boundary points share the highest gray level, then we choose the middle one. If the two nonadjacent points share the highest gray level, we choose arbitrarily.

In the noise-free image of a monotone spot, this algorithm will trace out the maximum gradient boundary; however, even small amounts of noise can send the tracking temporarily or hopelessly off the boundary. Noise effects can be reduced by smoothing the gradient image before tracking or by implementing a *tracking bug* [5]. Even so, boundary tracking does not guarantee closed boundaries, and the tracking algorithm can be lost and run off the border of the image.

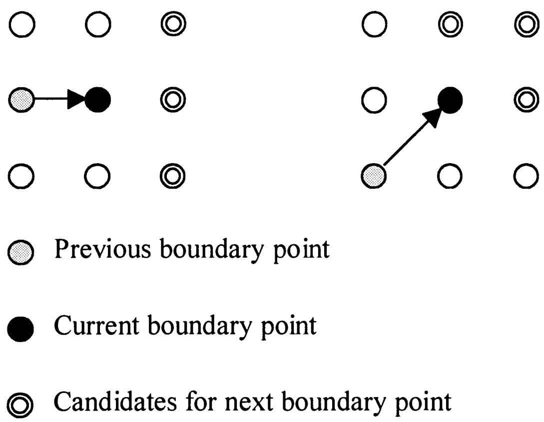


Figure 4-5. Boundary tracking.

Laplacian Edge Detection

The Laplacian is a scalar second-derivative operator for functions of two dimensions. It is defined as

$$\nabla^2 f(x,y) = \frac{\partial^2}{\partial x^2} f(x,y) + \frac{\partial^2}{\partial y^2} f(x,y). \tag{4-4}$$

As we see in Equation (4), the Laplacian $\partial^2 f / \partial x^2 + \partial^2 f / \partial y^2$ is an orientation-invariant derivative operator. Its analog for digital pictures is given by

$$\nabla^2 f(x,y) \equiv [f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1)] - 4f(x,y), \tag{4-5}$$

which is the digital convolution of f with

$$\begin{bmatrix} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{bmatrix}.$$

Also, it can be implemented digitally by the convolution kernels

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

It should be emphasized that if a noise-free image has sharp edges, the Laplacian can find them. However, the presence of noise imposes a requirement for lowpass filtering prior to using the Laplacian. A Gaussian lowpass filter is a good choice for this pre-smoothing. Since convolution is associative [5], we can combine the Laplacian and Gaussian impulse responses into a single *Laplacian of Gaussian* kernel [16,17].

EDGE DETECTION

Edge is one of the important local features which usually involve abrupt changes in gray level of an image. It takes special geometrical form: the gray level is relatively consistent in each of two adjacent, extensive regions, and changes abruptly as the border between the regions is crossed.

If a pixel falls on the boundary of an object in an image, its neighborhood will be a zone of gray-level transition. The two characteristics of principal interest are the slope and direction of that transition. These are the magnitude and direction of the gradient vector, respectively.

Edge detection operators examine each pixel neighborhood and quantify the slope, and often the direction as well, of the gray-level transition. There are several ways to do this, most of which are based on convolution with a set of directional derivative masks.

Roberts Edge Operator

One local differential operator for finding edges is the *Roberts edge detector* [18]. It is give by

$$g(x,y) = \left\{ \left[\sqrt{f(x,y)} - \sqrt{f(x+1,y+1)} \right]^2 + \left[\sqrt{f(x+1,y)} - \sqrt{f(x,y+1)} \right]^2 \right\}^{1/2}, \tag{4-6}$$

where $f(x,y)$ is the input image with integer pixel coordinates (x,y). The inner square roots make the operation resemble the processing that takes place in the human visual system.

Sobel Edge Operator

The *Sobel edge operator* [19] consists of two convolution kernels, shown in Figure 4-6. Each point in the image is convolved with both kernels. One kernel responds maximally to a generally vertical edge and the other to a horizontal edge. The maximum value of the two convolutions is taken as the output value for that pixel. The result is an edge magnitude image.

$$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \qquad \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}$$

Figure 4-6. Convolution kernels of the Sobel edge operator.

Prewitt Edge Operator

The two convolution kernels shown in Figure 4-7 form the *Prewitt edge operator*

[10]. Each point in the image is convolved with both kernels, which responds maximally to a vertical edge and a horizontal edge respectively, and the maximum determines the output. The Prewitt operator likewise produces an edge magnitude image.

$$\begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ +1 & +1 & +1 \end{bmatrix} \qquad \begin{bmatrix} +1 & 0 & -1 \\ +1 & 0 & -1 \\ +1 & 0 & -1 \end{bmatrix}$$

Figure 4-7. Convolution kernels of the Prewitt edge operator.

Kirsch Edge Operator

The eight convolution kernels shown in Figure 4-8 make up the *Kirsch edge operator* [20]. Each point in the image is convolved with all eight masks. Each mask responds maximally to an edge oriented in a particular general direction. The maximum value over all eight orientations is the output value for the edge magnitude image. The index of the maximally responding mask encodes the direction of the edge.

$$\begin{bmatrix} +5 & +5 & +5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \begin{bmatrix} -3 & +5 & +5 \\ -3 & 0 & +5 \\ -3 & -3 & -3 \end{bmatrix} \begin{bmatrix} -3 & -3 & +5 \\ -3 & 0 & +5 \\ -3 & -3 & +5 \end{bmatrix} \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & +5 \\ -3 & +5 & +5 \end{bmatrix}$$

$$\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ +5 & +5 & +5 \end{bmatrix} \begin{bmatrix} -3 & -3 & -3 \\ +5 & 0 & -3 \\ +5 & +5 & -3 \end{bmatrix} \begin{bmatrix} +5 & -3 & -3 \\ +5 & 0 & -3 \\ +5 & -3 & -3 \end{bmatrix} \begin{bmatrix} +5 & +5 & -3 \\ +5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$$

Figure 4-8. Convolution kernels of the Kirsch edge operator.

4.3 Feature Selection

INTRODUCTION

For a pattern recognition problem (recall Figure 4-1), one usually deals with the task of selecting which of the many available features should actually be measured and presented to the classifier. The feature selection problem has received considerable attention in the literature, but no clear-cut solution to it has emerged. In fact, the selection of an initial set of features, while crucial to the performance of the pattern recognition system, still remains largely empirical and draws to a large extent upon the designer's knowledge about the problem at hand. First of all, the number of features should be as low as possible in order to reduce the measurement cost and ease the implementation of the classifier. Secondly, features should be effective in discriminating between pattern classes while being invariant to irrelevant variations, such as limited amount of translation, rotation, scale change, etc.

Feature selection, then, may be viewed as the process of eliminating some features (starting with the poorest) and combining others that are related, until the feature set becomes manageable and performance is still adequate.

Good features have four characteristics [5]:

1. *Discrimination.* Features should take on significantly different values for objects belonging to different classes.
2. *Reliability.* Features should take on similar values for all objects of the same class.
3. *Independence.* The various features used should be uncorrelated with each other.
4. *Small Numbers.* The complexity of a pattern recognition system increases rapidly with the dimensionality (number of features used) of the system. More importantly, the number of objects required to train the classifier and to measure its performance increases exponentially with the number of features [21]. In some cases, it may be impractical to acquire the amount of data required to train the classifier adequately. Finally, adding more features that are either noisy or highly correlated with existing features can actually degrade the performance of the classifier, particularly in view of the limited size of the training set [22-24].

In practice, few or none of the available features are ideal in terms of the foregoing characteristics.

FEATURE SEPARABILITY

Usually, features chosen for classification are not of the same significance. Setting different weights to features according to the significance is important to improve the performance of classifier.

When considering the simple case of a two-class problem, a relevant measure of the ability of a feature to distinguish between two classes is the variance-normalized distance between class means. For feature x, this is given by

$$\hat{D}_{xij} = \frac{\left|\hat{\mu}_{xi} - \hat{\mu}_{xj}\right|}{\sqrt{\hat{\sigma}_{xi}^{2} + \hat{\sigma}_{xj}^{2}}}, \tag{4-7}$$

where the two classes are i and j. $\hat{\mu}_{xi}$ is the mean value of the feature x within class i, computed by

$$\hat{\mu}_{xi} = \frac{1}{N_i}\sum_{k=1}^{N_i} x_{ki}, \tag{4-8}$$

where N_i is the numbers of objects from class i. The feature of the kth object in class i is x_{ki}. $\hat{\sigma}_{xi}^{2}$ is the estimated variance of the feature x within class i, given by

$$\hat{\sigma}_{xi}^{2} = \frac{1}{N_i}\sum_{k=1}^{N_i} (x_{ki} - \hat{\mu}_{xi})^2 . \tag{4-9}$$

Clearly, the superior feature, among all given features, is the one producing the widest class separation.

DIMENSION REDUCTION

In essence, the goal of dimensionality reduction is to determine a mapping, say $x \to y(x)$, which transforms the feature space, Ω, into a lower-dimensional space in such a way as to retain as much as possible of the discriminatory information available in the original representation and to remove the redundant and irrelevant information which could have a derogatory effect on the classifier performance.

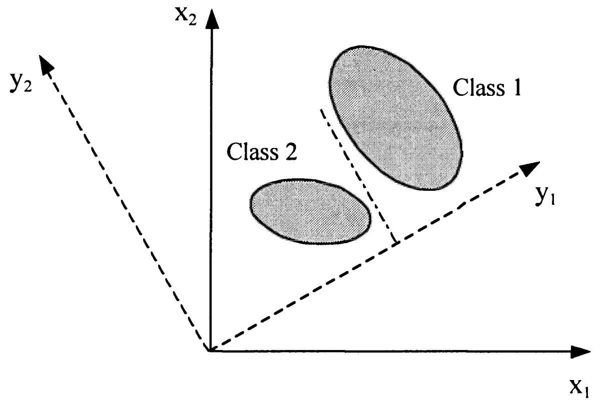


Figure 4-9. Dimension reduction.

There are many ways to combine the two features x and y into a single feature z. A simple way is to use a linear function [5]:

$$z = ax + by \tag{4-10}$$

Since classifier performance is not affected by scaling the magnitude of the feature, we can impose a restriction on the magnitude, such as

$$a^2 + b^2 = 1 \tag{4-11}$$

An illustrative example is shown in Figure 4-9. Clearly, after the transformation, one feature is sufficient for classification.

4.4 Pattern Classification

OVERVIEW

The concept of pattern classification may be expressed in terms of the partition of feature space (or a mapping from feature space to decision space). Suppose that N features are to be measured from each input pattern. Each set of N features can be considered as a vector, X, called a feature (measurement) vector, or a point in the N-dimensional feature space, Ω_x. The problem of classification is to assign each possible vector or point in the feature space to a proper pattern class. This can be interpreted as a partition of the feature space into mutually exclusive regions and each region will correspond to a particular pattern class.

DECISION SURFACES AND DISCRIMINANT FUNCTIONS

Patterns pertaining to different classes will fall into different regions in the pattern space (i.e., different classes of patterns will cluster in different regions and can easily be separated by separating surfaces). Separating surfaces, called *decision surfaces* [1], can formally be defined as surfaces in n dimensions which are used to separate known patterns into their respective categories and are used to classify unknown patterns. Such decision surfaces are called *hyperplanes* and are (N-1)-dimensional. When N=1, the decision surface is a point; when N=2, the decision surface becomes a line; when N=3, the surface is a plane; when n=4 or higher, the decision surface is a hyperplane represented by

$$w_1x_1 + w_2x_2 + \cdots + w_Nx_N + w_{N+1} = 0\,, \tag{4-12}$$

or expressed in matrix form as

$$W' \cdot X = 0\,, \tag{4-13}$$

where W and X are called the augmented weight vector and the augmented pattern vector, respectively.

A *discriminant function* is a function $D(x)$ which defines the decision surface. Mathematically, the problem of classification can be formulated in terms of discriminant functions. Let $\omega_1, \omega_2, \ldots \omega_m$ be designated as the m possible pattern classes to be recognized. Then the discriminant function $D_j(X)$ associated with pattern class ω_j, is such that if the input pattern represented by the feature vector X is in class ω_i, denoted as $X \sim \omega_i$, the value of $D_i(X)$ must be the largest, i.e., for all $X \sim \omega_i$,

$$D_i(X) > D_j(X), \quad i, j = 1,2,\ldots,m, i \neq j\,. \tag{4-14}$$

Thus, in the feature space Ω_x, the decision surface between regions associated with class ω_i and class ω_j, respectively, is expressed by the following equation [2].

$$D_i(X) - D_j(X) = 0\,. \tag{4-15}$$

Many different forms satisfying condition (14) can be selected for $D_i(X)$.

Linear Discriminant Functions

In this case, a linear combination of the feature measurements $x_1, x_2, \ldots, x_N$ is selected for $D_i(X)$, i.e.,

$$D_i(X) = \sum_{k=1}^{N} w_{ik}x_k + w_{i,N+1} \qquad i = 1,2,\ldots,m \tag{4-16}$$

To make the equation a valid vector multiplication, the input vector X has been augmented to become (N+1)-dimensional by adding $x_{N+1} = 1$, the discriminant function can be written as

$$D_i(X) = W_i' \cdot X \tag{4-17}$$

where

$$W_i = (w_{i1}, \ldots, w_{iN}, w_{i,N+1})' \tag{4-18}$$

$$X = (x_1, \ldots, x_N, 1)' \tag{4-19}$$

Piecewise Linear Discriminant Functions

A piecewise linear function is a function that is linear over subregions of the feature space. These piecewise linear discriminant functions give piecewise linear boundaries between categories as shown in Figure 4-10.

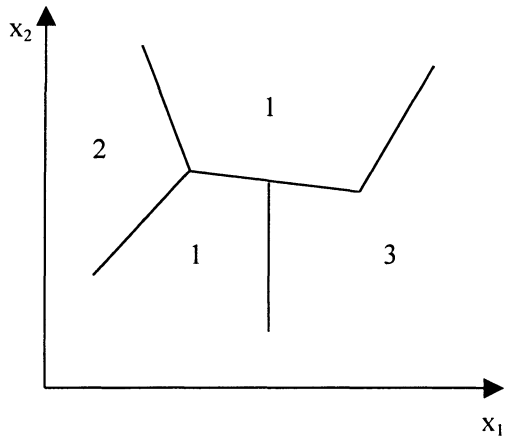


Figure 4-10. Piecewise linear decision surfaces.

The discriminant functions are given by

$$D_i(X) = \max_{k=1,\ldots,N_i} \{D_i^k(X)\} \qquad i = 1, \ldots, m \tag{4-20}$$

where N_i is the number of prototypes in class i and

$$D_i^k(X) = w_{i1}^k x_1 + w_{i2}^k x_2 + \cdots + w_{iN}^k x_N + w_{i,N+1}^k \tag{4-21}$$

Polynomial Discriminant Functions

A rth-order polynomial discriminant function can be expressed as

$$D_i(X) = w_{i1} f_1(X) + w_{i2} f_2(X) + \cdots + w_{iL} f_L(X) + w_{i,L+1} \tag{4-22}$$

where $f_i(X)$ is of the form

$$x_{k_1}^{n_1} x_{k_2}^{n_2} \cdots x_{k_r}^{n_r} \text{ for } k_1, k_2, \cdots k_r = 1, \cdots, N, \quad \textit{and} \quad n_1, n_2, \ldots, n_r = 0 \textit{ or } 1 \tag{4-23}$$

The decision boundary between any two classes is also in the form of a rth-order polynomial.

NONPARAMETRIC CLASSIFICATION

Minimum Distance Classifier

An important class of linear classifier is the one that uses the distances between the input pattern and a set of reference vectors or prototype points in the feature space as the classification criterion. Suppose that m reference vectors $R_1, R_2, \dots, R_m$ are given with R_j associated with the pattern class ω_j. A minimum-distance classification scheme with respect to $R_1, R_2, \dots, R_m$ is to classify the input X as from class ω_i, i.e.,

$$X \sim \omega_i \quad \text{if } |X - R_i| = \min_{1 \le k \le m} |X - R_k| \tag{4-24}$$

where $|X - R_k|$ is the distance defined between X and R_k.

It can be easily proved [2] that a minimum-distance classifier is also a linear classifier. The performance of a minimum-distance classifier is ,of course, dependent upon an appropriately selected set of reference vectors.

Nearest Neighbor Classification

The concept adopted in the previous section can be extended to the case of minimum-distance classification with respect to sets of reference vectors. Let $R_1, R_2, \dots, R_m$ be the m sets of reference vectors associated with classes $\omega_1, \omega_2, \dots, \omega_m$, respectively, and let reference vectors in R_j be denoted as $R_j^{(k)}$, i.e., $R_j^{(k)} \in R_j, k = 1, \dots u_j$, where u_j is the number of reference vectors in set R_j. Define the distance between an input feature vector X and R_j as

$$d(X, R_j) = \min_{k=1,\dots,u_j} \left|X - R_j^{(k)}\right| \tag{4-25}$$

That is, the distance between X and R_j is the smallest of the distances between X and each vector in R_j. The classifier will assign the input to a pattern class which is associated with the closest vector set.

BAYES (PARAMETRIC) CLASSIFICATION

For some classification problems, the PDF (Probability Density Function) of the feature vector is known for one or all classes. For example, the mean stature, which is the feature vector, of Chinese male is 1.68 m and the PDF is approximately Gaussian with a standard deviation of 6 cm. If the PDF of the feature vector is unknown, we might estimate it by measuring a large number of objects, plotting a histogram of their features, and computing the mean and variance. After normalization to unit area, this histogram can be taken as an estimate of the corresponding PDF.

Bayesian Decision

By Bayes' rule, we can write

$$P(\omega_i \mid X) = \frac{P(X \mid \omega_i)P(\omega_i)}{P(X)} = \frac{P(X \mid \omega_i)P(\omega_i)}{\sum_{k=1}^{m} P(X \mid \omega_k)P(\omega_k)} \tag{4-26}$$

where $P(X)$, is the probability that X occurs without regard to the category in which it belongs. $P(\omega_i)$ is the *a priori* probability of class ω_i, $P(X \mid \omega_i)$ is the likelihood function of class ω_i with respect to X, and $P(\omega_i \mid X)$ is the *a posteriori* probability that the object belongs to class ω_i.

Bayes' theorem allows us to combine the a priori probabilities of class membership, the conditional PDF, and the measurement made to compute, for each class, the probability that the measured object belongs to that class. Given this information, we might choose to assign each object to its most likely class. That is, if

$$P(\omega_i \mid X) > P(\omega_j \mid X) \quad \forall j \neq i \tag{4-27}$$

we would assign the object to class ω_i. Substituting Equation (26) into Equation (27), we have

$$P(X \mid \omega_i)P(\omega_i) > P(X \mid \omega_j)P(\omega_j) \qquad \forall j \neq i \tag{4-28}$$

Minimum Conditional Risk Classifier

Every time we assign an object to a class, we risk making an error. A quantitative way to account for this is with a cost function. Let L_{ij} be the loss incurred by the classifier if the decision $X \sim \omega_j$ is made when the input pattern is actually from ω_i. Usually, L_{ij} will take on the value zero for correct decisions, small values for harmless errors, and larger values for more costly mistakes.

Suppose we measure an object and assign it to class ω_i. The expected loss resulting from this assignment is the *conditional average risk*

$$R_i(X) = \sum_{k=1}^{m} L_{ki} P(\omega_k \mid X) \tag{4-29}$$

which is just the cost averaged over all m of the groups to which the object might actually belong. Thus, given the feature vector, there is certain risk involved in assigning the object to any group.

The job of the classifier is then to find an optimal decision that will minimize the average risk or cost. The decision rule will then consist of the following steps:

(1) Compute the expected losses, $R_i(X) \quad i = 1,2,\ldots,m$.

(2) Decide that $X \sim \omega_k$ if $R_k(X) \leq R_i(X) \quad \forall i$.

Thus, the corresponding discriminant function is

$$D_i(X) = -R_i(X) \tag{4-30}$$

Bayes' Discriminant Function

Substituting Equation (26) into Equation (29) for $R_i(X)$, we have

$$R_i(X) = \frac{1}{P(X)} \sum_{k=1}^{m} L_{ki} P(X \mid \omega_k) P(\omega_k) \tag{4-31}$$

Since $P(X)$ is common to all $R_i(X)$, we can eliminate it from the conditional average risk equation. Thus, we will get a new $D_i(X)$ by substituting Equation (31) into Equation (30), which is the *Bayes' discriminant function.* The corresponding classifier is called *Bayes' classifier*, which gives the optimum performance from the statistical point of view.

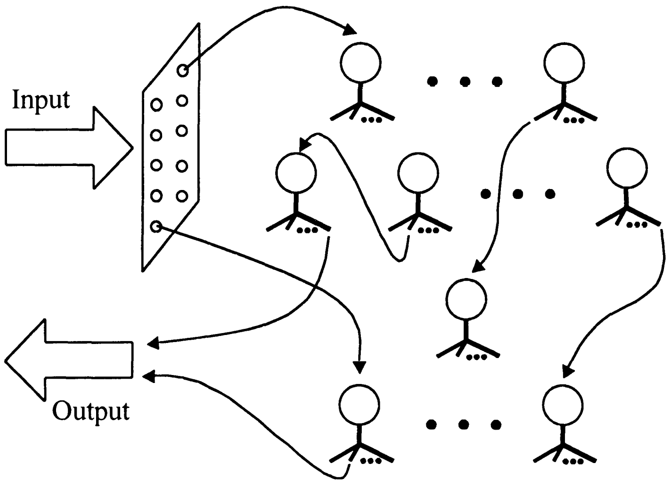


Figure 4-11. Neural networks architecture.

4.5 Neural Networks

INTRODUCTION

A different approach to pattern recognition that has attracted considerable interest in recent years comes out of the field of artificial neural network technology. Initially inspired by biological nervous systems, the development of artificial neural networks has more recently been motivated by their applicability to certain types of problems, such as static and dynamic pattern classification, and their potential for parallel-processing implementations that are especially useful in image understanding.

DEFINITION OF NEURAL NETWORKS

Artificial Neural Networks are relatively crude electronic models based on the neural structure of the brain. It is a parallel, distributed information processing structure consisting of processing elements (which can posses a local memory and can carry out localized information processing operations) interconnected via unidirectional signal channels called *connections*. Each processing element has a single output connection that branches into as many collateral connections as desired; each carries the same signal — the processing element output signal. The processing element output signal can be of any mathematical type desired. The information processing that goes on within each processing element can be defined arbitrarily with the restriction that it must be completely local; that is, it must depend only on the current values of the input signals arriving at the processing element via impinging connections and on values stored in the processing element's local memory [25]. Figure 4-11 shows a typical neural network architecture.

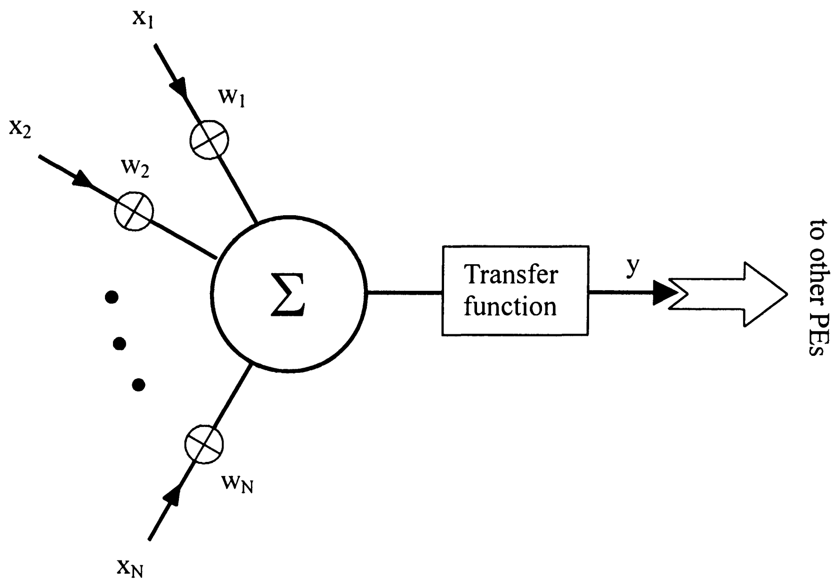


Figure 4-12. A generic processing element.

Processing Element

The fundamental processing element of a neural network is a neuron. It merely sums the product of its input vector and a weight vector, transforms the result according to a sigmoid transformation function, and outputs the scalar result. This result then passes on to become the input of one or many other PEs through the network interconnections. Figure 4-12 shows a typical representation of processing element.

The actual processing done by such a PE can be described as a function of a dot product; that is,

$$y = f\left[X^T \cdot W\right] = f\left[\sum_{i=1}^{N} x_i w_i\right] \tag{4-32}$$

where X is the input vector, W is the weight vector associated with the given PE, and f is the transfer function (activation function).

Transfer Function

The result of the summation function, the weighted sum, is transformed to a working output through an algorithmic process known as the *transfer function*. In the transfer function the summation total can be compared with some threshold to determine the neural output. If the sum is greater than the threshold value, the processing element generates a signal. If the sum of the input and weight products is less than the threshold, no signal (or some inhibitory signal) is generated. Both types of response are significant.

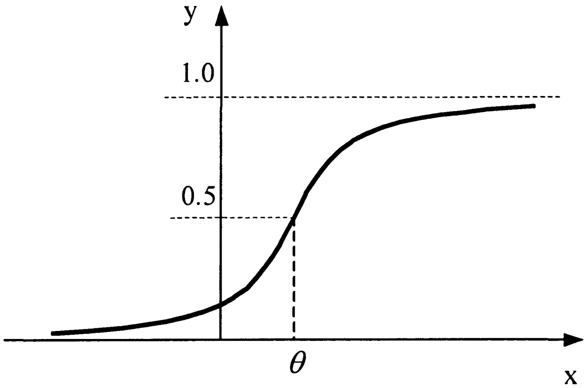


Figure 4-13. Sigmoid function.

The threshold, or transfer function, is generally non-linear. Linear (straight-line) functions are limited because the output is simply proportional to the input. Linear functions are not very useful. *Sigmoid function*, shown in Figure 4-13, is a very important non-linear transfer function.

THE BACKPROPAGATION NETWORK

The backpropagation neural network [26-35] is one of the most important historical developments in neurocomputing. It is a powerful mapping network that has been successfully applied to a wide variety of problems ranging from credit application scoring to image compression. Its greatest strength is in non-linear solutions to ill-defined problems.

The typical back-propagation network has an input layer, an output layer, and at least one hidden layer. There is no theoretical limit on the number of hidden layers but typically there is just one or two. Some work has been done which indicates that a minimum of four layers (three hidden layers plus an output layer) are required to solve problems of any complexity. Each layer is fully connected to the succeeding layer. The network topology is constrained to be feedforward: i.e. loop-free — generally connections are allowed from the input layer to the first (and possibly only) hidden layer; from the first hidden layer to the second, ... , and from the last hidden layer to the output layer.

The standard configuration for a two-layer backpropagation network is shown in Figure 4-14. The number of PEs in each layer varies with the application. More hidden layers can be added between the input hidden layer and the output layer.

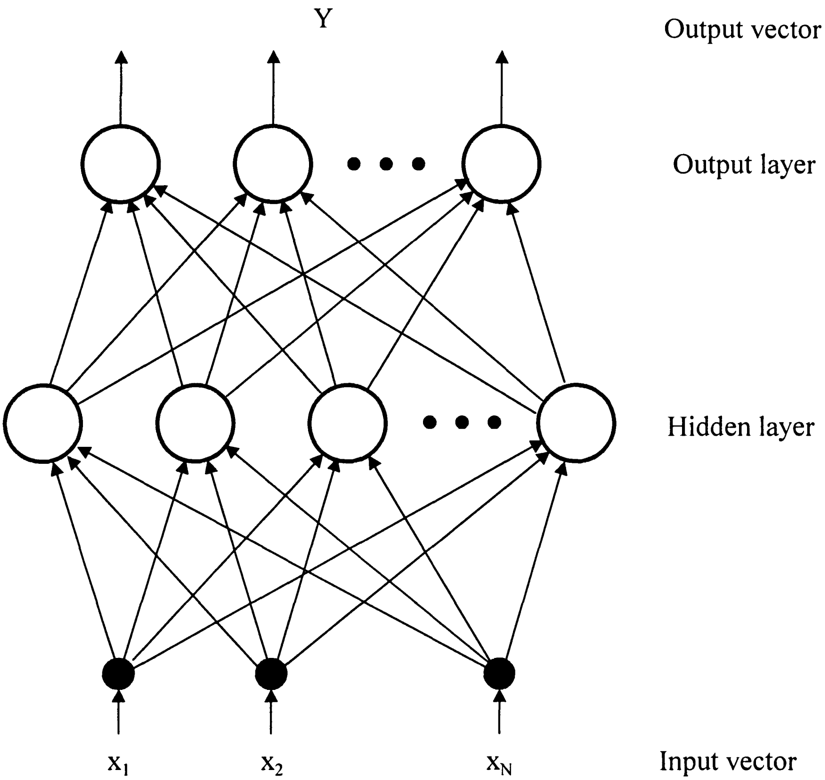


Figure 4-14. A typical two-layer backpropagation network.

THE COUNTERPROPAGATION NETWORK

The counterpropagation architecture [36] differs conceptually from the architecture of the backpropagation network. According to Robert Hecht-Nielsen [25], the counterpropagation network architecture is a combination of a portion of the self-organizing map and the instar / outstar structure of Stephen Grossberg [37].

The first counter-propagation network (CP net) consisted of a bi-directional mapping between the input and output layers. In essence, while data is presented to the input layer to generate a classification pattern on the output layer, the output layer in turn would accept an additional input vector and generate an output classification on the network's input layer. The network got its name from this counter-posing flow of information through its structure. Most developers use an uni-flow variant of this formal representation of counter-propagation. In other words, there is only one feedforward path from input layer to output layer.

The interconnection architecture for the *forward-only* [25] counter propagation network is the same as the backpropagation network (BP net) in Figure 4-14. However, whereas the BP net can have many layers, the CP net is limited to two. The main layers include an input buffer layer, a self-organizing Kohonen layer, and an output layer which uses the Delta Rule to modify its incoming connection weights. Sometimes this layer is called a Grossberg Outstar layer. Further, the activation function of each PE is linear rather than sigmoidal, and its output is merely the dot product of Equation (31).

The other important difference, compared with BP net, is the type of processing done in the hidden layer which is called a *competitive layer* because its nodes compete to generate an output value. The hidden node computing the largest result wins the competition and outputs a value of 1. All other hidden nodes output a value of 0.

Since only one output from the competitive Kohonen layer is active at a time and all other elements are zero, the resulting network output is actually a vector of connection weights that are connected to the output of the winning hidden layer node. As a result, the only weights adjusted for the output processing elements are the ones connected to the winning element in the competitive layer. In this way, the output layer learns to reproduce a certain pattern for each active processing element in the competitive layer. Moreover, the Kohonen layer is not effected.

The counterpropagation network is yet another topology to synthesize complex classification problems, while trying to minimize the number of processing elements and training time. Its pattern recognition capability is embodied in the computation done by the hidden layer's PEs. The counterpropagation network is capable of unsupervised learning. That is, the training set need not be preclassified. As training advances, the network will locate naturally occurring clusters of points in feature space and make classes out of them. In applications where the data are expected or known to occur in separate classes, but suitable preclassified input data are unavailable, this can be quite useful.

THE HOPFIELD NETWORK

John Hopfield first presented his crossbar associative network in 1982 at the National Academy of Sciences [38]. The network can be conceptualized in terms of its energy and the physics of dynamic systems.

The Hopfield network is shown in Figure 4-15. This *recurrent associative* network (that is, there is feedback in the network as opposed to *feedforward* nets which were the subject of the Backpropagation algorithm) has n processing elements, each of which is connected to every other node (but not to itself) and the connection strengths or weights are symmetric in that the weight from node i to node j is the same as that from node j to node i. That is, $w_{ij} = w_{ji}$ and $w_{ii} = 0$ for all i, j. Notice that the flow of information in this net is not in a single direction as it has been in the nets dealt with so far. It is possible for information to flow from a node back to itself via other nodes. That is, there is feedback in the network.

The Hopfield network processing element transfer function is given by

$$x_i^{new} = \begin{cases} 1 & if \sum_{k=1}^{N} w_{ki} x_k^{old} > T_i, \\ x_i^{old} & if \sum_{k=1}^{N} w_{ki} x_k^{old} = T_i, \\ -1 & if \sum_{k=1}^{N} w_{ki} x_k^{old} < T_i \end{cases} \tag{4-33}$$

for $i = 1,2,\ldots,N$. The values $T_1, T_2, \ldots, T_N$ are called *thresholds*. The processing elements of the network are updated one at a time. The only constraint on the scheduling of these updates is that all of the processing elements must be updated at the same average rate. A uniformly random updating schedule is often specified. The time evolution of the state of the network clearly depends on the updating schedule. The subsequent movement of the network's state vector $X = (x_1, x_2, \ldots, x_N)^T$ on cube $\{-1, 1\}^N$ is therefore not uniquely defined by the initial state.

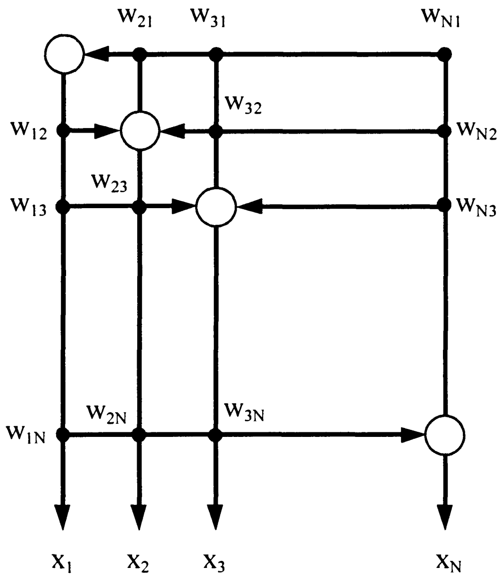


Figure 4-15. The Hopfield neural network.

The Hopfield network has an *energy function* associated with it. Whenever a processing element changes state, this energy function always decreases. Starting at some initial position, the system's state vector X simply moves downhill on the network's energy surface until it reaches a local minimum of the energy function. This convergence process is guaranteed to be completed in a fixed number of steps.

The Hopfield network does not have a learning law associated with its transfer function [25]. The $N \times N$ weight matrix $W = (w_{ij})$, which is therefore fixed, is assumed to be specified in advance.

Primary applications for this sort of network have included associative, or content-addressable, memories and a whole set of optimization problems, such as the combinatorial best route for a traveling salesman.

References

[1] S.T. Bow, *Pattern Recognition — Applications to Large Data-Set Problems*, 1984.
[2] K.S. Fu, *Applications of Pattern Recognition*, CRC Press, Florida, 1982.
[3] R.M. Haralick and L.G. Shapiro, "Survey: Image Segmentation," *Comput. Vision, Graphics, Image Proc.*, vol. 29, pp. 100-132, 1985.
[4] A. Rosenfield, "Connectivity in Digital Pictures," *Journal of the ACM*, vol. 17, pp. 146-160, 1970.
[5] K. R. Castleman, *Digital Image Processing*, Prentice Hall, 1996.
[6] R.J. Wall, "The Gray Level Histogram for Threshold Boundary Determination in Image Processing to the Scene Segmentation Problem in Human Chromosome Analysis," *Ph.D. Thesis, University of California at Los Angeles,* 1974.
[7] K. Castleman and R. Wall, "Automatic Systems for Chromosome Identification," *in T. Caspersso-n, ed., Nobel Symposium 23 — Chromosome Identification*, Academic Press, New York, 1973.
[8] K. Castleman and J. Melnyk, "An Automated System for Chromosome Analysis: Final Report," *Document No. 5040-30*, Jet Propulsion Laboratory, Pasadena, CA, July 4, 1976.
[9] J. Prewitt and M. Mendelsohn, "The Analysis of Cell Images," *Annals of the N.Y.*, Academy of Sciences, vol. 128, pp. 1035-1053, 1966.
[10] J. Prewitt, "Object Enhancement and Extraction", in B.Lipkin and A., Rosenfeld, eds., *Picture Processing and Psychopictorics*, Academic Press, New York, 1970.
[11] R.J. Wall, A. Klinger, and K.R. Castleman, "Analysis of Image Histograms," *Proc. 2nd. Int. Joint Conf. On Pattern Recognition* (IEEE Pub. 74CH-0885-4C), pp. 341-344, Copenhagen, August 1974.
[12] J. Weszka, "A Survey of Threshold Selection Techniques," *Computer Graphics and Image Processing*, vol. 7, pp. 259-265, 1978.
[13] M.E. Sieracki, S.E. Reichenbach, and K.L. Webb, "Evaluation of Automated Threshold Selection Methods for Accurately Sizing Microscopic Fluorescent Cells by Image Analysis," *Applied and Environmental Microbiology*, vol. 55, no. 11, pp. 2762-2772, November 1989.
[14] Y.H. Katz, "Pattern Recognition of Meteorological Satellite Cloud Photography", *Proc.3rd Symp. Remote Sensing of the Environment* (2d ed.), pp.173-190, Univ. Of Michigan, Ann Arbor, 1964.
[15] A. Rosenfeld and A.C. Kak, *Digital Picture Processing*, Academic Press, New York, 1982.
[16] D. Marr and E. Hildreth, "Theory of Edge Detection," *Proc. R. Soc. London*, Ser. B, vol. 207, pp. 187-217, 1980.
[17] D. Marr, *Vision*, Freeman, San Francisco, 1982.
[18] L.G. Roberts, "Machine Perception of Three-dimensional Solids," in J .T. Tippett, ed., *Optical and Electro-Optical Information Processing,* pp. 159-197, MIT Press, Cambridge, MA, 1965.
[19] L.S. Davis, "A Survey of Edge Detection Techniques," *CGIP*, vol. 4, pp. 248-270, 1975.
[20] R.A. Kirsch, "Computer Determination of the Constituent Structure of Biological Images," *Computers in Biomedical Research*, vol. 4, pp. 315-328, 1971.
[21] W. Meisel, *Computer-Oriented Approaches to Pattern Recognition*, Academic Press, New York, 1972.
[22] A.K. Jain and B. Chandrasekaran, "Dimensionality and Sample Size Considerations in Pattern Recognition Practice," *Handbook of Statistics*, vol. 2, pp. 835-855, North Holland Publishing Company, 1982.
[23] L. Kanal and B. Chandrasekaran, "On Dimensionality and Sample Size in Statistical Pattern Recognition," *Pat. Rec.*, vol. 3, pp. 225-234, 1971.
[24] I. T. Young, "Further Considerations of Sample Size and Feature Size," *IEEE Trans.*, vol. IT-24, no. 6, pp. 773-775, 1978.

[25] R. Hecht-Nielsen, *Neurocomputing*, Addison-Wesley Publishing Company, 1989.

[26] J.A. Anderson, and E. Rosenfeld, *Neurocomputing: Foundations of Research*, MIT Press, Cambridge MA, 1988.

[27] A.E. Bryson, and Y.C. Ho, *Applied Optimal Control, [Revised Printing of the 1969 Edition]*, Hemisphere Publishing, New York, 1975.

[28] D.B. Parker, "Optimal Algorithms for Adaptive Networks: Second Order Back Propagation, Second Order Direct Propagation, and Second Order Hebbian Learning," *Proc. of the Int. Conf. on Neural Networks*, vol. II, pp. 593-600, IEEE Press, New York, June 1987.

[29] D.B. Parker, "A Comparison of Algorithms for Neuron-Like Cells," in J. Denker, *Proc. Second Annual Conf. on Neural Networks for Computing*, vol. 151, pp. 327-332, Am. Inst. Of Physics, New York, 1986.

[30] D.B. Parker, "Learning-logic," *Technical Report TR-47*, Center for Computational Res. In Economics and Management Sci., MIT, April 1985.

[31] H. Robbins, and S. Monro, "A Stochastic Approximation Method," *Annals of Math. Stat.*, vol. 22, pp. 400-407, 1951.

[32] D.E. Rumelhart, and J.L. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, I & II*, MIT Press, Cambridge MA, 1986.

[33] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning Internal Representations by Error Propagation," in D.E. Rumelhart, and J.L. McClelland, *Parallel Distrbuted Processing: Explorations in the Microstructure of Cognition*, vol. 1, pp. 318-362, MIT Press, Cambridge MA, 1986.

[34] P.J. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, Doctoral Dissertation, Appl. Math., Harvard University, November 1974.

[35] H. White, "Learning in Artificial Neural Networks: A Statistical Perspective," *Neural Computation*, vol. 1, pp. 425-464, Winter 1989.

[36] R. Hecht-Nielsen, "Applications of Counterpropagation Networks," *Neural Networks*, vol. 1, pp. 131-139, 1988.

[37] S. Grossberg, *Studies Of Mind And Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control*, Reidel Press, Boston, 1982.

[38] J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proc. Natl. Acad. Sci.*, vol. 79, pp. 2554-2588, 1982.

5 FINGERPRINT SYSTEM

Fingerprint identification is the most widespread application of biometrics technology. In this chapter, we will firstly review the current fingerprint systems. Section 5.2 defines some different types of feature points and then systematically summarizes the rules of how to distinguish the true minutiae from the false ones. Fingerprint image processing and minutiae determination algorithms are described in Section 5.3 and 5.4, respectively. Fingerprint matching and the corresponding experiment results are finally given.

5.1 Review of Fingerprint Systems

Fingerprints are graphical flow-like ridges present on human fingers. They have been widely used in personal identification for several centuries [1]. The use of fingerprints as a biometrics is both the oldest mode of computer-aided, personal identification and the most prevalent in use today. This widespread use of fingerprints has been and still is largely for law enforcement applications. At the same time, There is expectation that a recent combination of factors will favor the use of fingerprint for much large market of personal authentication.

Fingerprint system can be separated into two categories: verification and identification [2]. Verification is comparison of a claimant fingerprint against an enrollee fingerprint, where the intention is that claimant fingerprint matches the enrollee fingerprint. To prepare for verification, a person initially enrolls his or her fingerprint into verification system. A representation of that fingerprint is stored in some compressed format along with the person's name or other identity. Subsequently, each access is authenticated by the person identifying him or herself, then applying the fingerprint to the system such that the identity can be verified. Verification is also termed, one-to-one matching. Identification is the traditional domain of criminal fingerprint matching. A fingerprint of unknown ownership is matched against a database of known fingerprints to associate a crime with an identity. Identification is also termed one-to-many matching. There is an informal third type of matching that is termed one-to-few matching. This is for the practical application where a fingerprint system is used by "a few" users, such as by family members to enter their house. A number that constitutes "few" is usually accepted to be somewhere between 5 and 20.

A number of methods are used to acquire fingerprints [3]. Among them, the inked impression method remains the most popular. It has been essentially a standard technique for fingerprint acquisition for more than 100 years. The first step in capturing an inked impression of a fingerprint is to place a few dabs of ink on a slab then rolling it out smoothly with a roller until the slab is covered with a thin, even layer of ink. Then the finger is rolled from one side of the nail to the other side over the inked slab which inks the ridge patterns on top of the finger appears on the paper. Obviously, this method is time-consuming and unsuitable for an on-line fingerprint verification system. Inkless fingerprint scanners are now available which are capable of directly acquiring fingerprints in digital form. This method eliminates the intermediate digitization process of inked fingerprint impression and makes it possible to build an on-line system.

5.2 Definitions and Notations

FINGERPRINT PATTERNS

The structure characters of the ridges in the center fingerprint area basically classify fingerprint patterns. Familiar fingerprint patterns have six main classes: arch, tented arch, right loop, left loop, whorl and twin loop, as shown in Figure 5-1. Besides, there is some pattern not included in above patterns as Figure 5-1(g). But these patterns occupy very small proportion, so they are usually classified as special patterns. Some specialists have done thorough research in the taxonomy of fingerprint patterns in the base of long-term experience. They often classify fingerprint patterns more finely according to some difference among the fingerprint ridges' characters. However some specialists in pattern recognition suggested various methods for fingerprint patterns' classification starting from the analyzing of patterns structure [18-19,24,29-30,33,36]. Automatic classification of fingerprint patterns is an important component in automatic fingerprint identification system and an important index of the its performance.

FINGERPRINT FEATURES

The singular points, namely the core and delta points (see Figure 5-2(a)) act as points of registration for comparing the fingerprints [4]. A core point is defined as the topmost point on the innermost upward recurving ridge. A recurving ridge means the ridge that curves back on itself. In mathematics the strict definition of core is very difficult, but fingerprint specialists can locate the core in each fingerprint very easily. However there may be different results for the fingerprints such as arch and tented arch. A delta point is defined as the point of bifurcation (in a delta-like region) on a ridge splitting into two branches extending to encompass the complete pattern area. The delta point's quantity and location to core point are very important for fine classifying of fingerprint patterns.

Despite singular points, the most important character for fingerprint recognition, especially for locale obscure fingerprint cognizance, is minutia [34]. The minutiae constitute ridge endings and ridge bifurcations. Examine each fingerprint image carefully, we can find each ridge has the following different characters and relations:

(1) Two ridges extend almost parallel in some areas and finally combine into one ridge, or one ridge bifurcates into two at some point.
(2) One ridge breaks abruptly between ridges at two sides of it when extending.

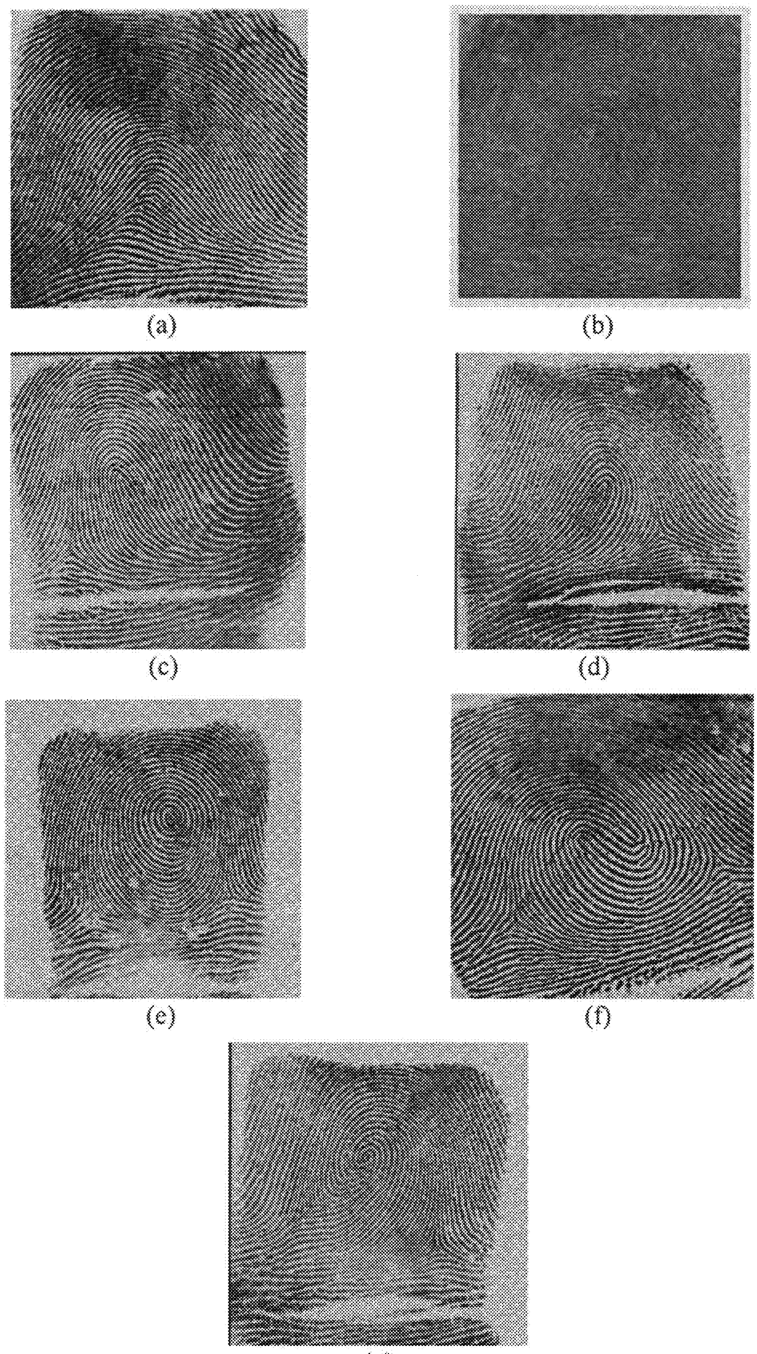
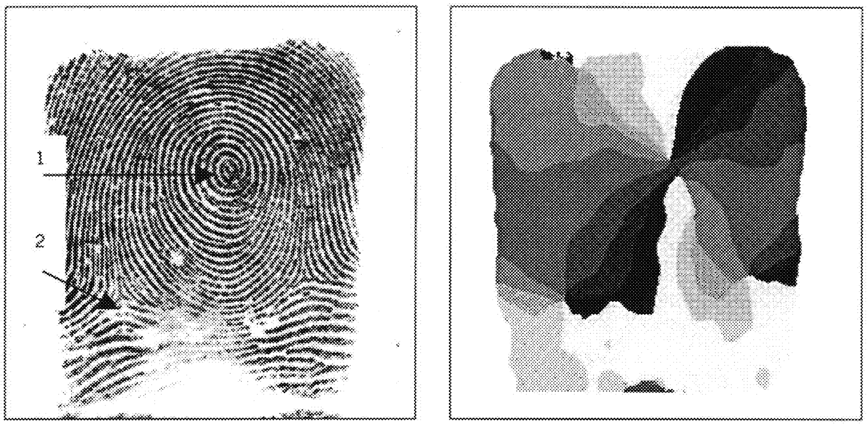


Figure 5-1. Fingerprint patterns. (a) Arch. (b) Tented arch. (c) Right loop. (d) Left loop. (e) Whorl. (f) Twin loop. (g) Special patterns.



(a) (b)

Figure 5-2. Singular points.
(a) Original image (1: core point; 2: delta point). (b) Direction image.

The above relation of two ridges form vary typical minutiae of fingerprint, and we classify the feature points into following eight types:

Type 1- *Long bifurcation.* Two ridges are emerged into one ridge or one ridge is split into two ridges, all of which have long length and their directions are near the same or change smoothly. In most cases it is a true and stable minutia (see Figure 5-3(a)).

Type 2 - *Long ending.* A ridge with sufficient length ends abruptly (see Figure 5-3(b)).

Type 3 - *Loop (lake).* Two bifurcations sharing two same ridges form a loop (see Figure 5-3(c)).

Type 4 - *Bridge (crossover).* Two parallel or near parallel ridges touch in midway by the third short ridge (see Figure 5-3(d)).

Type 5 - *Short bar (independent ridge or island).* A ridge ends at two endings and its length (pixel numbers) is less than a given threshold. In most cases a short ridge locates between two long parallel or near parallel ridges (see Figure 5-3(e)).

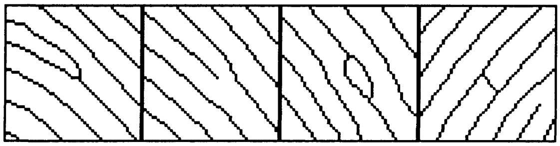
Type 6 - *Spur (bifurcation with short ridge).* A ridge is split into a short ridge ending at an ending and another longer ridge (see Figure 5-3(f)).

Type 7 - *Two adjacent bifurcations.* Two parallel ridges are emerged into one ridge with short length and then short ridge is split into two parallel ridges again (see Figure 5-3(g)).

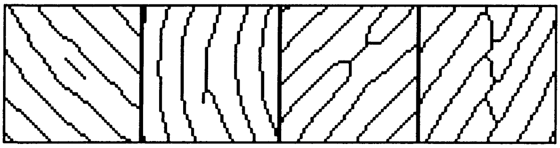
Type 8 - All other feature points. An example of this type is shown in Figure 5-3(h).

FINGERPRINT IMAGE

Fingerprint image is composed of many ridges [35]. So ideal fingerprint image should be a binary image alternate with black and white. However, after digitized by scanner or camera, a gray-lever image containing vary noises is got. If the skeleton of the ridges in binary image is extracted, a thinning image is defined. In addition, an important property of the ridges in a fingerprint image is that the gray-level values on ridges attain their local maxima along a direction normal to the local ridge orientation. The direction image represents the ridge orientation at each pixel of a fingerprint image. In fingerprint system, the images mentioned above are always used to extract the minutiae and singular points.



(a) (b) (c) (d)



(e) (f) (g) (h)

Figure 5-3. Eight types of feature points.
(a) Long bifurcation. (b) Long ending. (c) Loop. (d) Bridge.
(e) Short bar. (f) Spur. (g) Two adjacent bifurcations. (h) Other feature points.

WIDTH BETWEEN TWO RIDGES

When designing the filter for fingerprint image processing [31], the width between two ridges is an important parameter. For binary image, because pixel '1' indicates ridge and pixel '0' indicates valley, we can consider the number of continuous pixels '1' in the direction vertical to ridge as the width of this pixel. Similarly, we can get the width between two valleys. Because in actual application we often have to get the information about width before we get binary image, we usually get widths between two ridges or valleys of each pixel in the gray-level fingerprint image. The algorithm is given below:

(1) Input: Gray-level fingerprint image and the ridge directions of each pixel.
(2) Algorithm: Let (i_0, j_0) be the center, and choose $2n$ pixels' gray-level values in the direction vertical to the ridge to form the data collection, where n is generally two times lager than the maximum estimate, then calculate the correlation function in vary displacement:

$$R(s) = \sum_{m=1}^{2n-1} g(m)g(m+s) - \frac{1}{(2n-s)}\left(\sum g(m)\right)\left(\sum g(m+s)\right) \tag{5-1}$$

where $s = 1, 2, \ldots, k$. When $R(s)$ is maximum, the corresponding s is the width between two ridges or valleys at (i_0, j_0).

5.3 Fingerprint Image Processing

In order to extract the minutiae from the gray-level image, fingerprint image processing is a key operation in the current fingerprint identification systems [20,22-23,25-28,32]. It mainly consists of pre-processing (direction image estimation and filtering), binarization, and thinning.

DIRECTION IMAGE ESTIMATION

Direction image can be estimated directly from both fingerprint gray-level image and binary image. The following are some commonly used methods for getting ridge direction.

Method One

Input: Fingerprint gray-level image.

Algorithm:

(1) Let point (i, j) in the gray-level image as the center, at various directions, and symmetrically collect corresponding data in certain area. The different direction decide the different location of pixels, so if we limit fingerprint direction to the following discrete values $n\theta, n = 0,\cdots,15$, where $\theta = 180°/16$. If $n\theta \le 45°$ or $n\theta \ge 135°$, then

$$P_n = \left\{(i + k_1, j + k_2)\left| \begin{matrix} k_2 = -l, -l+1,\cdots,0,\cdots,l-1,l; \\ k_1 = \mathrm{int}[k_2 \tan(n\theta)] \end{matrix}\right.\right\} \tag{5-2}$$

If $45° \le n\theta \le 135°$, then

$$P_n = \left\{(i + k_1, j - k_2)\left| \begin{matrix} k_1 = -l, -l+1,\cdots,0,\cdots,l-1,l; \\ k_2 = \mathrm{int}[k_1 \tan(n\theta - 90°)] \end{matrix}\right.\right\} \tag{5-3}$$

The value of l should be lager than the width of the ridge, and should meet the restraint of ridge directions basically same in the $2l$ area.

(2) For $n = 0,\cdots,15$, calculate the gray variance,

$$V_n = \frac{1}{2l+1}\sum_{i,j \in P_n} f^2(i,j) - \frac{1}{(2l+1)^2}\Big(\sum_{i,j \in P_n} f^2(i,j)\Big)^2 \tag{5-4}$$

where the direction of V_n's minimum is the direction of the ridge at this point.

Method Two

We can get continuous ridge directions from fingerprint gray-level image. The following algorithm is given in the document [5]. Let pixel (i_0, j_0) be the center, and confirm an $w \times w$ area, where the value of w is determined as Method one. Let

$$\begin{aligned} i_0 &= i_0 - \frac{(w-1)}{2}, i_b = i_0 + \frac{(w-1)}{2} \\ j_0 &= j_0 - \frac{(w-1)}{2}, j_b = j_0 + \frac{(w-1)}{2} \end{aligned} \tag{5-5}$$

For $\theta = 0°$, calculate

$$\rho_{i_0,j_0}(\theta=0°)=\sum_{i=i_a}^{i_b}\left[\sum_{j=j_a+1}^{j_b}\left|f(i,j)-f(i,j-1)\right|\right] \tag{5-6}$$

For $\theta=60°$ and $120°$, calculate the sum of the mean value of the difference of corresponding pixels' gray-level.

The direction of each point's ridge direction is calculated as following method. Given $\rho_{i_0,j_0}(\theta_{\max})\geq\rho_{i_0,j_0}(\theta_{mid})\geq\rho_{i_0,j_0}(\theta_{\min})$, the ridge direction at pixel (i_0, j_0) will be

$$d_{i_0,j_0}=\left\{\begin{array}{ll}\theta_{\max}+\frac{1}{2}\times\dfrac{\rho(\theta_{mid})-\rho(\theta_{\min})}{\rho(\theta_{\max})-\rho(\theta_{\min})} & ,\text{if } (\theta_{\max}+60°)_\pi=\theta_{mid} \\ \theta_{\max}-\frac{1}{2}\times\dfrac{\rho(\theta_{mid})-\rho(\theta_{\min})}{\rho(\theta_{\max})-\rho(\theta_{\min})} & ,\text{if } (\theta_{\max}-60°)_\pi=\theta_{mid}\end{array}\right\}, \tag{5-7}$$

where $(\cdot)_\pi$ express model π calculation. Of course, using Method one can also get the continuous value of each point's direction by weighed mean, however for fingerprint image processing, 16 discrete directions are often enough. The above method can also be applied to binary image, but pixels should be only '1' (black) or '0' (white) to make the calculation easier.

For ridges in the thinner image, its ridge direction should be defined as the angle between the ridge from a_1 to a_2 and horizontal axle *j*, where a_1 and a_2 are the locations when (i_0, j_0) follows *n* points along the ridge's two sides respectively. Apparently, the value of *n* cannot be very large to avoid the fluctuation of local points in the thinner ridges to affect calculation and to make the angle's calculation resolution not high (If *n* is 1, then angle can only be eight values as $0°, 22.5°, 45°, \cdots, 157.5°$, etc.) When *n* is lager, to consider the other following points' affection to the direction calculation beside a_1 and a_2, we can use least-square method to get these 2*n*+1 points' matching line, to get the ridge direction of the midpoint.

Both an original fingerprint image and its direction image are shown in Figure 5-2(a) and Figure 5-2(b), respectively.

DESIGN OF FINGERPRINT IMAGE FILTER

Many experiments indicated that the filter based on Gabor function performed well to fingerprint gray-level images and it was widely applied in many fingerprint systems [6, 7].

The impulse response of Gabor filter is

$$h(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}e^{-j2\pi(\omega_x x+\omega_y y)}, \tag{5-8}$$

where *x* and *y* are the coordinate of pixel. Obviously, it is a complex sine formula modulated by a Gauss function whose variance is σ^2, and its frequency center is (ω_x, ω_y). So, the real part of its odd-even component is denoted as

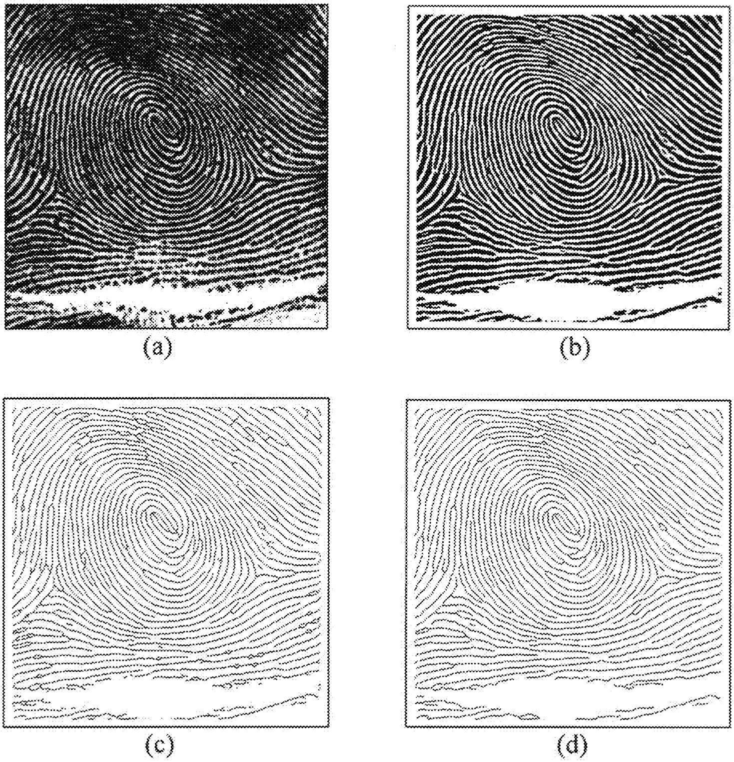


Figure 5-4. Original fingerprint image and experimental results.
(a) Original image. (b) Binary image. (c) Initial thinning image. (d) Revised thinning image.

$$\begin{cases} \Psi_s = e^{-\frac{1}{2}\times\frac{x^2+y^2}{\sigma^2}} \sin(\omega_x x + \omega_y y) \\ \Psi_c = e^{-\frac{1}{2}\times\frac{x^2+y^2}{\sigma^2}} \cos(\omega_x x + \omega_y y) \end{cases} . \tag{5-9}$$

Then $\Psi_c = g_1(x)g_3(y) - g_2(x)g_4(y)$, where $g_1(x) = e^{-\frac{x^2}{2\sigma^2}} \cos\omega_x$, $g_2(x) = e^{-\frac{x^2}{2\sigma^2}} \sin\omega_x x$, $g_3(y) = e^{-\frac{y^2}{2\sigma^2}} \cos\omega_y y$ and $g_4(y) = e^{-\frac{y^2}{2\sigma^2}} \sin\omega_y y$ can be applied to design the filters which are based on both width and direction of ridges. If the continuous variable (x, y) is replaced by the discrete variable (i, j), the operation which image $f(x, y)$ is convolved with Ψ_c can be calculated as

$$
\begin{aligned}
f(i,j) * \Psi_c(m,n) &= \sum_k \sum_l f(k,l)(g_1(m-k)g_3(n-l) - g_2(m-k)g_4(n-l)) \\
&= \sum_k g_1(m-k) \sum_l f(k,l) g_3(n-l) - \sum_k g_2(m-k) \sum_l f(k,l) g_4(n,l)
\end{aligned}
\tag{5-10}
$$

In our fingerprint system, the filter masks based on Gabor function can improve the fingerprint gray-level image's quality obviously.

BINARIZATION

The fingerprint histogram has achieved the clear bimodality after Gabor filtering so that the binarizarion of fingerprint image will be simple. Because the ridge direction has been calculated, it is easy to know the ridge direction at each pixel. Compute the threshold at some pixel, k, as follows:

(1) Let k be the center, then get several pixels along the vertical line of the ridge direction.

(2) Calculate the threshold

$$
T = \frac{\sum_{k=1}^{N} h_k}{\sum_{k=1}^{N} e_k}
\tag{5-11}
$$

where $e_k = (f_k\text{-}f_{k\text{-}2})$, $h_k = f_k(f_k\text{-}f_{k\text{-}2})$ and f_k is the gray level values of pixel k. N is the number of pixels which corresponds width of both a ridge and a valley. Kittler once proved that this algorithm has a nice robustness to Gauss noise [8].

Comparing f_k with T, we get a binary image (see Figure 5-4(b)). However, due to the presence of noise, breaks, and smudges, etc. in the gray-level image, the binary ridge map often contains holes and speckles. These holes and speckles may drastically change the skeleton of the ridges, so a hole and speckles removal procedure needs to be applied before ridge thinning.

THINNING

Thinning reduces the widths of the ridges down to a single pixel for minutiae detection. The algorithms about thinning can be classified into two categories: iterative thinning methods, which delete successive layers of pixels on the boundary of the pattern until only a skeleton remains, and noniterative thinning methods, which produce a skeleton by examination and deletion of contour pixels [9]. In the former, the deletion or retention of a (black) pixel p would depend on the configuration of pixels in a local neighborhood containing p. A pixel p is 8- (or 4-) deletable if its removal does not change the 8- (or 4-) connectivity of a picture. The pixels considered for deletion are contour pixels. It has an advantage that the relationships among the pixel and its 8-adjacnets can be represented by the look-up-table and the thinning refers to the comparison between the look-up-table and the pixel's neighborhood relation.

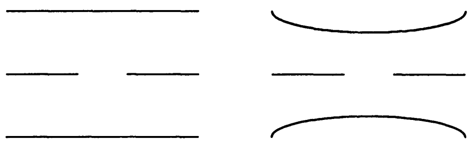
5.4 Minutiae Determination

In the thinning image (see Figure 5-4(c)), three kinds of points (pixels) can be determined by computing their connection numbers [9]. They are endings, intermediate points and bifurcations, which the connection numbers are 2, 4 and 6 (or 8), respectively. We call both endings and bifurcations as feature points. The true minutiae can be extracted from them. Not all feature points are true minutiae. Some of them are produced by uneven pressure in taking the fingerprints and some other feature points are obtained by the smear of ink, etc. In the following, we will discuss the difference between the true and false minutiae and propose the corresponding algorithms for their discrimination.

For all types of the feature points, we can define their orientations as follows: Starting from a feature point, trace the ridge (or ridges) and stop at nth step. If we denote the position of the first ending as (i_k, j_k) and the position of its nth tracing step as (x_k, y_k), then the orientation of the ending is defined as a direction from the point, (i_k, j_k), to the point, (x_k, y_k). Usually we use the angle, θ_k, to represent the orientation calculated clockwise with the j axis and the value of θ_k may vary between 0 and 360 degree. Obviously, for each ending there is only one orientation, but for each bifurcation, three orientations can be calculated.

BROKEN RIDGE CASE

One of typical encountered cases of the broken ridge is shown in Figure 5-5(a). However, in some cases, the break of ridges is an inherent property of the fingerprint and it should be reserved as shown in Figure 5-5(b).



(a) (b)

Figure 5-5. Broken ridge in the thinning image.
(a) Typical encountered case of broken ridge. (b) Inherent property of fingerprint.

Now we denote two orientations of these endings with positions (i_1, j_1) and (i_2, j_2) as θ_1 and θ_2, respectively. The orientation of a line starting from (i_1, j_1) and ending at (i_2, j_2) is denoted as θ_3. All these orientations are shown in Figure 5-6. Let θ_1 replaced by 180 degrees plus θ_1, and if it is greater than 360 degrees, 360 degrees should be subtracted from it. In any case, the absolute value of the difference of two orientations is within 180 degrees. This means that if $(\theta_2 - \theta_1)$ is greater than 180 degrees, it is replaced by 360 degrees minus $(\theta_2 - \theta_1)$ and if $(\theta_2 - \theta_1)$ is less than -180 degrees, it is replaced by 360 degrees plus (θ_2 - θ_1). The necessary conditions of connecting two endings are:

- $|\theta_2 - \theta_1|$ is less than 90 degrees.
- The product of (θ_3 - θ_1) and (θ_2 - θ_3) is greater than a given threshold which is usually a negative value.
- The line connecting the two endings is never allowed to cross another ridges.
- If there are two endings satisfying the above conditions and two adjacent ridges surround them, then we can calculate three distances between these adjacent ridges as shown in Figure 5-7, where a_1 and a_2 are points distant from endings about six pixels and a_3 is the midpoint between these two endings. If the distances, d_1 and d_2, exceed d_3 over a given threshold, then the break of ridges is inherent, the connection condition is not satisfied.

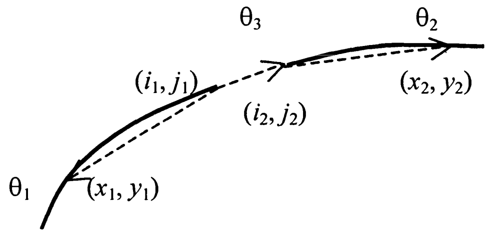


Figure 5-6. The orientations of two endings on a broken ridge.

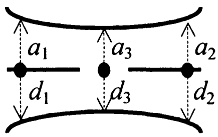


Figure 5-7. Three distances between the adjacent ridges surrounding a broken ridge.

For the real encountered cases, there are frequently several endings satisfying the necessary conditions for connecting to same ending (see Figure 5-8). Our algorithm must be able to process such situation.

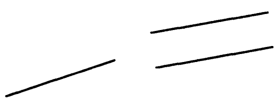


Figure 5-8. Two endings satisfying the necessary conditions for connecting to the same ending.

Algorithm 1

Input: Set of endings and thinning image.

(1) For every ending, find the set of candidates satisfying the conditions described above.

(2) If the set is empty, go to step 4. If there is only one candidate, connect it with the current ending. If the number of candidates is over one, then select only two candidates from set of candidates with the least values of $|\theta_2 - \theta_1|+|\theta_3 - \theta_2|+|\theta_3 - \theta_1|$.

(3) If the first candidate of one ending is just the other ending and vice versa, then connect them; otherwise, connect the candidate of one of the endings with this ending which has the minimum value of $|\theta_2 - \theta_1|+|\theta_3 - \theta_2|+|\theta_3 - \theta_1|$.

(4) Go to check the next ending.

LOOP CASE

Now we discuss how to identify the true or false loop minutiae. The most important approach determining loops as true minutiae is to compare the width of the loops with that of the ridges connected to the loop. The width comparison must be operated in the binary image. The algorithm can be described as follows:

Algorithm 2

Input: Set of bifurcations, thinning image and binary image.

(1) Starting from each bifurcation, trace its three ridges of the thinning image. The number of pixels of the tracing ridges is denoted by n. The larger n is, the greater loop can be detected. After that, the orientations of these three tracing ridges are calculated. The tracing is done for all bifurcations and the positions of three tracing ridges and their orientations are stored in corresponding arrays for later use.

(2) If two of the tracing ridges starting from a bifurcation point meet in another bifurcation, then a loop is found. If the difference between the orientation of the third tracing ridge and that of one of the meet ridges is greater than 90 degrees and less than 270 degrees, five characteristic points are determined according to Figure 5-9. In this figure, a_1 *and* a_3 *are bifurcations,* a_4 *and* a_5 *are four and eight pixels distant from* a_1*, respectively, and* a_2 *is the midpoint of point* a_1 *and* a_3.

(3) Calculate the widths of ridges at point, a_1, a_2, a_3, a_4 *and* a_5*, in the binary image. The widths are denoted as* w_1, w_2, w_3, w_4 *and* w_5 *respectively. Then we have the following rules to determine whether the loop is true or false: (i) if* w_1 *is greater than* $0.7{\times}w_2$ *or* w_3 *is greater* $0.7{\times}w_2$*, or (ii) if* w_4 *is greater than* $0.7{\times}w_2$ *and* w_5 *is greater than* $0.7{\times}w_2$*, then the loop is false. If the loop is false, we delete these two ridges and connect these two bifurcations with a straight line.*

(4) Go to check other bifurcations.

BRIDGE CASE

In many fingerprint images, we can frequently find false bridges produced by redundant ink. For most cases the bridge happens between two parallel or near parallel ridges. As a result, such structure can be found easily from checking the data stored in the arrays obtained from Algorithm 2. For each such ridge structure, we can decide whether the bridge between the two ridges is true or false. The orientation of the ridges and the original gray image are informative in making this decision. The following algorithm is to check bridge feature points.

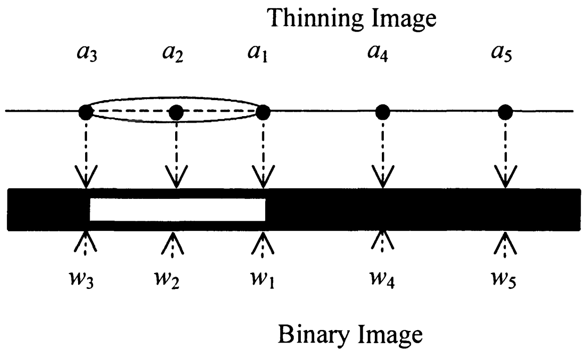


Figure 5-9. The width of ridge with a loop in the binary image.

<u>Algorithm 3</u>

Input: Set of bifurcations, thinning image and gray level image.

(1) From the set of bifurcations, find such bifurcation, of which two tracing ridges have near opposite orientations and the third ridge has the length shorter than a given threshold and ends at the second bifurcation. If the second bifurcation also has two longer ridges with near opposite orientations and its shorter ridge ends at the first bifurcation, this shorter ridge is a bridge. The ridge corresponding to the bridge is denoted by r_1.

(2) Compute the difference between the orientation of r_1 *and the orientations of the other four ridges. If all differences are more than 45 degrees and less than 135 degrees, then the bridge is false and should be deleted. Otherwise, calculate the average gray level of* r_1 *(denoted by* g_1*) with the average gray level of those pixels located perpendicular with* r_1 *and distant four to eight pixels from it (denoted by* g_2*) in the gray level image. If* (g_2-g_1) *is less than a given threshold, the bridge is noisy and should be deleted.*

(3) Go to check other bifurcations.

SHORT BAR CASE

The noises in the fingerprint image always cause the independent ridges, dots or islands in the thinning image. To check these short bars, which locate between two long parallel or near parallel ridges, we provide the algorithm as follows.

<u>Algorithm 4</u>

Input: Set of endings and thinning image.

(1) *From the set of endings, find the short bar. If the distance between its two endings is less than a given threshold, then the bar is considered as a noise and should be deleted.*
(2) *From the thinning image, find two adjacent ridges located at the two sides of the short bar as shown in Figure 5-10. Compute three distances, d_1, d_2 and d_3, of the two adjacent ridges at three points, a_1, a_2 and a_3. The point a_3 is the midpoint of the short bar. The other two points, a_1 and a_2, are at the extension of the ending of the short bar about six pixels, respectively. If the distance d_3 is longer than other two distances, d_1 and d_2, over a given threshold, then it is a true short bar; otherwise, it is a noise and should be deleted.*
(3) *Go to check other short bars.*

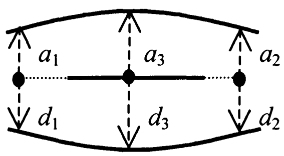


Figure 5-10. Three distances between the adjacent ridges surrounding a short bar.

SPUR CASE
Some short bar will become the spur feature point when it links to a ridge. On the other hand, the spur feature point is formed because the ridge is not smooth enough in the binary image. In order to eliminate these false minutiae, we give the algorithm for checking them.

Algorithm 5
Input: Set of bifurcations, set of endings, thinning image and binary image.
(1) *From the set of bifurcations, find such bifurcation with three ridges satisfying the following conditions: (i) a shortest ridge (denoted by r_1) ends at a ending with length less than a given threshold; (ii) the difference between the orientation of one of the other ridges (denoted as r_2) and that of r_1 being greater than 120 degrees and less than 240 degrees; and (iii) the difference between the orientation of the third ridge (denoted by r_3) and that of r_1 is less than 60 degrees or greater than 300 degrees.*
(2) *For the bifurcation found by step 1, calculate the perpendicular distance from the ending of r_1 to ridge r_3 (denoted by l_1). Find the average widths of ridge r_1, r_2 and r_3 at points distant about six pixels from the bifurcation point in the binary image and denoted by w_1, w_2 and w_3 (see Figure 5-11). If $(w_1+w_3+l_1-w_2)$ is less than a given threshold, then this short ridge is a false ridge and should be deleted as shown in Figure 5-12.*
(3) *Check other spur feature points.*

ADJACENT BIFURCATIONS CASE

The Type 7 feature points always emerge when two ridges are merged in the gray level fingerprint image. The following algorithm is for checking these feature points.

Algorithm 6

Input: Set of bifurcations, thinning image and binary image.

(1) From the set of bifurcations find such bifurcation of which two tracing ridges (denoted by r_1 and r_2) have the same or near same orientations (denoted by θ_1) and the third one (denoted by r_3) is opposite or near opposite to the orientation θ_1. Besides, the third tracing ridge must end at another bifurcation (obviously, this bifurcation must have a tracing ridge ending at the original bifurcation) which also has two tracing ridges (denoted as r_4 and r_5) with the same or near same orientations (denoted by θ_2) as shown in Figure 5-13. If the difference between θ_1 and θ_2 is greater than 150 degrees and is less than 210 degrees, then go to step 2, otherwise, go to step 3.

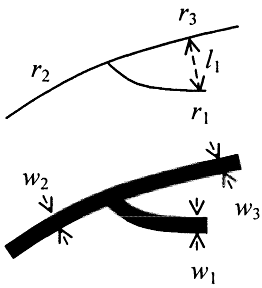


Figure 5-11. Definitions of some distances in thinning image and the widths in binary image for spur.

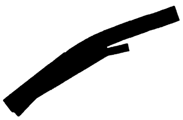


Figure 5-12. A spur with two false minutiae.

(2) Find the widths of r_1, r_2, r_3, r_4 and r_5 from the binary image. If the width of r_3 is near the widths of r_1, r_2, r_4 and r_5, then these two bifurcations are true minutiae. Otherwise, they are false and two line segments should replace the third tracing ridge as shown in Figure 5-14.

(3) Check other bifurcations.

The above six algorithms are mostly used to treat the main types of feature points. Nevertheless, there are still another false feature points, which cannot be processed by the above algorithms. For example, if the fingerprint is recorded across a printed straight line of the paper, then a lot of false bifurcation points produce (see Figure 5-3(h)). Fortunately, for each real case, the rules for discrimination can always be suggested. As an example, for the above case, we only need to find the printed straight line either from thinning image or from binary image and the bifurcation which locates or nearly locates in this line are considered as false bifurcation. We also can analyze the positions of the tracing ridges of the adjacent bifurcations and then easily make conclusion whether these adjacent bifurcations are produced by a straight line or not. In addition, much attention must be paid to the feature points in the neighborhood of the core point and the delta points. Since the orientation of the ridges in these regions may be changed greatly, some rules described above might not be useful for decision. In this case, the reliability of minutiae can be determined by analyzing the width and the length of the ridges by checking the binary image. Since there are a lot of different cases of Type 8 feature points and in most cases, the dominant minutiae are not in Type 8, we will not further discuss the algorithm for its treating.

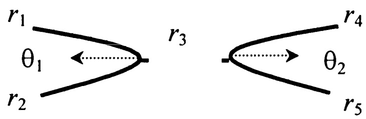


Figure 5-13. Definitions of the ridges and their orientations for two adjacent bifurcations.

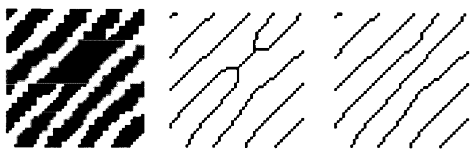


Figure 5-14. Example of two false adjacent bifurcations eliminated.

Based on the experiences of the fingerprint experts, we have developed a complete approach composed of the above algorithms and a large number of fingerprint images have been processed to show the effectiveness of our approach. For illustration, we here give some practical results of fingerprints with three different qualities (good, middle and poor). For each kind of fingerprints (see Figures 5-15, 5-16, and 5-17), the original thinning image, the improved thinning image and the original gray level image with minutiae extraction are shown in their (a), (b) and (c), respectively. The error rates are calculated in Table 5-1. It is necessary to point out for the region of great noises in which the ridge orientations are irregular, all endings and bifurcations are considered to be false without discrimination which is required by the users.

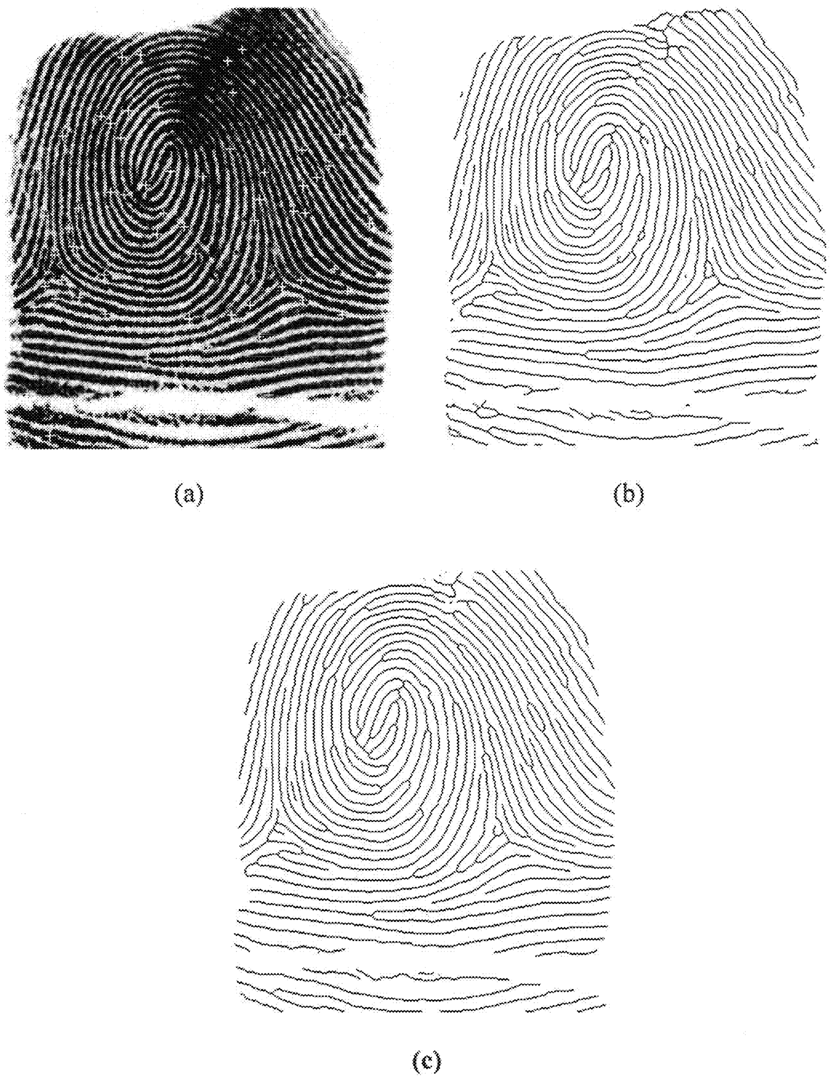


Figure 5-15. Good quality fingerprint image tested by our approach. (a) Original gray-level image with minutiae, (b) Thinning image, and (c) Thinning image after improvement.

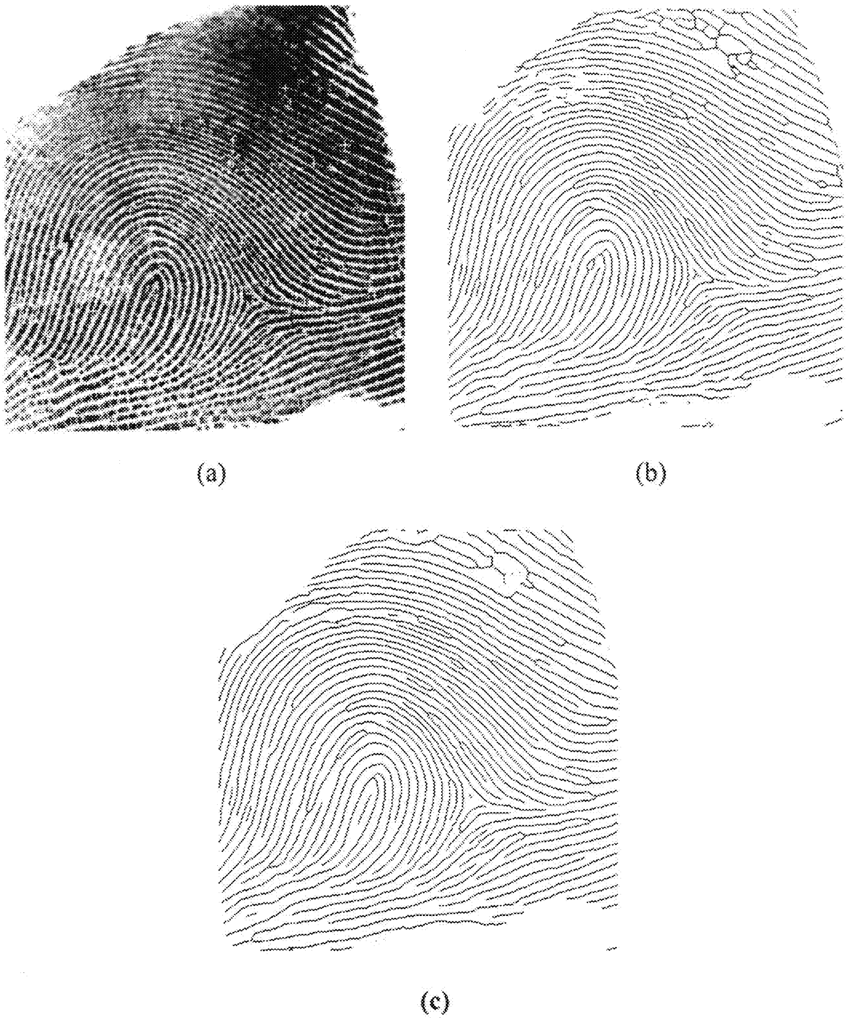


Figure 5-16. Middle quality fingerprint image tested by our approach. (a) Original gray-level image with minutiae, (b) Thinning image, and (c) Thinning image after improvement.

Table 5-1. Experimental results of three fingerprint images with different qualities

Quality	*Total extracted minutiae*	*False number*	*Lost minutiae*
Good (Figure 5-15)	60	2	3
Middle (Figure 5-16)	98	4	4
Poor (Figure 5-17)	97	7	0

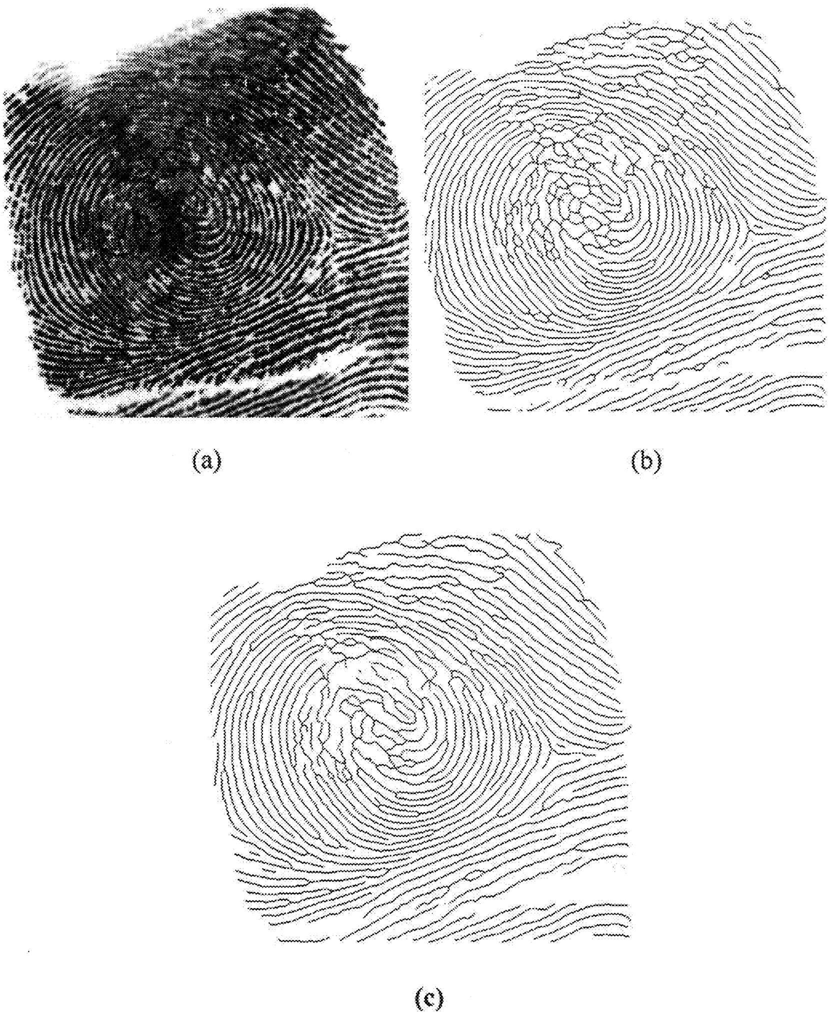


Figure 5-17. Poor quality fingerprint image tested by our approach. (a) Original gray-level image with minutiae, (b) Thinning image, and (c) Thinning image after improvement.

5.5 Fingerprint Matching

FINGERPRINT CLASSIFICATION

At the coarse level, fingerprints can be classified into seven classes: six main classes mentioned above and special pattern. Although the classification does not identify a fingerprint uniquely, it is helpful in determining when two fingerprints do not match. For example, a right loop image should be matched with only other right loop images in the database of fingerprints. So, fingerprint classification cannot only improve the recognition accuracy but also drastically reduce the proportion of database images to be matched at the fine level.

The fingerprint classification problem has been addressed by many researches. A syntactic method is presented by Rao et al. [10]. Wislson et.al. have used a neural network to classify fingerprint images [11, 12]. The approach taken by Srinivasan et al. extracts the core and delta points and performs classification based on the number and locations of the detected singular points [4]. K. Karu and A. K. Jain adopt a different method, which the Poincaré Index is used to located the singular points [13].

In many cases, determination of singular points is the key algorithm in fingerprint classification. The popular method is to locate the singular point in the direction image as several orientations meet (shown as Figure 5-2).

MATCHING POLICIES

Matching is a key operation in the current fingerprint identification system. It is to determine whether two fingerprints are from the same finger or search for a query fingerprint in a database [21]. This kind measure of similarity is always based on the matching of minutiae patterns [14].

In general, the relationship among the minutiae extracted from the fingerprint can be considered as a graph that is denoted as $G = (N, E, A, B, F_N, F_E)$, where $N = \{n_1, n_2, \ldots, n_m\}$ and it presents a finite set of nodes; m is the number of nodes, $E = \{e_1, e_2, \ldots, e_p\}$ and it is a node pair set; $e = (n_i, n_j)$, $1 \leq i, j \leq m$, and it is an edge between n_i and n_j, p is the number of node pairs; A is the attribute set of nodes and B is the attribute set of edges; F_N is the function (or the function set) which generates the node attributes and F_E is the function (or the function set) which generates the edge attributes. For instance, the fingerprint shown in Figure 5-18 can be described as the following graph, $G = (N, E, A, B, F_N, F_E)$, where $N = \{n_1, n_2, n_3, n_4\}$, $E = \{e_1, e_2, \ldots, e_6\}$ and $A = \{(c{:}\ x, y), (e{:}\ x, y), (b{:}\ x, y)\}$; c, e, b present the pattern of minutiae: core, ending and bifurcation, respectively; (x, y) is the coordinate of minutia; $B = \{n, n = 1, 2, \ldots\}$ and n is the number of ridges. $F_N : n_1 \rightarrow\ c{:}\ x_1, y_1, n_2 \rightarrow\ b{:}\ x_2, y_2, n_3 \rightarrow\ e{:}\ x_3, y_3, n_4 \rightarrow\ b{:}\ x_4, y_4$ and $F_E : e_1{=}(n_1, n_2) \rightarrow 1, e_2{=}(n_1, n_3) \rightarrow 3, e_3{=}(n_1, n_4) \rightarrow 1, e_4{=}(n_2, n_3) \rightarrow 1, e_5{=}(n_2, n_4) \rightarrow 2, e_6{=}(n_3, n_4) \rightarrow 1$. So, the matching of minutiae is that graph matching [15].

Let G_1 and G_2 be a pair of graphs, they are defined as isomorphic graphs if there exists a mapping, P_i $(i = 1, \ldots)$, and it makes $P_1(N_1) = N_2$, $P_2(E_1) = E_2$, ... The isomorphic subgraphs mean that there is an isomorphic mapping between subgraph G_1 and G_2. In a lot of fingerprint authentication systems, the matching is to find out the isomorphic subgraphs.

Although the most prevalent current model for automated fingerprint identification systems are based on minutiae, a fingerprint is also represented in the form of a graph whose nodes correspond to ridges and edges represent ridge adjacency information [16]. Hence the graph structure captures the topological relationships within the fingerprint.

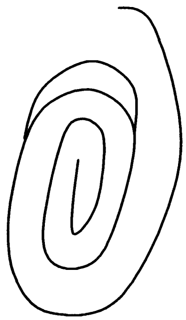


Figure 5-18. A subimage of fingerprint.

In addition, the methodology is based on a data model for fingerprint that is structural, rather than coordinate [17]. The structural model for fingerprint is defined as follows. For each feature on the fingerprint, a neighborhood of some specified radius R about the center of the feature is denoted. If we denote the types of features, such as mentioned above, by $T_1 \ldots T_8$ each occurring with frequency $f(T_1) \ldots f(T_8)$ in the neighborhood, and the type of central feature for the neighborhood by T, then we associate with the central feature a vector $(T, f(T_1) \ldots f(T_8))$ as a characterization of the neighborhood. With a suitable choice of R, a list of such feature vectors can be obtained for each fingerprint, containing information similar to that useful in a manual search; namely, distinctive local patterns of feature distribution, which can be serve as the basis for search and identification.

As a summary, the fingerprint system including fingerprint image processing and matching is introduced in this chapter and some proposed algorithms are used in a commercial system named CAFIS (China Automated Fingerprint Identification System). Its applications show that the system is acceptable.

References

[1] H.C. Lee and R.E. Gaensslen, editors, *Advances in Fingerprint Technology*, Elsevier, New York, 1991.

[2] A.K. Jain, R. Bolle and S. Pankanti (editors), *Biometrics: Personal Identification in a Networked Society*, Kluwer Academic Press, 1999.

[3] A. Jain, L. Hong and R. Bolle, "On-line Fingerprint Verification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 4, pp. 302-314, 1997.

[4] V.S. Srinivasan and N.N. Murthy, "Detection of Singular Points in Fingerprint Images", *Pattern Recognition*, vol. 25, no. 2, pp. 139-153, 1992.

[5] S. Marcelja, "Mathematical Description of the Responses of Simple Cortical Cells," *J. Opt. Soc. Am. A70*, pp. 1297-1300, 1980.

[6] M. Lindenbaum, M. Fischer and A. Bruckstein, "On Gabor's Contribution to Image Enhancement," *Pattern Recognition*, vol. 27, no. 1, pp. 1-8, 1994.

[7] A.K. Jain and F. Farrokhnia, "Unsupervised Texture Segmentation Using Gabor Filters," *Pattern Recognition*, vol. 24, no. 12, pp. 1167-1186, 1991.

[8] J. Kittler and J. Illingworth, "Minimum Error Thresholding," *Pattern Recognition*, vol. 19, no. 1, pp.41-47, 1986.
[9] L. Lam, S.W. Lee and C.Y. Suen, "Thinning Methodology - A Comprehensive Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, pp. 869-885, 1992.
[10] C.V.K. Rao and K. Black, "Type Classification of Fingerprints: A Syntactic Approach," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 2, pp. 223-231, 1980.
[11] J.L. Blue, G.T. Candela, P. J. Grother, R. Chellapa and C. L. Wilson, "Evaluation of Pattern Classifiers of Fingerprint and OCR Applications", *Pattern Recognition*, vol. 27, no. 4, pp. 485-501, 1994.
[12] C.L. Wilson, G.T. Candela and C.I. Waston, "Neural Network Fingerprint Calssification," *J. Artific. Nerual Networks*, vol. 1, no. 2, pp. 1-25, 1993.
[13] K. Karu and A.K. Jain, "Fingerprint Classification," *Pattern Recognition*, vol. 29, no. 3, pp. 389-404, 1996.
[14] A.K. Jain, L. Hong, S. Pankanti, and R. Bolle, "An Identity-Authentication System Using Fingerprint", *Proceedings of The IEEE*, vol. 85, no. 9, pp. 1365-1388, 1997.
[15] Z. Bian, D. Zhong (editors), *Intelligent Signal Processing and Intelligent Control*, Zhejiang Science & Technology Publishing House, 1998.
[16] D.K. Isenor and S.G. Zaky, "Fingerprint Identification Using Graph Matching," *Pattern Recognition*, vol. 19, no. 2, pp. 113-122, 1986.
[17] A.K. Herchak and J.A. McHugh, "Automated Fingerprint Recognition Using Structural Matching," *Pattern Recognition*, vol. 23, no. 8, pp. 893-904, 1990.
[18] A. Wahab, S.H. Chin, E.C. Tan, "Novel Approach to Automated Fingerprint Recognition," *IEE Proc., Vision, Image Signal Proc-ess*, vol. 145, no. 3, June 1998.
[19] B. Moayer, and K.S. Fu, "A Tree System Approach for Fingerprint Pattern Recognition," *IEEE Trans., PAMI*, vol. 8, no. 3, pp. 376-387, 1986.
[20] B.G. Sherlock, D.M Monro, and K. Millard, "Fingerprint Enhancement by Directional Fourier Filtering," *IEE Proc., Vision, Image, Signal Process.*, vol. 141, no. 2, pp. 87-94, 1994.
[21] K.R. Nalini, K. Karu, S. Chen and A.K. Jain, "A Real-time Matching System for Large Fingerprint Databased," *IEEE Trans., PAMI*, vol. 18, no. 8, pp. 799-813, 1996.
[22] B.G. Sherlock, D.M Monro and K. Millard, "Algorithm for Enhancing Fingerprint Images," *Electron. Lett.,* vol. 28, no. 18, pp. 1720-1721, 1992.
[23] Q. Xio and H. Raafat, "A Combined Statistical and Structural Approach for Fingerprint Image Postprocessing," *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, Los Angeles, CA, USA, pp. 331-335, Nov. 1990.
[24] R. Bahuguna, "Fingerprint Verification Using Hologram Matched Filterings," in *Proc. Biometrics Consortium Eighth Meeting*, San Jose, CA, June, 1996.
[25] L. Coetzee and E.C. Botha, "Fingerprint Recognition in Low Quality Images," *Pattern Recognition*, vol.26, no. 10, pp. 1441-1460, 1993.
[26] P.E. Danielsson and Q.Z. Ye, "Rotation-Invariant Operators Applied to Enhancement of Fingerprints," Proc. Eighth ICPR, pp. 329-333, Rome, 1988.
[27] L. Hong, A.K. Jain, S. Pankanti, and R. Bolle, "Fingerprint Enhancement," in *Proc. IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, pp. 202-207, 1996.
[28] D.C.D. Hung, "Enhancement and Feature Purification of Fingerprint Images," *Pattern Recognition*, vol. 26, no. 11, pp. 1661-1671, Nov. 1993.
[29] M. Kawagoe and A. Tojo, "Fingerprint Pattern Classification," *Pattern Recognition*, vol. 17, no. 3, pp.295-303, 1984.
[30] D. Maio and D. Maltoni, "A Structural Approach to Fingerprint Classification", in *Proc. 13th ICPR*, Vienna, Austria, pp. 132-138, 1996.
[31] L. O'Gorman and J. Nickerson, "An approach to fingerprint filter design," *Pattern Recognition,* vol.22, pp. 29-38, 1989.
[32] A. Sherstinsky and R.W. Picard, " Restoration and Enhancement of Fingerprint Images Using M-Lattice-A Novel Nonlinear Dynamical System," *Proc. 12th ICPR-B*, pp.195-200, Jerusalem, 1994.
[33] J. D. Stosz and L. A. Alyea, "Automated System for Fingerprint Authentication Using Pores and Ridge Structure," in *Proc. SPIE, Automatic Systems for the Identification and Inspection of Humans*, San Diego, CA, vol. 2277, pp. 210-223, July 1994.
[34] A.R. Roddy, J. D. Stosz, "Fingerprint Features – Statistical Analysis and System Performance Estimates, " *Proceedings of the IEEE*, vol.85, no.9, September 1997.

[35] B.M. Mehtre, "Fingerprint Image Analysis for Automatic Identification," *Machine Vision and Applications*, vol.6, pp. 124-139, 1993.
[36] D. Maio and D. Maltoni, "An Efficient Approach to On-lne Finger Print Verification," *Proc. VIII Int. Symposium on Artificial Intelligence*, 1995.

6 PALMPRINT VERIFICATION

Automatic palmprint verification is an important complement of biometrics authentication. As the first attempt of personal identification by palmprint, this chapter explores some new methods for three main processing stages in palmprint verification, including datum point registration, line feature extraction and palmprint classification. Various palmprint images have been tested to illustrate the effectiveness of the methods proposed.

6.1 A New Type of Biometrics: Palmprint

Palmprint, as well as fingerprint which has been utilized as a positive human identifier for more than 100 years, is still considered as one of the most reliable means distinguishing a man from his fellows because of its stability and uniqueness. Although law enforcement agencies use human prints (both fingerprints and palmprints) routinely for criminal identification, human prints can provide an unequalled method of personal identification for various other areas such as access control for high security installations, credit card usage verification and employee identification [1-3]. Fingerprint identification is the most widespread application of biometrics technology. However, it is a difficult task to extract some small unique features (known as minutiae) from the fingers of elderly people as well as manual labourers [4,5]. As a result, it is necessary to develop a new technique of automatic personal identification by palmprint.

Palm is the inner surface of the hand between the wrist and the fingers (see Figure 6-1). There are usually three principal lines made by flexing the hand and wrist in the palm, which are named as heart line, head line, and life line, respectively [6]. Two endpoints, ***a*** and ***b***, are obtained by the principal lines (life line and heart line) which intersect the both sides of palm. Owing to the stability of the principal lines, the endpoints and their midpoint, ***o***, remain unchanged day in and day out. Some significant properties can be defined in the following: (1) The locations of the endpoints and their midpoint are rotation invariant in a palmprint; (2) A sole two-dimensional right angle coordinate system can be established, of which the origin is the midpoint, ***o***, and the main axis passes through the endpoints; (3) A palm can be divided into three regions, including finger-root region (I), inside region (II) and outside region (III) by some connections between the endpoints and their perpendicular bisector; (4) The size of a palm can be uniquely estimated by both the Euclidean distance between two endpoints and the length of their perpendicular bisector in the palm (segment ***c-d*** in Figure 6-2). In this way, the pair of endpoints and their midpoint will act as the important datum points in the palmprint identification because of their invariable locations.

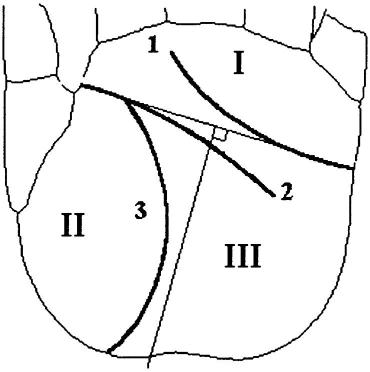


Figure 6-1. Definitions of a palmprint: principal lines (1–heart line, 2–head line and 3–life line), regions (I–finger-root region, II–inside region and III–outside region).

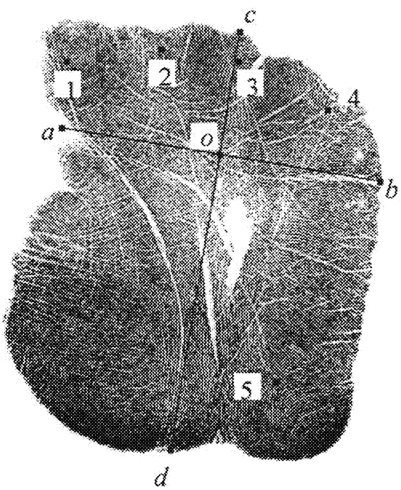
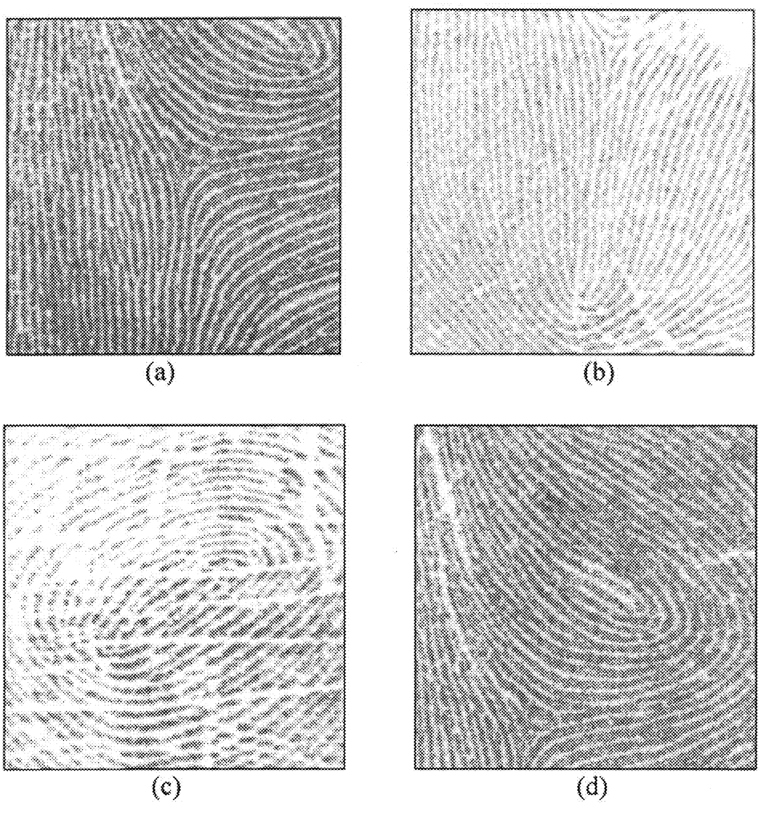


Figure 6-2. Geometry features and delta point features of a palmprint, where ***c-d*** is the perpendicular bisector of segment *a-b* and points 1-5 are delta points.

In addition to the datum points, there still exists rich and useful information in a palmprint [7]. Some kinds of features could be listed as follows:

- *Geometry Features*: According to the palm's shape, we can easily get the corresponding geometry features, such as width, length and area [29-30].

- *Principal Line Features*: Both location and form of principal lines in a palmprint are very important physiological characteristics to identify individual because they vary little from time to time.
- *Wrinkle Features*: In a palmprint, there are many wrinkles which are different from the principal lines in that they are thinner and more irregular. They are classified as coarse wrinkles and fine wrinkles so that more features in detail can be acquired.
- *Delta Point Features*: Delta point is defined as the center of a delta-like region in the palmprint. It is shown that there always are some delta points located in finger-root region and outside region (e.g., **1** ~ **5** in Figure 6-2). This makes us possible to obtain these features and well establish the stability and uniqueness.
- *Minutiae Features*: A palmprint is basically composed of the ridges (as an example, some classes of ridges in palmprint are shown in Figure 6-3), hence the minutiae features can be used as a significant measurement.



(a) (b) (c) (d)

Figure 6-3. Classes of ridges in palmprint: (a) Delta point; (b) Arch; (c) Whorl; and (d) Loop.

In general, geometry features, principal line features and wrinkle features can be derived from the image with low resolution, even low quality. Delta points can be located in the directional image and their features will be procured from a fine resolution palmprint image in case of the presence of noise in it [8,9]. Nevertheless, minutiae features are detected only from a high quality and fine resolution image.

Palmprint verification, which is to determine whether two palmprints are from the same palm, can apply the physical features mentioned above in principle. However, both delta point and minutiae features only can be obtained in fine resolution image. In addition, geometry features are easily captured so that a fake palm could be created [2]. As a new biometrics technology, we will adopt some significant features called line feature (such as principal lines and coarse wrinkles) in palmprint verification. Compared with fingerprint identification, which is a point matching process [4,25-28], this palmprint verification by using line feature matching is effective in the following respects: (1) It can be used on the image with low spatial resolution so that the size of a palmprint image can be reduced, even it is comparable to that of a fingerprint image; (2) Significant features can be determined in case of the presence of noise because of feature extraction relied on low spatial resolution; (3) Palmprint verification is a line matching process, which has several advantages over point matching since a line has more information than a point [10]. In addition, the singular points, namely the core and the delta points, are used as points of registration for comparing minutiae in fingerprint [8]. However, the effective method to detect these structural features makes use of directional histograms in the neighbourhood of these points [4,8,9], hence it will be difficult to provide accurate estimates of the singular points while fingerprint image is rotated and deformed. Therefore, the datum points of palmprint are more stable than those of fingerprint.

6.2 Datum Point Determination

The goal of datum point determination is to achieve the registrations in palmprint feature representation and matching. As a prime but important process in the palmprint verification, it is demanded as simple and effective as possible. The basic idea of datum point determination is to locate the endpoint of each principal line. According to the given regularity, the principal lines and their endpoints are accurately detected by using the directional projection algorithm.

DIRECTIONAL PROJECTION ALGORITHM

It is widely known that projection is a simple yet effective method of line segment detection along some orientation [11].

Let **F** be an $M \times N$ gray scale image and $f(i, j)$ be the gray level of pixel (i, j) ($i = 0, 1, \ldots, M\text{-}1$; $j = 0, 1, \ldots, N\text{-}1$). Without loss of generality, we consider that the projective angle α is measured clockwise from the i-axis and pixel (i', j') is a pixel in **F**. Then, an x-y right angle coordinate system can be established, of which pixel (i', j') is the origin and the orientation of x-axis is that of the projection (see in Figure 6-4). In the x-y coordinate system, a $(2m+1) \times (2n+1)$ pixels in size subimage, $\mathbf{F}_1$, can be obtained, and $f(x, y)$ is the gray level of pixel (x, y) ($x = -m, -m+1, \ldots, 0, \ldots, m-1, m$; $y = -n, -n+1, \ldots, 0, \ldots, n-1, n$). As a result, the correspondence between this pair of coordinate systems is denoted as follows:

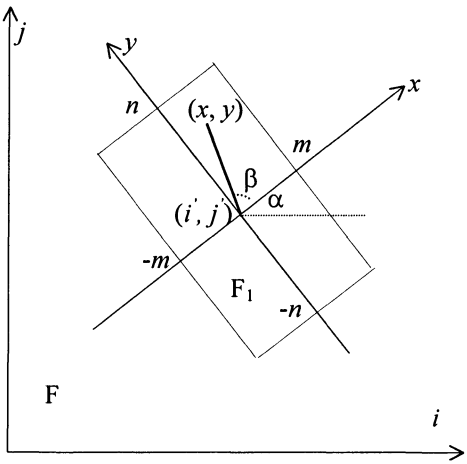


Figure 6-4. The correspondence of two coordinate systems in directional projection algorithm.

$$\begin{aligned} i &= i' + \cos(\alpha+\beta)\sqrt{x^2+y^2} \\ j &= j' + \sin(\alpha+\beta)\sqrt{x^2+y^2} \end{aligned} \tag{6-1}$$

where $\beta = \tan^{-1}(y / x)$.

In $\mathbf{F}_1$, obviously, the directional projection of this subimage is

$$p(y) = \sum_{i=-m}^{m} f(x, y)\,. \tag{6-2}$$

The smoothed set $q(y)$ is calculated as

$$q(y) = \frac{1}{2w+1}\sum_{k=-w}^{w} p(y+k)\,. \tag{6-3}$$

Then, pixel $(0, y_0)$, where

$$y_0 = \{k \mid q(k) = \max_{y} q(y)\}, \tag{6-4}$$

is detected and the corresponding pixel in **F** is defined as pixel (i_0, j_0), where

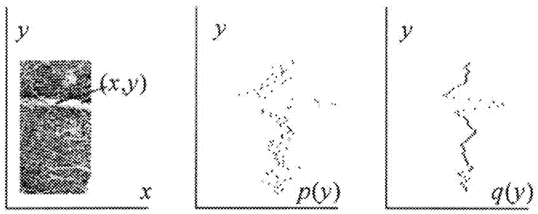
$$\begin{aligned} i_0 &= i' + y_0\cos(\alpha+\tfrac{\pi}{2}) \\ j_0 &= j' + y_0\sin(\alpha+\tfrac{\pi}{2}) \end{aligned} \tag{6-5}$$

Here, this basic algorithm is classified as four forms by the different projective orientations: horizontal projection ($\alpha=0^o$), projection with 45 degrees ($\alpha=45^o$), vertical projection ($\alpha=90^o$) and projection with 135 degrees ($\alpha=135^o$).

PROPERTIES FOR PRINCIPAL LINES

From the study of the principal lines, their properties can be obtained as follows: (1) Each principal line meets the side of palm at approximate right angle when it flows out the palm; (2) The life line is located at the inside part of palm, which gradually inclines to the inside of palm in parallel at the beginning; (3) Most of the life line and head line flow out the palm at the same point; (4) The endpoints are closer to fingers than wrist.

Based on the projective correspondence (see Figure 6-5), it is clear that the pixel (x, y) calculated by the basic algorithm belongs to the principal line if the orientation of directional projection follows the principal line and the parameter w in Equation (6-3) equals the half width of a principal line. However, the pixel on the other different kinds of the lines cannot be determined by means of the conditions given above. This is because the projection of line which is not in parallel with the projective orientation or is shorter than the principal line would be less than that of principal line. In addition, a thin line might map to the maximum value in the directional projection, $p(y)$, but it could be reduced after smoothing. So, the basic algorithm can be used to locate the pixel only on the principal line while the directional projection along the principal line is maximized in a palmprint subimage.



(a) (b) (c)

Figure 6-5. Computation of the directional projection: (a) Original image; (b) Directional projection along principal line; and (c) Directional projection after median filter.

To apply the above rules to get the endpoints, two processing stages, i.e., principal line detection stage and principal line tracing stage, should be defined. Without loss of generality, the images of person's right palm are employed and all palmprints are considered as fingers upturned so as to describe the method of datum point determination simply. The track of the datum point determination is shown in Figure 6-6.

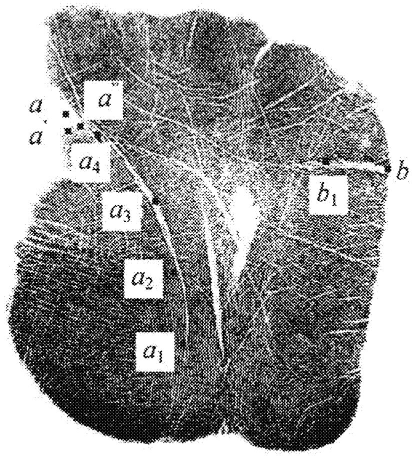


Figure 6-6. Track of datum point determined by using the directional projection algorithm.

PRINCIPAL LINE DETECTION

The purpose of this stage is to obtain the points, those belong to heart line or life line. The heart line detection is easier than that of life line because there is only one principal line in the outside part of palm. After the edges of outside and topside are detected, a pixel which has the suitable offsets to two edges is adopted in palmprint. Point b_1, which belongs to heart line, is located by the horizontal projection algorithm. However, the detection of life line cannot use the horizontal projection in the inside part of palm because the head line is determined sometimes instead of the life line while their flowing out palm at the different points. Therefore, another subimage close to wrist is processed by the vertical projection algorithm and point a_1 on the life line is calculated.

PRINCIPAL LINE TRACING

The main idea is to estimate a pair of endpoints by tracing principal lines.

Determining Datum Point, *b*

The heart line detection is easy because there is only one principal line in the outside part of palm. After the edges of outside and topside are detected, a pixel that has the suitable offsets to two edges is adopted in palmprint. The horizontal projection algorithm locates Point b_1, which belongs to heart line. According to the peculiarities of heart line, the horizontal projection algorithm is easy to locate endpoint b to a subimage where there is a pixel situated at the outside edge with the same level as pixel b_1.

Determining Datum Point, *a*

The detection of endpoint $\boldsymbol{a}$ is different from that of endpoint $\boldsymbol{b}$ because the life line is a curve from pixel $\boldsymbol{a}_1$ to point $\boldsymbol{a}$. The contour of inside region is shown in Figure 6-7 and it can be approached by a sequence. So, both the contour and projection are cared so that Point $\boldsymbol{a}$ can be located correctly. The details of the improved algorithm are shown below:

Step 1: Point $\boldsymbol{a}'$, $(\boldsymbol{i}_{00}, \boldsymbol{j}_0)$, is defined by the straight line, which is a connection between endpoint $\boldsymbol{b}$ and $\boldsymbol{b}_1$, intersecting the inside edge of palm.

Step 2: A sequence $\{(i_{y0}, j_y), y\in[-n, n]\}$ is given by

$$j_{y+1} = j_y+1,\ B(i_{y0}, j_y) = 0 \text{ and } B(i_{y0}-1, j_y) = 1, \tag{6-6}$$

where $B(i, j)$ is the value of pixel (i, j) in the binary image; 0 and 1 represents background and palm, respectively.

Step 3: Both the maximum value and the mean of the sequence are calculated. They are $\boldsymbol{i}_{max}$ and $\mathbf{E}_i$ and then a function is defined as

$$w(i_{y0}) = (\boldsymbol{i}_{max} - \mathbf{E}_i)\exp\{-\alpha(i_{y0} - \boldsymbol{i}_{max})\},\ -n \le y \le n, \tag{6-7}$$

where α is a constant.

Step 4: The improved horizontal projection is denoted as

$$\boldsymbol{p}(y) = w(i_{y0})\times \sum_{k=-m}^{m} f(i_{yk}, j_y),\ -n \le y \le n, \tag{6-8}$$

where $f(i, j)$ is the gray level of pixel (i, j).

Step 5: Smooth $\boldsymbol{p}(y)$ to $\boldsymbol{q}(y)$. Then, pixel (i_{Y0}, j_Y), where

$$Y = \{k \mid q(k) = \max_y q(y)\}, \tag{6-9}$$

is located.

Step 6: Confirm the datum point. In sequence $\{(i_{y0}, j_y), y\in[-n, n]\}$, let

$$\Delta i_1 = |i_{Y0-\Delta n} - i_{Y0}| \text{ and } \Delta i_2 = |i_{Y0+\Delta n} - i_{Y0}|, \tag{6-10}$$

where Δn is an offset. If both Δi_1 and Δi_2 are less than some threshold $\mathbf{T}_1$, pixel (i_{Y0}, j_Y) is point $\boldsymbol{b}$; if both Δi_1 and Δi_2 are more than $\mathbf{T}_1$, piexl $((i_{Y0-\Delta n}+ i_{Y0+\Delta n})/2, j_Y)$ is $\boldsymbol{b}$; if Δi_1 is more then $\mathbf{T}_1$ and Δi_2 is less than $\mathbf{T}_1$, pixel $(i_{Y0-\Delta n}, j_Y)$ is $\boldsymbol{b}$; otherwise, pixel $(i_{Y0+\Delta n}, j_Y)$ is $\boldsymbol{b}$.

6.3 Two Typical Features in Palmprint

LINE FEATURE

There are mainly two operations in the following identification, feature extraction and feature matching. Owing to line features adopted to identify the individual, line feature extraction and line matching are proposed. Obviously, line feature includes curves and straight lines. In fact, most line features can approximate straight line segments except the principal lines. Based on our observation, the curvature of the principal lines is small enough to be represented by several straight line segments. In this way, the representation and matching of straight lines are simple. As a result, line feature is considered as straight line segments. The measurement of palmprints matching will be also described.

$$\begin{array}{cccccc} -1 & -1 & 2 & 2 & -1 & -1 \\ -1 & -1 & 2 & 2 & -1 & -1 \\ -1 & -1 & \mathbf{2} & 2 & -1 & -1 \\ -1 & -1 & 2 & 2 & -1 & -1 \\ -1 & -1 & 2 & 2 & -1 & -1 \end{array} \qquad \begin{array}{ccccc} -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \\ 2 & 2 & \mathbf{2} & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \\ -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \end{array}$$

(a) (b)

$$\begin{array}{cccccccccc} -1 & -1 & 2 & 2 & -1 & -1 & & & & \\ & -1 & -1 & 2 & 2 & -1 & -1 & & & \\ & & -1 & -1 & \mathbf{2} & 2 & -1 & -1 & & \\ & & & -1 & -1 & 2 & 2 & -1 & -1 & \\ & & & & -1 & -1 & 2 & 2 & -1 & -1 \end{array} \qquad \begin{array}{cccccccccc} & & & & -1 & -1 & 2 & 2 & -1 & -1 \\ & & & -1 & -1 & 2 & 2 & -1 & -1 & \\ & & -1 & -1 & \mathbf{2} & 2 & -1 & -1 & & \\ & -1 & -1 & 2 & 2 & -1 & -1 & & & \\ -1 & -1 & 2 & 2 & -1 & -1 & & & & \end{array}$$

(c) (d)

Figure 6-7. Four templates: (a) Vertical; (b) Horizontal; (c) Left diagonal; (d) Right diagonal.

Line Feature Extraction

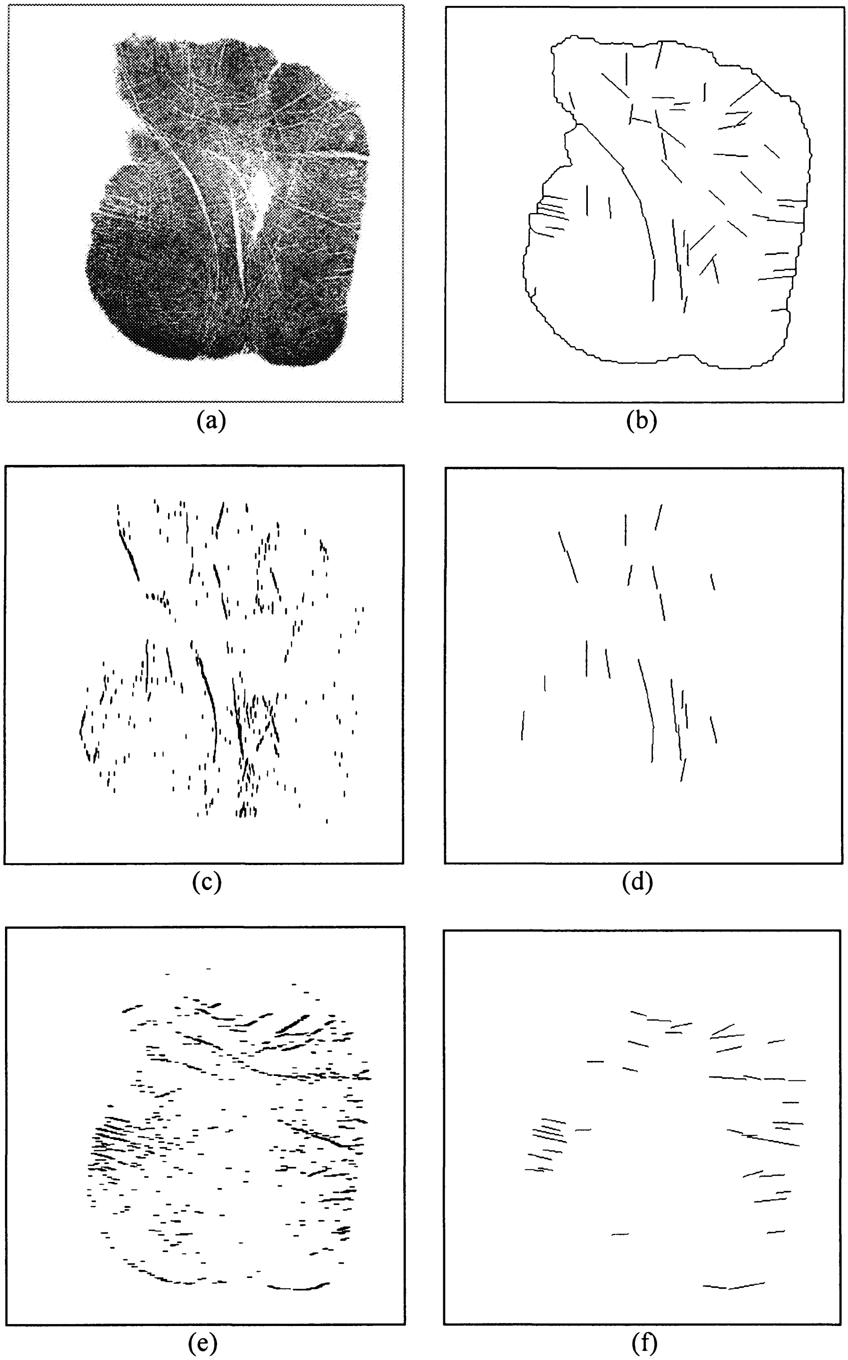
Line feature extraction is always an important yet difficult step in image verification and many line feature detection methods have been proposed [12]. Most of these methods cannot generate a precise line map of stripe images such as palmprint images except template algorithm. Nevertheless, many ridges and fine wrinkles have the same width as coarse wrinkles except that they are shorter in inked palmrprint image. As a result, the template algorithm is also difficult to acquire the line features from a palmprint because a mass of ridges and fine wrinkles could dirty the line features. We improve on this algorithm, which is called as improved template algorithm, so as to extract and post-process line segments of each orientation respectively, and then combine them. The improved template algorithm consists of the following main steps:

1) Determine vertical line segments by using the vertical templates (see Figure 6-7), and then thin and post-process these line segments. The rule of post-processing is to clear up the short segments.
2) Similarly, detect lines in other three directions.
3) Combine the results of four different directions, and then post-process once more to eliminate the segments overlapped.

The result of line feature extraction by the improved template algorithm is shown in Figure 6-8.

Line Feature Matching

In general, there are many ways to represent a line. One way, which is always possible, is to store the endpoints of every straight line segment [13]. In a two-dimensional right angle coordinate system which is uniquely established by the datum points, line segments can be described by endpoints: $(X_1(i), Y_1(i))$, $(X_2(i), Y_2(i))$, $i = 1$,



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

Figure 6-8. Results of line feature extraction.

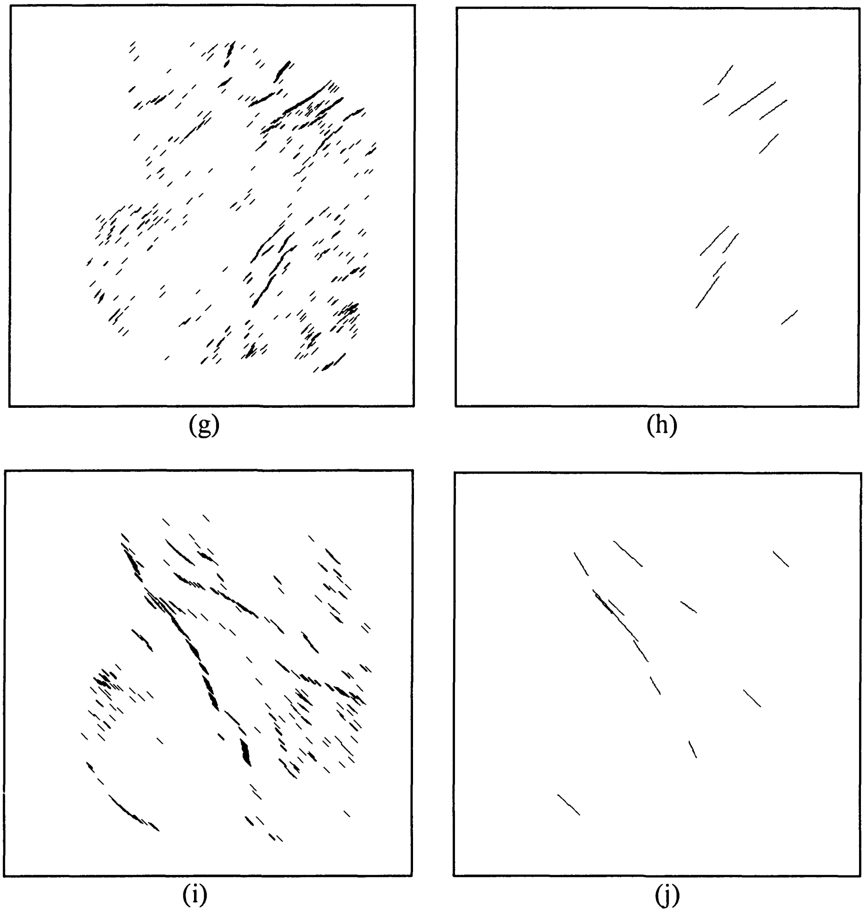


Figure 6-8. Results of line feature extraction (cont.):
(a) Input image; (b) Map with linear features extracted; (c),(e),(g) and (i) Line segments detection in four different directions by template algorithm, respectively; (d),(f),(h) and (j) The results thinned and post-processed.

..., I, where I is the number of line segments. Without loss of generality, exchange the endpoints of each line segment so that $X_1(i) \leq X_2(i)$, $i = 1, \ldots, I$. If $X_1(i) = X_2(i)$, exchange the endpoints so that $Y_1(i) \leq Y_2(i)$. Next, three parameters of each line segment, including slope, intercept and angle of inclination, can be calculated as follows:

$$slope(i) = (Y_2(i) - Y_1(i)) \,/\, (X_2(i) - X_1(i)), \tag{6-11}$$

$$intercept(i) = Y_1(i) - X_1(i) \times slope(i), \tag{6-12}$$

and

$$\alpha(i) = \tan^{-1}(slope(i)). \tag{6-13}$$

The object of matching is to tell whether two line segments from a couple of palmprint images are the same in a palmprint. The two-dimensional right angle coordinate system established by the datum points does act as the important registration in line feature matching. For example, two line segments from two images can be represented as $(X_1(i), Y_1(i))$, $(X_2(i), Y_2(i))$ and $(X_1(j), Y_1(j))$, $(X_2(j), Y_2(j))$, respectively. And the Euclidean distances between the endpoints of two line segments are denoted as

$$\nabla_1 = \sqrt{((X_1(i) - X_1(j))^2 + ((Y_1(i) - Y_1(j))^2}\,, \tag{6-14}$$

$$\nabla_2 = \sqrt{((X_2(i) - X_2(j))^2 + ((Y_2(i) - Y_2(j))^2}\,. \tag{6-15}$$

Without question, the following conditions for line segment matching can be proposed: (1) If both ∇_1 and ∇_2 are less than some threshold D, then it clearly indicates that two line segments are same. (2) If the difference of angle of inclination (difference between two line segments) is less than some threshold β and that of intercepts is less than some threshold B, then it shows that two line segments have the equal angle of inclination and intercept. Within class of equal angle of inclination and equal intercept, if one of ∇_1 and ∇_2 is less than D, then two line segments clearly belong to the same one. (3) While two line segments overlap, they are regarded as one line segment if the midpoint of one line segment is between two endpoints of another line.

Verification Function

By applying the above rules of line feature matching to a couple of palmprint images, we can achieve the corresponding pairs of lines (shown as Figure 6-9). The verification function, r, is defined as

$$r = N / (N_1 + N_2 - N), \tag{6-16}$$

where N is the number of these corresponding pairs; N_1 and N_2 are the numbers of the line segments determined from two palmprint images, respectively. In principle, it shows that two images are from one palm if r is more than some threshold T which is between 0 and 1.

TEXTURE-BASED FEATURE

Palmprint identification is one of the biometrics computing approaches to recognize a person from a large group of people. In a palmprint identification system, a person is represented by a set of palmprint features and the recognition process is to compare the newly captured sample against all templates in the database. When the database becomes large, the one-by-one comparison is very time consuming which can hardly meet the requirement for real-time on-line identification. It is essential to develop an effective indexing scheme to lead to quick search and comparison. However, little has been done on the palmprint identification systems.

In order to index the palmprint database, a suitable feature should be found first. A feature that can be used as index should exhibit a large variance between individuals and small variance between samples from the same palm. It also should be compact which means small size and comparison effective which means easy to compare.

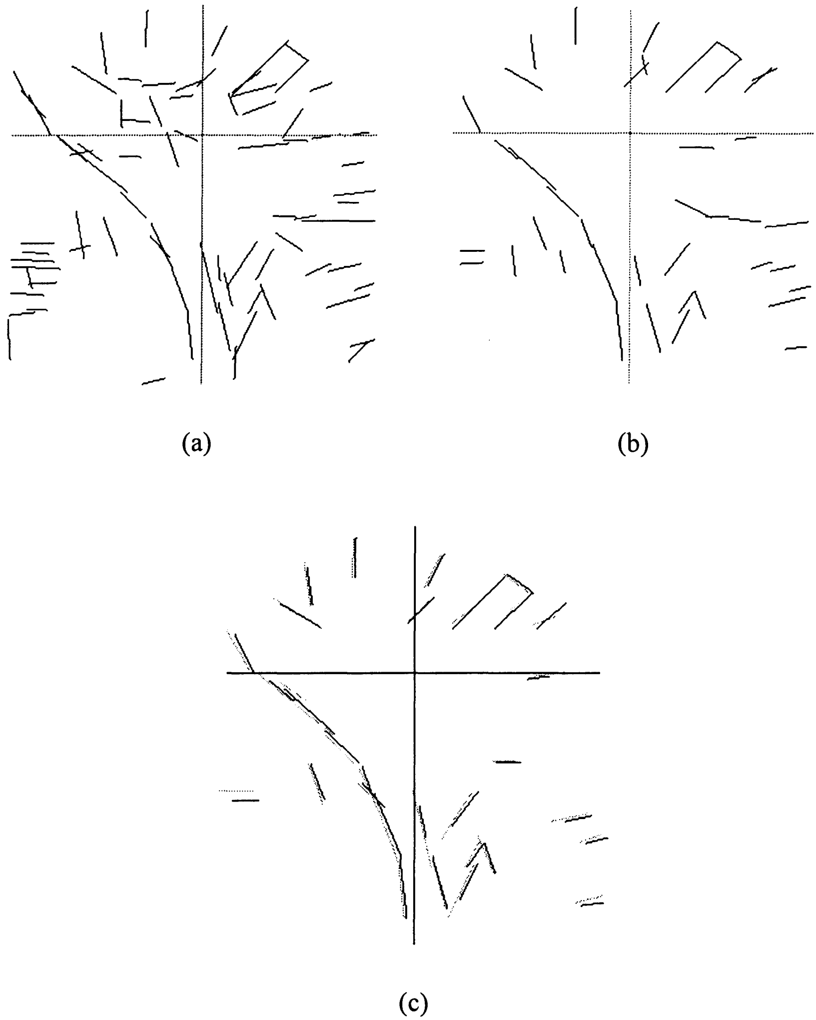


Figure 6-9. Results of a pair of palmprint matching by using line features: (a) Template linear feature set; (b) Input linear feature set; and (c) Matching result.

Unlike the existing techniques which are concentrate on local attributes such as points and lines, we proposed a texture-based feature extraction method which can describe the global attributes of a palm. These texture-based features can be represented by several numbers, therefore are compact, comparison effective and suitable for palmprint database indexing.

In terms of texture-based global features, some palms are so distinct that they can be easily discriminated from all others while some are so popular that they are similar to a set of palms. Considering of this characteristic, a dynamic selection scheme is introduced to ensure that the palmprint from the same palm won't be missed as well as the smallest candidate set will be given.

The following part of this section will explain the texture-based feature extraction and dynamic selection scheme in details as well as the experimental results are highlighted in the end.

Texture feature extraction

As mentioned before, a Palmprint consists of various lines with different widths and directions. Generally these lines can be classified into three categories, principle lines, wrinkles and ridges. Principle Lines are three strong major lines named heart line, head line and life line respectively. Wrinkles are thin and irregular lines and curves different from principle lines. Ridges are the tiny regular lines covered the palm skin surface.

The overall look of a palm does not change significantly with time. While palmprints from different palms may have totally different line patterns. Figure 6-10 shows four palmprints with different line patterns.

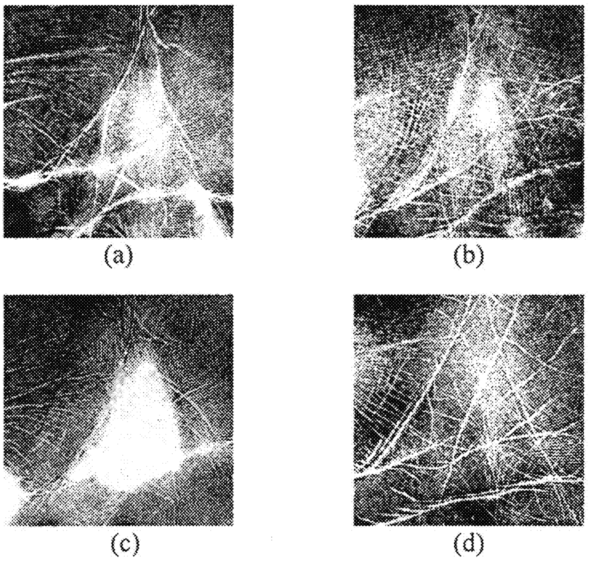


Figure 6-10. Palmprint images with different line patterns. (a) Strong principle lines, (b) Full of small wrinkles, (c) Less wrinkles, and (d) Strong wrinkles.

In order to tell how similar or how different a plamprint is from another, a reference system should be defined and the distance between two palmprints should be measured. Because a palmprint consist large amount of thin and short line segments represented in forms of wrinkles and ridges, it can be well characterized by texture. This section presents a new approach to represent palmprint features by texture measurement. The concept of texture energy is introduced to classify palmprint patterns. Historically, structural and statistical approaches have been adopted for texture feature extraction [16-18,22-24].

Global texture energy (GTE)

A palmprint can be marked as matrix $I_{n\times n}$ (here we suppose only a square area of the palmprint is considered) and a point at location (i,j) can be marked as $I(i,j)$. A small line pattern unit can be marked as a mask $A_{(2a+1)\times(2a+1)}$ and a point on the mask is $A(i,j)$. The following is the definition to global texture energy:

$$E_A(I)=\frac{1}{n\times n}\sum_{i=a+1}^{n-a}\sum_{j=a+1}^{n-a}F(i,j), \tag{6-17}$$

$$F(i,j)=\left|\sum_{k=-a}^{a}\sum_{l=-a}^{a}(I(i+k,j+l)\times A(k,l))\right|. \tag{6-18}$$

The larger $E_A(I)$ is, the more likely I is constructed by $A_{(2a+1)\times(2a+1)}$.

Considered of the line features on a plamprint, four masks named M_1,M_2,M_3,M_4 are chosen to calculate the global texture energy. Figure 6-11 shows the four masks. The four GTE values are named E_1,E_2,E_3,E_4 respectively. Because the four masks demonstrate the line directional features, the accordingly calculated GTE values may also reflect such features. Particularly E_1 reflects the horizontal line energy, E_2 reflects the vertical line energy, E_3 reflects the 135° line energy and E_4 reflects the 45° line energy.

-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)

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

(c)

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

(d)

Figure 6-11. Four masks used to calculate GTE
(a) Horisontal line, (b) Vertical line, (c) 135° angle line and (d) 45° angle line.

Characteristics of the GTE features

Using the above four masks, each palmprint may get four numerical global feature values. These global features have the following characteristics:

a) Insensitive to noise caused by small dirt dot on the palm.
b) Insensitive to shift.
c) Easy to compute.
d) Compact and comparison effective.

Dynamic Selection Scheme

Based on the global palmprint feature GTE, a fast and flexible selection algorithm will lead to a small set of the most similar candidates in the database. As a selection result, palmprints that are far from the being recognized one are excluded and a candidate list is generated. The selection algorithm first measured the distance between the being checked sample and each palmprints in the database using the GTE values. And then according to the popularity degree of the sample palmprint dynamically adjust the filter parameters. Because different filter parameters are used to different palmprint samples, the candidate list generated finally may contain different numbers of palmprints. The basic rule for filter parameter adjustment is to ensure the real wanted palmprint won't be missed and the candidate list should be as short as possible.

Definition and Notation

Suppose I_i and I_j represent two palmprints and their GTE features are $E_1(I_i)$, $E_2(I_i)$, $E_3(I_i)$, $E_4(I_i)$, $E_1(I_j)$, $E_2(I_j)$, $E_3(I_j)$, $E_4(I_j)$ respectively. The distance $D(I_i, I_j)$ is defined as follows:

$$D(I_i, I_j) = \sum_{k=1}^{4} | E_k(I_i) - E_k(I_j) | . \tag{6-19}$$

Owning to the shift, rotation, different pressure, and any other unpredictable variance of a palm at different capture time, palmprints captured from the same palm may vary from one another. But the overall look of a palm won't change significantly. So when given a palmprint, all samples from the same palm should be closed to the given one. This characteristic can be noted as follows: For a given I_A, $\forall I_i$ from the same palm as I_A, $D(I_A, I_i) < N$, N is a fixed value. The smaller N is, the better.

After calculated the GTE features to all palmprints in the database, the dispersion of global features and the average distance between two palms can be given. With any one GTE feature, the range that all palmprints fallen in can be noted as (A, B) and the average distance between two different palms is noted as M. Figure 6-12 shows the meanings of B, A, M and N.

Regards to fact that four GTE features are calculated, four groups of B, A, M, N will be calculated and they can be noted as (B_i, A_i, M_i, N_i) ($1 \le i \le 4$) respectively.

Of course, we hope that $|B-A|>>M>>N$. Also the global features of all palmprints from different palms are expected to disperse in the range (A, B) equally.

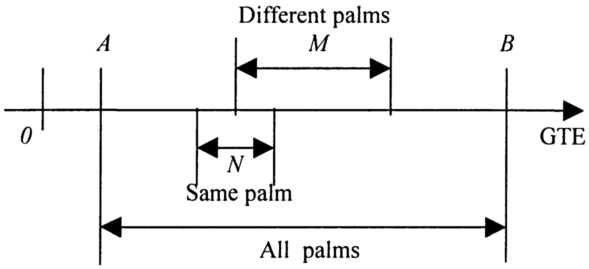


Figure 6-12. Meanings of A, B, M, N.

Implementation

In terms of global GTE feature, some palmprints are so distinct that they can be easily discriminated from others while some are so popular that they are similar to a group of palmprints. Our selection objective is to provide a candidate list with least amount while ensure the expected correct answer won't be missed. In order to achieve this goal, a dynamic selection scheme is developed. It is called dynamic because the selection rules are decided by the being recognized palmprint. The selection process is described as follows:

a) Compare $E_i(I_A)$ with A_i and B_i ($1 \le i \le 4$) to decide the important sequence of $E_i(I_A)$. The more closed to the boundary of range (A_i, B_i) $E_i(I_A)$ is, the better.
b) Let list R be the result candidate list and at first it includes all palmprints in the database. According to the significance of $E_i(I_A)$, perform four passes of filtering, at the end of each pass check the candidate number in R, if it is less than 10, Return R. The filter rule is as follows: if $|E_i(I_A) - E_i(I_x)| \rangle N_i$, delete I_x from R.
c) Adopt filter rule: if $D(I_A, I_x) > \sum_{k=1}^{4} N_k$, delete I_x from R, Return R.

Performance Analysis

The palmprint collection process is like this: first have the inked palmprint on a white paper and then scan it into the computer and perform a series of preprocessing such as histogram equalization to the scanned images. Only the central part of a palm is used to calculate the GTE features. Totally 100 palms are used and two palmprints from each palm are captured. For inspection to the compactness within one palm, 100 palmprints from one palm are captured. Four types of experiments are conducted.

The first is to test whether the GTE is insensitive to shift by shifting a palmprint in different degree. Table 6-1 shows that GTE is insensitive to tiny shift. The image size is 232×232 pixel. The number in the first row represents horizontal shift offset and the first column represents vertical shift offset. The value in the middle of the table represents the distance between the shifted image and the original image. The average distance between two palms is 51 units.

Table 6-1. Experimental results of insensitive to shift.

D	*-20*	*-10*	*0*	*10*	*20*
-20	4	3	3	1	2
-10	3	1	2	3	2
0	4	1	0	2	2
10	3	3	1	2	3
20	4	2	2	2	2

The second is to verify whether GTE is compact within one palm. The third is to inspect how well GTE disperses among all palmprints in the whole database. Figure 6-13 shows the GTE is compact within a palm and nearly equally dispersed in the whole database. Only one feature is showed, and other three are similar.

And the last one measures the efficiency of the texture-based dynamic selection scheme. Figure 6-14 shows the accuracy and efficiency of the texture-based dynamic selection scheme.

6.4 Palmprint Classification

Although the coarse classification does not identify a palmprint uniquely, it is helpful in determining when two palmprints do not match. There is only few research work of palmprint classification. Here, we are interested in the palmprint classification.

In general, the number of the principal lines is applied in the coarse-level palmprint classification. However, it is hard to detect some principal lines in inked palmprint image. As a result, the method by using the orientation property of ridges on the palmprint image is developed.

There are some classes of ridges, such as loop and whorl, in palmprint, but the techniques used in fingerprint classification [4,25-28] cannot be suitable for palmprint since there are always more than one ridge class in a palmprint image. One of the effective methods is classify them by the singular point number. So, the palmprint classification is changing into singular point determination.

SINGULAR POINT DETECTION

A point in the directional image is classified as an ordinary point, core or delta by computing the field the Poincaré index (PI) along a small closed curve around the point [14]. The PI is computed by summing up the changes in the direction angle around the curve. It is easily proved that the PI of delta point is positive and that of core point is negative when a full counter-clockwise turn is made around the curve in a directional image. So, the algorithm can be simplified as follows:

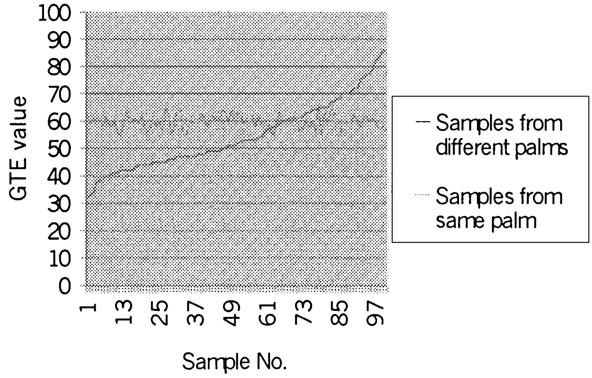


Figure 6-13. Dispersion of GTE feature.

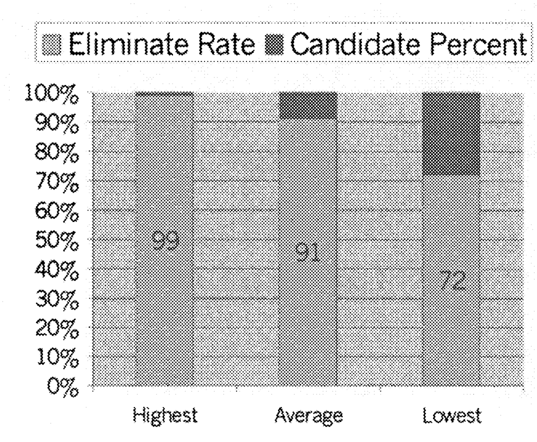


Figure 6-14. Elimination rate using texture-based dynamic selection scheme.

(1) Compute four-direction image. Four directions are vertical, horizontal, left diagonal and right diagonal. They represent 90, 0, 135 and 45 degrees, respectively. The method in reference [15] is used.
(2) Detect the singular point. In directional image, a pixel whose eight neighborhoods have more than two directions is a singular point. Then, according to their PI, core and delta points are determined.

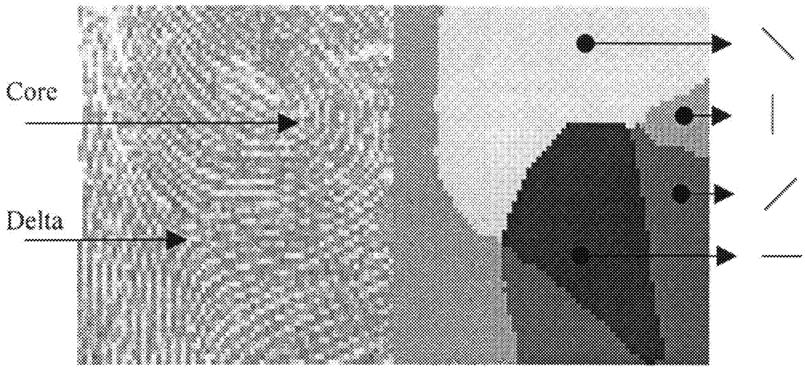


Figure 6-15. Singular points in both original image and directional image: (a) Original image; (b) Directional image.

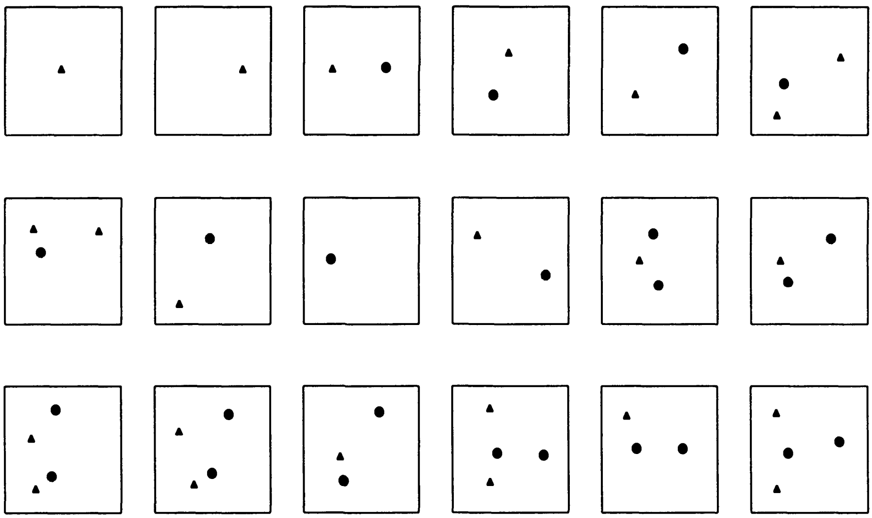


Figure 6-16. Typical singular points at outside region ("Δ" and "•" represent delta and core point, respectively).

(3) Eliminate the false singular points. If a pair of core points (or delta points) is too close, then they are considered as one core point (or delta point); if a core point nears a delta point, they are eliminated.

Figure 6-15 shows an original image and its directional image. The singular points can be easily found on it.

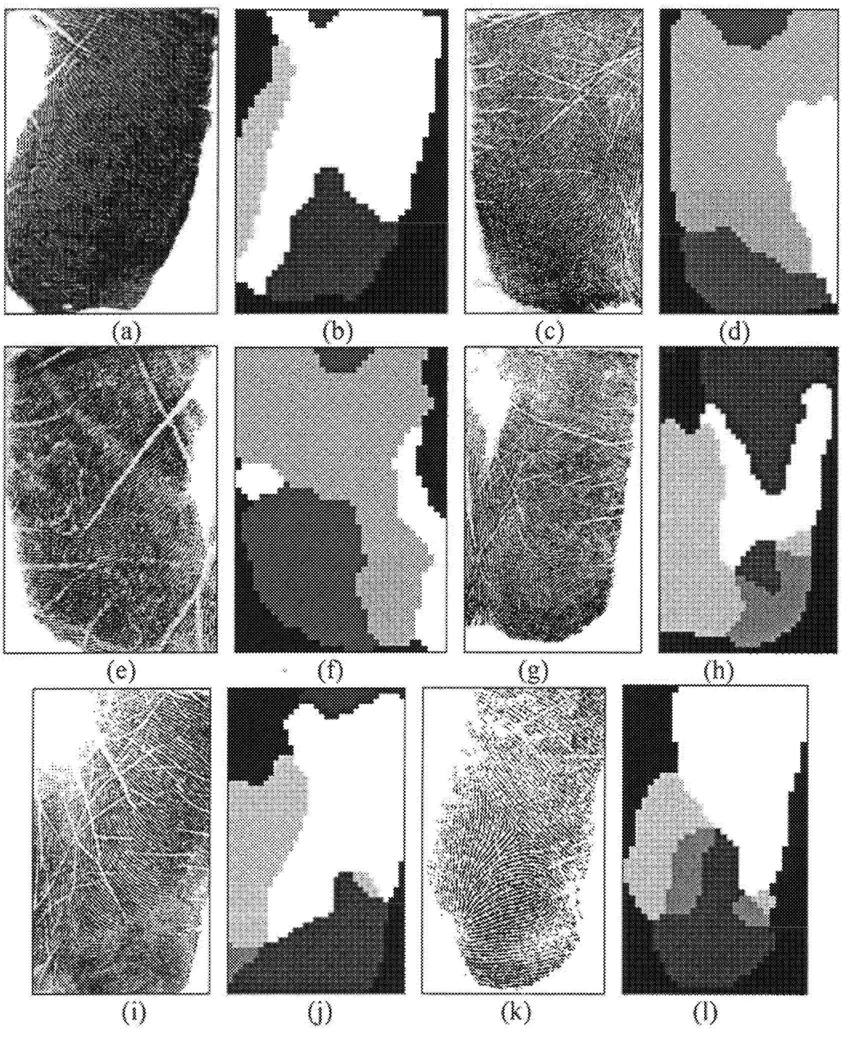


Figure 6-17. Six typical cases at outside region: (a), (c), (e), (g), (i) and (k) The original images; (b), (d), (f),(h), (j) and (l) Their directional images.

OUTSIDE REGION CLASSIFICATION

There are abundance shapes consisting of the singular points at the outside region. From hundreds of palmprint images, the typical cases of singular points at the outside region are shown in Figure 6-16.

As the first attempt of automatic palmprint classification, we only apply the singular points at the outside region. Obviously, we can classify the outside region as following six cases: (I) no singular point; (II) one delta point; (III) one core point;

(IV) one core point and delta point; (V) two delta points and one core point; (VI) others (see Figure 6-17).

6.5 Experimental Results

The algorithm of datum point determination has been tested by many 400 × 400 gray-scale inked palmprint images. Of the datum points calculated in the 300 palmprint images, 286 are found to be in excellent agreement with the manual estimate. Some typical palmprint images are shown in Figure 6-18 and the results of the datum points detection to these special images are listed in Table 6-2.

Table 6-2. Experimental results of datum points determination in special palmprint images.

Classification	*R*	*I*	*U*
Experiment images	20	35	5
Accurate determination images	19	33	4
The rate of accuracy (%)	95	94	80

Note: R = rotated image; I = incomplete image;
U = life line and head line unintersetion.

Table 6-3. Classification results in outside region.

Assigned class	*True class*					
	I	II	III	IV	V	VI
I	**132**	4	1	2	0	0
II	0	**149**	0	0	0	2
III	0	0	**15**	1	0	0
IV	0	0	0	**27**	1	0
V	0	0	0	0	**8**	0
VI	0	0	0	0	0	**12**

Next, the palmprint verification with both datum point invariance and line feature matching also has been tested by 200 palmprint images from 20 right palms. A slippery pair of statistics known as false rejection rate (FRR) and false acceptance rate (FAR) represents the measure of experimental result. In this experiment, some thresholds are adopted as follows: D = 5, β = 0.157 and B = 10. The results are shown in Figure 6-19 with the various thresholds T and the palmprint verification can obtain an ideal result while T is between 0.1 to 0.12.

In addition, 354 outside region palmprint images are tested. The resolution is 400 dpi, 256 gray levels. The classification results are shown in Table 6-3.

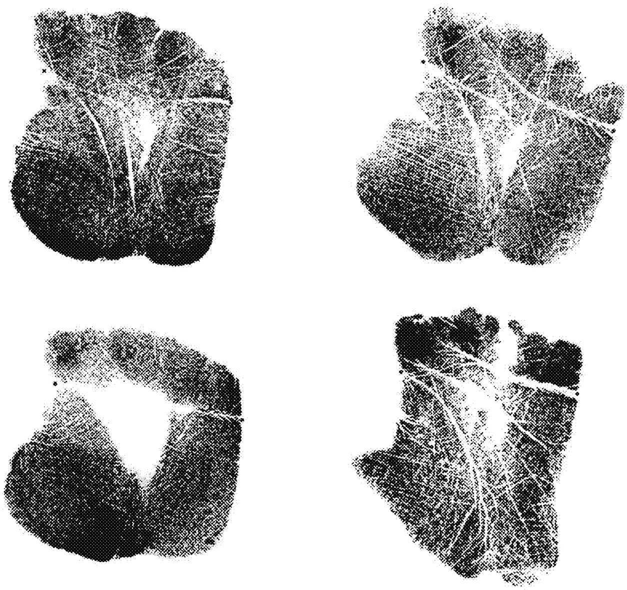


Figure 6-18. Examples of datum points determination.

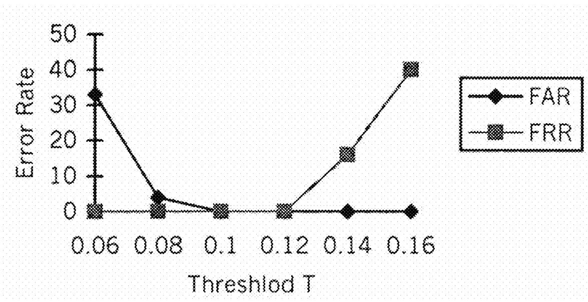


Figure 6-19. Experimental results for FAR and FRR ($D = 5$, $\beta = 0.157$, and $B = 10$).

In this chapter, we have implemented an automated biometrics-based verification by using both datum points and line features extracted from the palmprints to verify the identity of a live person.

The datum points are defined as the points of registration in palmprint matching because of their stability. The improved directional projection is designed to locate the principal line as well as to determine the endpoints accurately. A number of palmprint images have been tested and the experimental results show that most of datum points are in excellent agreement with the manual estimate. The computation of the basic algorithm is simple and logical.

The template algorithm of line feature extraction is presented due to its simplicity. The line features are considered as straight line segments and represented by their endpoints. Several rules are applied to match the line features so as to detect whether a couple of palmprints are from the same palm. 200 palmprint images from 20 persons are determined by both datum point invariance and line feature matching and the experimental result shows that the verification accuracy is acceptable. The palmprint verification proposed in this chapter is also foolproof because these significant physiological characteristics are unique, unchanging, and cannot be forged and transferred. As an important complement of automated biometrics-based authentication, the palmprint verification can be effectively used to identify individual.

References

[1] M. Eleccion, "Automatic Fingerprint Identification," *IEEE Spectrum*, vol. 10, no. 9, pp. 36-45, 1973.

[2] B. Miller, "Vital signs of identity," *IEEE Spectrum*, vol. 32, no. 2, pp. 22-30, 1994.

[3] D.R. Richards, "Rules of Thumb for Biometrics Systems," *Security Management*, vol. 39, no. 10, 1995, pp. 67-71.

[4] A. Jain, L. Hong and R. Bolle, "On-line Fingerprint Verification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 4, pp.302-313, 1997.

[5] L. Coetzee and E.C. Botha, "Fingerprint Recognition in Low Quality Images," *Pattern Recognition*, vol. 26, no. 10, pp. 1441-1460, 1993.

[6] Grolier Incorporated, *The Encyclopedia American*. Grolier, USA, 1995.

[7] W. Shu and D. Zhang, "Automated Personal Identification by Palmprint," *Optical Engineering*, Vol. 37, No 8, 1998, pp. 2359-2362.

[8] V.S. Srinivasan and N.N. Murthy, "Detection of Singular Points in Fingerprint Images," *Pattern Recognition*, vol. 25, no. 2, pp. 139-153, 1992.

[9] K. Karu and A.K. Jain,"Fingerprint Classification,"*Pattern Recognition*, vol. 29, no. 3, pp. 389-404, 1996.

[10] J.H. Mcintish and K.M. Mutch, "Matching Straight Lines," *Computer Vision, Graphics, and Image Processing*, vol. 43, pp. 386-408, 1988.

[11] T. Pavlidis, *Algorithms for Graphics and Image Processing*, Rockville, Md.: Computer Science Press, Inc., 1982.

[12] D. Zhang and W. Shu, "Two Novel Characteristics in Palmprint Verification: Datum Point Invariance and Line Feature Matching," *Pattern Recognition*, Vol. 32, No 4, pp. 691-702, 1999.

[13] P.S. Wu and M. Li, "Pyramid Edge Detection Based on Stack Filter," *Pattern Recognition Letter*, Vol. 18, No 4, pp. 239-248, 1997.

[14] J.F. Keegan, "How Can You Tell If Two Line Drawings are the Same?," *Computer Graphics and Image Processing*, vol. 6, no. 1, pp. 90-92, 1977.

[15] M. Kawagoe and A. Tojo, "Fingerprint Pattern Classification," *Pattern Recognition*, vol. 17, no. 3, pp.295-303, 1984.

[16] R. M. Haralick, "Statistical and Structural Approaches to Texture," *Proc. IEEE,* vol. 67, pp. 786-804, 1979.

[17] S. Peleg, J, Naor, R. Hartley and D. Avnir, "Multiple Resolution Texture Analysis and Classification", *IEEE Trans. PAMI,* vol. 6, pp. 518-527, 1984.

[18] H. Wechsler and T. Citron, "Feature Extraction for Texture Classification," *Pattern Recognition* vol. 12, pp. 301-311, 1980.

[19] P.W. Verbeek, L.J. van Vliet, "Line and Edge Detection by Symmetry Filters," in *Proceedings of* 11th IAPR International Conference on *Pattern Rec*, vol. III, pp. 749-753, 1992.

[20] A.R. Circuits and M.B. Sanbdler, "General Test Framework for Strait-line Detection by Hough Transforms," *1993 IEEE International Symposium on Circuits and Systems, ISCAS '93,* vol. 1, pp. 239-242, 1993.

[21] T.S. Chan and R.K.K Yip, "Line Detection Algorithm," in *Proceedings of the 13th International Conference on Pattern Recognition*, vol. 2 , pp. 126-130, 1996.

[22] A.E. Svolos, C.A. Hutton and A.T. Pokropek, "Co-occurrence Trees: A Dynamic Solution for Texture Feature Extraction," *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine., 18th Annual International Conference of the IEEE*, vol. 3, pp. 1142-1144, 1997.

[23] T.N. Tan and A.G. Constantinides, "Texture Analysis Based on a Human Visual Model," *in Proceedings of 1990 International Conference on Acoustics, Speech, and Signal Processing, ICASSP-90*, vol. 4, pp. 3137-2140, 1990.

[24] S.W.C. Lam, "Texture Feature Extraction Using Gray Level Gradient Based Co-occurrence Matrices," in *Proceedings of IEEE International Conferencr on Systems, Man and Cybernetics*, vol. 1, pp. 267-271, 1996.

[25] A. Wahab, S.H. Chin, E.C. Tan, "Novel Approach to Automated Fingerprint Recognition," *IEE Proc., Vision, Image Signal Proc-ess*, vol. 145, no. 3, June 1998.

[26] B. Moayer, and K.S. Fu, "A Tree System Approach for Fingerprint Pattern Recognition," *IEEE Trans., PAMI*, vol. 8, no. 3, pp. 376-387, 1986.

[27] Q. Xio and H. Raafat, "A Combined Statistical and Structural Approach for Fingerprint Image Postprocessing," *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, Los Angeles, CA, USA, pp. 331-335, Nov. 1990.

[28] D. Maio and D. Maltoni, "A Structural Approach to Fingerprint Classification," in *Proc. 13th ICPR*, Vienna, Austria, pp. 132-138, 1996.

[29] K. Ishibuchi, H. Takemura and F. Kishino, "Real Time Hand Shape Recognition for Man-machine Interfaces," in *Proceedings of IEEE International Conference on System, Man and Cybernetics*, vol. 2, pp. 1407-1412, 1992.

[30] K. Ishibuchi, H. Takemura and F. Kishino, "Real Time Hand Shape Recognition Using Pipe-line Image Processor," in *Proceedings of IEEE International Workshop on Robot and Human Communication*, pp. 111-116, 1992.

7 FACE RECOGNITION

In this chapter, we first introduce the face recognition system so as to give an outline of what face recognition system does and its structure. Then, two key steps in face recognition system, including detection / location of faces and feature extraction, are reviewed in detail in Section 7.2 and Section 7.3, respectively. In Section 7.4, a dual eigenspaces method for face recognition is given. At the end of this chapter, we discuss what we need to do in future.

7.1 Introduction to Face Recognition System

Face recognition from images has lower uniqueness than fingerprint recognition and iris recognition, while it provides a more direct, friendly and convenient identification way and is more acceptable compared with individual identification ways by other biometrics features. Thus the research on face recognition has become one of the most important parts in biometrics.

A lot of research works have been done on face recognition since 1960's, and many fruitful results have been obtained. One reason for which face recognition gets extensive research is that some practical applications are urgently required. For example, clerks of bank want to know whether a customer is using the right account. Policemen try to find whether a man in photo is a recidivist, or whether he is being pursuing. Custom officers wonder if a person passing the custom takes his own passport. Even in our daily life, we wonder who enter the office and building everyday, and whether they are permitted. All these practical and potential applications are driving the research on face recognition.

Face recognition system contains two key steps (see Figure 7-1), which are face detection and location together with features extraction and face recognition. The first step decides whether the input images or image sequences include faces, and if they do, figure out the position of the faces, then segments each face from background. The second step looks for face features which distinguish individuals, and judges whether the people in image is the given person or whether he or she is in database.

Face detection and location depend on the input type of a system very much. It is easy to detect and locate faces in some applications, such as login system on computers and passing custom, because the images or image sequences have uniform background. As a result, the orientation and pose of faces are well controlled and known. In these cases, we can usually acquire high detecting and locating accuracy easily.

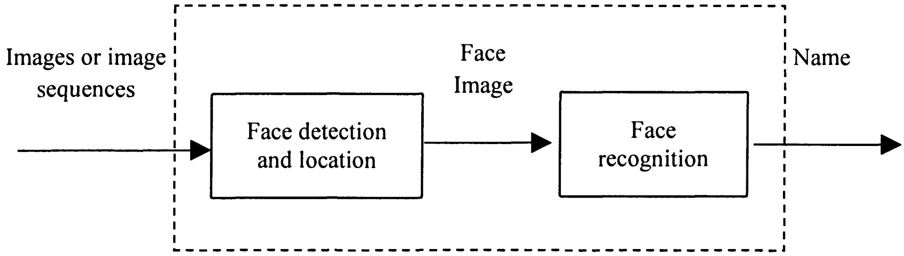


Figure 7-1. Flowchart of face recognition system.

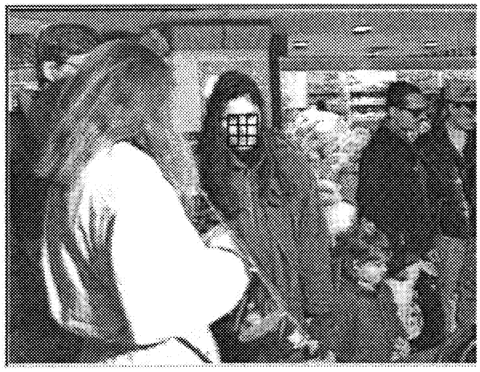


Figure 7-2. An example of the image with a number of faces, each with different size, location, orientation, expressing and luminance.

While in other universal applications, the input images perhaps contain a number of faces and a complex background. We could not know their sizes and location. Maybe the luminance on faces varies with illumination, even orientation and expressing could be various. In these cases, face detection and location are very difficult. Such an example is shown in Figure 7-2.

Both features extraction and face recognition are also various according to different applications. Sometimes, we want to know whether the person is whom he has declared, such as depositing in banks, which is called face verification. While in other conditions, we wonder which person in database the man is. In other words, what his name is, such as surveillance system in office entrance. We call it as face recognition. Sometimes we only want the system to give a candidate list sorted by the similarity to given people. Police department often uses this case.

Since various expressing, different orientation and age also can generate large differences between face in database and face being recognized, it is very difficult to extract these features and recognize such faces.

7.2 Detection and Location of Faces

The basic idea of detecting faces is to model face and compare every detected area with the face model, then calculate their similarity with the model to decide if there is such a face. We need to figure out its position in the image. According to the methods of modeling, face detection and location can be classified into two different types, statistics-based method and knowledge-based method.

STATISTICS-BASED METHOD

Subspace Method

The purpose of detecting faces is to check faces in images according to the common features owned by faces. So what features are the common features of faces? That is a problem we focused on.

Each face image is considered as a higher dimensional vector and each pixel corresponds to a component. If all the face images lie in the same subspace of the higher dimensional space, this subspace is a good representation of face images because it shows the common features of faces. So what detection of faces does is to find the subspace.

Pentland used Karhumen_Loève transformation into face detection and recognition, and made it one of the subspace methods for detecting faces [1]. This method is described as follows:

Suppose $A=[a_{ij}]_{r\times c}$ as an image, where r and c are the number of rows and columns of the images, respectively; a_{ij} is the gray value of the pixel in i th row and j th column. We re-arrange a_{ij} and make it a column vector:

$$X=\begin{bmatrix}a_{11} & a_{21} & \cdots & a_{r1} & a_{12} & a_{22} & \cdots & a_{r2} & \cdots & a_{1c} & a_{2c} & \cdots & a_{rc}\end{bmatrix}^T,$$

where X is a D dimension vector, $D=r\times c$.

One face image can be considered as a stochastic sample. Thus, giving a group of face image specimens, $G=\{X_0, X_1, \cdots, X_{M-1}\}$, the covariance matrix can be calculated as:

$$\Sigma=\frac{1}{M}\sum_{i=0}^{M-1}(X_i-m)(X_i-m)^T, \tag{7-1}$$

where m is the average vector of the training specimens set, and M is the number of images in the training specimens set,

$$m=\frac{1}{M}\sum_{i=0}^{M-1}X_i. \tag{7-2}$$

Let $\lambda_1, \lambda_2, \cdots, \lambda_d$ ($\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_d > 0$) and $u_1, u_2, \cdots, u_d$ be eigenvalues and corresponding eigenvectors of covariance matrix Σ, respectively. So every face image, X_i, can be represented by the liner combination of the eigenvectors.

According to the algebra theory, we know that $u_1, u_2, \cdots, u_d$ will be orthogonal one another and unit vector. Usually, $M < D$ can be satisfied because D is larger than the number of the specimens. Then $d < D$ is derived. In other words, the given face image can be represented by fewer base vectors (d vectors).

Some values, λ_i, in d eigenvalues are very small, whose corresponding eigenvectors give little contribution to represent the face image specimens, hence they can be ignored. Thus, we sort the eigenvectors according to the decreasing eigenvalues, and select the top k eigenvectors to represent the specimens. The space, S', spanned by them is a subspace of the D dimension space, S.

How to select k? We can choose k as a very big number, for example, $k = d$. But we know some eigenvectors have little contribution to face space. On the contrary, if we select k as a very small number (for example, $k = 1$), the subspace is not sufficient to represent the face image specimens. Usually, we can select the smallest k among which satisfies the expression,

$$\frac{\sum_{i=0}^{k} \lambda_i}{\sum_{i=0}^{M-1} \lambda_i} \geq \alpha, \tag{7-3}$$

where α may be a real number which is close to 1, such as 99%. It states that the top k axes possess 99% energy of all axes'.

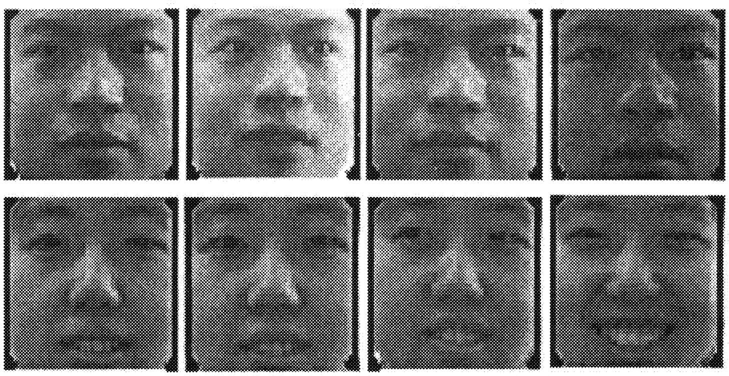


Figure 7-3. Some face images in training set.

Figure 7-3 shows some face images in the given training specimen set collected in the face recognition research. The training images are normalized as described below and used for calculating the eigenvectors and eigenvalues. Because every eigenvector represents a point of the D dimension space, they can be reverted or rebuilt into an image. Figure 7-4 shows the images rebuilt from the top 8 eigenvectors. Since these images are like faces, they are called "eigenfaces".

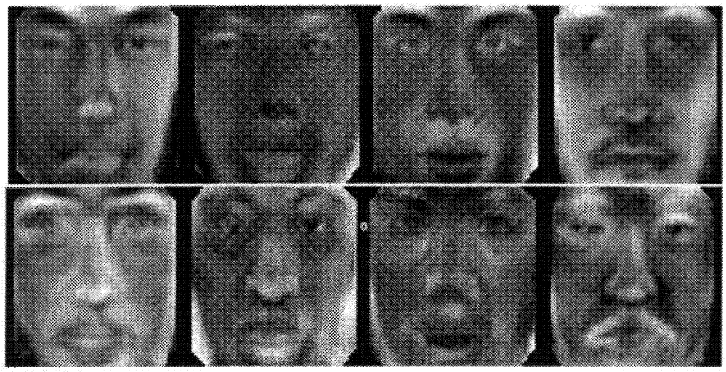


Figure 7-4. Several eigenfaces by Karhumen_Loève transformation based on the samples of the training set.

Every point of $S^{'}$ can be rebuilt to an image. The left figure of Figure 7-5(a) shows a face image, and the right one gives the projected and rebuilt image. We know from Figure 7-5(a) that $S^{'}$ represents face images inherently. It is interesting that Figures 7-5(b)-(f) give some special cases of the images and their projected and rebuilt images, which show that $S^{'}$ represents the common features of faces, and it can associate the original images itself to some extents. While the orthogonal supplemental space of $S^{'}$ shows more noise of training images, so it is useless to represent images.

Since the subspace, $S^{'}$, is a face space, its orthogonal supplemental space has little thing to do with the expressing face. Project a given image, X, to the orthogonal supplemental space of S, then we can get $X' = X - \overline{X}$, where $\overline{X}$ is the projection of vector, X, to space, $S^{'}$. If X' holds much energy, it indicates that X is not a face image; and if X' holds little energy, we can define X as a face image.

Apparently, the selection of *k* affects the accuracy of face detection directly. Figure 7-6 shows the experiment results of the relation between *k* and the accuracy.

Detecting faces with Karhumen_Loève transformation is an image-gray-value-based method, so all the training images must be normalized to make its subspace available. Usually, normalization includes scaling, setting a special point as base point (such as eyes), and rotating face image. Another important work is to normalize the gray values of images. If the training images and the test images are grabbed in different lighting condition, the face can not be directly detected by this subspace-based method. For example, as to the training image, the top right areas are bright, and it is dark in bottom left area. But for the test image, the bottom left area is bright, and the top right area is dark. In order to detect face in image by the subspace-based method, we have to normalize the image gray values. In general, we can normalize the gray values by some effective approaches, such as global gray value scaling, local gray value scaling and brightness compensation. It is shown that local gray value scaling can get better results than global gray value scaling. Different with gray value scaling method, brightness compensation analyzes the distribution of gray value and tries to restore the distribution of gray value of the original face image in the normal dispersion light. In the summary, it is necessary and important to normalize the images before detecting faces with Karhumen_Loève transformation.

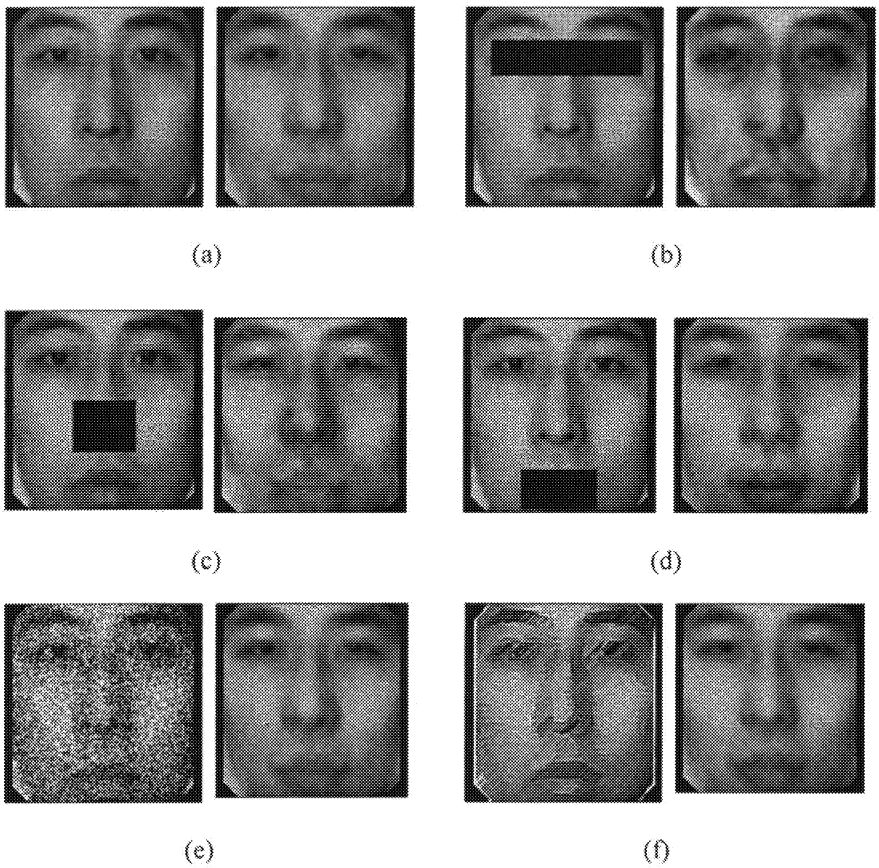


Figure 7-5. Some original images and their rebuilt images:
(a) Normal image, (b) Lose eyes, (c) Lose nose, (d) Lose mouth,
(e) Add uniform white noise, and (f) Sketch effect image.

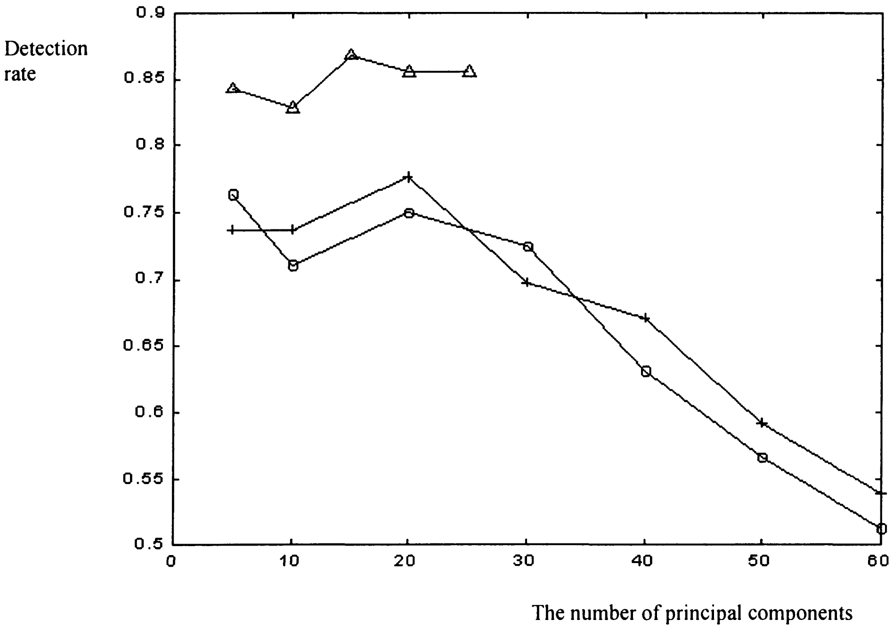


Figure 7-6. The relation diagram between *k* and detection accuracy. The lines with triangles, cross and circle characters are gotten with the methods in [1], [2] and [3], respectively.

Neural Networks Method

Detecting faces with neural network method is considered as a two-class classification problem: face class and non-face class. For example, K. Sung and T. Poggio collected 4150 face image specimens and 6189 non-face image specimens and studied them with elliptical k-means clustering algorithm [4,5]. There were six clusters obtained in either class, which were considered distribution-based face model. Then they trained a forward multi-layer neural network, whose inputs are the distances from specimens to each cluster. At last, the images were classified with this trained network, and the question whether there were faces was answered. In H. Rowley's work, great deals of face and non-face images were sent to a forward multi-layer neural network to be trained directly. Analysis to overlapped areas was necessary in detecting process, which helped to get the result [6].

Many other useful researches have been done in this field. With neural network method, we need enough face and non-face specimens provided to train neural network. That is a problem because there are many kinds of non-face images we can not collect all these non-face specimens. When there are lots of specimens or input nodes, it is difficult for the neural network to reach its global minimum point and also the rate of convergence slows down, which is the inherent defect of neural network.

Neural network method used for detecting and recognizing faces has its limitation. To get better effect, neural network method is often combined with other methods, which is worthy of researching [5].

KNOWLEDGE-BASED METHOD

Distribution Ruler of Gray-Value-Based

Although there are great differences among individual's appearances, there are some nearly universal distribution rulers of gray values in normal light condition. For example, the gray values of eyes' area are always lower than that of forehead and zygoma. Detecting faces using these rulers is efficient usually.

Mosaic Method [7] is a representative method. Even if the face area in an image is very small or the image is dim, human can easily tell whether there exist faces. Inspired with human experience of recognizing faces, Mosaic method divides the given image areas into image blocks of 4×4. These image blocks lie in eyes, mouth and cheeks, etc. If the given image area is a face area, the image blocks should satisfy some distribution rulers of gray values. Then we can get some candidate areas of faces according to these rulers. In other words, the rulers are set loose and only images that are definitely not faces according to the rulers are denied. After that it divides the rest image areas further into image blocks of 8×8, selects and decides the candidate areas as before. Finally, this method locates the position of eyes, mouth, etc. with edge detection algorithm in face areas.

Other blocking partition methods are used to make Mosaic method more efficient. For example, they divide the given image areas into blocks of 3×3 instead of 4×4, and design new distribution rulers of gray value again. Also other algorithms for calculating the gradient or edges of the image are used for the given images, and then apply the new Mosaic rulers to the edge or gradient image. This is because that the distribution rulers are sensitive to light condition of images.

Of course, we know that when detecting faces with Mosaic method, all above three detecting steps should be executed for image areas of various positions, sizes and directions in the given image. So this method is very slow. It is a good idea to combine Mosaic method with other fast scan methods. Some faster scan can be used to filter every detected area firstly. Then Mosaic method is adopted for the remaining fewer image areas. Generally, the combined method is more effective.

The integral projection is a simple and quick scan method. It needs to project and integrate the given gray images or gradient images or edges images in some directions. We know the projection values will obey some rulers because of the distribution of face parts. As a result, we can analyze the projection values and filter those areas that are not faces.

In fact, Mosaic method has considered a geometrical relationship between face parts. Another method, which has a bit similarity with Mosaic method, is also one of bottom-to-up methods, i.e. it detects face parts such as eyes, nose and mouth with other methods firstly. Then the scan rulers are drawn according to the geometrical relationship between face parts to decide whether there are faces.

Contour Ruler

Contour is an important feature of face. If the face contour can be extracted correctly and exactly, face detection will be more accurate. But we are not sure we can often detect and extract correct edges. Since the difficulty in image processing, it is not easy to extract face contour perfectly using existing edge detection algorithms. However, we can use face contour as an important feature in other ways.

The face contour can be usually modeled as ellipse [8,9]. Because the same faces can not be approximated well by an ellipse, one improved method models the face contour as two straight lines (cheek) and two arcs of ellipse [10,11]. In these methods, some edge detection algorithms are used to extract edges of face image. Then Hough transformation or size changeable ellipse mask is used to get the final face contour. Another kind of methods uses snake techniques to get the face contour [12-15]. In a snakes method for face contour extraction, an energy function of the given contour is established to present the energy of the contour. The energy function is the weighted sum of three components: the internal energy of the contour, the image force, and the external constraint forces. Then some optimal algorithms are used to minimize the energy function. The contour with the minimum of energy is the final result. When we use a snake method, in general, the real contour should be estimated roughly by other methods or prior knowledge in advance. Then we use the estimated contour as the initial contour. Another limitation of a snake method is that the final result may be the false edge, caused by shadows, beard and moustache.

Color Information

Although skin colors are various for individuals, they are different to background color often. Especially the distribution of the face colors in the same race is similar and the pixels in face areas are clustered in a small area in a color space. So color information is very helpful for detecting faces effectively [16].

Many colors in normal light condition can be clearly distinguished and easily by simple algorithms, but in very bright or very dark light condition, it is hard to distinguish different colors, even the colors are very different from people. So in some applications, if we can control the background of face images, design special background and control illumination, detecting and locating faces will become much easier.

Movement Information

In some applications, sequence images can be provided for detecting and recognizing faces. If persons are moving relative to background, faces can be segmented from background according to movement information easily and efficiently, vice verse [2, 16]. But if there are several faces moving towards different directions with different speed before vidicon and some faces overlaps each other, detecting faces with movement information will be not easy to do.

Symmetry Information

Face is symmetrical to some extent, which can be applied to detect faces. Before detecting, finding the symmetry of objects in images is needed.

Zabrodshky puts forward continuous symmetry detecting method that detects the symmetry of a circle area to decide whether it is a face [17]. Reisfeld adopts generalized symmetry transformation method that detects local strongly symmetrical points to locate face parts [18]. The other researchers define direction symmetry transformation, which considers symmetry in different directions [19,20]. It can not only find strongly symmetrical points, but also describe shape information of strongly symmetrical objects, which is helpful to locate faces.

Detecting faces in a complex background is usually difficult. It seldom attains good result with only one method above, so combining kinds of methods will be effective. If the background of imaging or environment surrounding a face recognition application can be well controlled, face detection and location will be much easier.

7.3 Features Extraction and Face Recognition

Face detection tries to find common features of faces so as to achieve the purpose of detecting faces. On the contrary, face recognition tries to find the differences of individuals' images to identify each person. But what are the differences to recognize individuals? Which parameters indicate the differences? How to attain these parameters? Are they reliable and robust? How to recognize persons with them? These are the problems that we focus on at the features extraction and face recognition step.

GRAY-VALUE-BASED METHOD

We describe above how to get a set of principal components of faces with Karhumen_Loève transformation in former section. Because there is enough information in the subspace spanned by this group of principal components, we can use them as features to recognize faces.

After principal components are gained by training step, each image in a training set is projected to eigenface subspace. Each training image can be represented by a k-dimension vector. In recognizing step, project a test images to the space and gain k-dimension vector, then compare it with training image vectors. The comparison can be done with nearest neighbor method, k-nearest neighbor method, or neighbor method with rejection to decide the class label that the image belongs to. Which method should be selected depends on the number of training images and whether the test image is in the database.

Although it is reported in the articles [3,21] that they have gained perfect result of 95% recognizing rate in 3000 images from 200 persons by eigenface method, this method has its defection. Before training and testing, the size of face images should be normalized and images should be rotated to right direction. And gray values of images should also be normalized. It is a bit of difficult to normalize the gray values. In addition, noise in training images does some effect to recognition. So not every eigenface we got is useful to recognize faces, especially the top a few eigenvectors are not always good features for recognizing individuals. How to select components of the k-dimension vectors or which features have much help to recognize individuals? That has become a problem to be solved in some research works, such as dual eigenspace method by Peng [22] and linear discriminating analysis method by Weng [23,24]. And Fisher face method in [23,24,25] is good and worth to be described here.

Liner discriminating analysis based on Fisher criterion is an important method in statistics pattern recognition. It tries to find a group of features by liner transformation to maximize the Fisher criterion,

$$J_F(\omega) = \frac{\left\| S_b^{'} \right\|}{\left\| S_w^{'} \right\|}, \tag{7-4}$$

where ω is the liner transformation to be found; $S_b^{'}$ and $S_w^{'}$ denote total between-class scatter matrix and within-class scatter matrix of training specimens after liner transformation, respectively. According to pattern recognition theories, if S_b and S_w denote between-class and within-class scatter matrix of original training specimens, the features to be searched will be the eigenvectors corresponded with the biggest eigenvalues of $S_w^{-1}S_b$

After transforming an image to a vector with subspace method discussed in former section, then we can get a group of feature vectors with Fisher criterion method. Each feature vector represents a feature with strong classifying capability. The linear space spanned by the feature vectors forms a feature subspace used for recognizing faces. Each feature vector can also be rebuilt to an image, and the rebuilt images are face images similar to eigenfaces, which are called Fisher faces. The method recognizing faces with Fisher criterion above is called Fisher faces method.

Having gained the subspace of Fisher face, it can project every training specimen to the subspace and use the projected coefficients as a group of features. Then the test specimen is projected to the subspace to gain another group of features. Finally a class label, which the test image belongs to, is determined by nearest neighbor method or other methods.

While executing Fisher faces method, the number of training specimens is often less than the dimension of image, so the within-class scatter matrix, S_w, is invertible. This problem can be solved with a double discriminating subspace method: do Karhumen_Loève transformation using the training specimens, so a group of feature vectors can be obtained. Then Fisher criterion is used to the projected features of training specimens, and a smaller group of feature vectors will be collected, which can be called the Most Discriminating Features (MDF).

It is helpful for our comprehending the Fisher faces method to compare eigenface method with Fisher face method. After Karhumen_Loève transformation, each eigenvalue shows the variance of all the specimens in the direction of its corresponding eigenvector. If the distribution of the specimens in some eigenvector, u_i, is leaded by noise and other factors, the eigenvalue in the direction is large. But the eigenvector, u_i, doesn't have any contribution to classification. But in Fisher face method, we want to find the features to make between-class distance big and within-class distance small. Since the direction u_i has no any contribution to classification, u_i vector can't be Fisher face or MDF. The experimental results we got support the explanation. If some of the training images are in such light condition that all the light comes from right, one of the eigenface will show the effect. But it doesn't do any effect to Fisher face. Also it tells us that eigenvectors with big eigenvalues in eigenface method may not have strong classification capability, and more features of eigenface may not improve recognition capacity efficiently.

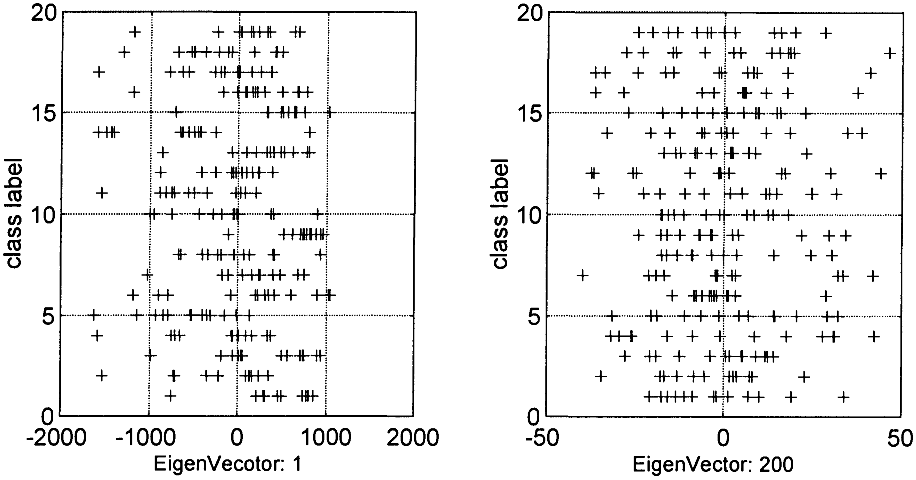
Theoretical analysis tells that Fisher face method has better recognition capacity than eigenface method. Experimental results support the conclusion. We have done experiments in a small database of 19 persons and 13 images each person, which include front-view face images of different light condition, different pose, different expression and various distance from camera. And all the images have been normalized. Recognition capacities of single feature with eigenface method and Fisher face method have been showed in Figures 7-7 and 7-8. The vertical axes show class labels of 19 classes, while horizontal axes show projection of specimens to one eigenvector. Every cross character in figures presents a specimen. We know from these figures that recognition capacity of single feature with eigenface method is worse in spite of its principal components or minor components. But recognition capacity with Fisher face method is good. In this training specimen set, recognition rate can reach 100% only using the former two MDFs. Further more, Fisher face method is better than eigenface method even in other bigger database [25].

Other methods have been derived from eigenface approach. For example, view-based eigenface [26] builds subspace for every group of specimens grabbed from one imaging angle. So they got as many subspaces as the number of possible imaging angles. Feature part method [26] builds subspaces of eyes, nose and mouth, respectively. Frequency domain eigenface [27] gets principal components from two-dimension spectrum of gray value image. Atick [28] uses eigenface method in three-dimension range images of faces and gets three-dimension eigenfaces.

GEOMETRY-FEATURE-BASED METHOD

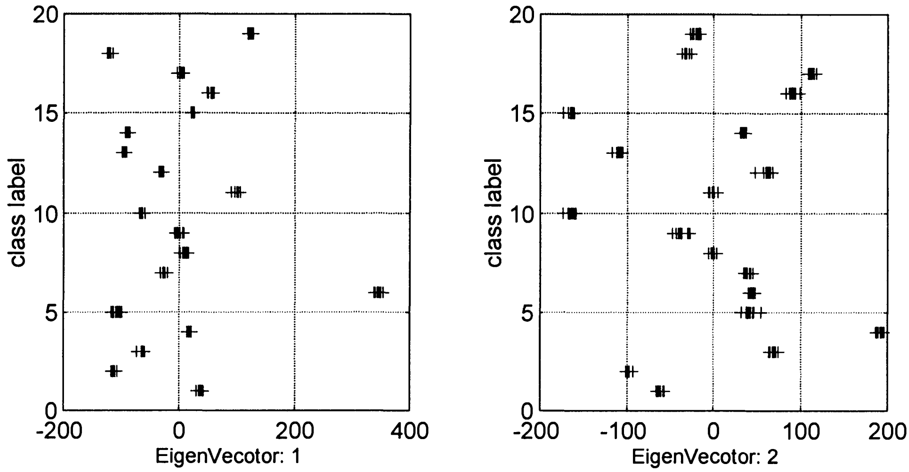
Position relationship between face parts, such as eyes, nose, mouth and chin, their shapes and sizes have strong contribution to classify faces. Extracting geometry features like shapes and construction of face parts to recognize faces are very different from subspace method.

What features should be derived? How to get these features and how about their classification capacity to faces are problems encountered in research of this kind of methods. In early research period, researchers mark face feature points such as eyes, nose and mouth manually, calculate all kinds of geometry features, such as distances, separation angles and areas of feature points, then study and observe the recognition rate according to these deduced features. Since then, more research works focused on automatically deriving geometry features. These works include automatically deriving some important features like eyes, mouth and nose by Kanade [29], geometry features of eyes by Nixson [30], figure shape of face by Young [31], and other 35 features such as width and length of nose, position of mouth and shape of chin, etc. [32]. Intuitively we know these geometry features are very useful for face recognition. It is also reported in article [29,33] that the recognition rate with some geometry features can reach 45%-95% in a small database. But it is more important that we can not calculate these geometry features accurately for the sake of well-known difficulty in image processing, which effects the recognition capacity directly.



(a) The first principal component (b) Some components

Figure 7-7. Recognition capacity of single feature with eigenface method.



(a) The first eigenvector (b) The second eigenvector

Figure 7-8. Recognition capacity of single feature with Fisher face method.

Deformation template method put forward by Yuille [33] can also be considered as a geometry features method, which tries to describe the shape of face more subtly. They designed a parameter adjustable model for every face part (deformation template, see Figure 7-9), and defined an energy function whose variables are peak-hollow gray value images and edge images of face parts together with model parameters. Then adjusted model parameters continuously to minimize the energy function. The model parameters gained ultimately are the geometry features of the face part. Based on the basic idea of template method mentioned above, its defect lies in the coefficients of energy function can only be defined by experience and can't be generalized. In addition, the process in optimizing progress of energy function is slow and the result may probably sink into local minimum.

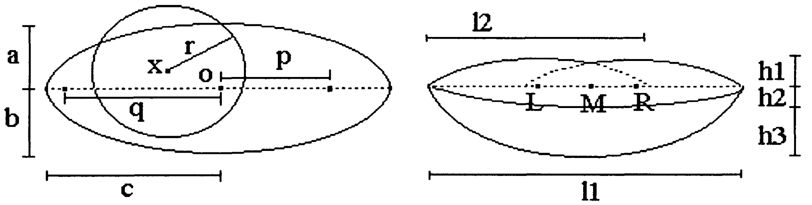


Figure 7-9. Models for face parts in deformation template method.

By previous research, it is difficult for the geometry-features-based method to reach higher recognition rate. So it is often combined with other features to get better results. For example, geometry features can be used to filter before using other recognition methods.

DEFORMATION MODELS

Lades presents dynamic link architecture model to recognize the distortion invariant object [34,35]. The distortion invariant object in the method can be expressed by a sparse graph whose vertices can be marked with multi-resolution description of local energy spectrum, and whose edges show topological relation between vertices, and edges have distance property. When using this method to recognize faces, a face in normal condition can be expressed by a uniform image. As sizes, angles or expression of faces vary, the face image also will be changed. So face recognition is transformed as a graphic matching problem. Elastic graphic matching algorithm is used to recognize faces in the method. 87 persons' images with various expression and different visual angle were used in Lades's experiments. And it took 25 second for a parallel computer composed of 23 transputers to calculate. After improving their methods, Wiscott did experiments with FERET image database, and its accuracy is 97.3% [36,37].

Moghaddam and Nastar also researched on distortion invariant object modeling. In their work, An image, $I(x,y)$, is considered as 3-D surface, $(x,y,I(x,y))$, where

$I(x, y)$ denotes the height of the surface at the position, (x, y). So the object's distortions in the image not only include distortions at the position, (x, y), but also include that at the height, $I(x, y)$. This model has considered distortion characters of faces. For example, smile can arouse the variety of position and gray value around mouth in the image. Then, the matching of two face images becomes the elastic matching of two 3-D distortion surfaces. Speaking about distortion of surfaces, both the distortion of a single person in different conditions and that between different persons are considered. When analyzing distortion of the surfaces, finite element analysis is adopted. At last, use maximum posterior rule to decide whether the two images are from the same person. The experimental results are better than eigenface method reported by the articles.

Except for these methods mentioned above, other methods include a template matching method by Brunelli [32], a local-features-based feature extraction and face recognition method by Atick [28], and a flexible appearance modeling method by Lanitis. Also there is a method that is to extract isolines as features to recognize faces from segmented images and matching them with isolines in face database.

NEURAL NETWORKS METHOD

Neural network methods can be divided into two types. The first type is to derive face features by other methods and then design neural network classifier. And the second one is that deriving features and classifying faces are completed with neural network.

Valentin [38] brings forward a method. He derives 50 principal components of a face firstly, and maps it to a five-dimension space with autocorrelation neural network, then makes decision using an ordinary multi-layer perception. It has good result for some simple testing images. Intrator [39] gives a mixed neural network to recognize faces, which uses non-supervisory neural network to derive features and supervisory one to do classification. Lee [40] describes face using six rules and locates the parts according to these rules, and then inputs the geometry distance among parts into fuzzy neural network for recognition, whose effect improves largely compared with Euclidean-distance-based method. Laurence [41] uses convolutional neural network to recognize faces. As correlation of adjacent pixels is integrated in convolutional neural network, shifting, rotating and local distortion of image are invariant to some extent. So we can get ideal recognizing effect with the method. Lin [41] advances Probabilistic-Decision–Based Neural Network (PDBNN). It uses positive and negative samples to train, and gets rather perfect statistical estimation result. Training process is quickened by modularized One Class One Net (OCON) structure. This method has good application in the steps of detecting faces, locating face and recognizing faces. There are also some other researches, such as Hopfield network to associate and recognize lower resolution faces by Dai [42], and mixed model which combines RBF and tree classification to recognition faces by Gutta [43].

Since it is difficult for us to describe rules or laws of face recognition explicitly, while the implicit expression of them can be acquired by neural network methods, neural network methods have more advantages and higher adaptability than other methods mentioned above.

7.4 An Dual Eigenspaces Method for Face Recognition

As described above, Traditional Eigenfaces Method (TEM) is to use principal components of an ensemble of face images and then complete the recognition procedure in an orthogonal "face space" [1]. However, its recognition rate is largely reduced when head posture, lighting conditions or facial expressions vary [24,44]. To solve this problem, in this section, we present a novel Dual Eigenspaces Method (DEM) to further analyze the features distribution in the "face space" and use coarse-to-fine matching strategy for face recognition [22]. It is shown that our method is superior to TEM in recognition rate. It is also demonstrated that DEM has insensitivity to the face posture, expressions and illumination conditions to a certain extent.

In addition, this method is an example to show the procedure of face recognition calculation.

ALGEBRAIC FEATURES EXTRACTION

As the most optimal orthogonal expansion for image compression, K-L transform can also be used to feature extraction and pattern recognition [45]. In TEM, the generating matrix of K-L transform is a total scatter matrix:

$$\Sigma = \frac{1}{M}\sum_{i=1}^{M}(x_i - m)(x_i - m)^T, \tag{7-5}$$

where x_i is a vector of length N^2 ($= N \times N$ face image); m is the average image of all the training samples; and M is the number of images in the training set.

In our scheme, in order to achieve higher computational simplicity without loss of accuracy, a between-class scatter matrix is adopted as the generating matrix:

$$S_b = \frac{1}{P}\sum_{i=1}^{P}(m_i - m)(m_i - m)^T = \frac{1}{P}XX^T, \tag{7-6}$$

where $X = [(m_1 - m), \cdots, (m_P - m)]$; m_i is the average image of the i^{th} person's training samples; and P is the number of people in the training set.

As mentioned above, the eigenvectors of S_b can span an algebraic eigenspace and provide optimal approximation for those training samples in the sense of mean-square error. Given a face image, it can be projected onto these eigenvectors and represented in terms of a weight vector regarded as its algebraic features.

However, determining the eigenvectors of the matrix, $S_b \in \Re^{N^2 \times N^2}$, is an intractable task. It can be solved by using SVD theorem [45].

Firstly, the following matrix is formalized as:

$$R = \frac{1}{P} X^T X \in \Re^{P \times P}. \tag{7-7}$$

Obviously, it is much easier to calculate both eigenvalues, $A = diag[\lambda_1, \cdots, \lambda_{P-1}]$, and orthonormal eigenvectors, $V = [v_1, \cdots, v_{P-1}]$, in this lower-dimensional matrix.

Then, the eigenvectors of S_b can be derived by SVD theorem:

$$U = XVA^{-\frac{1}{2}}, \tag{7-8}$$

where $U = [u_1, \cdots, u_{P-1}]$ denotes the basis vectors which span an algebraic subspace called unitary eigenspace of the training set.

Finally, we can obtain the following result:

$$C = U^T X = A^{\frac{1}{2}} V^T, \tag{7-9}$$

where $C = [c_1, \cdots, c_P]$ is referred to the standard feature vectors of each person.

In TEM, face recognition is performed only in above unitary eigenspace. However, some eigenvectors might act primarily as "noise" for identification because they mainly capture the variations due to illumination and facial expressions. It results in the reduction in recognition rate of TEM. In order to further characterize the variations among each person's face and analyze different distributions of their weight vectors in the unitary eigenspace, our method is to construct new eigenspaces for each person by carrying out another K-L transform. For the i^{th} person, its generating matrix is selected as a within-class scatter matrix of all the weight vectors of its training samples:

$$W_i = \frac{1}{M_i} \sum_{j=1}^{M_i} (y_i^{(j)} - c_i)(y_i^{(j)} - c_i)^T, \quad i = 1, \cdots, P, \tag{7-10}$$

where $y_i^{(j)} = U^T (x_i^{(j)} - m)$ is defined as the weight vector of the i^{th} person's training sample $x_i^{(j)}$; and M_i is the number of i^{th} person's images in the training set.

Note that the eigenvectors of each W_i are easily obtained. Here those minor components (MCs) are chosen to span each person's individual eigenspace denoted by $\widetilde{U}_i (i = 1, \cdots, P)$. In cooperation with the unitary eigenspace, the construction of dual eigenspaces has been completed.

FACE RECOGNITION PHASE

A two-layer classifier is built in this phase. In the top layer, a common minimum distance classifier is used in the unitary eigenspace. For a given input face image, f, its weight vector can be derived with a simple inner product operation:

$$y = U^T (f - m). \tag{7-11}$$

In this way, the coarse classification can be performed by the distance between y and each person's standard feature vector, $c_i (i = 1, \cdots, P)$. Then a few candidates who have the minimum distance are chosen for the finer classification.

In the bottom layer, the weight vector, y, is mapped separately onto each candidates' individual eigenspace to yield coordinate vectors:

$$\tilde{y}_i = \tilde{U}_i^T (y - c_i). \tag{7-12}$$

If $d_j = \min\{d_i :\ d_i = \|\tilde{y}_i\|\}$, the input image, f, can be recognized as the j^{th} person.

The scheme shown above has been implemented on a Sun Sparc20 workstation. We firstly set up a database of about 250 face images. Eighteen Chinese male students have been taken frontal photos under controlled conditions, but without any special restrictions to their posture. These images are different in lighting conditions, facial expressions, head orientation and the distance to the camera.

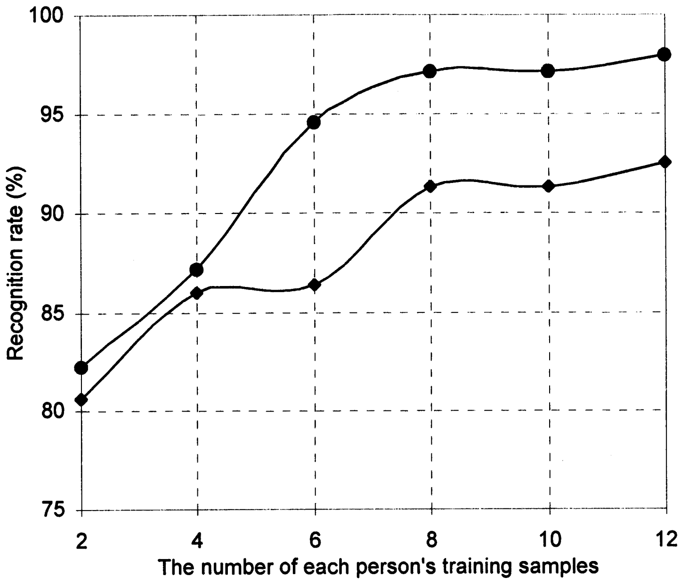


Figure 7-10. Comparison of performances between DEM and TEM

—●— Dual Eigenspaces Method (DEM)
—◆— Traditional Eigenfaces Method (TEM)

In our experiments, the number of each person's training samples varies from 2 to 12, while the remaining images constitute a test set. The recognition rates depicted in Figure 7-10 indicate that DEM is obviously better than TEM. For example, when six face images of each person are selected as training samples, there is a dramatic improvement in the recognition rate from 86.36% (TEM) to 94.63% (DEM). In particularly, when twelve images of each person are used as training samples, the recognition rate of DEM can be up to 97.93%. Considering the characteristics of the test images that contain the changes of head posture, facial expressions and illumination directions, it is obvious that our method is effective to these ambiguous images.

7.5 What Should We Do Next?

Although lots of important achievements have been gained in automatic face recognition field, especially in recent 10 years, the problem is far from being resolved. In fact, face recognition is one of the most difficult pattern recognition problems. And we have much work to do on it.

Not only are we interested in face recognition, but also neuroscientists and psychophysicists have done much for face recognition for many years. Differently, they care about and focus on human capacity for face recognition [46], and built models of the human capacity [47] and find characters of human recognition faces [48,49]. Many achievements among them have direct and indirect help for the research of automatic recognition of faces with computer. So let us see some results they got here.

The research results show that there are special units in human's brain, which have especial reaction to recognize faces. For example, there are cells to distinguish face contour and non-face contour quickly. And there are part detectors that can reflect the peculiar structure of face like eyes and hairstyle. The areas in brain that recognize acquaintances and strangers are different, so some patients whose brain are hurt, such as prosopagnosia, only lose transaction capability of video and audio information to acquaintances, but have the capacity of identifying strangers as the same person. The contrary instances are also available. To human beings, recognizing faces is not a special application of recognizing normal object [50]. It is very different from the recognition of normal objects. Human's recognizing faces have some characters:

1) Recognition accuracy to native race people is higher than that to other races. Psychophysicists tell us, there are average face in our brain, which are various from races.
2) The faces which are neither attractive nor unattractive are the most difficult to recognize by observers. Both global features and local features are important in recognizing process and the latter are always used for more meticulous and fine recognizing process. But if the local features are very marked and distinct, the global features are not transacted sometimes [51]. We have the similar experience when we view cartoon portraiture of well-known people.
3) When faces are rotated more than 90 degrees or are upside down, it is often difficult to recognize.
4) Appearance consciousness is sensitive to the tiny change of any feature part. Children often recognize persons by some factors that have nothing to do with faces, such as the glasses, clothes, caps and hairstyles, which disappears after adolescence. Recognition capacity to faces degenerates for old people.

In fact, one reason of difficulty to recognize face automatically comes from that human can't tell clearly how they recognize face and what features are used. So the achievements from neuroscience and psychophysics will be helpful for automatic face recognition.

Besides, as stated above, we have many methods and achievements to recognize the face with computer, but we still need new ideas, new theories and new methods to resolve the problem. We should communicate and exchange information in academe more widely and more deeply. The regular annual academic conferences are not only needed, special topic discussions on all aspects relative to face recognition are also needed. And we should do experiments with some same databases to make the results comparable. Now there are some databases available for experiments, including Mit, Yale, Carnegie-Mellon and FERET image database. FERET image database [52] includes images of multi-peoples, multi-ages, various expressions, various light conditions and different poses. The quantity of images and number of persons is also large. So it is a good database to test face recognition methods. We should also do further works on applications. The practical applications of face recognition will promote its research effectively.

References

[1] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp.71-86, 1991.

[2] Y. Chao, *Studies on Several Methods for Face Detection*, Master dissertation, Tsinghua University, Beijing, P.R.China, 1999.

[3] B. Moghaddam, A. Pentland, "Probabilistic Visual Learning for Object Representation," *IEEE Trans. PAMI*, vol. 19, pp. 696-710, 1997.

[4] K.K. Sung, and T. Poggio, "Learning a Distribution-Based Face Model for Human Face Detection," *CVPR'95*, pp.398-406, 1995.

[5] K.K.Sung and T. Poggio, "Example-balsed Learning for View-Based Human Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39-51, 1998.

[6] H.A. Rowley, S. Baluja, T. Kanade, "Neural Network-Based Face Detection," *IEEE Trans. PAMI*, vol. 20, pp. 23-38, 1998.

[7] G.Z. Yang and T.S. Huang, "Human Face Detection in a Complex Background," *Pattern Recognition*, vol. 27, pp. 53~63, 1994.

[8] T. Sakai, M. Nagao, T. Kanade, "Computer Analysis and Classification Photographs of Human Face," *Proc. 1st USA-Japan Computer Conf.*, pp. 55-62.

[9] S.A. Sirohey, "Human Face Segmentation and Identification," *CAR-TR-695*, Maryland University, 1993.

[10] V.A. Govindaraju, "Computational Model for Face Location," *CVPR'90*, pp. 718-721.

[11] V. Govindaraju, "Locating Human Faces in Photographs," *Int. J. Computer Vision*, vol. 19, pp. 129-146, 1996.

[12} J.B. Waite, W.J. Welsh, "Head Boundary Location Using Snakes," *Br. Telecom. Technol*, vol. J.8, no. 3, pp. 127-135, 1990.

[13] C.L.Huang, C.W.Chen, "Human Facial Feature Extraction for Face Interpretation and Recognitio-n," *Pattern Recognition,* vol. 25, no. 12, pp. 1435-1444, 1992.

[14] K.M. Lam, H.Yan, "Fast Algorithm for Location Head Boundaries," *J.Electronic Imaging* vol. 3, no. 4, pp. 351-359, 1994.

[15] C.W.Chen, C.L.Huang, "Human Face Recognition from a Signle Front View," *Int. J. Pattern Recognition and Artificial Intell,* vol. 6, no. 4, pp. 571-593, 1992.

[16] C.H. Lee, J.S. Kim, K.H. Park, "Automatic Human Face Location in a Complex Background," *Pattern Recognition,* vol. 29, pp. 1877-1889, 1996.

[17] H. Zabrodsky, S. Peleg and D. Avnir, "Symmetry as a Continuous Feature," *IEEE Trans. PAMI*, vol. 17, pp. 1154~1166, 1995.

[18] D. Reisfeld, H. Wolfson, Y. Yeshurun, "Context-Free Attentional Operators: The Generalized Symmetry Transform," *Int. J. Computer Vision*, vol. 14, pp. 119-130, 1995.

[19] C.Y. Lu and C.S. Zhang, "PCA-base Symmetry Detection," *Chinese Journal of Electronics*, in Chinese, vol.27, no.5, 1999.

[20] J. Zhou, C.S. Zhang, and Y.D. Li, "Dirctional Symmetry Transform for Human Face Location," *Optical Engineering*, vol. 38, no. 12, 1999.

[21] M.A. Turk, A.P. Pentland, "Face Recognition using Eigenfaces," *CVPR'91*, pp. 586-591, 1991.

[22] H. Peng, D. Zhang, "Dual Eigenspace Method for Human Face Recognition," *Electronics Letters*, vol. 33, pp. 283-284, 1997.

[23] D.L. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. PAMI*, vol. 18, pp. 831-836, 1996.

[24] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. PAMI*, vol. 19, pp. 711-720, 1997.

[25] C.Y. Lu, *Studies on Several Problems of Automatic Face Recognition System*, Ph.D Dissertation. Tsinghua University, 1998.

[26] A. Pentland, B. Moghaddam, and T. Starner, "View-based and Modular Eigenspaces for Face Recognition," *Proc. CVPR'94*, pp. 84-91, 1994.

[27] S. Akamatsu, T. Sasaki, H. Fukamachi, et. al., "A Robust Face Identification Scheme – KL Expansion of an Invariant Feature Space," *SPIE*, vol. 1607, pp. 71-84, 1991.

[28] J.J. Atick, P.A. Griffin, A.N. Redlich, "Statistical Approach to Shape from Shading," *Neural Computation*, pp. 1321-1340, 1996.

[29] T. Kanade, *Picture Processing by Computer Complex and Recognition of Human Faces*, Ph.D. dissertation, Kyoto Uni., Japan, 1973.

[30] M. Nixon, "Eye Spacing Measurement for Facial Recognition," *SPIE Proc.*, vol. 575, pp. 279-285, 1985.

[31] A.W. Young, H.D. Ellis, *Handbook of Research on Face Processing*, NORTH-HOLLAND, 1989.

[32] R. Brunelli, T. Poggio, "Face Recognition: Features Versus Template," *IEEE Trans. PAMI*, vol. 15, pp. 1042-1052, 1993.

[33] I.J. Cox, J. Ghosn, P.N. Yianilos, "Feature-based Face Recognition Using Mixture Distance," *CVPR'96*, pp. 209-216, 1996.

[34] M. Lades, J.C. Vorbruggen, J. Buhmann, et. al. "Distortion Invariant Object Recognition in the Dynamic Link Architecture," *IEEE Trans. Computer*, vol. 42, pp. 300-311, 1993.

[35] J. Zhang, H. Yan, M. Lades, Face Recognition: Eigenface, Elastic Matching, and Neural Nets, *Proc. IEEE*, vol. 85, pp.1423-1435, 1997.

[36] L. Wiskott, "Phantom Faces for Face Analysis," *Pattern Recognition*, vol. 30, pp. 837-846, 1997.

[37] L. Wiskott, J.M. Fellous, N. Kruger, et. al., "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. PAMI*, vol. 19, pp. 775-779, 1997.

[38] D. Valentin, H. Abdi, A.J. O'Toole, et. al., "Connectionist Models of Face Processing: A Survey," *Pattern Recognition*, vol. 27, pp. 1209-1230, 1994.

[39] N. Intrator, D. Reisfeld, Y. Yeshurun, "Face Recognition Using a Hybrid Supervised/Unsupervised Neural Network," *PR Letters*, vol. 17, pp. 67-76, 1996.

[40] S.Y. Lee, et. al., "Recognition of Human Front Faces Using Knowledge-Based Feature Extraction and Neuro-Fuzzy Algorithm," *Pattern Recognition*, vol. 29, pp. 1863-1876,1996.

[41] S. Lawrence, et. al., "Face Recognition: A Convolutional Neural-Network Approach," *IEEE Trans. NN*, vol. 8, pp. 98-113, 1997.

[42] Y. Dai, Y. Nakano, "Recognition of Facial Images with Low Resolution Using a Hopfield Memory Model," *Pattern Recognition*, vol. 31, pp. 159-167, 1998.

[43] S. Gutta, H. Wechsler, "Face Recognition Using Hybrid Classifiers," *Pattern Recognition*, vol. 30, pp. 539-553,1997.

[44] B. Moghaddam, and A. Pentland, "Face Recognition Using View-Based and Modular Eigenspaces," *Proceedings of the SPIE*, vol. 2277, pp. 12-21, 1994.

[45] E. Oja, *Subspace Method of Pattern Recognition*, Research studies Press, England, 1983.

[46] R.Mauro and M.Kubovy, "Caricature and Face Recognition," *Memory and Cognition*, vol.20, no. 4, pp. 433-441, 1992.

[47] S.S.Rakover and B.Cahlon, "To Catch a Thief with a Recognition Test: The Model and Some Empirical Results," *Cognitive Psychology*, vol. 21, no. 4, pp. 423-468, 1989.

[48] H.D.Elllis, D.M.Ellis, and J.A.Hosie, "Priming Effects in Children's Face Recognition," *British Journal of Psychology*, vol. 84, no. 1, pp. 101-110, 1993.

[49] P.Green, "Biology and Cognitive Development: The Case of Face Recognition," *Animal Behaviour,* vol. 43, no. 3, pp. 526-527, 1992.

[50] H.D.Ellis, "Introduction to Aspects of Face Processing: Ten Questions in Need of Answers," In H.Ellis, M.Jeeves, F.Newcombe, and A.Young, editors, *Aspects of Face processing*, pp.3-13, Nijhoff, 1996.

[51] D.C.Hay and A.W.Young, "The Human Face," In H.D.Ellis, editor, *Normality and Pathology in Cognitive Function*, pp. 173-202, Academic Press, London, 1982.

[52] J. Benarie, D. Nandy, "A Volumetric/Iconic Frequency-Domain Representation for Objects with Application for Pose Invariant Face Recognition," *IEEE Trans. PAMI*, vol. 20, pp. 449-457, 1998.

8 IRIS BIOMETRICS

Iris is an information density object, which is suitable for personal identification. In this chapter, we first give some definitions and notations for iris recognition. In Section 8.2, some current iris systems, including Daugman's approaches and others, are reviewed. Then, two novel methods, coordination system to solve head tilting problem and texture energy, are developed in Section 8.3 and 8.4, respectively. Their experimental results are shown in Section 8.5.

8.1 Introduction

Iris can be used as a biometrics signature, which was originally proposed by ophthalmologists [1-5,7-9]. According to their research, every iris has highly complex unique texture, which is unchanged over decades of life. The texture of iris is just like fingerprint but it includes over six times distinct features than fingerprint [10]. In terms of feature complexity, it is more suitable to be used for personal identification than fingerprint. Thus, iris is complex enough to be used as a biometic signature with probability that two irises would be identical by random chance is approximately 10^{-35} [6]. Observable features include the contraction furrows, striations, pits, collagenous fibers, filaments, crypts, serpentine vasculature, rings and freckles. Among the visible features in an iris, some of them can be observed by Figure 8-1. Iris is an internal organ, which is protected by cornea, so it cannot be surgically modified without damage to vision. Besides, it responds to light, which can be used to be a nature test against artifice.

Although iris has many good features, it has many difficulties using iris for personal identification. Cornea would cause high reflection. Some of parts of iris would be covered by eyelids and eyelash. Furthermore, many people wear contact lens. They cannot be removed when one uses iris recognition system. Besides, everyone trends to protect his or her eyes carefully. If the user is required to touch the equipment or flash used in the system, it would cause serious psychologically uncomfortable or even hurt the user's eyes. Thus, the system should allow that there is a certain distance between equipment and the user. The size of iris would be changed depending on different light condition. In real world situation, normally, we cannot control the light. Also, if the iris recognition system cannot treat with this problem, it cannot be used to test false or dead iris.

COMMERCIAL SITUATION

Almost all-current commercially available systems for iris recognition are based on John Daugman's algorithms. Some of them are shown in Figure 8-2. These systems are provided by: British Telecom (UK), NCR Corp (UK), Sensar/Sarnoff Inc (USA), GTE Corp (USA) Electronic Data Systems (USA), Spring Technologies (USA), IriScan Inc (USA), Oki Electric Co. (Japan), LG Electronics (Korea) and Garny AG (Germany). This technology has been applied to different areas which include bank Automatic Teller Machines, telecommunications, Internet security, portal entry control, nuclear power station security, computer login validation, prison controls and electronic commerce security.

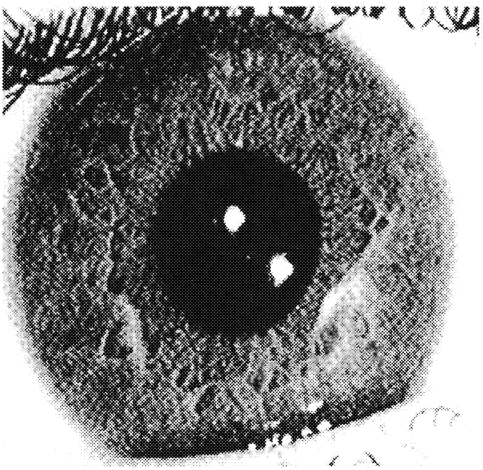


Figure 8-1. An iris which has highly complex unique texture.

IRIS SYSTEM ARCHITECTURE

Basically, existing iris recognition approaches [11-19] have a similar architecture which includes image acquisition, iris localization, transformation, filtering, feature extraction and matching (see Figure 8-3). Image acquisition is to yield an image of user's eye region [20]. Iris localization is used to point out the position of iris from the acquired image. Then, a transformation is utilized to overcome variation of iris's size and different filters are designed to capture texture information from the iris image. This information can be built an iris signature; such a process is called feature extraction. The last step is to compare two different iris signatures and make a final decision. In the coming sections, we will mention two different approaches.

DEFINITIONS AND NOTATIONS

An iris image is modeled by two circles, the limbus and pupil. An eye model is showed in Figure 8-4. In order to explain our method effectively, some denotations used in this chapter are defined as follows.

1) $I[x,y]$ Original image intensity at x, y position
2) (x_p, y_p) Center of pupil
3) r_p Radius of pupil
4) (x_l, y_l) Center of limbus
5) r_l Radius of limbus

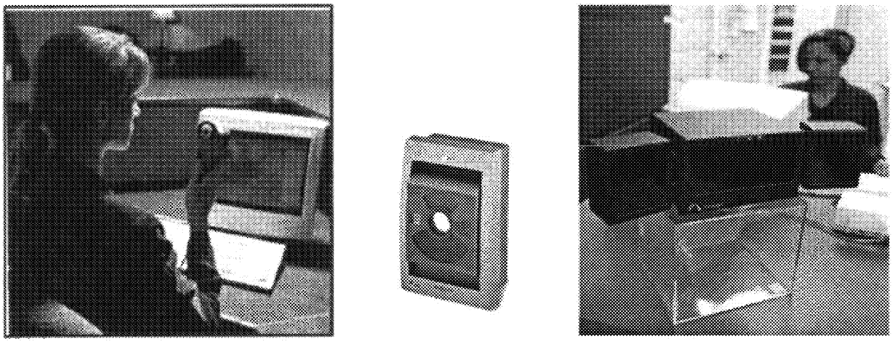


Figure 8-2. Three kinds of iris input devices.

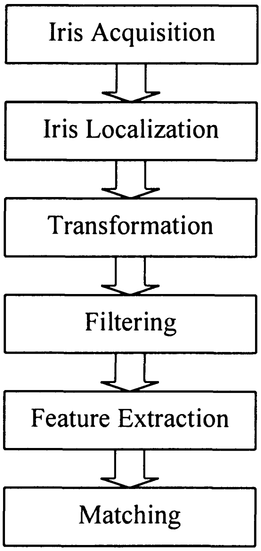


Figure 8-3. Basic architecture of iris recognition.

8.2 Iris Recognition

This method can be divided into four parts: 1) Iris location, 2) Doubly Dimensionless Projection, 3) Iris Code, and 4) Comparison. The detail of the method can be found in [11-13]. Daugman's experiment results were obtained from the images, which each size is 480 × 640 and the outer diameter of iris is over 60 pixels. They were usually between 100 to 200 pixels.

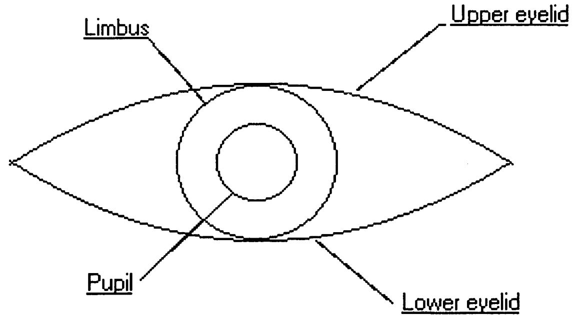


Figure 8-4. Eye model with pupil, limbus upper and lower eyelid.

IRIS LOCATION

Iris is a circular organ that is bounded by inner and outer boundaries (pupil and limbus). Daugman proposed an integrodifferential operator to search these two boundaries over the image domain. For every circular boundary can be defined by center (x_0, y_0) and radius *r*. Thus, these two circular located by the maximum in the blurred partial derivative, with respect to increasing *r*, of the normalized contour integral of *I(x,y)* along the circular arc *ds* of radius and center coordinates (x_0, y_0).

$$\max_{(r,x_0,y_0)} \left| G_\sigma(r) * \frac{\partial}{\partial r} \oint_{r,x_0,y_0} \frac{I(x,y)}{2\pi r} ds \right| \tag{8-1}$$

where * and G_σ denote convolution and Gaussion function of scale σ, respectively. The Gaussion function is used to reduce the computation time. Since upper and lower parts of iris are always covered by eyelids, the angular arc of contour integration *ds* is restricted on range to two opposite cones with 90 degree interior angle. Also, they are centered on the horizontal meridian. At the six o'clock position is allocated a refection point from cornea.

Since image *I(x,y)* is a digital image, Equation (8-1) should be converted to discrete form. The partial derivative is replaced by finite difference approximation. Additionally, interchange the order of the convolution and differentiation. The discrete form of equation has been got.

$$\max_{(n\Delta r,x_0,y_0)} \left| \frac{1}{\Delta r} \sum_k \{(G_\sigma((n-k)\Delta r) - G_\sigma((n-k-1)\Delta r) \sum_m I[k\Delta r\cos(m\Delta\theta) + x_0), (k\Delta r\sin(m\Delta\theta) + y_0)]\} \right| \cdot$$

(8-2)

In Equation (8-1), any point on the contour (x, y) now is replaced by $(k\Delta r\cos(m\Delta\theta) + x_0, k\Delta r\sin(m\Delta\theta) + y_0)$. Although Equation (8-2) is very effective to search limbus, it is not suitable for inner boundary because for "dark-eye people", the colour of iris would be similar to one of pupil. Thus, nonlinear enhancement of Equation (8-2) is necessary. If Equation (8-2) is divided by a sum of contour with smaller radius $(k-2)\Delta r$, when parameters $(n\Delta r, x_0, y_0)$ become well matched to inner boundary, its value is magnified. Thus, Equation (8-3) is obtained by

$$\max_{(n\Delta r, x_0, y_0)} \left| \sum_k \left\{ \frac{(G_\sigma((n-k)\Delta r) - G_\sigma((n-k-1)\Delta r)) \sum_m I[k\Delta r\cos(m\Delta\theta + x_0), k\Delta r\sin(m\Delta\theta + y_0)]}{\Delta r \sum_m I[(k-2)\Delta r\cos(m\Delta\theta + x_0), (k-2)\Delta r\sin(m\Delta\theta + y_0)]} \right\} \right| \tag{8-3}$$

DOUBLY DIMENSIONLESS PROJECTION

The size of iris would change because of pupillary constriction. Besides, distance between camera to eye and video zoom factor would affect size of iris in an image. Thus, Daugman proposed a doubly dimensionless non-concentric polar coordinate system to normalize all effects. Therefore, iris in an image from raw coordinates (x, y) maps to doubly dimensionless non-concentric polar coordinate (r, θ) where r is between 0 and 1 and θ is between 0 to 2π. It can be represented as

$$I(x(r,\theta), y(r,\theta)) \rightarrow I(r,\theta), \tag{8-4}$$

where $x(r,\theta)$ and $y(r,\theta)$ are defined as linear combinations of both the set inner boundary points $(x_{\min}(\phi), y_{\min}(\phi))$ and outer boundary points $(x_{\max}(\phi), y_{\max}(\phi))$. It can be represented as

$$x(r,\theta) = (1-r)x_{\min}(\theta) + rx_{\max}(\theta), \tag{8-5}$$

$$y(r,\theta) = (1-r)y_{\min}(\theta) + ry_{\max}(\theta). \tag{8-6}$$

When the raw image is transformed to doubly dimensionless non-concentric polar coordinate system, it would be demarcated into several zones. It is illustrated in Figure 8-5. The upper part of iris would not be used since it is always covered by eyelid. Moreover, 45° notch at the bottom where would be a desirable refection point so it would not be used in the zone analysis.

IRIS CODE

Iris code can be produced by a 2-D Gabor filter [21-22]. 2-D Gabor filters are defined in the doubly dimensionless polar coordinate system as follow

$$G(r,\theta) = e^{-iw(\theta-\theta_0)} e^{-(r-r_0)^2/\alpha^2} e^{-(\theta-\theta_0)^2/\beta^2}. \tag{8-7}$$

Both real and imaginary members of such quadrature filters are used in order to project image to complex plane. Also, the real part of 2-D Gabor filters is adjusted in order to ignore strength of illumination. The parameters α and β are free parameters which co-vary in inverse proportion to ω. Besides, position of the filters is specified by (θ_0, r_0) which is across the zones of analysis of the iris.

Each bit h in an iris code can be regarded as a point of the vertices of a logical unit square in the complex plane. It is computed by following condition for a particular Gabor filter.

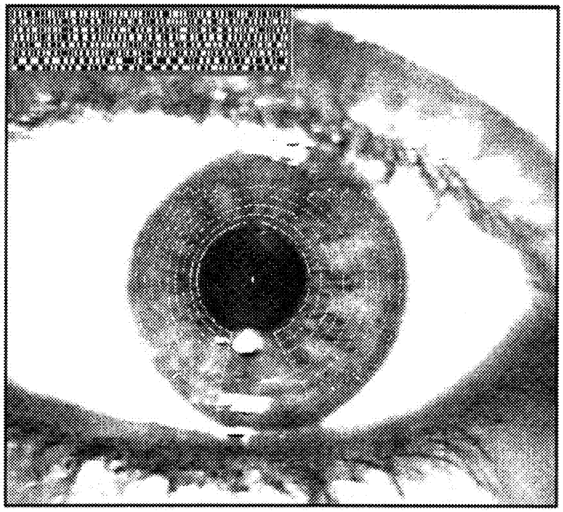


Figure 8-5. Demarcated zones of analysis and illustration of a computed iris code [12].

$$h_{\mathrm{Re}} = 1 \quad if \quad \mathrm{Re} \int_{\rho}\int_{\phi} e^{-iw(\theta_0-\phi)} e^{-(r_0-\rho)^2/\alpha^2} e^{-(\theta_0-\phi)^2/\beta^2} I(\rho,\phi)\rho d\rho d\phi \geq 0 \tag{8-8}$$

$$h_{\mathrm{Re}} = 0 \quad if \quad \mathrm{Re} \int_{\rho}\int_{\phi} e^{-iw(\theta_0-\phi)} e^{-(r_0-\rho)^2/\alpha^2} e^{-(\theta_0-\phi)^2/\beta^2} I(\rho,\phi)\rho d\rho d\phi < 0 \tag{8-9}$$

$$h_{\mathrm{Im}} = 1 \quad if \quad \mathrm{Im} \int_{\rho}\int_{\phi} e^{-iw(\theta_0-\phi)} e^{-(r_0-\rho)^2/\alpha^2} e^{-(\theta_0-\phi)^2/\beta^2} I(\rho,\phi)\rho d\rho d\phi \geq 0 \tag{8-10}$$

$$h_{\mathrm{Im}} = 0 \quad if \quad \mathrm{Im} \int_{\rho}\int_{\phi} e^{-iw(\theta_0-\phi)} e^{-(r_0-\rho)^2/\alpha^2} e^{-(\theta_0-\phi)^2/\beta^2} I(\rho,\phi)\rho d\rho d\phi < 0 \tag{8-11}$$

In Figure 8-5, an iris is demarcated into eight zones and it is cut in two 256 angular column. Thus, for one iris, it can be represented 2048 such paired bits.

COMPARISON

Comparing each pair of iris codes ***A*** and ***B*** bit by bit is based on Hamming distance (HD) and the Boolean operator (XOR). The Boolean operator equals to 1 if and only if the two bits A_j and B_j are different; otherwise, it equals to 0. The normalized Hamming distance equation is given below,

$$HD = \frac{1}{2048}\sum_{j=1}^{2048} A_j (XOR) B_j \,. \tag{8-12}$$

To compensate for possible tilt of head and eye rotation, comparisons between iris codes should be made several degrees of shifts along their angular axis. According to [12], when threshold is set to 0.342, the odds of false accept and odds of false reject are 1 in 1.2 million.

8.3 Coordinate System

In the following sections, we will discuss a novel method that can overcome some drawbacks of existing approaches such as head location problem. Other researcher does not consider this problem. This method can be divided into three parts. The first is to produce a coordination system which can solve head location problem including rotation and translation. Then normalization process for different size of iris and feature extraction are mentioned. Lastly, Mean of Correlation Coefficients is imposed to make a final decision.

HEAD TILTING PROBLEM

So far, a few different iris recognition approaches have been developed [11-19]. All of them need to adjust user's position in order to allow the camera to capture his or her iris, which easily cause tilting of head. In addition, relative head rotation would be produced when a user tilts handhold camera. However, the existing approaches cannot solve these problems well. Daugman's and Boles's approaches convert an iris signature to a set of digital signals which can be divided into several parts, each from one orbit of an iris. They locate head rotation by shitting the digital signals in same orbit of two iris's images. This comparison process is a "*best of n*" test of agreement. This technique can solve head location problem; however, it would lose accuracy. For Wildes's approach, his research team proposes an image-registration technique to solve these problems but this technique would increase correlation between two images. If degree of head rotation increases, the correlation between two different iris's images would increase. It means that the accuracy of the system reduces when tilting of head increase. This drawback is same as the previous approaches. Besides, it is commented that this image-registration needs high computational power that is depended on the degree of head tilting. In order to solve the above problems, we propose a new coordinate system which is computed from both the upper eyelid and center of pupil.

BASIC EYE MODEL

In general, an eye would be modeled by two circles and two parabolas which is shown in Figure 8-4. The circles are used to describe boundary of both pupil and limbus, and the parabolas are used to model the upper and lower eyelids. The two circles, pupil and limbus, can be defined by following equation.

$$(x - x_i)^2 + (y - y_i)^2 = r_i^{\,2}, \tag{8-13}$$

where (x_i, y_i) is a center and r_i is its radius, (p – pupil and l – limbus). We only use upper eyelid, which can be modeled by the following parabolic equation

$$(-(x-h)\sin\theta+(y-k)\cos\theta)^2 = a((x-h)\cos\theta+(y-k)\sin\theta)\,, \tag{8-14}$$

where a (<0) is imposed to control the curvature of a parabola; (h, k) is a vertex of the parabola and θ is a principle angle between x-axis and principle axis of the parabola, which measures with anti-clockwise. Figure 8-6 gives three parabolas with same principle axis and vertex but with different a's.

Our eye shape does not change too much when we rotate our head. Thus, we can impose eye shape to develop a method detecting the angle of rotation. In Equation (8-14), a principle angle is measured and it would be used to compare another principle angle that is calculated from registered image. Thus, we know the different head rotational angles between these two images. Besides, based on the position of pupil, the translation of an iris in an image could be normalized. Both the principle axis and pupil form a coordinate system to solve head location problem including rotation and translation.

MODIFIED PARABOLIC EQUATION

In Equation (8-14), we have shown a generally parabolic equation and given some explanation. If the vertex of eyelid appears on the pupil, the eyelid covers lot of parts iris and this image does not suitable for iris recognition. Thus, this property motivates to utilize polar coordinate system with origin at pupil's center (x_P, y_P) to represent the vertex. Besides, it produces a Hough space that is fully used. The transformation of vertex is based on the following equations

$$h = r\cos\alpha + x_p\,, \tag{8-15}$$

$$k = r\sin\alpha + y_p\,, \tag{8-16}$$

where r and α are two independent variables. The shape of eye is suitable for letting α equal to θ. Substitute Equations (8-15) and (8-16) into Equation (8-14) and let α equal to θ, Equation (8-14) becomes

$$r = (x-x_p)\cos\theta+(y-y_p)\sin\theta-\frac{1}{a}(-(x-x_p)\sin\theta+(y-y_p)\cos\theta)^2\,. \tag{8-17}$$

In Equation (8-17), only three variables are used, which is one variable less than in Equation (8-14). Thus the computational complexity is reduced.

SEARCHING ALGORITHM

In this section, we will discuss how to search the principle angle θ which is used to compare the rotation between two images. Since Equation (8-17) involves (x_P, y_P), we need to detect the pupil. In fact, all existing approaches essentially compute the pupil center and limbus center for normalizing vertical and horizontal translation. Thus, using pupil center does not increase computational time for existing approach.

Fitting the contour of pupil and limbus can be divided into two steps. First of all, an image would be convoluted by a lowpass filter, such as a two-dimensional Gaussian. Then, gradient operator $(\nabla \equiv (\partial/\partial x\ \partial/\partial y))$ is imposed to compute the image intensity. Mathematically, it can be represented by $|\nabla G(x,y) * I(x,y)|$, where *G(x,y)* is a two-dimensional lowerpass filter. If any point in the magnitude of the image intensity gradient is greater than a certain threshold, it is considered as an edge point. In order to find out the three parameters, (x_p, y_p, r_p), Hough transform is applied. However, one of fundamental drawbacks of Hough transform is slow in computation time so we use directional information about edge points to increase the speed. However, directional information could not be used in parabolic equation since gradient information is not stable for the eyelids and it is contaminated by hair.

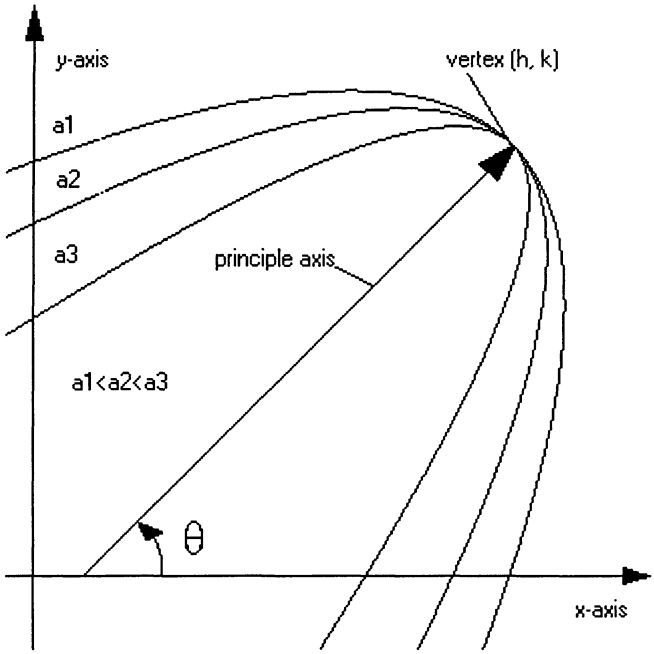
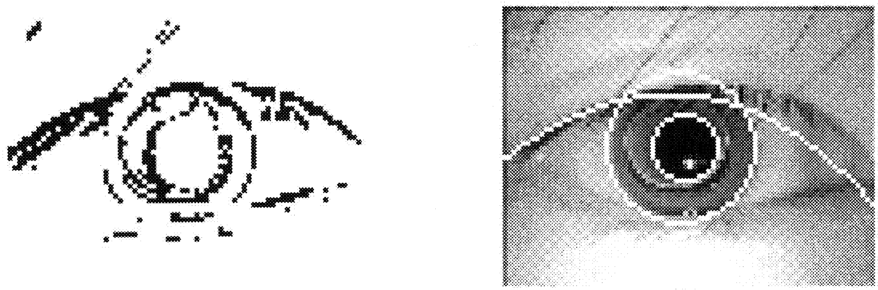


Figure 8-6. Parabolic equation explanation.

Using above method, searching a limbus and pupil should be faster but finding an eyelid still take time since Equation (8-17) has three variable *(r, θ, a)*. Therefore, we need to utilize other technique to speed up the searching process which is multi-resolution Hough transform [23]. The searching process is divided into two levels. The size of original images is 640×480 which is downsampled by four and eight for the two levels. Thus, edge detection becomes

$$(|\nabla G(x,y) * I(x,y)|)_{\downarrow j}, \tag{8-18}$$

where $(\,)_{\downarrow j}$ represents downsampling at j (j = 4, 8). For convenient, j = 8 is called Level 1; another is called Level 2.



(a) (b)

Figure 8-7. Edge points and result image on Level 1.

In the first level, approximate pupil, limbus and upper eyelid are obtained then in Level 2, the useless edge points are removed and the range of the parameters is restricted based on the approximate result from Level 1. According to this method, the number of edge is greatly reduced so that the computation time should decrease dramatically. However, we cannot directly apply Hough transform on Equation (8-17) to search upper eyelid since an image has other parabolic curves, such as limbus and pupil, so they should be deleted when the upper eyelids are being search. Figure 8-7(a) and Figure 8-8(a) show an edge points from Level 1 and Level 2 respectively. In addition, the result figure of Level 1 and Level 2 are displayed on Figure 8-7(b) and Figure 8-8(b), respectively. Figure 8-9 shows an image with principal axis which is used to solve head location problem including rotation and transition.

8.4 Texture Energy Feature

FOUR DIRECTIONAL FILTERING

In our method, texture information in iris image [24, 25] is captured first and the real circularly symmetric Gabor filters are chosen as [26]

$$f_{mn}(x,y) = \frac{1}{2\pi\sigma_m^2}\exp\left\{-\frac{x^2+y^2}{2\sigma_m^2}\right\}\times\cos(2\pi(u_m x\cos\theta_n + u_m y\sin\theta_n)), \tag{8-19}$$

where m (n) is index for the scale (orientation). To reduce computation time, Equation (8-19) can be decomposed into four one-dimensional filters. Suppressing the subscripts m and n, they can be described by

$$f(x,y) = h_1(x)h_2(y) - h_3(x)h_4(y), \tag{8-20}$$

and

$$h_1(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-x^2/2\sigma^2\right\}\cos(2\pi vx), \tag{8-21}$$

$$h_2(y) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-y^2/2\sigma^2\right\}\cos(2\pi wy), \tag{8-22}$$

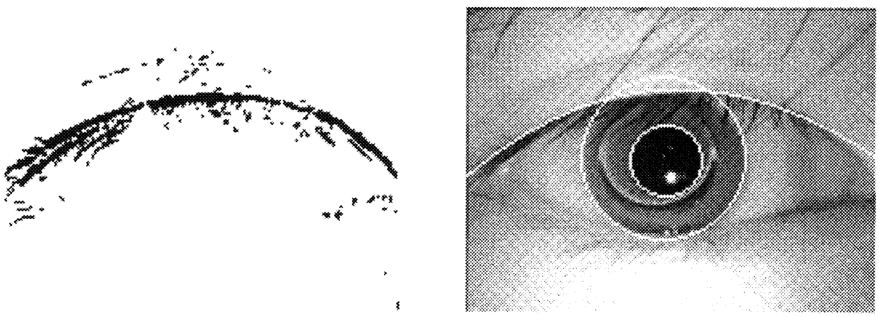
$$h_3(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-x^2 / 2\sigma^2\right\} \sin(2\pi v x), \tag{8-23}$$

$$h_4(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-y^2 / 2\sigma^2\right\} \sin(2\pi w y), \tag{8-24}$$

where $v = u\cos\theta$ and $w = u\sin\theta$. We cannot directly impose the filters to capture iris's texture information because of size variation of irises, so it is necessary to adjust the parameters, σ_m and u_m, which are directly and inversely proportion to radius of limbus, respectively. Thus, Equation (8-20) becomes

$$f_{mn}(x, y, r_l) = \frac{1}{2\pi(c_\sigma r_l)^2} \exp\left\{-\frac{x^2 + y^2}{2(c_\sigma r_l)^2}\right\} \times \cos(2\pi(\frac{c_u}{r_l} x\cos\theta_n + \frac{c_u}{r_l} y\sin\theta_n)), \tag{8-25}$$

where c_σ and c_u are the proportional constants of σ_m and u_m, respectively. Since iris's texture has different directions, we can define the total four orientations in our method.



(a) (b)

Figure 8-8. Edge points and result image on Level 2.

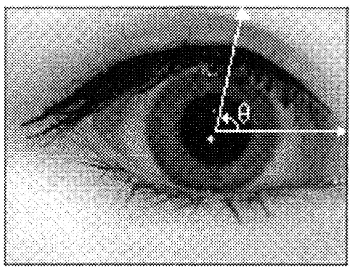


Figure 8-9. An image with principal axis.

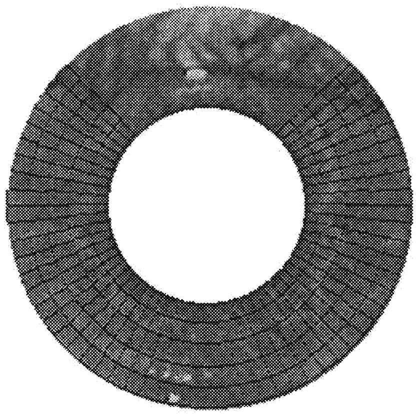


Figure 8-10. Illustration of regional decomposition image.

REGIONAL DECOMPOSITION

To compensate size variation of the iris which is due to different distance between camera and user, variation of lighting condition and different zoom factor of video camera, a convoluted iris image should be decomposed into some small regions which are represented by their mean energy. Firstly, a convoluted iris image is decomposed into several rings and then each ring is separated into a lot of small regions. We propose five rings so four circles are inserted in the image and employ linear interpolation to get the four circle's parameters (x_i, y_i, r_i), i = 1, 2, 3, 4. The parameters can be computed by the following equations

$$x_i = x_0 + i(x_l - x_p)/5, \tag{8-26}$$

$$y_i = y_0 + i(y_l - y_p)/5, \tag{8-27}$$

$$r_i = r_0 + i(r_l - r_p)/5. \tag{8-28}$$

Besides, each ring is decomposed to 270 small regions. However, the upper part of the iris would not be used since it is always covered by eyelids and disturbed by eyelashes. In addition, the outermost ring is not useful because different irises have similar texture energy in this ring. For the other rings, the texture energy will be considered as an iris feature. A decomposed image is shown in Figure 8-10. Notice that this figure does not be convoluted with any filter and only has 54 regions in each ring.

MEAN OF LOCAL ENERGY

For a small region, we take mean energy to represent it. Let I_k (k = 1, 2, 3, 4) be one of convoluted images by using one of filters and R be one of small region. Also, let n_R be number of pixels in this region. The mean energy of this small region, E_R, is computed as

$$E_R = \frac{\sum_{(x,y)\in R} I_k[x,y]^2}{n_R}. \tag{8-29}$$

The size of feature vector is $270\times4\times4$ = 4320. Then, each feature vector from different images can be stored in the database to compare with other feature vectors.

MATCHING BASED ON MEAN OF CORRELATION

As mentioned above, each convoluted image is cut down into five rings but only four of them will be utilized in matching process. Every feature vector from a ring is divided into three parts equally which are called ring sector. A ring sector has 90 elements. Two iris images are compared ring by ring. The similarity of two corresponding ring sectors is measured by correlation coefficient, *C* [-1, 1]. Let X and Y be two corresponding ring sectors from two different images. The correlation coefficient is shown as

$$C = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\mu_x}{\sigma_x}\right)\left(\frac{y_i-\mu_y}{\sigma_y}\right), \tag{8-30}$$

where n is the number of elements in X or Y, μ_x and σ_x (μ_y and σ_y) are sample mean and sample standard deviation of X (Y), respectively. If the correlation coefficient is equal to one, this means that all the points, (x_i, y_i), are on a straight line with position slop. However, if it equals to zero, the points, (x_i, y_i), have weak linear relationship. The certain number of elements in ring sector should be shifted in order to compensate for head and eye rotation. From all the ring sectors, totally there are $3 \times 4 \times 4 = 48$ correlation coefficients. To make a robust system, we take the mean over these correlation coefficients, which is called mean of correlation coefficients (MCC) to make the final decision.

8.5 Experimental Results

Four experiments mentioned in this section. The first experiment is used to test the ability of our method for solving rotation problem. Then a general experiment shows the accuracy of our method. Next, we investigate the robustness of the accuracy of our method. Some error is introduced in the position of iris's center to investigate the effect of such error. Finally, some noises are added to the images to test the robustness to noise.

ROTATION TEST

Various eye images from different persons are used to test our method. All experimental results are very satisfied. The parabola can fit upper eyelid well. As an example, an image is rotated by the different degree (ω = -20°, -15°, -10°, -5°, 0°, 5°, 10°, 15° and 20°). The same as pervious experiments, we apply our method to these rotated images and the typical results are shown in Figure 8-11.

If our method was perfect, the difference between θ (in Equation (8-17)) of the rotated images and θ of the original images would be same as the degree of the artificial rotation. The errors are measured by $|\theta_o - \theta_\omega - \omega|$, where θ_o is the original θ which is 77.9° and ω represents the degree of the artificial rotation. In addition, θ_ω is the corresponding result from the rotation image (see Table 8-1). The errors are mentioned in the last column of Table 8-1. The greatest error is smaller than 2° and the mean of the errors is 0.37°. It illustrates the accuracy of our method.

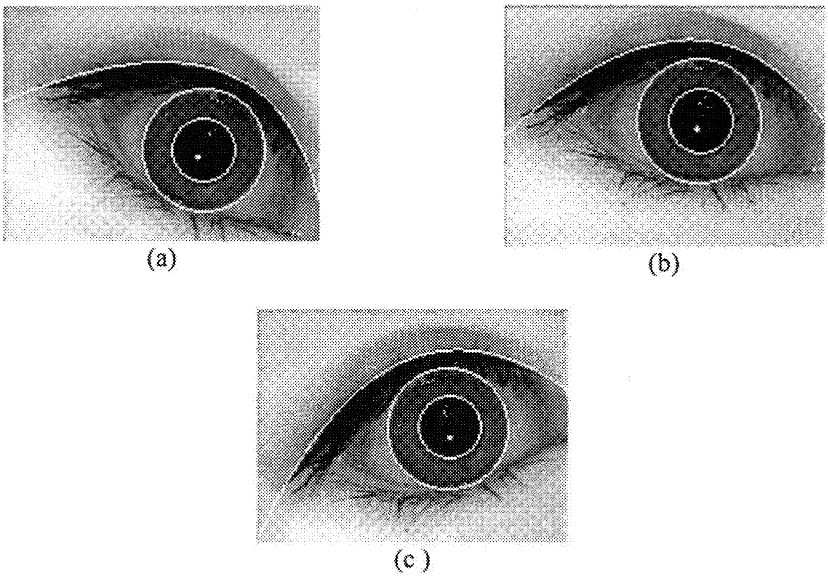


Figure 8-11. Three rotated images with fitting parabola.

Table 8-1. Experimental results in rotation test.

ω	θ_ω	*Error*
-20	57.0	0.9
-15	62.6	0.3
-10	68.5	0.6
-5	74.5	1.6
5	82.9	0
10	88.8	0.9
15	92.9	0
20	98.0	0.1

ACCURACY TEST

In this experiment, we want to know the accuracy of our method. Our database contains 15 iris images downloaded from http://www.neurotechnologija.com. The database contains 5 different irises and each iris has three images. The size of the image is 442 × 640 and the radius of limbus is about 170. All images are poor in contrast. The standard deviations of the iris's pixels are about 0.03. Besides, the images have different degree of head and eye rotation. Each iris in our database is compared with all other irises and the total number of comparison is 105. The parameters of filters are $c_u = 14.94$ and $c_\sigma = 0.03446$. The size of the mask in the filters is 35 by 35 and the four orientations are $(0, \pi/6, \pi/3, \pi/2)$.

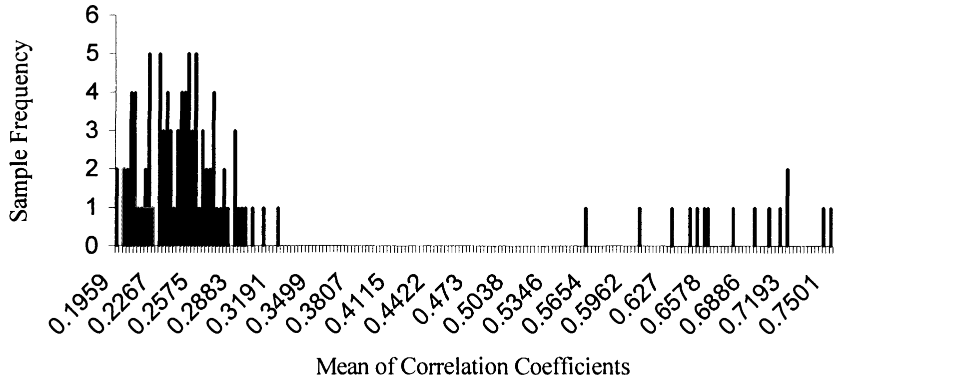


Figure 8-12. Distribution of mean of correlation coefficients.

In this experiment, we compare each iris image with all other iris's images in our database. The number of correct accepts and correct rejects are 15 and 90, respectively. The histogram in Figure 8-12 shows the distribution of the MCCs, which can be easily distinguished into two groups. The minimum distance between these two groups is about 0.23. This experimental result is shown in Table 8-2, where the correct accepts are highlighted. The first column and row are the images that are denoted by a number and an alphabet. For same irises, there are labeled same alphabet.

TEST FOR ROBUSTNESS OF IRIS' CENTRE POSITIONAL ERROR

We try to test the robustness of center positional error. Every iris's limbus center, (x_5, y_5), is shifted to $(x_5\text{-}5, y_5)$. In addition, the pupil center, (x_0, y_0), is moved to $(x_0\text{-}5, y_0\text{+}5)$. According to the error centers, we compute the iris feature using our proposed method again. The image has such error, that is labeled by a '*p*' after the name of original image, such as A1*p*, which is from A1. Combined with the original database, totally 30 images are tested in this experiment. Every original image is compared to all other error images. The experimental result is shown in Table 8-3 and the distribution of MCCs is displayed in Figure 8-13. Same as Table 8-1, the results from same eye's images are highlighted. The MCCs from same eyes are shifted to left because of positional error but the value is not too large. The mean of shifting distance is 0.105, which does not include the diagonal elements in two tables. In Table 8-1, all diagonal elements are one since they are obtained from comparing same image. The minimum distance between two groups is about 1.2, which is still enough to distinguish the two groups. From this experiment, we can see that our method is very robust to the localization error.

Table 8-2. Experimental results using proposed local texture energy recognition method.

	A1	*A2*	*A3*	*B1*	*B2*	*B3*	*C1*	*C2*	*C3*	*D1*	*D2*	*D3*	*E1*	*E2*	*E3*
A1	1.00	0.56	0.64	0.23	0.27	0.27	0.25	0.25	0.23	0.31	0.30	0.28	0.26	0.21	0.19
A2	0.56	1.00	0.66	0.25	0.28	0.28	0.22	0.21	0.23	0.30	0.29	0.30	0.26	0.24	0.21
A3	0.64	0.66	1.00	0.27	0.25	0.29	0.22	0.24	0.23	0.28	0.27	0.26	0.29	0.25	0.20
B1	0.23	0.25	0.27	1.00	0.68	0.61	0.25	0.27	0.26	0.26	0.24	0.26	0.26	0.25	0.27
B2	0.27	0.28	0.25	0.68	1.00	0.66	0.23	0.23	0.26	0.27	0.27	0.25	0.22	0.22	0.25
B3	0.27	0.28	0.29	0.61	0.66	1.00	0.29	0.28	0.32	0.25	0.25	0.25	0.26	0.21	0.25
C1	0.25	0.22	0.22	0.25	0.23	0.29	1.00	0.70	0.72	0.20	0.21	0.21	0.22	0.22	0.19
C2	0.25	0.21	0.24	0.27	0.23	0.28	0.70	1.00	0.69	0.24	0.23	0.25	0.23	0.22	0.20
C3	0.23	0.23	0.23	0.26	0.26	0.32	0.72	0.69	1.00	0.21	0.23	0.21	0.24	0.22	0.20
D1	0.31	0.30	0.28	0.26	0.27	0.25	0.20	0.24	0.21	1.00	0.71	0.75	0.24	0.25	0.21
D2	0.30	0.29	0.27	0.24	0.27	0.25	0.21	0.23	0.23	0.71	1.00	0.63	0.24	0.23	0.23
D3	0.28	0.30	0.26	0.26	0.25	0.25	0.21	0.25	0.21	0.75	0.63	1.00	0.26	0.24	0.21
E1	0.26	0.26	0.29	0.26	0.22	0.26	0.22	0.23	0.24	0.24	0.24	0.26	1.00	0.75	0.65
E2	0.21	0.24	0.25	0.25	0.22	0.21	0.22	0.22	0.22	0.25	0.23	0.24	0.75	1.00	0.72
E3	0.19	0.21	0.20	0.27	0.25	0.25	0.19	0.20	0.20	0.21	0.23	0.21	0.65	0.72	1.00

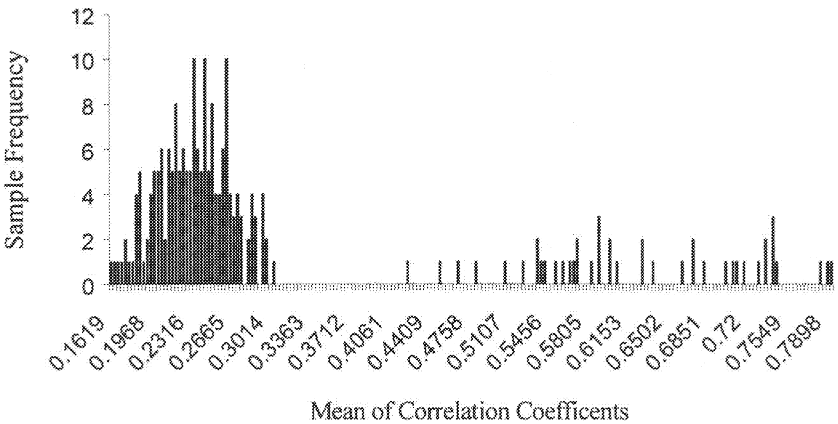


Figure 8-13. Distribution of mean of correlation coefficients for positional error image.

Table 8-3. Experimental results for positional error image.

	A1	*A2*	*A3*	*B1*	*B2*	*B3*	*C1*	*C2*	*C3*	*D1*	*D2*	*D3*	*E1*	*E2*	*E3*
A1p	0.67	0.54	0.51	0.24	0.25	0.23	0.21	0.21	0.22	0.29	0.26	0.28	0.22	0.19	0.18
A2p	0.42	0.71	0.45	0.20	0.22	0.21	0.21	0.20	0.21	0.30	0.27	0.27	0.25	0.21	0.17
A3p	0.52	0.63	0.71	0.24	0.25	0.22	0.23	0.23	0.22	0.30	0.25	0.25	0.24	0.21	0.18
B1p	0.26	0.26	0.29	0.70	0.57	0.47	0.26	0.24	0.27	0.26	0.25	0.26	0.27	0.24	0.26
B2p	0.27	0.26	0.23	0.56	0.73	0.48	0.25	0.26	0.27	0.30	0.27	0.29	0.24	0.24	0.26
B3p	0.25	0.28	0.27	0.54	0.56	0.72	0.28	0.28	0.29	0.25	0.23	0.25	0.24	0.22	0.22
C1p	0.24	0.23	0.22	0.24	0.22	0.26	0.79	0.67	0.67	0.19	0.18	0.21	0.17	0.17	0.16
C2p	0.22	0.22	0.22	0.25	0.23	0.27	0.59	0.78	0.58	0.23	0.23	0.25	0.20	0.21	0.18
C3p	0.23	0.25	0.22	0.25	0.26	0.26	0.60	0.63	0.79	0.20	0.22	0.21	0.20	0.18	0.19
D1p	0.29	0.30	0.25	0.24	0.23	0.20	0.19	0.20	0.19	0.74	0.59	0.68	0.24	0.25	0.18
D2p	0.29	0.30	0.27	0.26	0.26	0.23	0.22	0.22	0.22	0.66	0.74	0.60	0.24	0.24	0.22
D3p	0.29	0.28	0.26	0.23	0.24	0.20	0.19	0.19	0.19	0.59	0.54	0.75	0.24	0.24	0.21
E1p	0.25	0.28	0.27	0.24	0.24	0.26	0.20	0.20	0.19	0.25	0.23	0.23	0.74	0.61	0.54
E2p	0.23	0.23	0.25	0.24	0.22	0.23	0.22	0.21	0.21	0.25	0.23	0.23	0.64	0.73	0.57
E3p	0.20	0.24	0.26	0.29	0.26	0.26	0.20	0.20	0.20	0.22	0.21	0.20	0.55	0.57	0.73

IMAGE NOISE TEST

Our method is used to test robustness to noise. Different levels of Gaussian white noise is added to each original image. A noise image is named by a number, which represents the level of signal to noise in *db*. For example, $A1_{30}$ comes from the original image, A1, and the signal to noise ratio is 30 *db*. Every original image produces 6 images with different noise levels, i.e., 30db, 25db, 20db, 15db, 10db and 5db. The original images are then compared with all noise images; and therefore, one original image tries to match other 80 images. In this experiment, we do not apply any lower pass filter or median filter to do pre-processing. The experimental result illustrates that when signal to noise ratio is larger than 10db, the noise does not effect the match value in the first three significance digits. When the signal to noise ratio is 5db, MCCs do not reduce more than 0.004, which is not enough to influence the decision. The MCCs from both the noise image (with 5db signal to noise ratio) and original image is displayed in Table 8-4. In this experiment, we have shown that our method is very robust to noise.

There are still other advantages in our method. This is because that it only utilizes the same image for iris recognition and does not need any extra hardware to solve the rotation problem. Of course, the existing approaches can do it by decreasing the accuracy to increase the degree of rotation. However, the accuracy of our method does not depend on such a degree. Only upper eyelid is imposed in our method; therefore it can save the computational time. Most importantly, θ and the center of the pupil form a very good coordinate system to solve the transition and rotation problems, which can be pluged in any existing iris recognition approaches without modifying them.

Table 8-4. Experimental results for noise image with 5db signal to noise ratio.

	A1	*A2*	*A3*	*B1*	*B2*	*B3*	*C1*	*C2*	*C3*	*D1*	*D2*	*D3*	*E1*	*E2*	*E3*
$A1_5$	1.00	0.56	0.65	0.23	0.27	0.27	0.25	0.25	0.23	0.31	0.30	0.28	0.26	0.21	0.19
$A2_5$	0.56	1.00	0.65	0.25	0.28	0.28	0.22	0.22	0.23	0.30	0.29	0.30	0.26	0.22	0.21
$A3_5$	0.64	0.66	1.00	0.27	0.26	0.29	0.22	0.24	0.23	0.28	0.27	0.26	0.29	0.25	0.20
$B1_5$	0.23	0.25	0.27	1.00	0.68	0.61	0.25	0.27	0.26	0.26	0.24	0.26	0.27	0.25	0.27
$B2_5$	0.27	0.28	0.26	0.68	1.00	0.66	0.23	0.23	0.26	0.27	0.27	0.26	0.22	0.22	0.25
$B3_5$	0.27	0.28	0.29	0.61	0.66	1.00	0.29	0.28	0.32	0.25	0.25	0.25	0.26	0.21	0.25
$C1_5$	0.24	0.22	0.22	0.25	0.23	0.29	1.00	0.70	0.72	0.20	0.21	0.21	0.22	0.22	0.20
$C2_5$	0.25	0.22	0.24	0.27	0.24	0.28	0.70	1.00	0.69	0.25	0.23	0.25	0.23	0.22	0.20
$C3_5$	0.23	0.23	0.23	0.26	0.26	0.33	0.72	0.69	1.00	0.21	0.23	0.20	0.24	0.22	0.20
$D1_5$	0.31	0.30	0.28	0.26	0.27	0.25	0.20	0.25	0.21	1.00	0.72	0.75	0.24	0.25	0.21
$D2_5$	0.30	0.29	0.27	0.24	0.27	0.25	0.21	0.23	0.23	0.72	1.00	0.63	0.24	0.23	0.23
$D3_5$	0.28	0.29	0.26	0.26	0.25	0.25	0.21	0.25	0.21	0.75	0.63	1.00	0.26	0.24	0.21
$E1_5$	0.26	0.26	0.29	0.26	0.22	0.26	0.22	0.23	0.24	0.24	0.25	0.26	1.00	0.75	0.65
$E2_5$	0.21	0.24	0.25	0.25	0.22	0.21	0.22	0.22	0.22	0.25	0.23	0.24	0.76	1.00	0.72
$E3_5$	0.19	0.21	0.20	0.27	0.25	0.25	0.20	0.21	0.20	0.21	0.23	0.21	0.65	0.72	1.00

Comparing with the existing approaches, our method has other advantages. The real circularly symmetric Gabor filters have an efficiency implementation. The convolution working on the four one-dimensional filters is faster than that on the two-dimensional filter. Furthermore, the proposed method does not need any transformation to handle rotation of head and size variation of an iris because of possible changing distance between camera and user, different lighting and changing the zoom factor of video camera. In order to solve such a problem, Wildes employed image registration technique, which has two disadvantages. It not only requires high computational power, but also depends on the pixel intensity. Nevertheless, the pixel intensity depends on light. This means that high refection from cornea affects the accuracy of normalization process. Our normalization processes neither require computational demanding nor depend on pixel intensity. In addition, Boles's approach requires to use a closed ring in iris, which is considered as a periodic signal to obtain a wavelet signal representation. Notice that it is independent from starting point. Nevertheless, this approach does not easily apply to real-life situation because iris may cover by upper or lower eyelids. Thus, a closed ring is not obtained. Moreover, Boles's approach uses concentric circles to model the inner and outer boundaries. This is not true when the iris is very large in the image. Our model has considered non-concentric cases. Also all irises are modeled by non-concentric circles in our database. Each convoluted image is decomposed into five rings and the width of each ring is about 20 pixels. Thus, it is robust to the error occurred from localization process.

References

[1] R.P. Wildes, "Iris Recognition: An Emerging Biometrics Technology," *Proceedings of the IEEE*, vol.85, no.9, September 1997.

[2] F. Bouchier, J. S. Ahrens, and G. Wells, "Laboratory Evaluation of the Irisscan Prototype Biometrics Identifier", *Sandia National Laboratories, Albuquerque, NM, Tech. Rep. SAND'96-1033*, 1996.

[3] R.G. Johnson, "Can Iris Patterns be Used to Identify People," Los Alamos National Laboratory, CA, *Chemical and Laser Sciences Division, Rep. LA-12331-PR*, 1991.

[4] Daugman J. G, "Recognizing Persons by Their Iris Pattern," in *Biometrics: Personal Identification in Networked Society*, 1998.

[5] Daugman J, "Recognizing Persons by Their Iris Patterns," in: *Biometrics: Personal Identification in Networked Society.* Amsterdam: Kluwer, pp 103-121, 1998

[6] P.W Hallian, "Recognizing Human Eyes," *Geometric Methods Computer Vision*, vol. 1570, pp. 214-216, 1991.

[7] F.H Adler, *Physiology of the Eye: Clinical Application* (fourth edition), The C.V. Mosby Company, London , 1965.

[8] L. Flom and A. Safir, *Iris Recognition System, U.S. Patent 46413149, 1987.*

[9] J. Rohen, "Morphology and Pathology of the Trabecular Meshwork," *in The Structure of the Eye*, ed. Smelser, pp. 335-341, Academic Press, New York, 1961.

[10] J. E. Siedlarz, "Iris: More Detailed than a Fingerprint," *IEEE Spectrum*, vol. 31, pp. 27, Feb. 1994.

[11] G.O. Williams, "Iris Recognition Technology," *IEEE Aerospace and Electronics Systems Magazine*, vol. 124, pp. 23 –29, April 1997.

[12] J. Daugman, "High Confidence Visual Recognition of Persons a Test of Statistical Independence," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15 pp. 1148-1161, 1993.

[13] J.G. Daugman, *Biometrics Personal Identification System Based on Iris*, United Stated Patent No. 5,291,560, US. Government Printing Office, Washington, D.C. 1994.

[14] R.P. Wildes, J.C Asmuth, G.L. Green, S.C. Hse, R.J. Kolczynski, J.R. Matey, and S.E. McBride, "A Machine Vision System for Iris Recognition," *Mach Vision Applicat.*, vol. 9, pp. 1-8, 1996.

[15] R.P. Wildes, J.C Asmuth, G.L. Green, S.C. Hse, R.J. Kolczynski, J.R. Matey, and S.E. McBride, "A System for Automated Iris Recognition," *Proc. IEEE Workshop on Applications of Computer Vision*, Sarasota, *FL*, pp121-128, 1994.

[16] R.P. Wildes, J.C Asmuth, G.L. Green, S.C. Hse, R.J. Kolczynski, J.R. Matey, and S.E. McBride, "Iris Recognition for Security Access Control: Final Repot," National Information Display Laboratory, Princeton, *NJ Tech. Rep.*, 1992.

[17] W.W. Boles, "A Security System Based on Human Iris Identification Using Wavelet Transform," *1997 First International Conference on Knowledge-Based Intelligent Electronic Systems*, pp. 533-541, Adelaide, Australia, 21-23 May 1997.

[18] W.W. Boles, "A Human Identification Technique Using Images of the Iris and Wavelet Transform," *IEEE Transactions on Signal Processing*, vol. 46, no 4, pp. 1185-1188, April 1998.

[19] R.P. Wildes, J.C. Asmuth, S.C. Hsu, R.J. Kolczynski, J.R. Matey, and S.E. McBride, Aut*omated, Noninvasive Iris Recognition System and Method*, U.S. Patent 5 572 596, 1996.

[20] K. Hanna, R. Mandelbaum, L. Wixson, D. Mishra, and V. Paragana, "A System for Nonintrusive Human Iris Acquisition," *in proc. Int. Association for Pattern Recognition Workshop on Machine Vision Applications*, pp. 200-203, Tokyo, Japan, 1996.

[21] J. G. Daugman, "Uncertainly Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized By Two-dimensional Visual Cortical Filters," *Journal of the Optical Society of America A*, vol. 2, pp. 1160-1169, 1985.

[22] J. G. Daugman, "Uncertainly Relation Spectral Analysis of Cortical Receptive Field Profiles," *Vis. Res* vol. 20, pp. 847-856, 1980.

[23] M. Atiquzzaman, "Multiresolution Hough Transform - An Efficient Method of Detecting Patterns In Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, pp.1090-1095, 1992.

[24] D. McMordie, "Texture Analysis of The Human Iris for High Security Authentication," *http://www.ee.mcgill.ca/~mcmordie/iris/iris.ps*, 1997.
[25] J.G. Daugman, "High Confidence Personal Identification by Rapid Video Analysis of Iris Texture," in *Proc. IEEE Int. Carnahan Conf. Security Technology*, pp. 1-11, 1992.
[26] A. Jain and G. Healey, "A Multiscale Representation Including Opponent Color Features for Texture Recognition," *IEEE Trans. on Image Processing*, vol. 7, no. 1, pp. 124-128, January 1998.

9 SPEAKER RECOGNITION

In this chapter, we first introduce the background of speaker recognition and some useful concepts associated with it. The principles of speaker recognition, including production of speech, basic structure of speaker recognition systems, feature parameters, speaker modeling, likelihood normalization and speaker recognition methods, are briefly reviewed in Section 9.2. A novel and effective speaker verification method is presented in Section 9.3. The experimental results in Section 9.4 illustrate the effectiveness of our methods proposed by this chapter.

9.1 Introduction

Speech, a most natural and convenient tool for human communications, is capturing the focus of researchers. As humans, we have developed the ability to identify people by merely hearing their voices. We can do this if they are in the same room with us, down the hall, on the telephone, or even talking through a personal address system. What makes this possible? What does our brain use to identify one person's voice from another?

It is easy to understand how to differentiate a male speaker from a female speaker because in most cases, the male's voice has a lower pitch. The problem becomes more difficult when trying to identify one particular male from a group of all male speakers. Maybe we can use the fact that one speaker has a Southern accent while the others do not. It could also be the case that the speaker pronounces certain words differently than other speakers. We have developed this discriminative ability and used it without giving it much thought. However, programming a computer or machine to perform the same task has been difficult.

Speaker recognition is the process of automatically recognizing who is speaking by using speaker-specific information included in speech waves [1]. The potential for application of speaker recognition exists whenever speakers are unknown and their identities are important. It can be used to verify the identity claimed by people accessing systems; *i.e.*, it enables access control of various services by voice. Applicable services include voice dialing, banking over a telephone network, telephone shopping, database access services, information and reservation services, voice mail, security control for confidential information, and remote access of computers. Besides, in meetings, conferences, or conversations, this technique makes machine identification of participants possible. If used in conjunction with continuous speech recognizers, automatic transcriptions could be produced containing a record of who said what. This capability can serve as the basis for information retrieval technologies from the vast quantities of audio information produced daily. The law enforcement community is also user group for speaker recognition, which can be used to help identify suspect [2].

There exist two branches in the research of speaker recognition: speaker identification and speaker verification. In automatic speaker identification, there is no prior identity claim, and the system decides who the person is, or what group the person belongs to [3]. While speaker verification is to analyze an utterance from an unknown speaker and compare it with the model of the speaker whose identity is claimed (e.g., by entering an employee number or presenting his smart card) [4]. The fundamental difference between identification and verification is the number of decision alternatives. In identification, the number of decision alternatives is equal to the size of the population size. Therefore, speaker identification performance decreases as the size of the population increases, whereas speaker verification performance approaches a contrast, independent of the size of the population, unless the distribution of physical characteristics of speakers is extremely biased [5].

Speaker recognition methods can also be divided into text-dependent and text-independent methods. The former requires the speaker to provide utterances of key words or sentences that are of the same text for both training and recognition, whereas the later do not rely on a specific text being spoken. Both text-dependent and independent methods have a serious weakness. That is, these systems can easily be defeated, because someone who plays back the recorded voice of a registered speaker uttering key words or sentences into the microphone can be accepted as the registered speaker. To cope with this problem, some methods use a small set of words, such as digits, as key words, and each user is prompted to utter a given sequence of key words that is randomly chosen every time the system is used [6]. Yet even this method is not reliable enough, since it can be defeated with advanced electronic recording equipment that can reproduce key words in a requested order.

9.2 Principles of Speaker Recognition

PRODUCTION OF SPEECH

The speaker-specific characteristics of speech are due to differences in physiological and behavioral aspects of the speech production system in humans. The main physiological aspect of the human speech production system is the vocal tract shape. The vocal tract is generally considered as the speech production organ above the vocal folds, which consists of laryngeal pharynx (beneath the epiglottis), oral pharynx (behind the tongue, between the epiglottis and velum), oral cavity (forward of the velum and bounded by the lips, tongue, and palate), nasal pharynx (above the velum, rear end of nasal cavity), and nasal cavity (above the palate and extending from the pharynx to the nostrils).

The vocal tract modifies the spectral contents of an acoustic wave as it passes through it, thereby producing speech. Hence, it is common in speaker recognition systems to make use of features derived only from the vocal tract. In order to characterize the features of the vocal tract, the human speech production mechanism is represented as a discrete-time system of the form depicted in Figure 9-1.

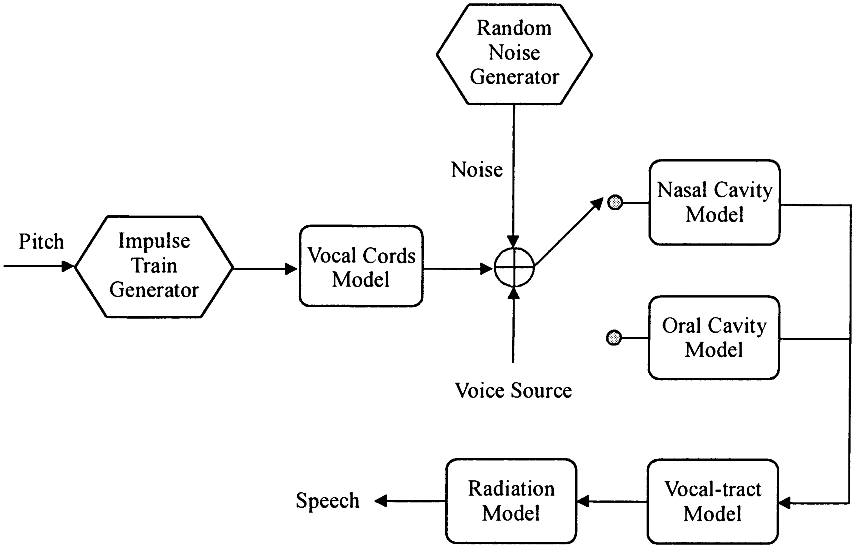


Figure 9-1. The human speech production mechanism.

The acoustic wave is produced when the airflow from the lungs is carried by the trachea through the vocal folds. This source of excitation can be characterized as phonation, whispering, friction, compression, vibration, or a combination of these. Phonated excitation occurs when the airflow is modulated by the vocal folds. Whispered excitation is produced by airflow rushing through a small triangular opening between the arytenoid cartilage at the rear of the nearly closed vocal folds. Friction excitation is produced by constrictions in the vocal tract. Compression excitation results from releasing a completely closed and pressurized vocal tract. Vibration excitation is caused by air being forced through a closure other than the vocal folds, especially at the tongue. Speech produced by phonated excitation is called voiced, that produced by phonated excitation plus friction is called mixed voiced, and that produced by other types of excitation is called unvoiced.

It is possible to represent the vocal-tract in a parametric form as the transfer function H(z). In order to estimate the parameters of H(z) from the observed speech waveform, it is necessary to assume some form for H(z). Ideally, the transfer function should contain poles as well as zeros. However, if only the voiced regions of speech are used, an all-pole model for H(z) is sufficient. Furthermore, linear prediction analysis can be used to efficiently estimate the parameters of an all-pole model. Finally, it can also be noted that the all-pole model is the minimum-phase part of the true model and has identical magnitude spectra, which contains the bulk of the speaker-dependent information.

Other aspects of speech production that could be useful for discriminating between speakers are learned characteristics, including speaking rate prosodic effects, and dialect (which might be captured spectrally as systematic shift in formant frequencies).

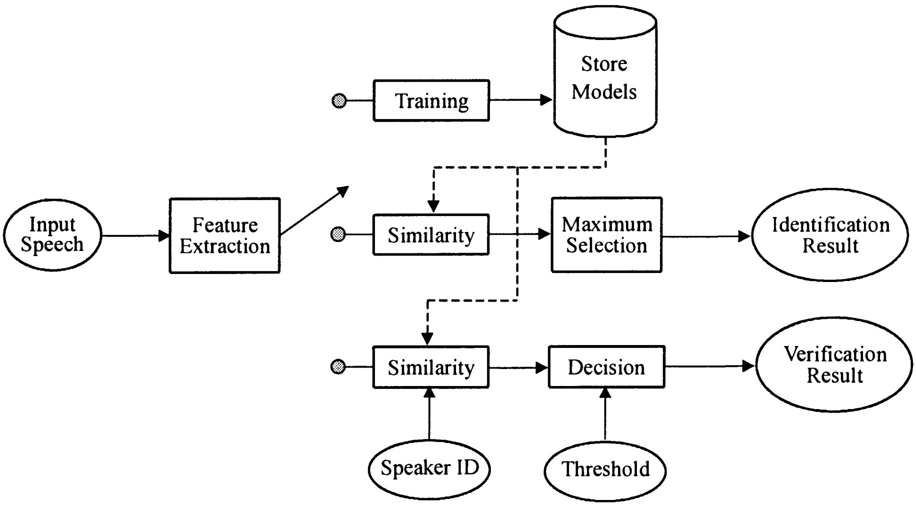


Figure 9-2. Basic structure of speaker recognition systems.

BASIC STRUCTURE OF SPEAKER RECOGNITION SYSTEMS

Figure 9-2 is the basic structure of a speaker recognition system. During the training phase, the speaker model is produced according to the input utterances. During the recognition phase, the distance between input speech and the stored speaker model is computed and the recognition decision is made.

In the speaker identification task, a speech utterance from an unknown speaker is analyzed and compared with speech models of known speakers. The unknown speaker is identified as the speaker whose model best matches the input utterance. In speaker verification, an identity claim is made by an unknown speaker, and an utterance of this unknown speaker is compared with the model for the speaker whose identity is claimed. If the match is good enough, that is, above a threshold, the identity claim is accepted. A high threshold makes it difficult for impostors to be accepted by the system, but at the risk of falsely rejecting valid users. Conversely, a low threshold enables valid users to be accepted consistently, but at the risk of accepting impostors. To set the threshold at the desired level of customer rejection and impostor acceptance, it is necessary to know the distribution of customer and impostor scores [5].

The effectiveness of speaker verification systems can be evaluated by using the receiver operating characteristics (ROC) curve adopted from psychophysics. The ROC curve is obtained by assigning two probabilities, the probability of correct acceptance and the probability of incorrect acceptance, to the vertical and horizontal axes respectively, and varying the decision threshold. The equal-error rate (ERR) is a commonly accepted overall measure of system performance. It corresponds to the threshold setting at which the false acceptance rate is equal to the false rejection rate.

FEATURE PARAMETERS

Speaker identity is correlated with the physiological and behavioral characteristics of the speech production system for each speaker. These characteristics exist both in the spectral envelope (vocal tract characteristics) and in the supra segmental features (voice source characteristics) of speech. Although it is impossible to separate these kinds of characteristics, and many voice characteristics are difficult to measure explicitly, many characteristics are captured implicitly by various signal measurements. Signal measurements such as short-term and long-term spectra and overall energy are easy to obtain. These measurements provide the means for effectively discriminating among speakers. Fundamental frequency can also be used to recognize speakers if it can be extracted reliably [7].

The LPC features were very popular in the early speech-recognition and speaker-recognition systems. However, comparison of two LPC feature vectors requires the use of computationally expensive similarity measures such as the Itakura-Saito distance and hence LPC features are unsuitable for use in real-time systems. Furui suggested the use of the cepstrum, defined as the inverse Fourier transform of the logarithm of the magnitude spectrum, in speech-recognition applications. The use of the cepstrum allows for the similarity between two cepstral feature vectors to be computed as a simple Euclidean distance. Furthermore, Atal has demonstrated that the cepstrum derived from the LPC features results in the best performance in terms of false acceptance rate (FAR) and false rejection rate (FRR) for a speaker verification system. Moreover, the derivatives of the cepstral coefficients capture the temporal information in speech that is essential for text-dependent tasks. It should be noted that some post-processing techniques are applied to the cepstrum in order to maintain the performance of a speaker recognition system under such diverse conditions of speech as noise- or channel-corrupted speech [8]. These techniques include cepstral weighting [9], cepstral mean subtraction (CMS) [10], pole-filtered cepstral mean subtraction (PFCMS) [11], adaptive component weighted cepstrum (ACWC) [12], post-filter cepstrum [13], and etc.

SPEAKER MODELING

Using cepstral analysis as described in the previous section, an utterance may be represented as a sequence of feature vectors. Utterances spoken by the same person but at different times result in similar yet a different sequence of feature vectors. The purpose of voice modeling is to build a model that captures these variations in the extracted set of features. There are two types of models that have been used extensively in speaker verification and speech recognition systems: stochastic models and template models. The stochastic model treats the speech production process as a parametric random process and assumes that the parameters of the underlying stochastic process can be estimated in a precise, well-defined manner. The template model attempts to model the speech production process in a non-parametric manner by retaining a number of sequences of feature vectors derived from multiple utterances of the same word by the same person. Template models dominated early work in speaker verification and speech recognition because the template model is intuitively more reasonable. However, recent work in stochastic models has demonstrated that these models are more flexible and hence allow for better modeling of the speech production process.

A very popular stochastic model for modeling the speech production process is the Hidden Markov Model (HMM). HMMs are extensions to the conventional Markov models, where in the observations a probabilistic function of the state, i.e., the model is a doubly embedded stochastic process where the underlying stochastic process is not directly observable (it is hidden). The HMM can only be viewed through another set of stochastic processes that produce the sequence of observations. Thus, the HMM is a finite-state machine, where a probability density function $p(x \mid s_i)$ is associated with each state s_i. The states are connected by a transition network, where the state transition probabilities are $a_{ij} = p(s_i \mid s_j)$.

For speech signals, another type of HMM, called a left-right model or a Bakis model, is found to be more useful. A left-right model has the property that as time increases, the state index increases (or stays the same)-- that is the system states proceed from left to right. Since the properties of a speech signal change over time in a successive manner, this model is very well suited for modeling the speech production process. Gaussian mixture model is also a popular speaker model, which provides a probabilistic model of the underlying sounds of a person's voice, but unlike HMM does not impose any Markovian constraints between the sound classes. More specifically, the distribution of feature vectors extracted from a person's speech is modeled by a Gaussian mixture density [14].

SPEAKER RECOGNITION METHODS

In the completed speaker recognition systems, the methods usually used include the traditional pattern matching methods, statistical model methods, neural networks, speech-recognition-based methods, and the hybrid methods of the above ones. Different from speech recognition, there is no knowledge-based speaker recognition system in the research of this field. This is limited to the level of perception science, since the expert knowledge about how a person can identify other persons by speech can not be concluded systematically till now.

Pattern Matching Methods

This kind of methods is dominant in the early research about text-dependent speaker recognition. Typical pattern matching methods include Dynamic Time Warping (DTW), Vector Quantization (VQ) source modeling as well as nearest neighboring method.

DTW method is the most popular method that can make up for the variability of speaking rate in pattern-based systems [15]. In this approach, each utterance is represented by a sequence of feature vectors. Generally, short-term spectral feature vectors, and the trail-to-trial timing variation of utterances of the same text is normalized by aligning the analyzed feature vector sequence of a test utterance to the template feature vector sequence using a DTW algorithm. The overall distance between the test utterance and the template is used for recognition decision. VQ source modeling uses multiple patterns to represent a frame of speech [16]. In this approach, each speaker is represented by a code-book of spectral templates representing the phonetic sound clusters in his/her speech. Nearest neighboring method integrates DTW and VQ-based methods [17,18].

Statistical Model Methods

Compared with pattern matching methods, statistical model methods are more flexible, and the distribution of likelihood scores is more reasonable in theory. These methods use stochastic model, and then the likelihood score of an observation is computed by this model. An observation is a random vector, whose conditional probability density function depends on the speaker. Given the density function, the probability produced by a speaker is decided. A popular statistical model method is the HMM model [19]. HMM models are not only the underlying speech sounds, but also the temporal sequencing among these sounds. Therefore, the HMM models are suitable for the text-independent and text-dependent speaker recognition.

Neural Networks

Rather than training individual models to represent particular speakers, neural networks (NNs) are trained to model the decision function which best discriminates speaker within a known set. Since no common opinion on the extraction of human speech characteristics exists, neural network methods show some advantages. In the static cases, neural networks can perfectly identify a speaker. In order to adapt to the dynamic characteristics of speech signals, research is mainly focused on time-delay neural network (TDNN) and recurrent networks, as well as hybrid networks. Generally, NNs require a smaller number of parameters than text-independent speaker models and have produced good speaker recognition performance comparable to that of VQ systems. The major drawback to many of the NN techniques is that the complete network must be retrained when a new speaker is added to the system. Besides, auditory model is also a topic on speaker recognition [13].

Speech-Recognition-Based Methods

Gauvain et al. investigated a statistical modeling approach, where each speaker was viewed as a source of phonemes, modeled by a fully connected Markov chain. Maximum a *posteriori* (MAP) estimation was used to generate speaker-specific models from a set of speaker-independent seed models. The lexical and syntactic structures of the language were approximated by local phonotactic constraints. The unknown speech is recognized by all of the speakers' models in parallel, and the hypothesized identity is that associated with the model set having the highest likelihood [21].

Since phonemes and speakers are simultaneously recognized by using speaker-specific Markov chains, this method can be considered as an extension of the ergodic-HMM-based method. The experimental results using the BREF corpus showed that this method clearly out-performed a simpler Gaussian mixture (single-state HMM) model. It was also found that text-independent and text-dependent verification ERRs were about the same [5].

LIKELIHOOD NORMALIZATION

In a broader sense, pattern recognition may be considered as a problem of estimating density functions in a high-dimensional space and dividing the space into the regions of categories or classes. Normalization can be viewed as a very important process for a successful classifier design [22], because it not only refines the input samples to help the further handling, but also removes the unimportant or common information contained in the samples. As a result, the distinguishing ability of the classifier will be improved.

Generally speaking, speaker verification is a classifying problem, although it seems related only to a person himself at first glance. In speaker verification tasks, the absolute likelihood score of an utterance from a speaker model is affected by many speakers' vocal characteristics, the linguistic content and the speech quality. These factors make it very difficult to set a decision threshold for absolute likelihood values to be used over different verification tests. The likelihood ratio normalization produces a relative score, which is more a function of the utterance speaker and less volatile to non-speaker utterance variations. Obviously, it is expected that the separability between speaker will be more obvious and the threshold will be more easily set by normalization [20,34].

The most poplar likelihood normalization techniques are basically utterance-level. Higgins et al proposed a normalization method based on the likelihood ratio [6]. Li and Porter proposed normalizing the likelihood values by subtracting the mean of imposter models and then dividing by their standard derivation [23]. Chi-Shi Liu et al used anti-speaker models to represent the impostors of reference speakers and to normalize the likelihood score [24]. Theoretically, the likelihood values between the input utterance and the models of a large number of speakers must be calculated. Unfortunately, the amount of calculation becomes enormous when the number of reference speakers is large. In order to reduce the amount of calculation, cohort speakers were selected from the speakers in the database for each customer. The summation of the likelihood values over all the reference speakers was approximated by that over the cohort speakers [6,25]. However, it is difficult to select a proper set of cohort speakers.

Frame level likelihood normalization technique has also been applied successfully in speaker recognition, which involved in the non-linear transformation on the frame likelihood scores [26]. This makes likelihood normalization technique apply successfully in the speaker identification task. For the speaker verification, the combination of frame and utterance level likelihood normalization was also successful.

9.3 GSMSV Method

In the GSMSV (Speaker Verification based on the Global Speaker Model) method we proposed [27], the global speaker model represents all of the common information contained in the speech of multiple speakers, such as the pronunciation characteristics, background noises, and common features of texts with different contents. Hence, it is utilized to normalize the likelihood score so that the score distribution is much more concentrated, and the difference between reference speakers and impostors is increased.

GSMSV is different from both the speaker verification methods with the conventional likelihood score (noted as CSV method in the following) [28] and the normalization method proposed in [24]. As we know, CSV method has some limitations, *i.e.*, the loose distribution of likelihood scores leading to the vague boundaries between speakers and the burden to set a proper threshold which is a direct consequence of scattered likelihood scores, as well as the low system adaptability to protean input utterances with different duration and different content. The proposed GSMSV method perfectly solves these limitations of CSV method by employing the global speaker model to normalize the likelihood score.

The verification method proposed in [24] employs anti-speaker models to normalize the likelihood score (called ASMSV here). Although this method can also significantly improve the performance of CSV method, it is faced with the conflict between the scale of the anti-speaker model (*i.e.*, L, the number of speaker models included in an anti-speaker model) and verification speed. With increase in the value of L, the verification speed is slowed down. On the other hand, L can not be too small; otherwise, the equal error rates are very high. Experiments in [24] show this dilemma. When L is 0, 6, 8 and 19, the equal error rates for closed set test are 11.65%, 7.24%, 6.44% and 3.65%, and the equal error rates for open set test are 10.99%, 8.48%, 8.46% and 8.22%, respectively. Therefore, ASMSV method has to make a trade-off between these two factors. Furthermore, ASMSV method can not well distinguish the outside impostors. This limitation is unsuitable and intolerable for safety applications. The establishment of anti-speaker models is also difficult and time-consuming. GSMSV method avoids the conflict facing ASMSV method. The global speaker model is easy to obtain and the verification speed is very fast. For verification, all computations needed by GSMSV are two calculations of the likelihood score, while ASMSV needs L+1 likelihood score calculations.

The global speaker model is similar to the speaker world model described in [29]. However, they are established in different ways. In [29], the speaker world model is an average of all reference speaker models. While the global speaker model is obtained during training procedure by using all training data from all reference speakers. Therefore, the global speaker model is much more accurate than the speaker world model. Besides, the global speaker model can also be quickly established by using its quick adaptive and real-time estimation.

Figure 9-3 gives the flow chart of the verification system with GSMSV method. In the registration phase, the reference speaker model is established and the global speaker model is modified. In the verification phase, the likelihood scores produced by the claimed speaker model and the global speaker model are computed, respectively, and the conventional score is normalized. Then the normalized score is compared to some threshold to decide whether to accept or reject the claimed identity.

In order to employ GSMSV method to real-time applications, an adaptive re-estimation approach for modifying the parameters of the global speaker model is proposed to shorten the waiting time for a new registration. By employing this technique, GSMSV method can perfectly meet the real-time requirements, thus its practicability is ensured.

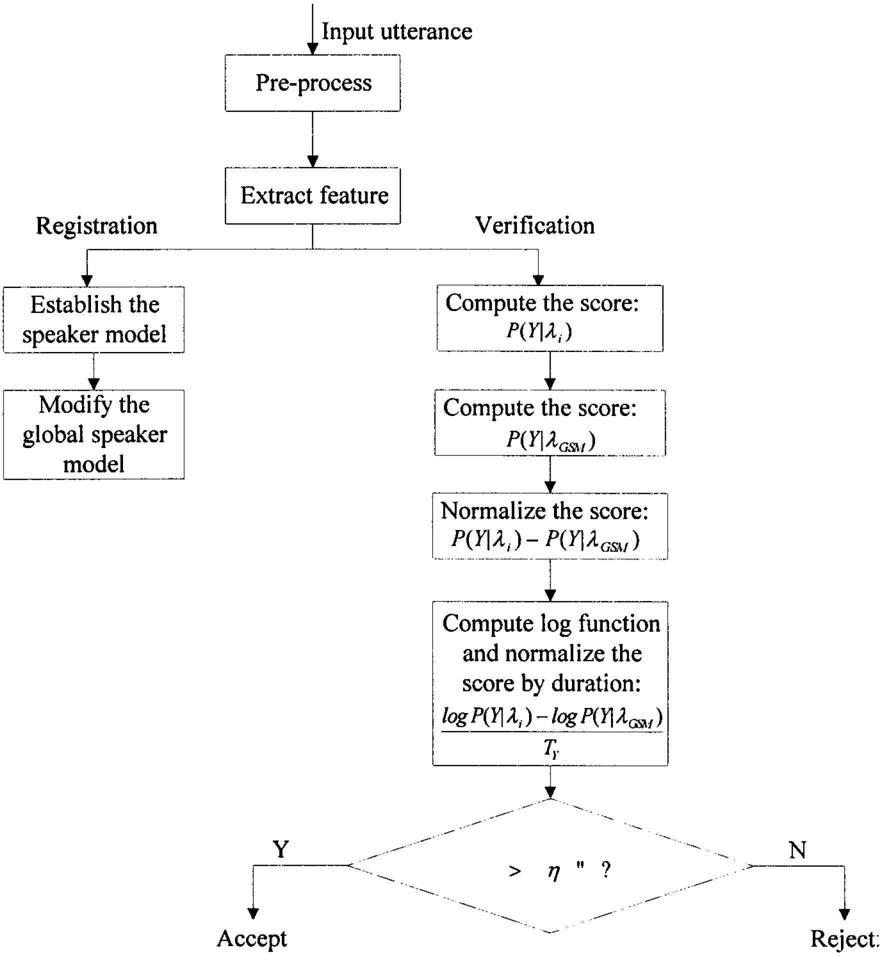


Figure 9-3. Flow chart of the verification system with GSMSV method.

DEFINITIONS AND NOTATIONS

Given N reference speakers, whose models are $\lambda_1, \cdots, \lambda_i, \cdots, \lambda_N$, in which λ_i is obtained by maximizing the likelihood score $P(Y_i \mid \lambda_i)$, and Y_i is the training data of reference speaker i. In GSMSV method, the global speaker model, λ_{GSM}, is established besides the N reference speaker models. λ_{GSM} is acquired by maximizing $\prod_{i=1}^{N} P(Y_i \mid \lambda_{GSM})$. Thus there are totally $N+1$ speaker models, of which λ_{GSM} is a universal one, and part of the speech characteristics for each speaker has its reflection in λ_{GSM}. If the speech Y to be verified is uttered by some reference speaker i, the following formula should be satisfied for ideal cases

$$P(Y \mid \lambda_i) > P(Y \mid \lambda_{GSM}). \tag{9-1}$$

It is because that λ_i is obtained from the i-th speaker's own training data, it can describe the speech distribution of the i-th speaker in the acoustic space much more properly than λ_{GSM}. At the same time, the following formula should be satisfied:

$$P(Y \mid \lambda_{GSM}) > P(Y \mid \lambda_j), \ j = 1, 2, \cdots, N, and \ \ j \neq i \tag{9-2}$$

since λ_{GSM} includes the characteristics of many speakers, and the information about the i-th speaker contained in λ_{GSM} should be more than that contained in the speech of other reference speakers. If Y is not uttered by any reference speaker, *i.e.*, it is the speech of an outside impostor, Equation (9-3) should be satisfied for ideal cases because λ_{GSM} is a universal speaker model

$$P(Y \mid \lambda_i) < P(Y \mid \lambda_{GSM}), \ i = 1, 2, \cdots, N. \tag{9-3}$$

From Equations (9-1)-(9-3), the following decision strategy may be obtained:

$$P(Y \mid \lambda_i) \begin{cases} > P(Y \mid \lambda_{GSM}), \textit{accept the claim to reference speaker } i \\ \leq P(Y \mid \lambda_{GSM}), \textit{reject the claim to reference speaker } i. \end{cases} \tag{9-4}$$

$P(Y \mid \lambda_{GSM})$ can be regarded as a dynamic threshold, changing with different input utterances. Since Equations (9-1)-(9-3) are satisfied only for ideal cases, Equation (9-4) should not be directly used to speaker verification.

From another point of view, the training data for the global speaker model is so plentiful that the produced model, λ_{GSM}, contains not only the universal speech characteristics of multiple speakers, but also the environmental features related to the recording and speaking background. According to the principle idea of pattern recognition, *i.e.*, removing the common information is helpful to improve the discriminating ability of a classifier. It is anticipated that the differences between speakers will be emphasized if the common speech characteristics are found out and obliterated from speech. Based on this consideration, GSMSV method is designed as follows:

Let $S_{GSM}^{(i)}(Y)$ be the normalized likelihood score for an input utterance, Y, claimed to be uttered by the i-th reference speaker, we can get

$$S_{GSM}^{(i)}(Y) = P(Y \mid \lambda_i) - P(Y \mid \lambda_{GSM}). \tag{9-5}$$

By subtracting the score obtained from λ_{GSM}, the common information of both pronunciation characteristics and environmental features is obliterated. As a result, the interference of unimportant factors is avoided, and the differences between speakers are brought into prominence. Therefore, the decision rule for GSMSV method can be defined as:

$$S_{GSM}^{(i)}(Y)\begin{cases} > \eta, \textit{accept the claim to reference speaker } i \\ \le \eta, \textit{reject the claim to reference speaker } i \end{cases} \tag{9-6}$$

where η is a threshold.

The value of η represents the conditional limitation of acceptance and rejection. It is allowed to be negative or positive according to certain requirement. The greater the value of η, the more strict the conditional limitation.

Since λ_{GSM} is a universal speaker model, it contains all of the common information related to speakers, such as the pronunciation characteristics, background noises and common features of texts with different contents. λ_{GSM} is suitable to represent the above information for both reference speakers and outside impostors. In this way, GSMSV has a powerful distinguishing ability.

In order to avoid overflow in computation, logarithm likelihood score is utilized and the decision rule of Equation (9-6) becomes:

$$\log P(Y \mid \lambda_i) - \log P(Y \mid \lambda_{GSM})\begin{cases} > \eta', \textit{accept the claim to reference speaker } i \\ \le \eta', \textit{reject the claim to reference speaker } i \end{cases} \tag{9-7}$$

where η' is a threshold.

To further improve the system adaptability and alleviate the influence of speaking rate, the likelihood score is normalized again by duration in the following:

$$\frac{\log P(Y \mid \lambda_i) - \log P(Y \mid \lambda_{GSM})}{T_Y}\begin{cases} > \eta'', \textit{accept the claim to reference speaker } i \\ \le \eta'', \textit{reject the claim to reference speaker } i \end{cases} \tag{9-8}$$

where T_Y is the number of frames, and η'' is a threshold.

GENERAL ESTIMATION OF THE GLOBAL SPEAKER MODEL

In GSMSV, the employed speaker model is Gaussian mixture model (GMM), *i.e.*, the distribution of the training speech data in the acoustic space for each reference speaker is represented by mixture Gaussian probability density functions. It is similar to "semi-continuous" probability distribution or "tied mixture" technique for representing speech segments in hidden Markov based speech recognition [30]. The parameters of a GMM can be represented as:
$\lambda = ((c_1, \mu_1, \Sigma_1), \cdots, (c_k, \mu_k, \Sigma_k), \cdots, (c_M, \mu_M, \Sigma_M))$, in which μ_k and Σ_k are the mean vector and the covariance matrix for the k-th Gaussian density function, respectively; c_k is the corresponding weight; and M is the number of mixture

components. Let $Y=\{y_1,\cdots,y_k,\cdots,y_T\}$ be the sequence of feature vectors for an input utterance, thus its likelihood score produced by λ is obtained as:

$$P(y_k \mid \lambda)=\sum_{m=1}^{M} c_m \cdot \frac{1}{(\sqrt{2\pi})^{D/2}\cdot(|\Sigma_m|)^{1/2}} \cdot exp(-\frac{1}{2}(y_k-\mu_m)^T \Sigma_m^{-1}(y_k-\mu_m)) \tag{9-9}$$

and

$$P(Y \mid \lambda)=\prod_{k=1}^{T} P(y_k \mid \lambda) \tag{9-10}$$

where D is the dimensionality of feature vectors.

Assume the current verification system has N users, whose training data is represented as $Y_i=\{y_1^{(i)},y_2^{(i)},\cdots,y_k^{(i)},\cdots,y_{T(i)}^{(i)}\}$ ($i=1,2,\cdots,N$), after being transformed to feature vectors, in which i denotes the i-th speaker and $T(i)$ denotes the total number of feature vectors for the i-th speaker. The training data for a new system user, the $(N+1)$-th reference speaker, is noted as $Y_{N+1}=\{y_1^{(N+1)},y_2^{(N+1)},\cdots,y_k^{(N+1)},\cdots,y_{T(N+1)}^{(N+1)}\}$.

Let the parameters of λ_{GSM} be:
$\lambda_{GSM}=((c_1^{GSM},\mu_1^{GSM},\Sigma_1^{GSM}),\cdots,(c_k^{GSM},\mu_k^{GSM},\Sigma_K^{GSM}),\ \cdots,(c_M^{GSM},\mu_M^{GSM},\Sigma_M^{GSM}))$, in which c_k^{GSM} is the weight of the k-th Gaussian density function; μ_k^{GSM} and Σ_k^{GSM} are the corresponding mean vector and covariance matrix, respectively. As the previous description, the parameters of λ_{GSM} are obtained by maximizing $\prod_{i=1}^{N+1} P(Y_i \mid \lambda_{GSM})$ with Maximum Likelihood criterion [31] in the general re-estimation method, which is an iterative procedure, starting from the initial values set by Segmental K-Means Procedure [32]. Thus the re-estimation formulas for λ_{GSM} are as follows.

$$\hat{c}_j^{GSM}=\frac{\sum_{n=1}^{N+1}\sum_{t=1}^{T(n)}\theta_j^{(n)}(t)}{\sum_{n=1}^{N+1}\sum_{t=1}^{T(n)}\alpha_t^{(n)}\cdot\beta_t^{(n)}} \qquad j=1,2,\cdots,M \tag{9-11}$$

$$\theta_j^{(n)}(t)=\begin{cases} c_j^{GSM} p_j[y_1^{(n)}]\beta_1^{(n)} & t=1 \\ c_j^{GSM} p_j[y_t^{(n)}]\alpha_{t-1}^{(n)}\beta_t^{(n)} & t=2,3,\cdots,T(n) \end{cases} \tag{9-12}$$

$$\alpha_t^{(n)}=\begin{cases} p[y_t^{(n)}]\alpha_{t-1}^{(n)} & t=2,3,\cdots,T(n) \\ p[y_1^{(n)}] & t=1 \end{cases} \tag{9-13}$$

$$\beta_t^{(n)} = \begin{cases} p[y_{t+1}^{(n)}]\beta_{t+1}^{(n)} & t = 1, 2, \cdots, (T(n) - 1) \\ 1 & t = T(n) \end{cases} \quad (9\text{-}14)$$

$$\hat{\mu}_j^{GSM} = \frac{\sum_{n=1}^{N+1}\sum_{t=1}^{T(n)} \theta_j^{(n)}(t) y_t^{(n)}}{\sum_{n=1}^{N+1}\sum_{t=1}^{T(n)} \theta_j^{(n)}(t)} \qquad j = 1, 2, \cdots, M \quad (9\text{-}15)$$

$$\hat{\Sigma}_j^{GSM} = \frac{\sum_{n=1}^{N+1}\sum_{t=1}^{T(n)} \theta_j^{(n)}(t) \cdot (y_t^{(n)} - \hat{\mu}_j^{GSM})(y_t^{(n)} - \hat{\mu}_j^{GSM})^T}{\sum_{n=1}^{N+1}\sum_{t=1}^{T(n)} \theta_j^{(n)}(t)} \qquad j = 1, 2, \cdots, M$$

(9-16)

In Equations (9-11)-(9-16), $p[y_t^{(n)}]$ is same as $p[y_t^{(n)} \mid \lambda_{GSM}]$, and $p_j[y_t^{(n)}]$ is computed as:

$$p_j[y_t^{(n)}] = p_j[y_t^{(n)} \mid \lambda_{GSM}]$$
$$= \frac{1}{(\sqrt{2\pi})^{D/2} \cdot (|\Sigma_j^{GSM}|)^{1/2}} \cdot exp(-\frac{1}{2}(y_t^{(n)} - \mu_j^{GSM})^T (\Sigma_j^{GSM})^{-1} (y_t^{(n)} - \mu_j^{GSM})),$$

(9-17)

where $\hat{c}_j^{GSM}$, $\hat{\mu}_j^{GSM}$ and $\hat{\Sigma}_j^{GSM}$ are the latest values, c_j^{GSM}, μ_j^{GSM} and Σ_j^{GSM} are their corresponding values by the end of the last iteration. The above re-estimation procedure for a new user is time-consuming, especially when the system has a large number of users. The best usage of this re-estimation method is to utilize the system idle time, if the registration time is not one of the main concerns of system users.

REAL-TIME APPLICATION

In GSMSV method, the global speaker model is a critical factor, directly influencing its system performance and practical applications. Therefore, the method to establish the global speaker model is an important issue worthy of discussing.

Since λ_{GSM} is obtained by using all the training data of current users, the training procedure will take a long time, especially when the system has a large number of users. A real-time speaker verification system employing GSMSV method is developed, where the parameters of λ_{GSM} are modified by Equations (9-11)-(9-16). The computer used is P-II 233. The speech duration of each registration is 10 seconds.

When the system users are fewer, the waiting time for registration is acceptable. For instance, if there are 10 users, the waiting time for a new user is about 4.5 minutes. However, with the increase in the number of system users, the registration time increases proportionally. If the current system has 100 users, the waiting time for a new registration is about 40 minutes. This is unacceptable and intolerable for real-time applications. Since the training time is consumed mainly on re-training λ_{GSM}, the aim of the adaptive re-estimation method is to shorten the registration time for a new user so that GSMSV can be efficiently employed to real-time applications.

Different from the general re-estimation method, the adaptive one updates the parameters of λ_{GSM} in a one-shot step. The initial values are set as those modified by the last registration. The adaptive re-estimation formulas for updating λ_{GSM} parameters are as follows:

$$\hat{c}_j^{GSM} = \frac{(1-\rho)\cdot\sum_{n=1}^{N}\sum_{t=1}^{T(n)}\theta_j^{(n)}(t) + \rho\cdot\sum_{t=1}^{T(N+1)}\theta_j^{(N+1)}(t)}{(1-\rho)\cdot\sum_{n=1}^{N}\sum_{t=1}^{T(n)}\alpha_t^{(n)}\cdot\beta_t^{(n)} + \rho\cdot\sum_{t=1}^{T(N+1)}\alpha_t^{(N+1)}\cdot\beta_t^{(N+1)}} \qquad j=1,2,\cdots,M \tag{9-18}$$

$$\hat{\mu}_j^{GSM} = \frac{(1-\rho)\cdot\sum_{n=1}^{N}\sum_{t=1}^{T(n)}\theta_j^{(n)}(t)y_t^{(n)} + \rho\cdot\sum_{t=1}^{T(N+1)}\theta_j^{(N+1)}(t)y_t^{(N+1)}}{(1-\rho)\cdot\sum_{n=1}^{N}\sum_{t=1}^{T(n)}\theta_j^{(n)}(t) + \rho\cdot\sum_{t=1}^{T(N+1)}\theta_j^{(N+1)}(t)} \qquad j=1,2,\cdots,M \tag{9-19}$$

$$\hat{\Sigma}_j^{GSM} = \frac{(1-\rho)\cdot A + \rho\cdot B}{(1-\rho)\cdot\sum_{n=1}^{N}\sum_{t=1}^{T(n)}\theta_j^{(n)}(t) + \rho\cdot\sum_{t=1}^{T(N+1)}\theta_j^{(N+1)}(t)} \qquad j=1,2,\cdots,M \tag{9-20}$$

where

$$A = \sum_{n=1}^{N}\sum_{t=1}^{T(n)}\theta_j^{(n)}(t)\cdot(y_t^{(n)} - \hat{\mu}_j^{GSM})\cdot(y_t^{(n)} - \hat{\mu}_j^{GSM})^T \tag{9-21}$$

$$B = \sum_{t=1}^{T(N+1)}\theta_j^{(N+1)}(t)\cdot(y_t^{(N+1)} - \hat{\mu}_j^{GSM})\cdot(y_t^{(N+1)} - \hat{\mu}_j^{GSM})^T \ . \tag{9-22}$$

In Equations (9-18)-(9-22), $\theta_j^{(n)}(t)$, $\alpha_t^{(n)}$ and $\beta_t^{(n)}$ are computed as Equations (9-12)-(9-14). ρ ($0<\rho<1$) is a weighting coefficient measuring the contribution of the new registration speech to updating the global speaker model. The greater the value of ρ, the more the contribution owing to the new training data. Since the adaptive re-estimation procedure starts from the last modified parameter values, the setting of ρ will determine the verification performance after the system scale is augmented.

Without ρ (or it is too small), the system would not adapt to the new user. For example, when the number of system users is large, the contribution of the new speech will be overwhelmed by that of the existing users. Thus the updated global speaker model is similar to the last one, and the system is not suitable to verify the claim to the new user. If ρ is too large, the global speaker model is changed exceedingly to accommodate the new user, but it will not be applicable to the old ones.

ρ may be set as a specific value according to the experience. Or, it may be set dynamically according to the change of the number of reference speakers. For example, ρ can be decided by $\rho = \varepsilon \cdot R(N_{users})$, where ε is a coefficient showing the greatest portion of the new speaker's contribution to all the old speakers' contribution, and $R(N_{users})$ is a function whose values increase with the number of valid users, N_{users}, and whose limitation value is 1, e.g., $R(N_{users})$ can be $\frac{N-1}{N}$, sigmoid function, or some other functions.

Two main points lead to the decrease on computation overhead. First, the parameter modification is no longer an iterative procedure as the general one. The adaptive updating is completed in one phase. The other is that updating starts from much better initial values so as to avoid the step of setting proper initial values, which is important and time-consuming for getting high performance systems.

9.4 Experimental Results

DATABASE AND EXPERIMENTAL SETTINGS

Data used in the following experiments come from a Mandarin speech database *863Bag* provided by the *State Education Commission* of China. Speech data of 50 persons (25 females and 25 males) is used. Each person uttered 50 sentences, 15 of which are used as the training data, and the other 35 sentences are used as the test data. Each test is on one sentence. Sentence duration ranges from 5.6 seconds to 1.2 seconds. The average duration of training data for each speaker is about 60 seconds, and that of each test utterance is about 3.5 seconds.

30 selected speakers (15 females and 15 males) are regarded as reference speakers. Tests on data of reference speakers consist of closed set test, in which the speech of one reference speaker makes up the disguised utterance to other reference speakers. Tests on data of 20 other speakers (10 females and 10 males) who are regarded as outside impostors constitute open set test.

Equal error rate is used to measure the performance of different speaker verification methods. The equal error rate is a *posterior* error rate, and at this equal error rate, the decision boundary is set to make the error rate of false rejection be equal to that of false acceptance. The *posterior* equal error rate is a convenient measure of the degree of separation between true and false speaker scores and, therefore, a useful predictor of speaker verification performance.

A feature vector consists of 16 cepstrum coefficients acquired from auto-relation analysis, 16 dynamic cepstrum coefficients and a dynamic energy [33]. GMM speaker model has 64 mixtures. ρ is set to be a specific value 0.2. The computer used for experiments is P-II 233.

STATISTICAL ANALYSIS

In this experiment, the statistical likelihood scores of both CSV and GSMSV methods are analyzed and compared. The likelihood scores of closed set test are recorded and the corresponding histograms are shown in Figure 9-4. The results of statistical analysis are also listed in Table 9-1.

The statistical results are calculated as follows: Let $\{s_1, s_2, \cdots, s_Q\}$ be the score to be analyzed, in which Q is the number of total scores. Some score values may appear more than one occurrence. Let $\{s'_1, s'_2, \cdots, s'_{Q'}\}$ be the unique score set, in which Q' is the number of unique scores. The appearing frequency of each unique score is computed and represented as $f(k)$. Thus the mean, m, and variance, v, can be computed by

$$m = \sum_{k=1}^{Q'} s'_k \cdot f(k) \tag{9-23}$$

and

$$v = \sqrt{\sum_{k=1}^{Q'} (s'_k - m)^2 \cdot f(k)} \ \cdot \tag{9-24}$$

The likelihood score difference, D, between valid users and impostors is obtained as:

$$D = (m_u - v_u) - (m_i + v_i) \tag{9-25}$$

in which m_u and m_i are the mean values of valid users and impostors, respectively, and v_u and v_i are their corresponding variances. It is obvious that D is a measurement of the system capability. The greater the value of D, the more powerful the distinguishing ability. If $D > 0$, the likelihood scores of valid users are generally higher than those of impostors; if $D = 0$, the likelihood scores of valid users and impostors are usually overlapped; and if $D < 0$, the likelihood scores of valid users are generally lower than those of impostors; this shows the distribution irrationality and the system unusability.

The following interesting observations can be obtained from both Figure 9-4 and Table 9-1:

1) For the speech of either valid users or impostors, the variance of GSMSV likelihood scores is much smaller than that of CSV method. This illustrates that GSMSV makes the distribution of the likelihood score more compact.
2) The difference between the likelihood scores of GSMSV valid users and impostors is greater than that of CSV method. It demonstrates that GSMSV enlarges the distance between valid users and impostors, so its distinguishing ability is more powerful than that of CSV method.
3) The GSMSV likelihood score overlap between valid users and impostors is smaller than that of CSV method, therefore the boundary between valid users

and impostors is more explicit and the threshold can be more conveniently set by GSMSV.

Table 9-1.
Statistical analysis result of the likelihood scores for CSV method and GSMSV method.

	Valid speakers		*Impostors*		*Likelihood score difference between valid users and impostors*
Method	Mean	Variance	Mean	Variance	D
CSV	3354.79	584.36	1762.85	730.89	276.69
GSMSV	285.32	188.13	-1312.71	641.47	768.43

Table 9-2. Performance comparison (equal error rates and verification speed) of different methods.

Method	*Closed set test (%)*	*Open set test (%)*	*Average verification Speed (s)*
CSV	6.19	1.69	0.57
ASMSV (L=29)	0.19	1.06	17.26
GSMSV	0.59	0.51	1.15
ASMSV (L=1)	7.80	2.16	1.15

COMPARISON OF DIFFERENT METHODS

ASMSV method has the lowest equal error rates when an anti-speaker model consists of all of other reference speakers [24], thus in this experiment L is set to be 29. For the sake of comparison convenience, the results of the case ($L = 1$) are also given. Table 9-2 lists the equal error rates of CSV, ASMSV and GSMSV methods.

In Table 9-2, the equal error rates of both ASMSV ($L = 29$) and GSMSV methods are all significantly lower than those of CSV method. This shows the necessity to normalize the likelihood score of the conventional scoring method. For closed set test, the equal error rate of GSMSV is higher than that of ASMSV, but for open set test the equal error rate of GSMSV is much lower than that of ASMSV. One point should be noted that the equal error rates under the case ($L = 29$) are the best results that ASMSV can reach. Therefore when L is smaller, the superiority of GSMSV over ASMSV will be much more prominent.

Table 9-2 also lists the average time needed by each method to verify an input utterance. It costs ASMSV over 17 seconds to verify an utterance, while GSMSV spends only about 1 second. Although for the closed set test under the case of $L = 29$, ASMSV has a lower equal error rate than that of GSMSV, its verification speed is so slow that the system may be unusable for practical applications. If ASMSV spends 1 second to verify an utterance, *i.e.*, an anti-speaker model consists of one reference model, the equal error rates for closed set test and open set test are 7.80% and 2.16%, respectively. They are much higher than the corresponding values of GSMSV method.

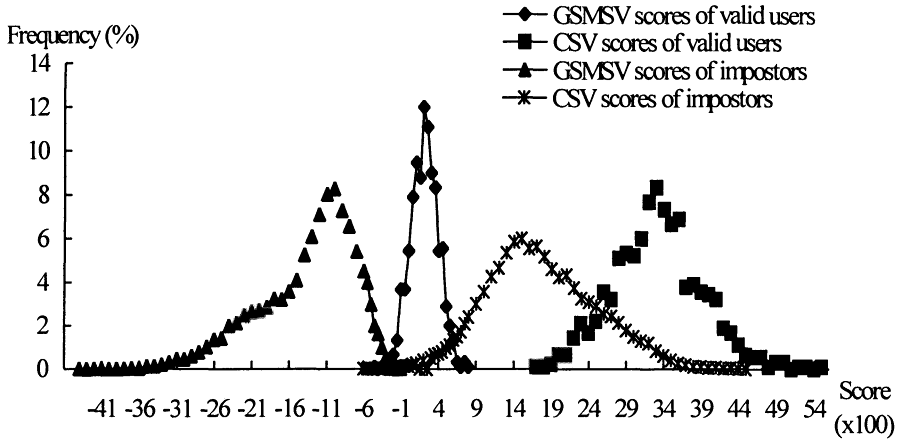


Figure 9-4. Likelihood score histograms of the closed set tests for CSV and GSMSV methods.

The above experiments show that GSMSV method has the advantages on both low equal error rates and fast verification speed. Either CSV method or ASMSV method can not keep low equal error rates and fast verification speed at the same time.

EXPERIMENTAL RESULTS OF GENERAL AND ADAPTIVE GSMSV

A series of experiments are performed on different number of reference speakers. Experiments start from 2 reference speakers (1 female, 1 male). At this point, the parameter values (2 reference speakers) of the adaptive re-estimation method are same as those of the general method. And then in the following experiments, two reference speakers (1 female and 1 male) are added each time. In each experiment, λ_{GSM} is updated two times, by firstly using the training data of the new female user, and then using those of the new male user. After these modifications, λ_{GSM} is used for verification tests. The test results of GSMSV (with adaptive re-estimation and general re-estimation methods) and CSV are depicted in Figure 9-5 and Figure 9-6 for closed set test and open set test, respectively.

The following observations can be obtained from the test results. For both closed set test and open set test, GSMSV with either the general re-estimation or the adaptive re-estimation method has much lower equal error rates than CSV method. The equal error rates of the adaptive re-estimation approximate to those of the general re-estimation, and are a littler higher on only a few cases. When the system has 30 users, the equal error rates of GSMSV with the general re-estimation method are about 0.59% and 0.51% for the closed set test and the open set test, while those of GSMSV with the adaptive re-estimation method are about 0.67% and 0.66%, respectively.

When two new users (1 female and 1 male) are added into the system, the training time for GSMSV is recorded and depicted in Figure 9-7. The training time for the adaptive re-estimation increases a little with extension of the system, while that for the general re-estimation increases proportionally and significantly. When the verification system has 30 users, GSMSV with the general re-estimation method needs about 148 minutes for the registrations of two new users, but GSMSV with the adaptive re-estimation method needs only 11 minutes.

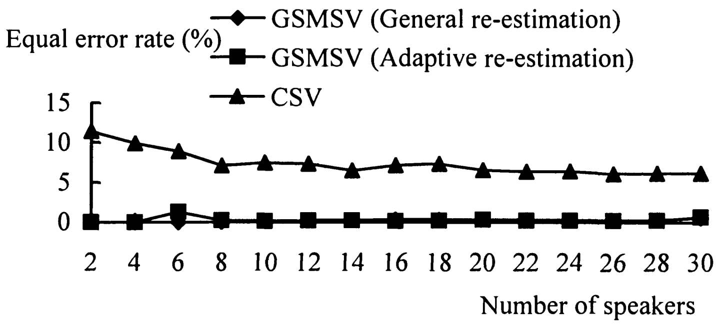


Figure 9-5. GSMSV test results of the closed set test with different re-estimation methods to update the global speaker model.

The effectiveness and practicability of the adaptive re-estimation method has been fully illustrated by these serial experiments. Compared to the general method, the adaptive one decreases the registration time significantly without increasing the equal error rates. Sometimes the adaptive method even has a lower equal error rate than the general one.

One may ask how about the experimental results when the number of users is larger than 30 such as 200 or 500 or more than 500. The following is our opinion. Since the equal error rates of speaker verification systems are insensitive to the number of users [35], we believe that when the number of users is very large, the equal error rates of GSMSV method will not increase greatly, but rather increase little or even keep steadily low. As a result, the equal error rates using GSMSV with adaptive re-estimated method and the general re-estimated method respectively are still comparable, while the adaptive one has a significant advantage on registration time.

This chapter begins with introduction to background and basic concepts for speaker recognition. Then, some principles of speaker recognition are described. Moreover, we propose a novel speaker verification method, GSMSV. GSMSV has the following characteristics:

1) The separation between speakers is large and explicit;

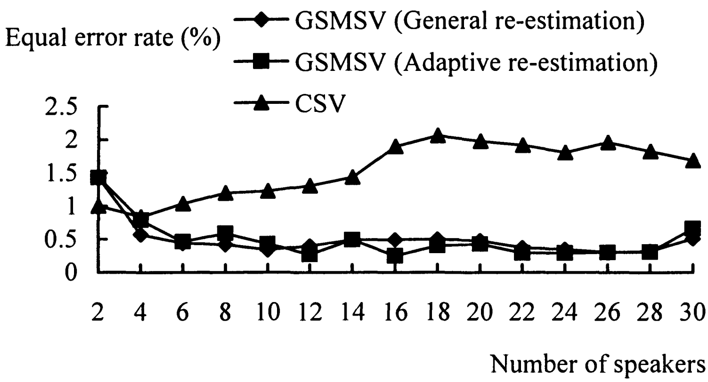


Figure 9-6. GSMSV test results of the open set test with different re-estimation methods to update the global speaker model.

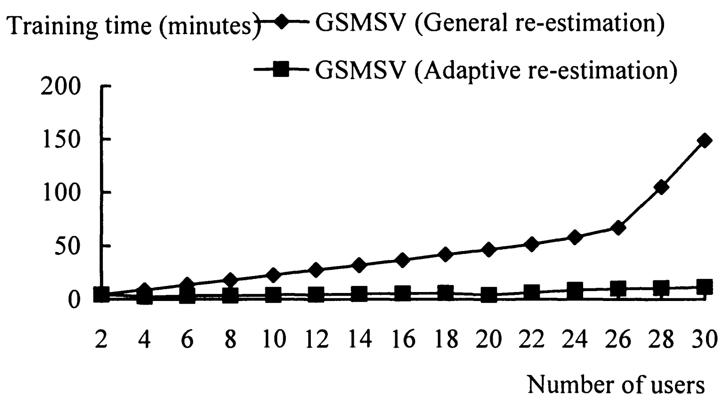


Figure 9-7. Training time for GSMSV with different re-estimation methods to update the global speaker model when two new users register into the verification system.

2) The distinguishing ability of the system is powerful;
3) Verification speed is fast;
4) It is adaptable to speaking speed.

In addition, since an adaptive re-estimation approach to updating the global speaker model is suggested to shorten the waiting time for a new user, GSMSV method can also be perfectly utilized to real-time applications. The method and techniques presented here provide a promising way for realizing practical speaker verification systems with high performance.

There have been many recent advances and successes in speaker recognition technology. However, there are still many problems for good solutions remaining to be found. Most of these problems arise from variability, including speaker-generated variability and variability in channel and recording conditions. It is very important to investigate the feature parameters that are stable over a long period, insensitive to variations in voice quality such as those due to voice disguise or colds. It is also important to develop a method to cope with the problems of distortion due to telephone sets and channels and background and channel noises [5].

References

[1] D.O. Shaugnessy, "Speaker Recognition," *IEEE ASSP Magazine*, vol. 3, no. 4, pp. 4-17, 1986.

[2] H.J. Kunzel, "Current Approaches to Forensic Speaker Recognition," *ESCA Workshop on Automatic Speaker Recognition, Identification and Verification*, pp. 135-141, 1994.

[3] Rosenberg, "Automatic Speaker Verification: A Review," *Proc. of IEEE*, vol. 64, pp. 475-487, Apr. 1976.

[4] J.P. Campbell, "Speaker Recognition: A Tutorial," *Proc. of IEEE*, vol. 85, no. 9, pp. 1437-1462, Sep. 1997.

[5] A. Furui, "Recent Advances in Speaker Recognition," *Proc. of the first International Conference on Audio- and Video-Based Biometrics Perso Authentication*, pp. 237-252, 1997.

[6] Higgins, L. Rahler and J. Porter, "Speaker Verification Using Randomized Phrase Prompting," *Digital Signal Processing*, vol. 1, pp. 89-106, 1991.

[7] T. Matsui and S. Furui, "Text-independent Speaker Recognition Using Vocal Tract and Pitch Information," *Proc. of the International Conference on Spoken Language Processing*, Kobe, pp. 137-140, 1990.

[8] J.M. Richard, X.Y. Zhang, and R. Ramachandran, "Robust Speaker Recognition - A Feature-Based Approach," *IEEE Signal Processing Magazine*, pp. 58-71, Sep. 1996.

[9] L.R. Rabiner and B.H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[10] S. Furui, "Cepstral Analysis Technique for Automatic Speaker Verification," *IEEE Trans. Acoust., Speech, Signal Process.*, 29:254-272, 1981.

[11] D. Naik, "Pole-filtered Cepstral Mean Subtraction," *Proc. of ICASSP*, vol. 1, pp. 157-160, 1995.

[12] K.T. Assaleh and R.J. Mammone, "New LP-Derived Features for Speaker Identification," *IEEE Trans. Speech and Audio Processing*, vol. 2, no. 4, pp. 630-638, 1994.

[13] M.S. Zilovic, R. P. Ramachandran and R.J. Mammone, " Speaker Identification Based on the Use of Robust Cepstral Features Obtained from Pole-zero Transfer Functions," *Submitted to IEEE Trans. Speech, Audio Processing*, 1995.

[14] D.A. Reynolds and R.C. Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models," *IEEE Trans. on Speech and Audio Processing*, vol. 3, no. 1, pp. 72-83, Jan. 1995.

[15] H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-26, no. 1, pp. 43-49, 1978.

[16] F.K. Soong, A.E. Rosenberg, L.R. Rabiner, and B.H. Juang, "A Vector Quantization Approach to Speaker Recognition," *AT&T Tech. J.*, vol. 66, no. 2, pp. 14-26, 1987.

[17] Higgins, "YOHO Speaker Verification," Presented at the *Speech Research Symposium*, Baltimore, MD, 1990.

[18] Higgins, L. Bhaler, and J. Porter, "Voice Identification Using Nearest Neighbor Distance Measure," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 375-378, 1993.

[19] E. Rosenberg, C.H. Lee, and F.K. Soong, "Sub-word Unit Talker Verification Using Hidden Markov Models," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 1, pp. 269-272, Apr. 1990.

[20] X. Jiang, Z.Y. Gong, F. Sun, *et al*, "A Hybrid Speaker Recognition System Based on the Auditory Path Model," *Proceedings of WCNN*, pp. 598-601, 1993.

[21] M. Savic, "Variable Parameter Speaker Verification System Based on Hidden Markov Modeling," *Proceedings of 1990 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 281-284, 1990.

[22] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Werner Rheinboldt (Ed.), pp. 10-15. San Diego, California, 1990.

[23] K.P. Li and J.E. Porter, "Normalization and Selection of Speech Segments for Speaker Recognition Scoring," In *Proc. of ICASSP*, New York, 595-598, 1988.

[24] C.S. Liu, H.C. Wang, and C.H. Lee, "Speaker Verification Using Normalized Log-Likelihood Score," *IEEE Trans. on Speech and Audio Processing*, vol. 4, no. 1, pp. 57-60, Jan. 1996.

[25] E. Rosenberg, J. Delong, C.H. Lee, B.H. Juang, and F.K. Soong, "The Use of Cohort Normalized Scores for Speaker Recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 1, pp. 599-602, Oct. 1992.

[26] K.P. Markov and S. Nakagawa, "Text-Independent Speaker Recognition Using Non-Linear Frame Likelihood Transformation," *Speech Communication*, vol. 24, pp. 193-209, 1998.

[27] Y.Y Zhang, D. Zhang, and X.Y Zhu, "Speaker Verification by Removing Common Information", to appear in *IEE Electronics Letters*.

[28] N. Tishby, "On The Application of Mixture AR Hidden Markov Models to Text Independent Speaker Recognition," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 39, pp. 563-570, Mar. 1991.

[29] T. Matsui, "Speaker Adaptation of Tied-Mixture Based Phoneme Model for Text-Prompted Speaker Recognition," *Proceedings of 1994 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 125-128, 1994.

[30] X.D. Huang and M.A. Jack, "Semi-continuous Hidden Markov Models for Speech Signals," *Computer Speech and Language*, vol. 3, pp. 239-251, 1989.

[31] L.A. Liporace, "Maximum Likelihood Estimation for Multivariate Observations of Markov Sources," *IEEE trans. on Information Theory*, vol. IT-28, no. 5, pp. 729-734, Sep. 1982.

[32] L.R. Rabiner, "Recognition of Isolated Digits Using Hidden Markov Models with Continuous Mixture Densities," *AT&T Tech J.*, vol. 64, no. 6, pp. 1211~1222, July-Aug. 1986.

[33] C.H. Lee, "On Robust Linear Prediction of Speech," *IEEE Trans. on Acoustics, Speech and Signal Processing*, pp. 642-650, 1988.

[34] S. Furui, *Digital Speech Processing, Synthesis, and Recognition*, Chapter 9, "Speaker recognition," pp. 291-309, Marcel Dekker Inc., 1989.

10 SIGNATURE SYSTEM

Handwritten signature is one of the most popular ways to verify one's identity. In Section 10.1, we introduce some basic principles and methods of signature verification systems. Two kinds of signature systems, off-line and on-line, are discussed in Section 10.2 and 10.3, respectively. An on-line signature verification system based on Dynamic Time Warping (DTW) is described in Section 10.4. Section 10.5 will present an on-line signature verification application in the Internet/Intranet.

10.1 Signature Verification

Currently, security practice always involves PIN number, password, and access card. However, these tokens are not very reliable as they may be forgotten or lost. Nobody can restrict or prevent any unauthorized person from using them in automatic machine verification environment.

Automatic signature verification is one of the most practical ways to verify human's identity. Signature cannot be lost, stolen or forgotten and signature has a fundamental advantage in that it is the customary way of identifying an individual in daily operations. Signature verification can be used in many applications such as security, access control, or financial and contractual matters.

Till now, the mechanism of the generation of human's signature is not very clear, one of the models was described like this [1]: some central nervous mechanisms within the brain is fire, with the predetermined intensity and duration, the nerve network activates the proper muscles in a predetermined order. The motion of the pen on the paper, resulting from muscle contraction/relaxation, leaves a partial trace of the trajectory of the pen tip. The design of a signature system is based on that people do not write according to a standard penmanship, and the deviation from the norm is individual dependent.

There are two types of signature verification systems: on-line signature verification system and off-line signature verification system. In off-line systems, people write the signature on paper, digitize it through an optical scanner or a camera, finally the systems determine the genuineness of the signature through examining the overall shapes of the signatures. But in on-line systems, the signature trace is acquired in real time with a digitizing tablet or an instrumented pen or other specialized hardware. All these devices capture the static and dynamic information of the signature during the signing process. Since an on-line system can utilize not only the shape information of the signature but also the dynamic information, it is also named as dynamic signature verification system, and an off-line system is named as static signature verification system.

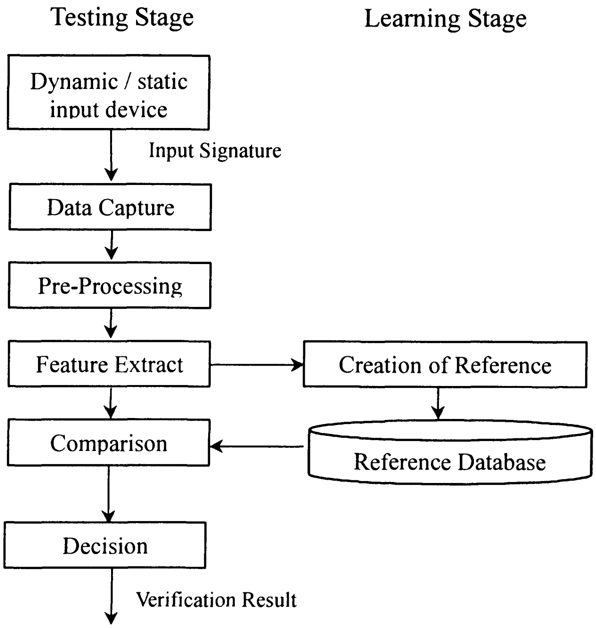


Figure 10-1. Signature verification system flowchart.

The process of signature verification often consists of a learning stage and a testing stage, as shown in Figure 10-1. In the learning stage, the verification system uses the features extracted from one or several training samples to build a reference signature database. Each signer gets its own ID in the database. This ID is used as the unique identity of this signer, and linked to the signer's reference in the database. In the testing stage, the user inputs the ID and signs on the input device. Then the system uses this ID information to extract the reference in the database, and compares the features extracted from the input signature with the reference. Finally the decision process checks out whether the test signature is genuine or not.

It is necessary to have an ability to prevent various forgeries for a signature verification system. Depending on testing condition and environment, these forgeries can be divided into three types [1]:

a. 'Simple' forgery: where the forger makes no attempt to simulate or trace a genuine signature.
b. 'Substitution' or 'random' forgery: where the forger uses his/her own signature instead of the signature to be tested.
c. 'Freehand' or 'skilled' forgery: where the forger tries and practices imitating as closely as possible the static and dynamic information of a signature.

In the research area of signature verification problem, Type I and Type II error rates are used to evaluate a verification system (see Chapter 1.3). Type I error rate and Type II error rate is usually called False Reject Rate (FRR) and False Acceptance Rate (FAR) respectively [1]. To minimize the Type II errors, which represent the acceptance of the counterfeited signatures (forgeries) will normally increase the Type I errors, which are the rejections of genuine signatures. In most cases, Type II error rate is considered to be more important, but it is not a must. This will depend on the purpose, design, characteristics and application of the verification systems. If the system requests a high security, False Accept Rate should reduced to its lowest; if the security is not so strict, the system can be adjust to its lowest Average False Rate. The Type I and Type II errors can be adjusted by the changing of threshold, as it is shown in Figure 10-2.

There are two ways to distinguish the relations between Type I Error Rate and Type II Error Rate, called Error tradeoff curve of FAR *vs.* FRR (Figure 10-2) and Error Rate *vs.* Threshold (see Figure 1-8 in Chapter 1).

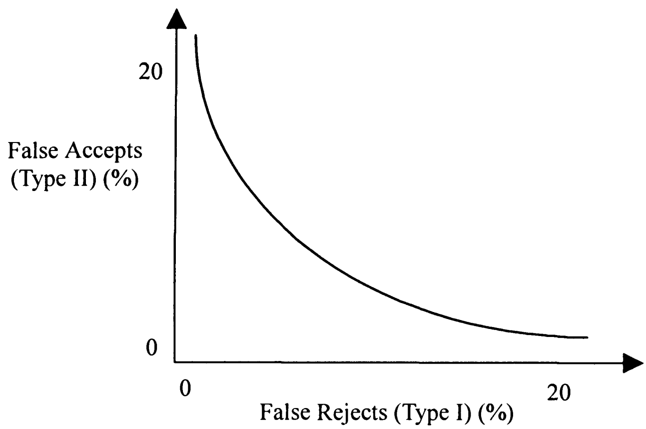


Figure 10-2. Error tradeoff curve (FAR *vs.* FRR).

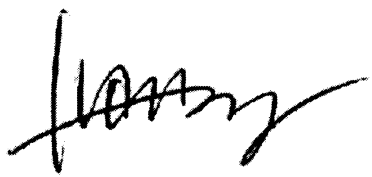


Figure 10-3. Off-line signature image.

10.2 Off-line Signature

In an off-line system, the signature is written on the paper, and a scanner or a camera acquires the signature data. The input data can be seen as a signature image $L(x,y)$. Figure 10-3 is such a sample of off-line signature image.

Since the acquisition of off-line signature is somewhat complex, and the distinguishability is weak, the off-line signature systems are usually used as an aid in the legal cases to identify criminals. In order to overview a basic off-line signature system, three main stages in the system are described as follows.

PREPROCESSING

In this stage, some standard preprocessing algorithms are used, such as smoothing, thinning and skeleton generation, segmentation, and normalization. In a practical system, not all these algorithms are needed. It depends on what features we wants to work with.

a. *Thresholding*: The grayness of each pixel is compared to a threshold, and the pixel is assigned to one of two categories depending on whether the threshold is exceeded or not. After thresholding a gray scale image is changed to a binary one.
b. *Smoothing and Normalization*: In order to get rid of isolated noises, smoothing technique is usually involved. Normalization technique is also needed. After normalization, the image range is changed to a certain scale. Normalization methods include two types: linear and non-linear. In signature system, the linear method is usually used since the shape of the signature should be reserved.
c. *Thinning:* It is the technique to simplify pictorial forms. After thinning, the width of each stroke in signature image is changed to one pixel, and only the skeleton is left.

FEATURE EXTRACTION

The features in off-line signature verification system can be classified into two types: text-insensitive and text-sensitive features.

The so-called text-insensitive features are the features that exist no matter what people write. Text-insensitive features extraction is usually regarded as texture analysis, some transform approaches and histograms are used here. It was reported [2] that the low frequency band of Fourier spectrum is useful, and the features can be extracted from the frequency distribution of different global or local properties. Text-insensitive features are often used in the off-line writer identification cases. The description of the signature in these approaches is rough, and the discrimination between two different signatures is very weak.

Text-sensitive features quite depend on what people write. Most text-sensitive features are geometric and topologic features. Several approaches have been studied to find some close loops and special points, such as edge points and cross points to represent the signature [3,4]. Some approaches use the outline of the signature and these features include skeleton cursive, upper envelops and lower envelops [5]. The signature shape is also described with some grid information features and texture features [3]. Besides of these geometric features, a set of high-pressure features was also developed to extract timing and dynamic information indirectly from the signature image [5,6].

Comparing with text-insensitive features, text-sensitive features present a higher discrimination between signatures.

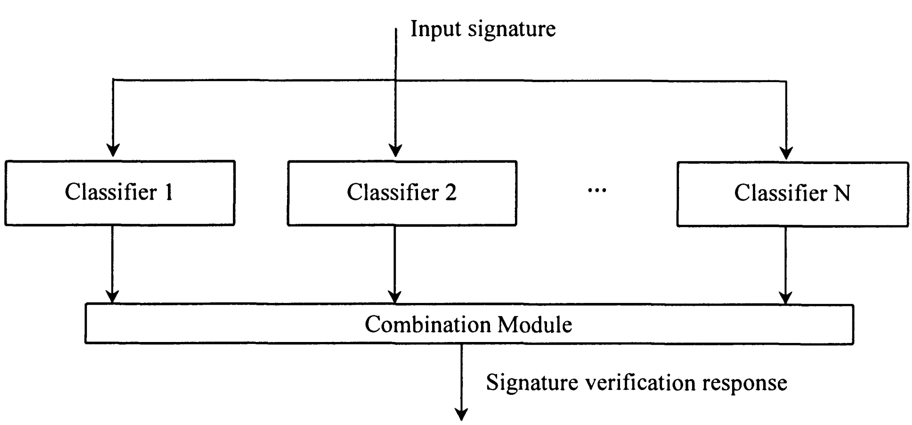


Figure 10-4. Architecture of multi-classifier-based approach.

COMPARISON ALGORITHM

Since most of the off-line signatures are represented as feature vectors, the comparison algorithms are mainly based on statistical method and artificial neural network (ANN). The statistical methods that were used to verify off-line signature include nearest neighbor classifier, K-nearest neighbor classifier, linear classifier, and threshold classifier. Neural networks include fuzzy ARTMAP, back propagation network, and multi-layer perception neural network. A few related methods are here introduced as

a. *Minutiae matching*: In this method, the signature image is firstly thinned, and then some special points are extracted, such as end points, turn points, cross points, *etc.* Finally relaxation matching is used to calculate the similarity between test and reference signatures.
b. *Neighbor classifier*: Feature vector is an ordinary way to represent off-line signatures, and two classic statistical algorithms: nearest neighbor classifier and K-nearest neighbor classifier, are usually used in this case. The test signature is compared with each signer's reference (or references), and classified to the nearest signer (or the signer with maximum number in the nearest K references).
c. *Neural network*: The multi-layer perception neural network is a commonly used method. As for the input of neural network, some approaches use the feature vector extracted in the feature extraction stage, others input the signature image directly, and extract the features within the network.

In order to improve the system's performance, many approaches are multi-classifier-based [3,5,7,8]. Figure 10-4 shows the architecture of these systems. In such a multi-classifier-based system, each classifier verifies the signature independently, and outputs its similarity measure respectively. The combination module educes the final conclusion according to some algorithm.

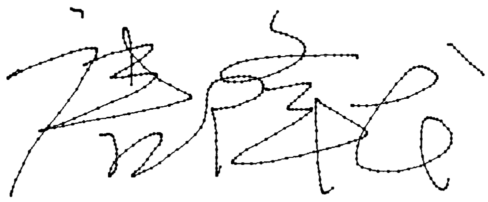


Figure 10-5. On-line signature trace.

The basic rule for combination is majority vote [7]. The test signature is sent into each classifier, which decides to accept or reject it. If the majority of classifiers are positive responses, then this signature is accepted; otherwise it is rejected. Some systems adopted the fuzzy integral vote method to do the combination [5]. The classifiers not only decide whether accept or reject the test signature, but also the acceptance degree, and then the combination module uses fuzzy theorem to calculate the integral acceptance degree, and decide its authenticity according to the comparison with threshold.

10.3 On-line Signature

On-line signature verification differs from off-line by data acquisition method. In on-line system, the signature trace is acquired in real time with a digitizing tablet or special pen. For this reason, the on-line system acquires not only the image of the signature, but also the dynamic features. It was noticed that these dynamic features reflect the unique habits of the signer, and are difficult to be imitated [9]. The on-line signatures are usually considered as, or represented by, mathematical time functions $F(t)$. Figure 10-5 shows an on-line signature sample in the trace, where the dots represent the sampling points of the pen motion, and the line connects these consecutive dots. How to utilize these dynamic features to detect forgeries is an important issue for the design of on-line signature system.

PREPROCESSING AND FEATURE EXTRACTION

The features of on-line signature can be divided into two classes: static features and dynamic features. They are also regarded as parameters and functions.

The parameters are usually represented as a feature vector: $P = (p_1, p_2, \cdots, p_m)^T$, where m is the dimension of the vector P and the number of parameters. These features are usually extracted from the whole process of signing, such as average of writing speed, max writing speed, curvature measurements, ratio of long to short stroke, segments length, *etc.*

The dynamic features regard the signature as the function of time, and concern with every instant of the signing. These features can be represented as a set of time function, include position *x(t)*, *y(t)*, velocity *v(t)*, acceleration *a(t)*, pressure *p(t)*, *etc.* In some method, the pen-tilt features are also used [10]. The pen tip motion during pen-up to pen-down can also be recorded and used. This duration is called virtual stroke. These parts of trace are noticed harder to imitate, since they are totally invisible, but they are also variable, and less repeatable [11].

The preprocessing methods highly depend on what features the system wants to use. If the static features are used, the preprocess term is relatively simple, and its major work is to reduce spurious noise, detect gaps in the pen-down signals, and normalize the signature in some ways when it is necessary. But if the dynamic features are used, besides all these works, another important problem must be solved. Since almost every function comparison technique is dynamic, and relies directly on the segmentation of signature, the motion of signature must be firstly separated into a serial of segments. These segments are often defined as portions of signals concurrent to pen-down movements and are deduced from pressure signals, position signals or velocity signals [1,9].

There are many features that can be extracted from the signature. In order to improve the performance of the verification system, and reduce the complexity of the design of classifier, a feature selection stage is necessary. Statistic method is a regular way to select features from a large set of parameters [12,13]. Genetic algorithm is also adopted in some systems [11,14]. Experiment results showed that using an optimum individualized subset of features is better than using the whole feature set.

VERIFICATION

In the verification stage, before comparison, user must input an ID number into the system. The system extracts a set of references, which are linked to the input ID, from the reference database: $R = \{R_1, R_2, \cdots, R_N\}$, which means that the user enrolled N references in the database. And when the user writes on a special device, the system acquires the signing trace, and represents it as S. Then, a similarity measurement, $d(S, R_i)$, is calculated, and compared with a predefined threshold, T, to decide whether to accept or reject the input signature, S. Some rules are defined as

If one of the references R_i in R satisfies $d(S, R_i) > T$, then S is accepted;

Otherwise, if every reference R_i in R satisfies $d(S, R_i) < T$, then S is rejected.

Two major problems in a verification stage are how to measure the similarity between S and R_i, and how to define the threshold T.

Comparison Methods

There are several methods to measure the similarity of the input and reference signatures.

Weighted distance methods are usually used in the verification systems that are using parameters [12,15]. The input and reference signatures are expressed as feature vectors: S and R. And the similarity is evaluated with the weighted distance between S and R:

$$d(S, R) = (S - R)^T W^{-1} (S - R), \tag{10-1}$$

where W is a diagonal weight matrix. This method is the most direct way to evaluate the similarity of the input and reference, and the weights are obtained from a training set of genuine signatures.

Statistical models are also introduced in [16]. These methods deal with the parameter features, and the probability of the reference R on the condition of input S: $P(R \mid S)$, is regarded as the similarity of input and reference. With Bayes's theorem, this posterior probability can be calculated with prior probability $P(S \mid R)$.

$$P(R \mid S) = \frac{P(S \mid R)P(R)}{P(S)}. \tag{10-2}$$

The probability models are pre-established with the help of a training set of signatures.

Neural network approaches are presented both in parameter and function feature systems [17-19]. Compared with the weighted distance and statistic methods, the neural network possesses high adaptability, and its non-linear ability can contribute to lower the misclassification error rate greatly.

Dynamic Time Warping (DTW) and Hidden Markov Model (HMM) are two popular approaches to process time sequence signals. When introduced from speech system, they achieved great success in on-line character recognition systems. They can be also used in on-line signature systems [10,20-25]. DTW and HMM methods are mainly applied to the function features. They present a high precision in describing the signature and can tolerate some slight distortion of the genuine.

Compared with other approaches, these two methods also posses another advantage. Other approaches always need many training samples to build the reference base, so it is very bothersome for a person to write many times on a tablet. DTW and HMM methods can deal with all kinds of training sample number cases. Even if there is only one sample, they also can construct a reliable reference for the testing stage. This will be beneficial to practical applications.

Before a Hidden Markov Model is used, the signature must firstly be segmented into a sequence, and the elements in the sequence are attributed with several features of the corresponding segment. And then using each writer's training signature to construct a particular HMM, All these HMM are regarded as the references. In the testing stage, the input signature's attributed sequence will input to a particular HMM, and its generation probability can be calculated. Based on a threshold, a decision on whether to accept the signature as authentic or to reject as forgery is made. There are many model structures in HMM. Usually, a 'left-to-right' transition model gives the best performance. Figure 10-6 shows such a structure.

Threshold

The threshold problem is very important to a signature system, since it is in large part the determining factor in the performance [1]. If the threshold T is loosened, Type II error (false accept rate) will increase; Type I error (false reject rate) decrease, and *vice versa*. The selection of threshold intensively depends on the different application. In a general system, thresholds with the similar Type I and Type II error rate are proposed. This is because the total error rate is lowest in this situation. But on some security occasions, false acceptance is very serious and the system must select a threshold to make the Type II error lowest.

In most systems, threshold T is a single value for all users, and it can be called a global threshold. But some researches have selected an individual threshold T_i for each user i, and it was found that personalized adaptive thresholds can produce good performance of the system [9,26-28].

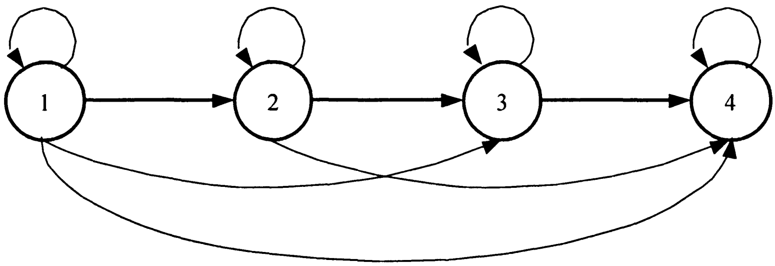


Figure 10-6. HMM model structure with left-to-right transition.

10.4 A Signature Verification System

In this section, we will introduce an on-line signature verification system based on Dynamic Time Warping (DTW).

On-line signature verification can be seen as the comparison of time signals, and each signature people write can be regarded as a function of time or a time varying sequence. The problem of comparison of time signals is the signals may be different in length, or have compressions, omissions or additional parts. DTW has the inherent robustness against local distortions, and its core technique *Dynamic Programming* (DP) is considered as an especially suitable algorithm for signature verification [20-24].

Figure 10-7 shows the flowchart of this system. Signature data is acquired from a digital table, the sample rate is 100 samples per second and spatial resolution is 1,200 points per inch. In order to reduce the computation complexity, a segmentation stage is adopted between preprocessing and feature extraction (Segmentation method will be discussed later). Instead of global features, this system extracts the shape and motion features in each segment, and then relaxation matching between reference segment sequence and testing segment sequence will be done with an improved DP algorithm. Finally the decision stage compares the matching distance and a predefined threshold to decide whether to accept the input or to reject.

DEFINITIONS AND NOTATIONS

The signature data is defined as a stroke sequence $S=(s^1,\cdots,s^N)$, where N is the stroke number of the signature. One stroke refers to the duration from the first pen down to pen up, and the *i*th stroke is defined as $s^i=(p_1^i,\cdots,p_{N_i}^i)$, N_i is its point number. One point is considered as *X*-*Y* coordinate: $p_j^i=(x_j^i,y_j^i)$, $j=1,\cdots,N_i$, which denotes the *j*th sample point in *i*th stroke. The sampling process is uniform, and time duration between the consecutive points is the same.

In other applications, some more information of the signature is considered, such as pressure and the interval during the pen is up. The collection of the pressure information highly depends on the instrument, till now most pressure sensors in the instrument pen are not very stable, and variable during the usage. The pen up duration is also instable, and is less precise. Since when the pen is in the air, if it is near the tablet, the movement can be captured; if it is a little far from the tablet, the movement will not be able to be induced. This system will only use the coordinates of the sampling points during the pen down period as the original information of the signature.

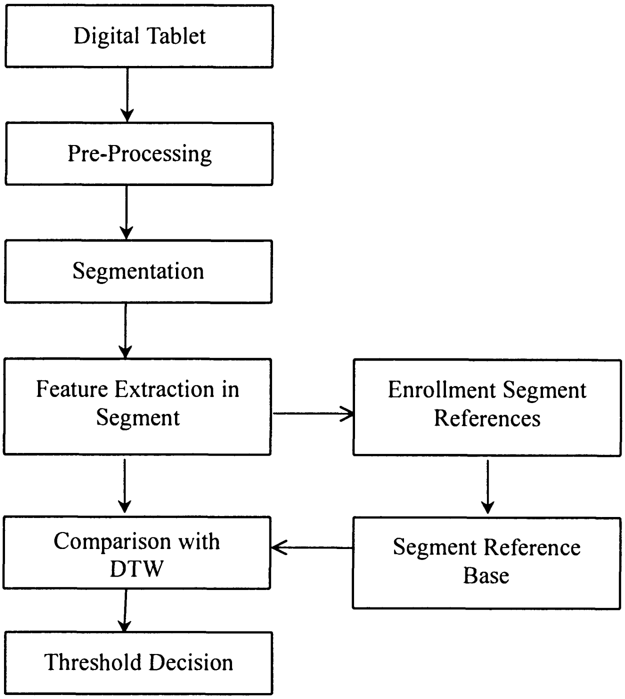


Figure 10-7. System flowchart.

Some more information of one stroke can be derived from the original point sequence.

The *i*th stroke's velocity can be defined as $V^i = (v_1^i, \cdots, v_{N_i}^i)$, where v_{i_j} is the velocity of the *j*th sampling point in the *i*th stroke:

$$v_j^i = \frac{1}{2}(\| p_{j-1}^i - p_j^i \| + \| p_j^i - p_{j+1}^i \|), \ j = 2, \cdots, N_i - 1,$$

$$v_1^i = \| p_1^i - p_2^i \|, \ v_{N_i}^i = \| p_{N_i - 1}^i - p_{N_i}^i \|. \tag{10-3}$$

The acceleration sequence can be defined as $A^i = (a_1^i, \cdots, a_{N_i}^i)$, where a_j^i is the acceleration of the jth sampling point in the ith stroke:

$$a_j^i = \frac{1}{2}(|v_{j-1}^i - v_j^i| + |v_j^i - v_{j+1}^i|),\ i = 2, \cdots, N_i - 1,$$
$$a_1^i = |v_1^i - v_2^i|,\ v_{N_i}^i = |v_{N_i-1}^i - v_{N_i}^i|. \tag{10-4}$$

The angular sequence can be defined as $\mathrm{B}_i = (\beta_{i_i}, \cdots, \beta_{i_{N_i-1}})$, where

$$\beta_j^i = \left[\operatorname{atan}\left(\frac{y_j^i - y_{j-1}^i}{x_j^i - x_{j-1}^i}\right)\right] \bmod \pi. \tag{10-5}$$

The curvature sequence can be defined as $C^i = (c_1^i, \cdots, c_{N_i}^i)$, where

$$c_j^i = \frac{1}{2}(|\beta_{j-1}^i - \beta_j^i| + |\beta_j^i - \beta_{j+1}^i|),\ i = 2, \cdots, N_i - 2,$$
$$c_1^i = |\beta_1^i - \beta_2^i|,\ c_{N_i-1}^i = |\beta_{N_i-2}^i - \beta_{N_i-1}^i|. \tag{10-6}$$

PREPROCESSING AND SEGMENTATION

Since the quality of the tablet using in this system is nice, there is less noise in signature signal. So the preprocessing stage is relatively simple and only an average filter is used,

$$x_i = \frac{1}{3}(x_{i-1} + x_i + x_{i+1}),\ y_i = \frac{1}{3}(y_{i-1} + y_i + y_{i+1}). \tag{10-7}$$

On-line signature is usually represented with dynamic features. These features can be seen as several functions of time, such as velocity, acceleration, *etc.* For example, let $s(t)$ and $r(t)$ be the function feature extracted from the test and reference signature, respectively. The most direct way to calculate the similarity of the two functions is to calculate the correlation of them,

$$d(s(t), r(t)) = \frac{\sum_t [s(t)r(t)]}{\left[\sum_t (s(t))^2 \sum_t (r(t))^2\right]^{\frac{1}{2}}}. \tag{10-8}$$

If the test signature is equal to the reference one, the correlation gets its maximum. But unfortunately, there are always some distortions in the signatures that people write twice, these distortions can cause a cumulative error. The best way to eliminate the cumulative error is to introduce DTW algorithm to calculate the similarity of the two signatures.

Many applications use the points in the signature trace as the basic components of DTW algorithm. Since the number of points in a signature can be very large, dynamic programming among these points will be time consuming. In order to reduce the time complexity, a signature is always separated into several segments, and then the dynamic programming is done within each segment. The process to separate one signature into several segments is called segmentation. Segmentation stage is very important in a DTW algorithm. The separation points must have some stability. If the separation points are various in the two signatures, the DTW algorithm will be failed.

The most popular segmentation method is to find points with some special characteristic, such as stroke starting points, stroke ending points, and turning points. The starting and ending points are easy to find in the input point sequence. To find the turning points will be somewhat difficult. It was studied that the motion speed of the pen is different between the linear section and the turning point, so some applications regard the local speed minima as segmentation points [9]. The most distinct characteristic of the turning points in a signature is its curvature variety. The signature trace appears some more curving on the turning points than other places. In this system, a segment method is introduced, which is directly based on this geometrical characteristic, and the local curvature maxima are regarded as the segmentation points.

After segmentation, one combined stroke s^i may be broken into several segments, and then the signature representation is changed to a segment sequence:

$$S = (z^i, \cdots, z^M)\,,\ M \geq N\,. \tag{10-9}$$

Compared with Figure 10-5, its on-line signature after segmentation is shown in Figure 10-8. Thick dots represent the separation points.

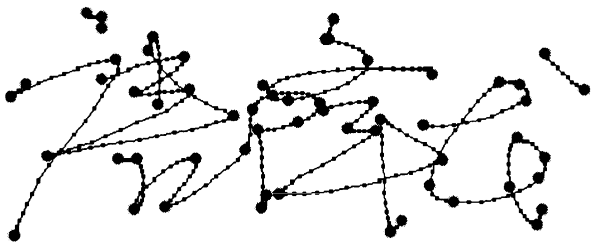


Figure 10-8. Segmentation of signature.

VERIFICATION

The task of the verification stage is to compare the similarity of the test signature's segment sequence $U = (u^1, \cdots, u^N)$, and reference signature's segment sequence $R = (r^1, \cdots, r^M)$. Since the length of the two sequences may be not equal, the segmentation result should have some distortion. Some segments would be omitted and some would be unnecessary. Also a dynamic time warping on the time axis is necessary.

The DTW algorithm aims at finding a time warping function, $i = w(j)$, and uses this function to map the input sequence's time axis to the reference sequence's time axis nonlinearly. The warping function, $w(j)$, should satisfy this condition:

$$D = \min_{w(j)} \sum_{j=1}^{M} d(u^{w(j)}, r^j)\,. \tag{10-10}$$

There are two major problems in using the DTW algorithm to signature verification: one is to define a distance function, $d(u^i, r^j)$, to measure the similarity of two segments, u^i and r^j, which are from the test sequence and reference sequence, respectively; one is the method to calculate the accumulative warping distance.

Distance Function

In order to define the distance function between two segments, one segment is firstly defined as a feature vector. The features include the segment's geometrical features and average dynamic features:

- Segment starting point and ending point coordinates (x_0^i, y_0^i), (x_1^i, y_1^i);
- Segment average curvature $AveC^i$;
- Segment average speed $AveV^i$;
- Segment average acceleration $AveA^i$;
- Segment max and min curvature $MaxC^i, MinC^i$;
- Segment max and min speed $MaxV^i, MinV^i$;
- Segment max and min acceleration $MaxA^i, MinA^i$.

These features constitute a 13-dimension vector for each segment. The Mahalanobis distance is used to evaluate the similarity of two segments:

$$d(u^i, r^j) = (\mathrm{u}^i - \mathrm{v}^j)^T \Sigma^{-1} (\mathrm{u}^i - \mathrm{v}^j), \tag{10-11}$$

where Σ is the covariance matrix, u^i and v^j are the feature vectors of test segment and reference segment, respectively.

Dynamic Programming

The DTW algorithm uses the Dynamic Programming (DP) to calculate the warping distance and to find the warping path. The traditional Dynamic Programming algorithm is an optimization method, and uses an iterative procedure to find the optimal path from the starting point to the ending point. In the segmentation stage, one integrated segment should falsely be separated into several short segments. In the system, an improved Dynamic Programming algorithm is adopted, which can merge some consecutively segment while programming. The algorithm is described as follows:

Improved Dynamic Programming algorithm:

1. Define a 2-dimensional array $g(N, M)$ as cumulative difference matrix;
2. Initial $i = 1, j = 1$;
3. While $i \le N$ and $j \le M$ do
4. If $i = 1$ and $j = 1$ then $g(1,1) = d(u^1, r^1)$;
5. Else $g_1 = g(i, j-1) + d(u^i, r^j)$

   $g_2 = g(i-1, j) + d(u^i, r^j)$

   $g_3 = g(i-1, j-1) + d(u^i, r^j)$

$$g_4 = g(i-1, j-2) + d(u^i, r^{j-1} + r^j)$$
$$g_5 = g(i-2, j-1) + d(u^{i-1} + u^i, r^j)$$
$$g(i,j) = \min(g^1, g^2, g^3, g^4, g^5);$$

6. Warping distance $D = \dfrac{g(N,M)}{N+M}$

This algorithm modifies the traditional DP at the fifth step, which not only calculates the minimum of g_1, g_2 and g_3, but also two additional value, g_4 and g_5. While calculating g_4 and g_5, '+' means merge the two consecutive segments, and the feature vector of the new combined segment should be re-calculated.

LEARNING

In the learning stage, three things need to be learned. One is the reference for each signer. In this system, a simple method is used. Let $S = (s_1, \cdots, s_L)$ be the training sample set for one user, and the reference $r = s_i$, where s_i satisfies the condition of minimizing this value: $\sum_{j=1}^{L} d(s_i, s_j)$. Another one needs to be learned is the threshold to accept or reject a testing signature. In this system, an individual threshold is chosen to minimize the average false rate for each signer. The last one is the covariance matrix used in the Mahalanobis distance. In order to simplify the problem, all the features are regarded to be independent, and the matrix changes to a diagonal one. The elements on the diagonal are the feature's variances that are calculated using statistical method.

EXPERIMENT

The experiment is based on a database containing 1716 signatures, which was written by 33 signers. Each signer was asked to sign his (or her) own signature 20 times as genuine ones, and other people's signatures 1 times as forgeries. When signing other people's signature, a genuine one is presented as reference to be imitated, so some of them can be regarded as skilled forgers. In the learning stage, 6 genuine signatures are used as training samples and others and all forgeries are used as test samples. The experiment result shows: Type I error is below 10%, and Type II error is about 16%.

This system is a relatively primary one. There are many improvements to be done, especially in the learning stage. Now the learning algorithm is very simple because it just chooses an optimal one from the training sample set. The improved method should do some optimization among the training set, and make the reference to minimize the total distance between each training signature.

10.5 Internet/Intranet Signature Verification Application

This section will present an on-line signature verification approach in the Internet and Intranet using Java technology. We will first discuss the superiors of on-line signature verification in the network comparing to the traditional password approach. A demo system then is to be presented with a new stroke based given algorithm. Finally a practical on-line signature verification system in Internet/Intranet will be proposed.

MAIN PROBLEMS

As mentioned above, signature verification is one of the active areas in pattern recognition research as well as application fields. Technologies and applications for automatic off-line and on-line signature recognition and verification are facing with a lot of real challenges. For example, how to achieve as low as false acceptance rate and false rejection rate, and how to reach the highest performance with least time and cost or how to make the application commercial... However, it has never rested long since more and more new ideas and technologies were employed into and deployed out of this area.

Powered by neural network, parallel processing, distributed computing, multimedia technology, the Internet and Java technology, and driven by more and more wired and wireless business demand, automatic signature verification is being more and more acceptable in the business area. Such as, off-line (static) signature recognition can applied in automatic bank process, document recognition and filing system... and on-line (dynamic) signature verification can be applied in automatic personal identity authentication, E-commerce, online banking...

Today, with the astonishing growth of the Internet and the Intranet, E-commerce and E-finance becomes one of the hottest topics in this planet. Doing business through the public network makes the network security more and more important as well. How to protect the private identification from being pirated is the key issue that the Internet and Intranet clients would concerned before such E-business could be widely accepted.

In the current practices, to identify a person in the Internet and Intranet is to use password, an encrypted combination of ASCII characters. However, no mater how strong such security system is, such as SSL (Secure Socket Layer: 40-100bit) or newer SET (Secure Electronic Transaction: 1024bit) with digital certificate (e-wallet), there are still fetal shorts of such approaches, such as easy to forget and easy to be stolen. However, with the nature time and speed features which are never forgot and difficult to be stolen, on-line human signature verification has its obvious advantages in the network personal identification.

There are a lot of aspects to be considered when to implement such kind of system. For example,

1) Signature verification algorithm, such as stroke-base, neural network or fuzzy logic;
2) Client/Server software, such as Java and Database technology;
3) Network data security, such as encryption/decryption and digital signature;
4) Business concern, such as legal issues and service contracts.

We will only discuss the first two key technical topics here below with a demo system presented.

DATA ACQUISITION

More than 90% of the automatic on-line signature verification processes use the similar model as seen in Figure 10-9 [29]. In most of the on-line signature applications, a special pen and tablet is selected for data acquisition to collect a lot of dynamic signature information with high resolutions, such as pen-tip position, speed, acceleration, pressure and angle. However, in the Internet/Intranet applications, asking all clients over different locations to use the same set of special pen and tablet will has unavoidable limitation of availability, calibration and cost. The network response is also an important consideration. Therefore, the practical Internet/Intranet application requires a generic common pen and tablet, and a pen independent methodology for data acquisition. With our previous study, the writing pressure and pen angle information would generate a lot of unexpected noises than what we expected during comparison. Furthermore, the speed and acceleration information can be easily calculated from the pen-tip position versus time.

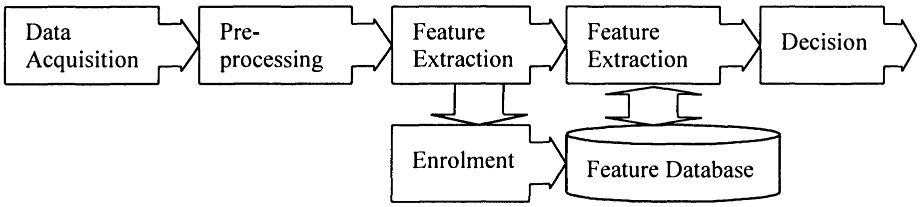


Figure 10-9. The signature verification process.

In this section, the signature input can be captured with any commercial mouse-like pen pad, such as Motorola WisePen™. The sample speed depends on the client machine instead of the server in order to override the network speed limitations. A generalization and normalization is applied to these acquired signature pen-tip data for server process. Since the client browsers and platforms vary a lot, the platform independent Java technology will fit well for such applications.

As an example, an original signature is captured in our text (see Figure 10-10). Every registered user needs to sign in the registration window for 5 times and the (X, Y) location of the pen tip together with its correlated time of signature will be recorded and stored for training. The training will generate a feature database for each signature. When the user wants to log in the system next time, he/she needs to sign again in the test window. If the features of the signature match those in the database, the signature is considered to be a genuine one. Otherwise, the access will be denied. In addition to this model, an analysis system can be designed to help develop, test and evaluate the verification methodology. Such system can display the original signatures and features, such as (X, Y) locations, speeds, accelerators, up and down (Z) status.

VERIFICATION ALGORITHM

To distinguish how good an on-line signature verification is nothing more important than the verification algorithm. In this section, a single stroke timing based algorithm has been introduced here below.

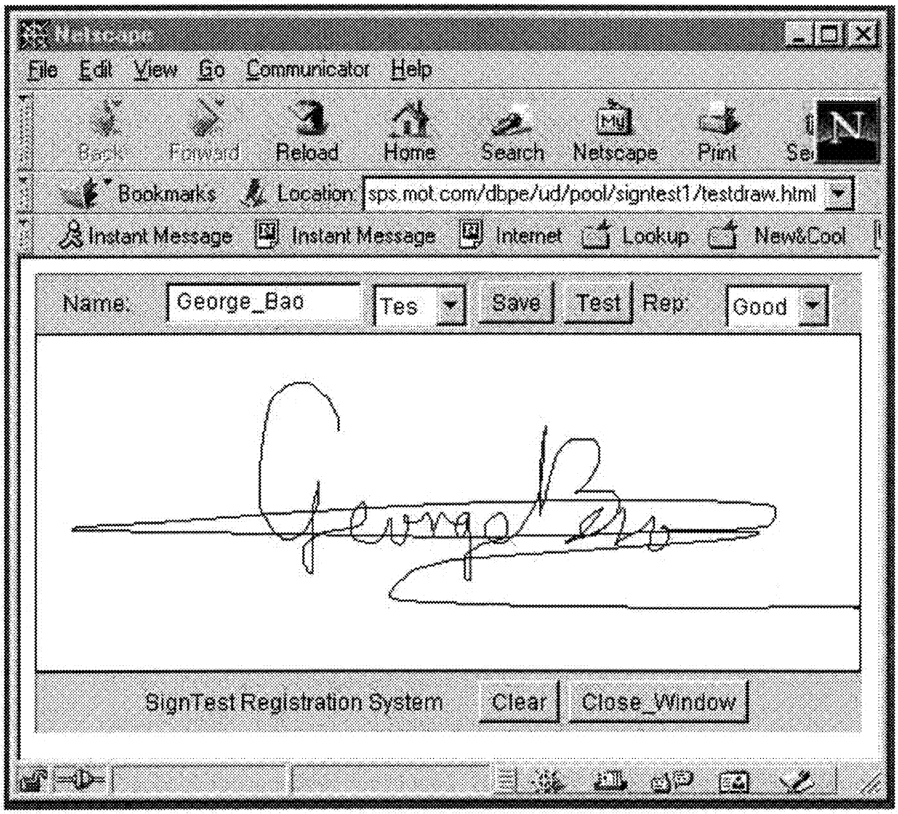


Figure 10-10. An original signature in our text.

- Record the curve of X, Y coordinates versus time T based on the client machine. The resolution depends on the client screen setup (see Figure 10-11).
- Generalize and normalize the curve by minimum 0.02 seconds time step. Preprocess uses an average filter.
- Abstract the time dependent features from the curve (as shown in Figure 10-12).
  - Select the peeks for feature points.
  - Find the basic point which has the biggest distance to its neighbor points at both sides.
  - Calculate the distance between basic point and every feature points and record these values as feature value.
- Define the range (maximum and minimum) of the feature values for each

correlated point from five signatures.

- Multiply the range by a factor as a threshold for judging other coming new signatures.

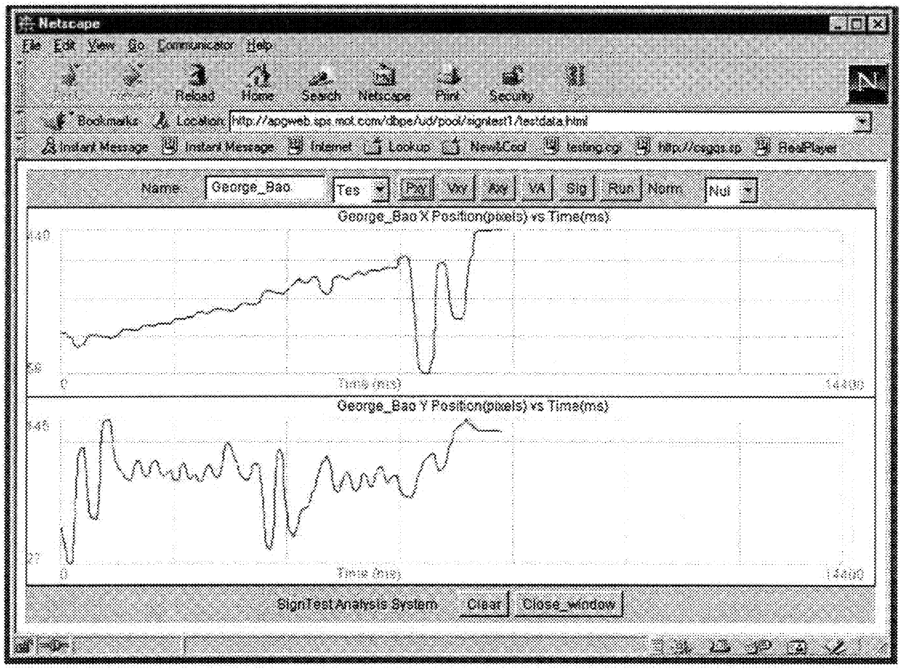


Figure 10-11. The waveforms, X and Y, coordinates versus time T.

DEMO SYSTEM

With the above algorithm, a demo system has been developed without considering the network communication securities concern (see Figure 10-13). Java applets and CGI-Perl scripts have been used for client application and server application respectively. Disk file system has been used for data storage. This demo must be compiled with JDK1.1 and up and must be installed in a Unix or NT server with CGI-Perl 5.0 and up. Any browsers supporting Java can be used as a client (see Figure 10-14). The total compiled code size is less than 50Kbytes.

We chose Java because the advantages of Java make it deemed to be the number one selection for Internet application [30]. For example:

- Java portability features make it easy to be running in various platforms.
- Java compact code size makes it fast to be download through the network.
- Java network libraries make it super for network programming.
- Java reusability features make it easy to be tailored and implant into the bigger business client/server system.

- Java scalability features make it easy to expand in the future.

More than 100 signatures has been collected and analyzed to determine the threshold. The experiment results showed that less than 10% Equal Error Rate could be obtained.

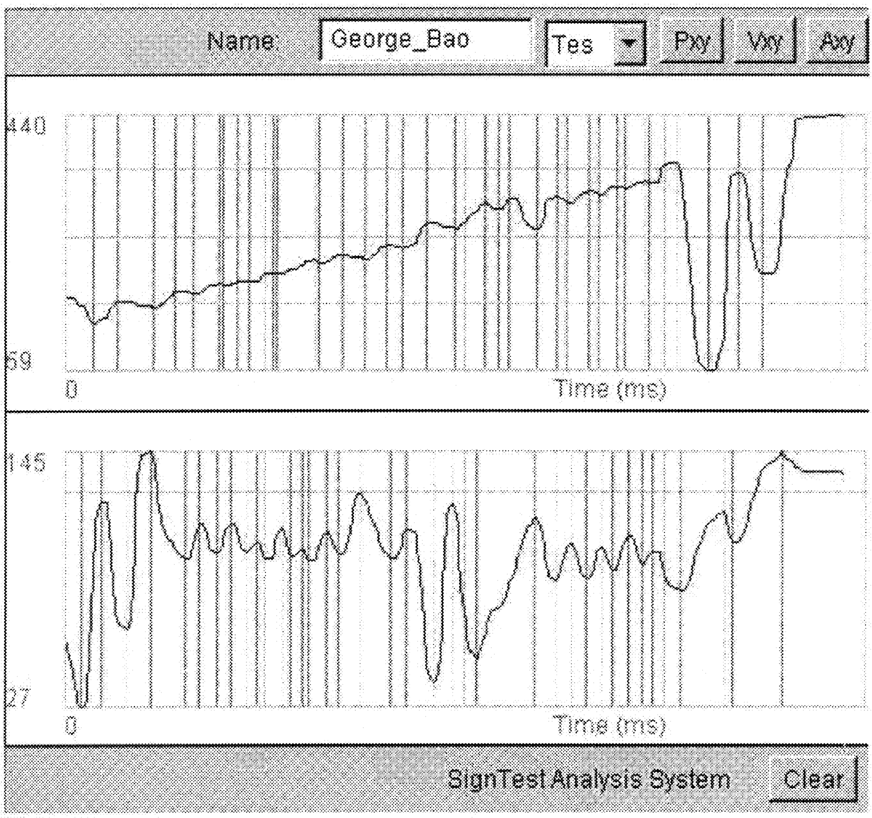


Figure 10-12. Feature points in the normalized curve of X, Y coordinates vs. time T.

The major error might come from the mismatch of the features among the five training signatures though the X/Y curve had been normalized. A possible improvement is to estimate the sequence of feature points by calculating the relative position in order to avoid the redundant points mismatch. As another issue, this demo system developed with only one kind of features may not enough for the practical use to detect skilled forges effectively. The improvement is to adopt multiple features including general ones (such as total time and total points average speed) and local ones (like speed between turning points and the one in this section).

PROPOSED PRACTICAL SYSTEM

As discussed above, the fast growing E-Commence worries about the Internet Security issues, especially the Credit Card verification because not only the lost card itself, but also the illegal uses of the stolen, faked information of the credit card would be even more harmful. Therefore, such a web based on-line signature verification system through Internet/Intranet has been proposed (see Figure 10-15).

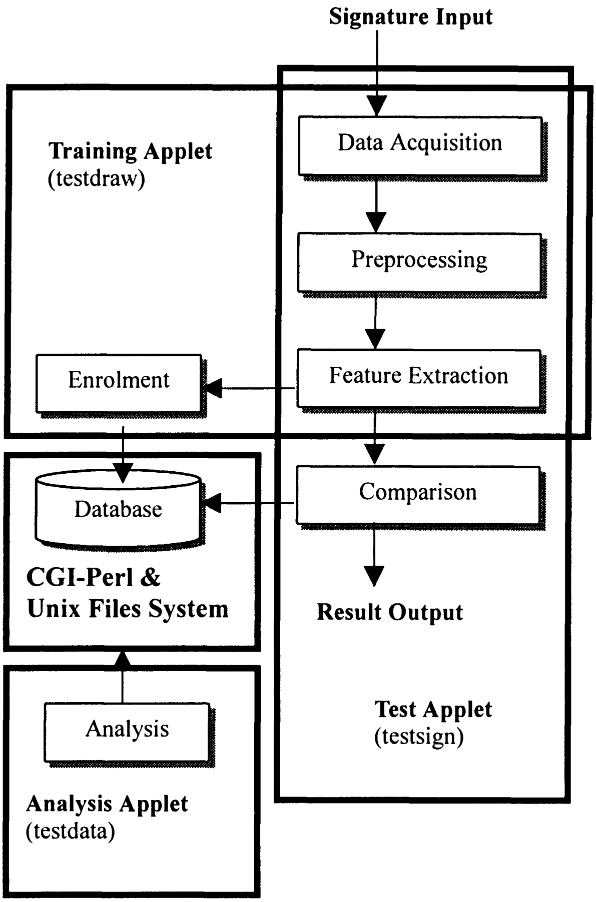


Figure 10-13. Block Diagram of the On-line signature verification demo in Internet/Intranet.

Pre-process:

1. The standards and contracts of the signature verification approaches are agreed among all parties.
2. The Internet Stores enroll and grant the access to bank Card Center server.
3. The Card Holder trains and stores his/her dynamic signature information in database in the Bank Card Center Server.

Process:

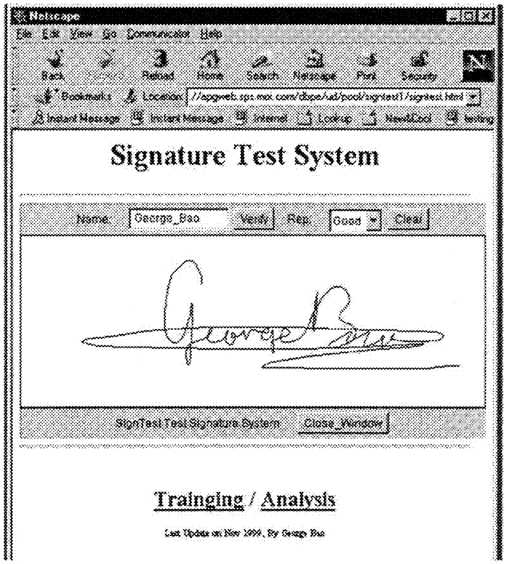


Figure 10-14. Appearance of the On-line signature verification demo in Internet/Intranet.

1. The Card Holder inputs order into the Store server.
2. The Card Holder inputs the card information into the Store server.
3. The Internet Store sends the signing window by Java Applet to the Cardholder.
4. The Card Holder signs on the window of his/her browser.
5. The Java Applet sends the captured signature to the Card Center.
6. The Card Center routes the confirmation or rejection information to the Card Holder.
7. The Card Center routes the confirmation or rejection information to the Store as well.
8. The Internet Store accepts or rejects the orders based on the signature verification result.

In the market, it is not difficult to find the products based on automatic signature verification. But most of these systems are standalone. Signer and the system must use the same hardware and software interface at the same place. With the dramatic exploration of the Internet, such methodology with the domain limitation can hardly work when a customer and supplier are located at different regions over the world. The above proposed system for the E-Commence will be able to deal with such demand because there are a lot of advantages in such system with Java Applet, CGI, Java Servlet, JDBC and DBS over the Internet/Intranet.

- Different customer can use different input interface, such as pen, mouse or touch screen, for signing, as long as he/she can keep on using this tool after registered in the signature server.

- Internet stores can serve customers worldwide without worrying about the correctness of the credit card information.
- Such system can be ported and installed on any Internet servers and running on any popular browsers supporting Java, such as NS Communicator 4 or MS Internet Explorer 5.

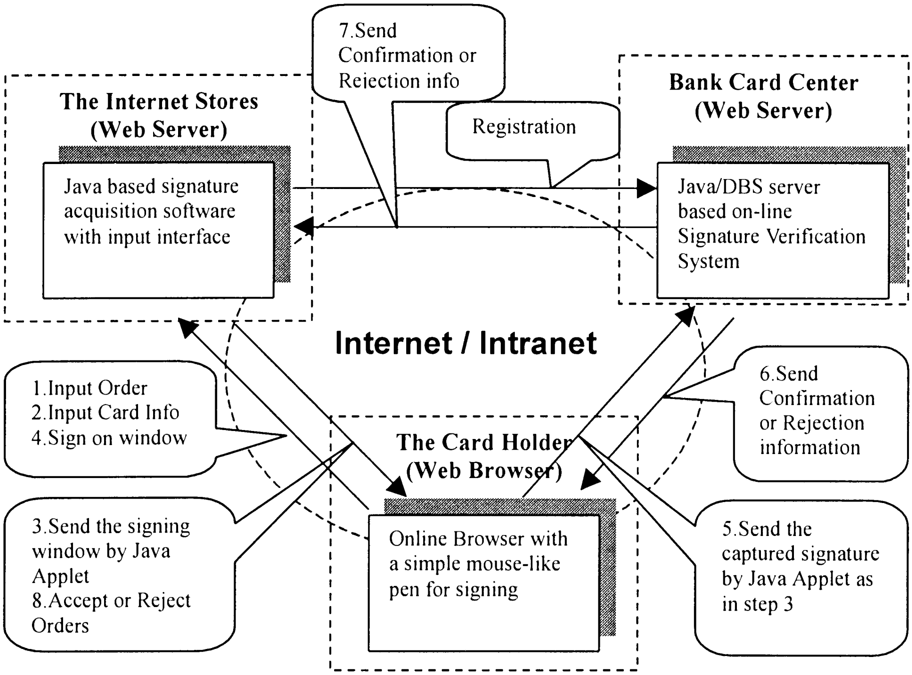


Figure 10-15. Proposed application process for signature verification in the Internet/Intranet.

However, there is still an unsolved issue - security in the Internet. This is a side topic that we believe will be either balanced with the Internet business demands or settled in the near future.

With a demo system presented and a practical system proposed, this section is just trying to trigger more improvement for human identification in the Internet and Intranet. In terms of human privacy, live signature is by no mean a better protection than the traditional password comprised of characters and digits. But it is still a challenge to the current computer software as well as the network technologies. But definitely it may help more and more E-business in the future.

Signature involves the behavior feature of human being, and signature verification is an important part of biometrics. Many research works have been done in this field. This chapter is an overview of these applications and methods, both on-line and off-line. Finally, a practical on-line signature verification system is presented in detail.

The signature verification process can be divided into several stages, preprocessing, feature extraction, comparison and decision. Till now, most of the methods used to signature verification are introduced from character recognition, and the features are almost the same. But there's a distinct difference between signature verification and character recognition, the former is concerning the differences between people's signing, the latter is concerning the similarities of the same character written by different people. So the methods and features that are used should be different. There are still many works need to be done in the signature verification.

References

[1] R. Plamondon and G. Lorette, "Automatic Signature Verification and Writer Identification – The State of the Art," *Pattern Recognition*, vol. 1, no. 2, pp. 107-131, 1989.

[2] J. Duvernoy, "Handwritting Synthesis and Classification by Means of Space-Variant Transform and Karhunen-Loeve Analysis," *J. Opt. Soc. Am.* 65, pp. 1331-1336, 1975.

[3] N. Papamarkos and H. Baltzakis, "Off-line Signature Verification Using Multiple Neural Network Classification Structures", *Proceedings of 13th International Conference on Digital Signal Processing*, 1997.

[4] K. Han and I.K. Sethi, "Handwritten Signature Retrieval and Identification," *Pattern Recognition Letters*, vol. 17, pp. 83-90, 0167-8655/96, 1996.

[5] M. Dehghan, K. Faez and M. Fathi, "Signature Verification Using Shape Descriptors and Multiple Neural Networks," *TENCON '97, IEEE Region 10 Annual Conference, Speech and Image Technologies for Computing and Telecommunications, Proceedings of IEEE*, vol. 1, pp.415-418, 1997.

[6] J. Lin and J.G. Li, "Off-line Chinese Signature Verification", *Proceedings of International Conference on Image Processing*, 1996.

[7] G. Dimauro, S. Impedovo, G. Pirlo and A. Salzo, "A Multi-expert Signature Verification System for Bankcheck Processing," *Automatic Bankcheck Processing*, pp.364-381, Wold Scientific Publishing Co. Pte. Ltd., Singapore, 1997.

[8] R. Sabourin and G. Genest, "An Extended-Shadow-Code Based Approach for Off-Line Signature Verification: Part -II- Evaluation of Several Multi-Classifier Combination Strategies," *Proc. ICDAR, Ulm, IEEE*, 1995.

[9] K. Huang and H. Yan, "On-line Signature Verification Based on Dynamic Segmentation and Global and Local Matching," *Optical Engineering*, vol. 34, no. 12, pp. 3480-3487, December 1995.

[10] J.G.A. Dolfing, E.H.L. Aarts, V. Oosterhout and J.J.G.M., "On-line Signature Verification with Hidden Markov Models", *Proceedings of 14th International Conference on Pattern Recognition*, 1998.

[11] X.H. Yang, T. Furuhashi, K. Obata and Y. Uchikawa, "Constructing a High Performance Signature Verification System Using a GA Method", *Proceedings of the Second New Zealand Two-Stream Int'l Conference on Artificial Neural Networks and Expert Systems (ANNES '95)*, 0-8186-7174-2/95 IEEE, 1995.

[12] F. Bauer and B. Wirtz, "Parameter Reduction and Personalized Parameter Selection for Automatic Signature Verification," *Proc. ICDAR, Ulm, IEEE*, 1995.

[13] L.L. Lee, T. Berger, and E. Aviczer, "Reliable On-line Human Signature Verification Systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, 0162-8828/96, IEEE 1996, June 1996.

[14] G.S.K. Fung, J.N.K. Liu, and R.W.H. Lau, "Feature Selection in Automatic Signature Verification Based on Genetic Algorithms," *Proceedings of the International Conference on Neural Information Processing, Progress in Neural Information Processing, Amari*, et al (Eds), pp. 811-815, Springer-Verlag, 1996.

[15] R. Martens and L. Claesen, "On-line Signature Verification: Discrimination Emphasised," *Proc. ICDAR, Ulm, IEEE*, 1997.

[16] B. Herbst and D. Richards, "On an Automated Signature Verification System," *Proceedings of IEEE International Symposium on Industrial Electronics*, 1998.

[17] L.L. Lee, "Neural Approaches for Human Signature Verification," *Proc. ICDAR, Ulm, IEEE 1995. 3rd International Conference on Signal Processing*, 1996.

[18] G.B. Hesketh, "COUNTERMATCH: A Nueral Network Approach to Automatic Signature Verification," *IEE Colloquium on Neural Networks for Industrial Applications* (Digest No. 1997/014).

[19] N. Mohankrishnan, W.S. Lee and M.J. Paulik, "Multi-Layer Neural Network Classification of On-line Signatures," *IEEE 39th Midwest symposium on Circuits and Systems*, 1996.

[20] R. Martens and L. Claesen, "On-line Signature Verification by Dynamic Time Warping," *Proceeding of 13th International Conference on Pattern Recognition*, 1015-4651/96, 1996.

[21] R. Martens and L. Claesen, "Dynamic Programming Optimization for On-line Signature Verification," *Proceeding of 4th ICDAR '97.* 0-8186-7898-4/97, 1997.

[22] B. Wirtz, "Stroke-Based Time Warping for Signature Verification," *Proc. ICDAR, Ulm, IEEE,* 1995.

[23] W.S. Lee, N. Mohankrishnan and M.J. Paulik, "Improved Segmentation Through Dynamic Time Warping for Signature Verification Using Neural Network Classifier," *Proceedings of 1998 International Conference on Image Processing*, 1998.

[24] B. Wirtz, "Average Prototypes for Stroke-Based Signature Verification," *Proceedings of 4th ICDAR '97.* 0-8186-7898-4/97,1997.

[25] C.C. Hsu, L.F. Chen, P.C. Chang and B.S. Jeng, "On-line Chinese Signature Verification Based on Multi-expert Strategy," *Proceedings of 32nd Annual 1998 International Carnahan Conference on Security Technology*, 1998.

[26] R. Plamomdon, "A Model-Based Dynamic Signature Verification System," *Fundamentals in Handwriting Recognition, Proceedings of the NATO Advanced Study Institute on Fundamentals in Handwriting Recognition, France: Springer-Verlag*, pp. 417-434, 1993.

[27] M.J. Paulik, N. Mohankrishnan and M. Mikiforuk, "A Time Varying Vector Autoregressive Model for Signature Verification," *The Proceeding of The IEEE*, 0-7803-2428-5/95, 1995.

[28] F. Leclerc and R. Plamondon, "Automatic Signature Verification: The State Of The Art – 1989-1993," *Progress in Automatic Signature Verification, Singapore: World Scientific*, pp. 3-20, 1994.

[29] G. Pirlo, "Algorithms for Signature Verification, Fundamentals in Handwriting Recognition," *Proceedings of the NATO Advanced Study Institute on Fundamentals in Handwriting Recognition, France: Springer-Verlag*, pp. 435-455, 1993.

[30] B. Daniel, *Advanced Techniques for Java Developers*, Wiley, New York: [c1998], c1997.

11 OTHER BEHAVIORAL BIOMETRICS

In this chapter, we begin by introducing the background for behavioral biometrics and reviewing briefly the voice and signature biometrics. Then, keystroke, gesture and gait biometrics are described in Section 2, 3 and 4, respectively. Section 5 gives the concluding remarks.

11.1 Behavioral Biometrics System

A biometrics is a unique, measurable characteristic or trait of a human being for automatically recognizing or verifying identity. So, biometrics technologies are concerned with the physical parts of the human body or the personal traits of human beings. The most common physical biometrics are the eye (iris and retina), face, fingerprint, palmprint and hand; while voice, signature, keystroke, gesture and gait are behavioral biometrics. Behavioral biometrics systems now come in many shapes and sizes. This can range from hardware and software to software development kits or complete solutions. Systems may be marketed and sold by vendors directly or through various distribution channels, such as system integrators, strategic partners or value added resellers.

Like physical biometrics, behavioral biometrics are also the classical pattern recognition problems: how to recognize instances of a behavior pattern given a stored library of models. They have the principles of acquisition, feature extraction, matching and decision in common. Figure 11-1 is the basic structure of the behavioral biometrics systems. Yet, behavioral biometrics technologies focus on different aspects of the human behavior, so the workings of each technology and each vendor's specific system will differ.

The effectiveness of behavioral biometrics systems can be evaluated mainly in terms of false rejection rate (FRR) and false acceptance rate (FAR), cost of recognition system, and time to access identity verification. The two important measures are FRR and FAR. The FRR of a verification system gives an indication of how often an authorized individual will not be properly recognized. The FRR is important since it describes the amount of user frustration in using the security system. The FAR describes how often an unauthorized individual will be mistakenly recognized and accepted by the system. It is generally more indicative than the FRR since it describes the degree to which the security measure may be breached by intruders. In addition, the equal error rate (ERR) is a commonly accepted overall measure of system performance. It corresponds to the decision threshold at which the FAR is the FRR.

The voice and signature biometrics described in the previous two chapters are the most important behavioral biometrics technologies. Hence, here we have necessity to review briefly the voice and signature biometrics.

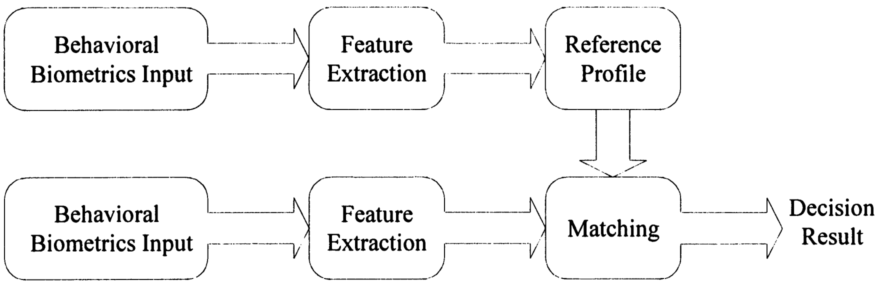


Figure 11-1. Basic structure of behavioral biometrics systems.

Voice biometrics focus on the sound of the voice. It is therefore important to distinguish this technology from voice recognition one that recognize words and act on commands. To avoid any confusion, the terms: speaker recognition, speaker verification and speaker identification should be used when referring to the biometrics. The sound of a human voice is caused by resonance in the vocal tract. The length of the vocal tract, the shapes of the mouth and nasal cavities are all important. Sound is measured, as affected by these specific characteristics. The technique of measuring the voice may use either text-dependent or text-independent methods. In other words, the voice may be captured with the user uttering a specifically designated password combining phrases, words or numbers (text-dependent) or any form of phrase, words or numbers (text-independent). Currently text-dependent techniques are dominant in commercially available speaker recognition systems. Speaker recognition biometrics are particularly useful for telephone-based applications. We are all used to speaking on the telephone and biometrics systems can be easily incorporated into private or public telephone networks. However, environmental background noise and interference over telephone networks can affect the performance of speaker recognition systems.

The popular text-dependent speaker recognition techniques include dynamic time warping (DTW) and hidden Markov model (HMM). The techniques for text-independent speaker recognition can be categorized into three major approaches [1]. The first and earliest approach is to use long-term averages of acoustic features, such as spectrum representations or pitch. The second approach is to model the speaker-dependent acoustic features within the individual phonetic sounds that comprise the utterance, which can be accomplished using explicit or implicit segmentation of the speech into phonetic sound classes prior to speaker model training or recognition. Explicit segmentation may be performed using a HMM-based continuous speech recognizer as a front-end segmenter for speaker recognition systems. Implicit segmentation relies on some form of unsupervised clustering to provide implicit segmentation of the acoustic features during both training and recognition, such as vector quantization (VQ), HMM and Gaussian mixture model (GMM). The third and most recent approach is the use of artificial neural network (ANN), including multilayer perceptron (MLP), time-delay neural network (TDNN) and radial basis function network (RBFN).

Signature biometrics are often referred to as dynamic signature verification (DSV) and look at the way we sign our names. Since DSV put emphasis on the signing rather than the finished signature, DSV can be differentiated from the study of static signatures on paper. A number of characteristics can be extracted and measured by DSV. For example the angle at which the pen is held, the time taken to sign, the velocity and acceleration of the signature, the pressure exerted when holding the pen and the number of times the pen is lifted from the paper, can all be extracted as unique behavioral characteristics. DSV is not based on a static image, so even if a signature is traced a forger would need to know the dynamics of that signature. This makes forgery very difficult. In additions, the signature is one of the most accepted means of asserting identity and used in a number of situations to legally bind an individual, such as the signing of a contract. These factors have taken signature biometrics to a number of diverse markets and applications, ranging from the legal binding of documents, to checking welfare entitlement, document management and pen-based computing.

Similar with text-dependent speaker verification, DSV must concern with the temporal sequencing among features extracted from signature signals. Therefore, DTW and HMM are the very effective approaches for DSV.

Table 11-1 is intended to give a general guide of voice and signature biometrics systems. It is extremely important to note that the nature of each application will in turn affect the performance of all biometrics technologies. Also, some of the factors in many cases are known to correlate with each other. For example, the accuracy of a biometrics system will affect its resilience as a barrier to attack and may be determined by the long-term stability of the biometrics characteristic. Each of the factors has been separated for simplicity and is ranked as low, medium, high or very high.

In addition to voice and signature biometrics, other important behavioral biometrics include keystroke, gesture and gait. They will be involved in the following sections.

11.2 Keystroke Biometrics

As the beginning of the 20^{th} century, it has been demonstrated that human actions are predictable in the performance of repetition and routine tasks. For example, each telegraph operator had a distinctive pattern of keying messages over telegraph lines so that an operator can find out who was sending information simply by listening to the characteristic pattern of dots and dashes [2]. So far as keyboard is concerned, it has been established that keyboard characteristics are rich in cognitive qualities and hold promise as an individual identifier. When a computer user types on the keystroke of a computer, a digital signature is left in the form of keystroke latencies (elapsed time between keystrokes and hold times). Experiment for keystroke characterization showed that a high degree of correlation could be obtained if the same person typed both the reference keystroke and the test ones [3].

Table 11-1. Overview of voice and signature biometrics.

	Signature	*Voice*
Level of Accuracy	High	High
Ease of Use	High	High
Barrier to Attack	Medium	Medium
Public Acceptability	Very High	High
Long-term Stability	Medium	Medium
Standards		Speaker Verification API
Possible Interference	Illiteracy; signatures that constantly change or are easily imitated	Background and network noise; colds and other factors can change the voice

In developing a scheme using keystroke dynamics for identity verification, it is very necessary to determine which keystrokes characterize the individual's key pattern. The common-use keystroke features are interkey times and hold times. The interkey times mean the inter-character time intervals measured as user types. The hold times represent the time duration between the moment every key button is hit to the moment it is released. Both features can be obtained by computer programming. Based on these features extracted from input keystrokes, classical pattern recognition (CPR) techniques and ANN techniques can be used to authenticate personal identity. The keystroke dynamics using interkey times and hold times as features have been successfully applied in the practical authentication systems.

Obaidat et al. used some ANN paradigms along with CPR techniques using the interkey times as features to identify the computer users [4,5]. The employed dataset consisted of interkey times while users are typing a known sequence of characters. For example, when password "DAVID" were entered, then the time duration between the letter pairs (D, A), (A, V), (V, I) and (I, D) would be computed. The participants in the experiment were asked to enter the same phrase correctly. The data were collected from six different users over a six-week period. This helped not only in averaging out the effect of uncorrelated sources of noise that could be introduced by instruments and participants, but also to gather data that represent the different modes of the participants. For training proposes, the raw data were separated into two parts: all of the odd-numbered patterns of each class, and all of the even-numbered patterns. In any given simulation run, only half of the data were used to form the training set. After each network was trained, the entire pattern set was presented to the network for classification. During the recognition phase, various combinations of these patterns were created to test the learning abilities of the three different ANN paradigms. After experimentation determined the best ANN architecture for this application, the network was incorporated into an "on-line" system that would collect the character time intervals from users in real-time and perform a classification immediately.

Bleha and Obaidat experimented with the Perceptron algorithm as a classifier to verify the identity of computer users [6]. By performing the real-time measurements of the time duration between keystroke entered in the user's password, data were collected from 10 valid users and 14 invalid users over a period of 8 weeks. Half of the data are training data and the other is test data. The 9% FRR and 8% FAR were achieved, respectively. The Perceptron algorithm was found to be robust with respect to the choice of the initial weight vector.

Obaidat evaluated the performance of five CPR algorithms as applied to the identification of computer users using the interkey times as features. These algorithms are potential function, Bayes classifiers, minimum distance and the cosine measure. The most and least successful algorithms are the potential function algorithm and the cosine measure, respectively. Obaidat and Sadoun [7] evaluated the performance of a newly devised ANN scheme, called Hybrid-Sum-Of-Product (HSOP) [8] for computer users verification using interkey times as features. The performance of HSOP to the Sum-Of-Products and Backpropagation (BP) neural networks is compared. They found that HSOP performs better than the other two paradigms.

Also, Obaidat and Sadoun verified computer users using hold times as features to authenticate computer users [9]. The participants in the experiment were asked to enter their login user ID during an eight-week period, which averaged out the effects of fatigue and stress as well as the uncorrelated sources of noise. The login routine was modified so that each time a login attempt was made, and the hit and release time was stored for analysis. The forgery attempts were collected in one session from each of the 15 invalid users who attempted each of the 15 ID's 15 times. CPR and ANN techniques were used for the classification process. Based on the FRR and FAR, it was found that hold times were more effective than interkey times and the combined hold times and interkey times based approaches gave the least misclassification error. Such results suggest that hold times may in general provide better characterization of the typing skills than the interkey times. The most successful ANN paradigms for the verification task are the fuzzy ARTMAP, RBFN, and learning vector quantization (LVQ) network. They basically gave a zero misclassification error for both the FRR and FAR. Also, it was found that the most successful ANN paradigm provides better authentication accuracy than the best CPR schemes. Table 11-2 shows these findings.

11.3 Gesture Recognition

Gesture is one of the typical methods of nonverbal communication for human beings and we naturally use various gestures to express our intentions in everyday life. The recognition process on gesticulative information in verbal communication remains unexplained, but it is obvious that gesture helps persons with easy understanding in communication. On the other hand, the ability to follow objects moving through space and recognizing particular motions as meaningful gestures is essential if computer systems are to interact naturally with human users. Currently, however, this information is largely unavailable to computing machines. Therefore, an important new application of machine vision is to extend the interface between man and machine, allowing the machine to directly perceive what its user is doing.

Table 11-2. Classification results using CPR and ANN techniques for keystroke biometrics.

Features and results(%) / *Methods*		*Hold times*		*Interkey times*		*Combined hold times and interkey times*	
		FRR	*FAR*	*FRR*	*FAR*	*FRR*	*FAR*
CPR Techniques	K-means	13	11	17	13	14	10
	Cosine Measure	28	19	37	33	33	21
	Min. Distance	12	7	17	14	15	10
	Bayes' Rule	2	1	5	2	5	2
	Pol. Function	2	1	5	2	4	2
ANN Techniques	BP, Sigmd	2	2	2	3	0	0.5
	CPNN	40	32	66	52	33	24
	Fuzzy ARTMAP	0	0	0	3	0	0
	RBFN	0	0	0	3	0	0
	LVQ	0	1	0	0	0	0
	RN	50	36	70	66	17	19
	SOP	6	5	6	5	4	2.5
	HSOP	4	2	4	3	1	0.5

Many methods for gesture recognition from image sequence have been proposed. In these methods, the time-varying features extracted from a gesture image sequence are recognized by using DTW [10,11], HMM [12] and ANN [13].

Darrell and Pentland [10] developed a method for learning, tracking and recognizing human gestures using a view-based approach to model articulated objects. Objects are represented using sets of view models, rather than single templates. Sterotypical space-time patterns, i.e., gestures are then matched to stored gesture patterns using DTW. Real time performance is achieved by using special-purpose correlation hardware and view prediction to prune as much of the search space as possible. Both view models and view predictions are learned from examples. Results show tracking and recognition of human hand gestures at over 10Hz.

Nagaya et al. [11] proposed a new appearance-based feature for real-time gesture recognition from motion images. The feature is the shape of the trajectory caused by gesture in pattern space defined by the inner product between patterns on frame images. The gesture trajectory feature has three merits, 1) invariant for the target human's position, size and lie, 2) gesture recognition without interpreting frame image contents, 3) no costly statistical calculation. Its effectiveness for gesture recognition is shown from the results of several experiments.

Brand et al. [12] presented algorithms for coupling and training HMMs to model interacting processes, and demonstrate their superiority to conventional HMMs in a vision task classifying two-handed actions. In the conventional HMM, the Markovian framework makes strong restrictive assumptions about the system generating the signal that it is a single process having a small number of states and an extremely limited state memory. Coupled HMMs provide an efficient way to resolve many of these problems, and offer superior training speeds, model likelihoods, and robustness to initial conditions.

Watanabe et al. [14] presented a new method for real-time gesture recognition. The key idea of this method is that multiple template models for specific feature parts of body such as arms are used and the pose of this part by considering similarities of an input image to the models is estimated. To calculate the similarity for the part, maskable template model (MTM) with which both masking and template matching are done simultaneously is developed. As an application of this method, a real time interactive system, which can activate the characters of video game by gestures in real time, has been developed.

A common feature of the previous approaches except [12] is to recognize gesture using two- dimensional information from only one image sequence. Since gesture is usually performed in 3D space, gesture recognition using only 2D information may not be able to recognize correctly a complicated gesture that may have a self-occlusion and confusion in the single input image. A glove-free method for tracking hand movements using a set of 3D models has been proposed [15]. In this approach, the hand is represented by five cylindrical models that are fit to the third phalangeal segments of the fingers. Six 3D-motion parameters for each model are calculated that correspond to the movement of the fingertips in the image plane. Trajectories of the moving models are then established to show the 3D nature of the hand motion.

The methods described above put emphasis on classifying a kind of gesture, and they cannot recognize the degree information of gesture that is whether the gesture is made quickly or slowly and in a big way or small way. Since the degree information often represents users' attitude, emotion and so on, the interactive system should recognize not only the kind of gesture but also the degree information. Watanabe and Yachida [16] proposed a new method of real time gesture recognition for human oriented interactive systems. This method obtains the approximate 3D information of gesture in real time using the eigenspace without a reconstruction of 3D structure, which is constructed by KL expansion from multi-input image sequences captured from different directions. Using the information, it can not only recognize the complicated gestures correctly but also obtain the degree information of gesture, such as speed and magnitude. In this gesture recognition method, the human part from each input image captured from different directions in real time is first segmented. The each input image segmented at the same time is connected, and a linked human part image is constituted. Figure 11-2 shows the process constituting a linked human part image. Next, the gesture eigenspace is constructed by KL expansions of linked image sequences of model gesture. Finally, input image sequences are represented in the eigenspace, so the gesture is recognized and the degree information is also obtained in real time.

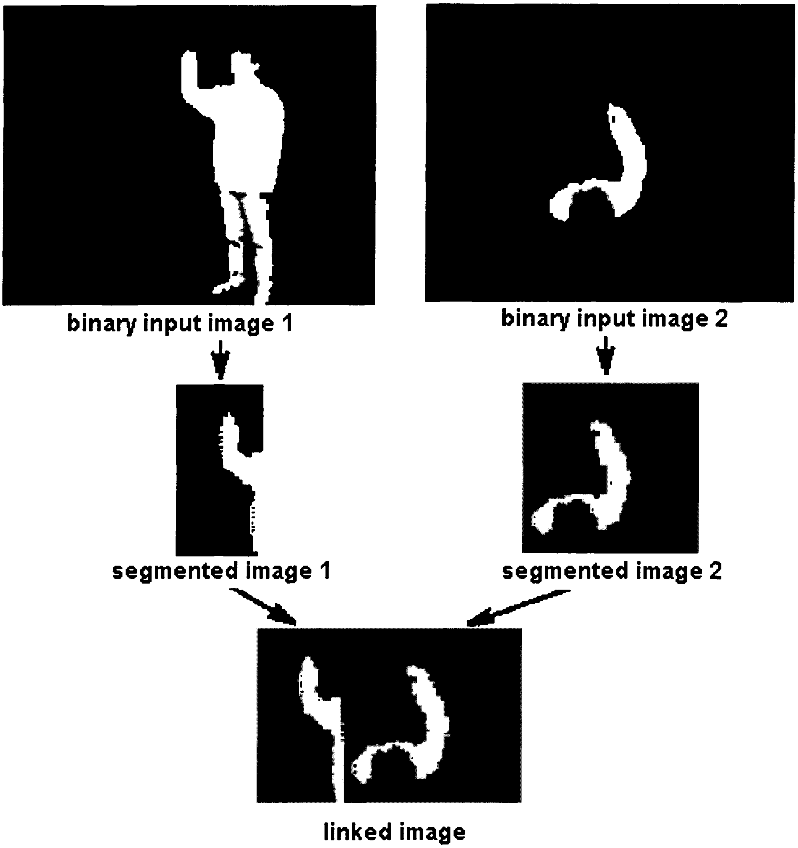


Figure 11-2. Making process of a linked human part image (Selected from Reference [16]).

11.4 Gait Biometrics

Using gait as a biometrics has recently attracted interest. In many applications of person identification, many established biometrics can be obscured such as the fingerprint at low resolution. However, people need to walk, so their gait is usually apparent and hard to disguise. Also, it requires no subject contact. Gait biometrics concern its derivation by computer vision, for this is only way it can satisfy its purpose. Some insight into gait as a biometrics can be drawn from psychology.

In the earliest studies of gait perception participants were presented with images produced from points of light attached to body joints [17]. When the points were animated, they were immediately perceived as representing a human in motion. Bingham has shown that point light displays are sufficient for the discrimination of different types of object motion and that discrete movements of parts of the body can be perceived [18]. Indeed, the redundancy involved in the light point display might provide an advantage for motion perception [19] and could perhaps offer improved performance over video images.

Surprisingly, research into the psychology of gait has indeed not received much attention; especially using video, in contrast with the enormous attention paid to face recognition. One recent study [20], using video rather than point light displays, has been that humans can indeed recognize people by their gait, and can learn their gait for purposes of recognition. The study confirmed that, even under adverse conditions, gait could still be used as a cue to identity. Clearly, psychological studies support the view that the gait can indeed be used for recognition. Prior to study of automatic recognition, we shall consider some approaches for modeling human body and motion, for these are of potential benefit in recognition. Indeed, some of the approaches have found deployment in automatic gait recognition.

Many studies have considered human motion extraction and tracking. The selection of good body models is important to efficiently recognize human shapes from images and analyze human motion properly. Stick figure models and volumetric models are commonly used for three-dimensional tracking and the ribbon models and blob models are also used but are not so popular. Stick figure models connect sticks at joints to represent the human body. Akita proposed a model consisting of six segments: two arms, two legs, the torso and the head [21]. Lee and Chen's model uses 14 joints and 17 segments [22]. Guo et al. represent the human body structure in the silhouette by a stick figure model which has ten sticks articulated with six joints [13].

Volumetric models are used for a better representation of the human body. One model consists of 24 segments and 25 joints and those segments and joints are linked together into a tree-structured skeleton [23]. The "flesh" of each segment is defined by a collection of spheres located at fixed positions within the segment's co-ordinate system. Concurrently, angle limits and collision detection are incorporated in the motion restrictions of the human model. Among the different volumetric models, generalized cones [24] are the most commonly used, which is the surface swept out by moving a cross-section of constant shape but smoothly varying size along an axis. Generalized cylinders are the simplified case of generalized cones that have a cross-section of constant shape and size. Figure 11-3 shows examples for a stick figure model and a cylinder model.

However, these structural models need to be modified according to different applications and are mainly used in human motion tracking. Human motion can be defined by the different gestures of body motion, different athletic sports or human walking or running. There are two main methods to model human motion. The first is model-based: after the human body model is selected, the 3D structure of the model is recovered from image sequence with [22] or without [13,21,25] moving light displays. The second emphasizes determining features of motion fields without structural reconstruction [19,26].

Ideas from human motion studies can be used for modeling the movement of human walking. Hogg [25] and Rohr [27] use flexion/extension curves for the hip, knee, shoulder and elbow joints in their walking models. Duo et al. [13] use joint angles between different sticks as features of different walking persons. Akita [21] uses a sequence of stick figures, called the key frame sequence, to model rough movement of the body. In his key frame sequence of stick figures, each figure represents a different phase of body posture from the point view of occlusion. Other approaches that are different from above consider the properties of the spatio-temporal pattern as a whole. Polana and Nelson [28] define temporal textures to be the motion patterns of indeterminate spatial and temporal extent, activities to be motion patterns which are temporally periodic but are limited in spatial extent, and motion events to be isolated simple motions that do not exhibit any temporal or spatial repetition.

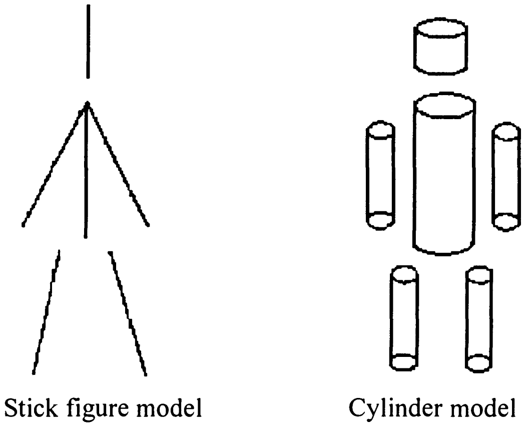


Figure 11-3. Examples for human body models.

There have been a number of approaches to tracking human in scenes. A model-based approach was used in one of the earliest tracking studies [25]. The WALKER model mapped images into a description in which a person was represented using a series of hierarchical levels. There has been much progress recently. Gavrila and Davis [29] presented a vision system for the 3-D model-based tracking of unconstrained human movement. Multiple view images were employed, to avoid using markers, to recover 3D-body pose. A new method for the 3D model-based tracking of human body uses multiple views to avoid occlusion of body parts [30]. Available parts are then tracked between frames of a video sequence in a model aimed to minimize the difference between the human model and imaged views. Initial results were presented showing how humans could be tracked in the presence of severe occlusion.

Parameterized optical flow has actually been used to track articulated motion in an image sequence [31]. Limbs were represented as a set of connected cardboard patches where analyzed motion was constrained to enforce articulated motion. The approach was demonstrated to track human walking over long image sequences. One system used 3D planar projections to achieve better tracking than contemporaneous 2D trajectory-based systems [32]. The system was based on detecting and segmenting optical flow from within a central region. Then 3D planar geometry was used with an active camera system to ensure focus on the central region. Extended Kalman filters were used to analyze the trajectories and the system was shown to successfully track moving objects, including people, and pursuit performance was shown to improve on 2D performance.

Most tracking approaches naturally lack the accuracy required for recognition since that was not their original purpose. However, it would seem reasonable to assume that tracking procedures could be deployed to develop a gait signature. Much of this work has been of benefit to the approaches to automatic gait recognition.

Indeed, a number of approaches have already shown that it is possible to recognize people by their gait. The majority of current approaches are motion-based, combining the image sequence by its motion or by statistical analysis. Only one technique is feature-based and its results can clearly be identified with the data from which they were derived.

In what perhaps was the earliest approach for automatic gait recognition, the gait signature was derived from the spatio-temporal pattern of a walking person [33]. Here, in the XT dimensions (translation and time), the motions of the head and of the legs have different patterns. These patterns were processed to determine the body motion's bounding contours and then a five-stick model was fitted. The gait signature was derived by normalizing the fitted model for velocity and then by using linear interpolation to derive normalized gait vectors. This was then applied to a database of 26 sequences of five different subjects, taken at different times during the day. Depending on the values used for the weighting factors in a Euclidean distance metric, the correct classification rate varies from nearly 60% to just over 80%, a promising start indeed.

Later, optical flow was used to derive a gait signature [34]. This did not aim to use a model of a human walking, but to describe features of an optical flow distribution. The optical flow was filtered to produce a set of moving points together with their flow values. The geometry of the set of points was then measured using a set of basic measures and further information was derived from the flow information. Then the periodic structure of the sequence was analyzed to show several irregularities in the phase differences; measures including the difference in phase between the centroid's vertical component and the phase of the weighted points were used to derive a gait signature. Experimentation on a limited database shown how people could be discriminated with these measures, appearing to classify all subjects correctly.

Another approach was aimed more at generic object-motion characterization [26], use gait as an exemplar. The approach was similar in function to spatio-temporal image correlation, but used the parameter eigenspace approach based on principal component analysis (PCA) to reduce computational requirement and to increase robustness. The approach first derived body silhouettes by subtracting adjacent images, with further processing to reduce noise. Then the images were projected into eigenspace. Eigenvalue decomposition was then performed on the sequence of silhouettes where the order of the eigenvectors corresponds to frequency content. Recognition from a database of 10 sequences of seven subjects showed classification rates of 100% for 16 eigenvectors and 88% for temporal correlation approach. Further, the approach appears robust to noise in the input images. However, PCA based on the global covariance matrix of the full set of image data is not sensitive to class structure in the data. Combining PCA with canonical analysis (CA) can reduce data dimensionality and optimizes class separability of different gait sequences simultaneously [35-37].

An alternative approach for collecting the motion information in an image sequence is to find features and collect their motion information. In the only model-based approach, the gait signature is derived from the spectra of measurements of the thigh's orientation, as extracted by computer vision techniques. This was demonstrated to achieve a recognition rate of 90% on a database of 10 subjects. Contemporaneously, the nature of gait has been recognized by "probabilistic decomposition of human dynamics at multiple abstractions" where the dynamics of gait in video trajectories of tracked body parts. The feature vector was extracted from optical flow and from trajectory information and then classified by use of HMM, showing good gait discrimination. As such, feature based metrics can be used to provide gait signatures in a way which agrees with human insight, allowing the results to be validated visually.

Clearly, gait biometrics provides lots of research opportunities, which exist not only in development of basic techniques, but also in application and as a potential contributor to multi-biometrics systems. So far as technique is concerned, we require techniques to isolate moving articulated objects. Naturally, these will focus more on the legs but there is also potential for extension to the thorax. Equally, these techniques could be aimed at extracting the generic shape of the human body. As such, this requires extraction of arbitrary moving articulated shapes, as required for human motion analysis and recognition. In terms of gait recognition, we require to derive a signature from an image sequence. So far, this has been derived by approaches that integrate motion across the sequence, and by one, which is feature based. A common paradigm in many statistical approaches has been to use binary images for recognition. There are many other approaches that offer similar capabilities, such as Fourier descriptors of an object's shape. Extension of feature-based measurement (or a model-based approach) would naturally focus on development of a technique that was mentioned previously.

Like all biometrics, gait research will benefit from an established database for purposes of development, preferably with a separate database for test purposes and hopefully with the stringency of the FERET test [38]. Clearly any database will need to include variation in factors, which can affect the perception of gait. These include variety in clothing, and in footwear and with subjects carrying common articles such as handbags or shopping bags. Also, we will require subjects walking with a wide variety of trajectories relative to the camera together with normal views as used in preliminary studies. Such a database will allow establishment of the properties and limits of signatures derived from gait. As such, they will provide an estimate of the confidence that can be associated with the use of gait to buttress other biometrics measures. As stated earlier, gait may be evident where other biometrics can be assessed with limited precision, or be obscured. Clearly, a good database can only serve to evidence the uniqueness or otherwise of automatic gait recognition.

In this chapter, after introducing the background for behavioral biometrics and briefly reviewing voice and signature biometrics, other important behavioral biometrics, including keystroke, gesture and gait, are described.

Keystroke dynamics can be used to authenticate access to computer systems and networks. The keystroke dynamics of a computer user's login string provide a characteristic pattern that can be used for verification of the user's identity. Keystroke patterns combined with other security schemes can provide a very powerful and effective means of authentication and verification of computer users. Further research into reliable methods for handling typographical errors is needed in order to make keystroke-based authentication systems non-irritating and widely accepted by the computing and network security community [39]. Finally, it is found that ANN paradigms are more successful than CPR algorithms in the classification of users.

Gesture recognition is essential if computer systems are to interact naturally with human users. After the time-varying features are extracted from image sequence, gesture can be recognized by using DTW, HMM, MTM and ANN. Further research puts emphasis on usefulness of 3D information and recognition of degree information of gesture.

Gait represents a potential biometrics as humans can perceive it. It is attractive because it requires no contact and is less likely to be concealed. Psychological studies support the view that the gait can be modeled and is unique. Indeed, lots of approaches have shown that it is possible to recognize people by their gait. The majority of current approaches are motion-based, combining the image sequence by its motion or by statistical analysis. Only one technique is feature-based and its results can clearly be identified with the data from which they were derived. There is great scope for future research effort, both in applications and developments. These developments could be improvements in recognition procedure or in automated technique. As such, future work will establish more precisely the results that can be achieved by this new biometrics. If its performance can be equivalent to that of other biometrics, then by its practical advantages it could indeed become a pragmatist's choice.

References

[1] D.A. Reynolds and R.C. Rose, "Robust Text-independent Speaker Identification Using Gaussian Mixture Speaker Models," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 1, pp. 72-83, 1995.

[2] W.L. Bryan and N. Halter, "Studies in The Physiology and Psychology of the Telegraphic Language," *The Psychology of Skill: Three Studies*, (E.H. Gardener and J.K. Gardner. Editors), pp. 35-44, NY Time Co., NY 1973.

[3] D. Umpress and G. Williams, "Identity Verification Through Keyboard Characteristics," *International Journal of Man-Machine Studies*, vol. 23, pp. 263-273, Academic Press, 1985.

[4] M.S. Obaidat and D.T. Macchairolo, "An On-line Neural Network System for Computer Access Security," *IEEE Trans. Industrial Electronics*, vol. 40, no. 2, pp. 235-241, April 1993.

[5] M.S. Obaidat and D.T. Macchairolo, "A Multilayer Neural Network System for Computer Access Security," *IEEE Trans. System, Man and Cybernetics*, vol. 24, no. 5, pp. 806-813, May 1994.

[6] S. Bleha and M.S. Obaidat, "Computer User Verification Using the Perceptron, " *IEEE Trans. System, Man and Cybernetics*, vol. 23, no. 3, May/June 1993.

[7] M.S. Obaidat and B. Sadoun, "An Evaluation Simulation Study of Neural Network Paradigm for Computer User Identification," *Information Science Journal-Applications*, Elsevier, vol. 102, no. 1-4, pp. 239-258, November 1997.

[8] M.S. Obaidat and B. Sadoun, "HSOP: A Neural Network Paradigm and Its Applications," *Neural Computing & Application Journal*, Springer, vol. 2, pp. 89-96, 1994.

[9] M.S. Obaidat and B. Sadoun, "Verification of Computer User Using Keystroke Dynamics," *IEEE Trans. System, Man and Cybernetics*, vol. 27, no. 2, pp. 261-269, 1997.

[10] T. Darrell and A. Pentland, "Space-time Gestures," in *Proc. of International Conference on Compuer Vision and Pattern Recognition*, pp. 335-340, 1993.

[11] S. Nagaya, S. Seki and R. Oka, "A Theoretical Consideration of Pattern Space Trajectory for Gesture Spotting Recognition," in *Proc. of International Conference on Automatic Face and Gesture Recognition*, pp. 72-77, 1996.

[12] M. Brand, M. Oliver and A. Pentland, "Coupled Hidden Markov Models for Complex Action Recognition," In *Proc. of International Conference on Computer Vision and Pattern Recognition*, pp. 994-999, 1997.

[13] Y. Guo, G. Xu and S. Tsuji, "Understanding Human Motion Patterns," In *Proc. of International Conference on Pattern Recognition*, pp. 325-329, 1994.

[14] T. Watanabe, C.W. Lee and A. Tsukamoto and M. Yachida, "A Method of Real Time Gesture Recognition for Interactive Systems," In *Proc. of International Conference on Pattern Recognition*, pp. 473-477, 1996.

[15] J. Davis and M. Shah, "Towards 3D Gesture Recognization," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 13, no. 3, pp. 381-393, 1999.

[16] T. Watanabe and M. Yachida, "Real Time Recognition of Gesture and Gesture Degree Information Using Multi Input Image Sequence," In *Proc. of International Conference on Pattern Recognition*, pp. 1855-1858, 1998.

[17] G. Johansson, "Visual Perception of Biological Motion and a Model for Its Analysis," *Perception and Psychophysics*, vol. 14, pp. 201-211, 1973.

[18] G.P. Bingham, P.C. Schmidt and L.D. Rosenblum, "Dynamics and Orientation of Kinematic Forms in Visual Event Recognition," *Journal of Experimental Psychology: Human Perception and Perfor-mance*, vol. 21, no. 6, pp. 1473-1493, 1995.

[19] G.L. Pellechia and G.E. Garrett, "Assessing Lumbar Stabilization from Point Light and Normal Video Displays of Lumbar Lifting," *Perceptual and Motion Skills*, vol. 85, no. 3, pp. 931-937, 1997.

[20] S.V. Stevenage, M.S. Nixon and K. Vince, *Visual Analysis of Gait as a Cue to Identity*, AT Press, 1998.

[21] K. Akita, "Image Sequence Analysis of Real World Human Motion," *Pattern Recognition*, vol. 17, no. 1, pp. 73-83, 1984.

[22] H.J. Lee and Z. Chen, "Determination of 3D Human Body Postures from a View," *Computer Vision, Graphics and Image Processing*, vol. 30, pp. 148-168, 1985.

[23] J.O'Rourke and N. Badler, "Model-based Image Analysis of Human Motion Using Constant Propagation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 2, no. 6, pp. 522-536, 1980.

[24] D. Marr and H.K. Nishihara, "Representation and Recognition of the Spatial Organization of Three Dimensional Shapes," in *Proc. of Royal Society London*, vol. B:200, pp. 269-294, 1978.

[25] D. Hogg, "Model-based Vision – A Program to See a Walking Person," *Image and Vision Computing*, vol.1, no. 1, pp. 5-20, 1983.

[26] H. Murase and R. Sakai, "Moving Object Recognition in Eigenspace Representation: Gait Analysis and Lips Reading," *Pattern Recognition Letters*, vol. 17, pp. 155-162, 1996.

[27] K. Rohr, "Towards Model-based Recognition of Human Movements in Image Sequences," *Computer Vision, Graphics and Image Processing*, vol. 59, no. 1, pp. 94-115, 1994.

[28] R. Polana and R. Nelson, "Detecting Activities," in *Proc. of International Conference on Computer Vision and Pattern Recognition*, pp. 2-7, 1993.

[29] D.M. gavrila and L.S. Davis, "3D Model-Based Tracking of Humans in Actions: A Multi-View Approach," in *Proc. of International Conference on Computer Vision and Pattern Recognition*, pp. 73-80, 1996.

[30] A. Kakadiaris and D. Metaxas, "Model-based Estimation of 3D Human Motion with Occlusion Based on Active Multi-Viewpoint Selection," *Proc. of International Conference on Computer Vision and Pattern Recognition*, pp. 81-87, 1996.

[31] S.X. Ju, M.J. Black and Y. Yacoob, "Cardboard People: A Parameterization Model of Articulated Image Motion," *in Proc. of International Conference on Automatic Face and Gesture Recognition*, pp. 38-44, 1996.

[32] K.J. Bradshaw, I.D. Reid and D.M. Murray, "The Active Recovery of 3D Motion Trajectories and Their Use in Prediction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 3, pp. 219-233, 1997.

[33] S.A. Niyogi and E.H. Adelson, "Analyzing and Recognizing Walking Figures in XYT," in *Proc. of*

International Conference on Computer Vision and Pattern Recognition, pp. 469-474, 1994.
[34] J. Little and J. Boyd, "Describing Motion for Recognition," in *Proc. of International Symposium on Computer Vision*, Coral Gables, FL, USA, pp. 235-240, 1995.
[35] P.S Huang, C.J. Harris and M.S. Nixon, "Human Gait Recognition in Canonical Space Using Temporal Templates," *IEE Proceedings: Vision, Image and Signal Processing*, vol. 146, no. 2, pp. 93-100, 1999.
[36] P.S Huang, C.J. Harris and M.S. Nixon, "Recognizing Humans by Gait Using a Statistical Approach for Temporal Templates," in *Proc. of International Conference on System, Man and Cybernetics*, vol. 5, pp. 4556-4561, 1998.
[37] P.S Huang, C.J. Harris and M.S. Nixon, "Statistical Approach for Recognizing Humans by Gait Using Spatial-Temporal Templates," in *Proc. of International Conference on Image Processing*, vol. 3, pp. 178-182, 1998.
[38] M.S. Nixon, J.N. Carter, D. Cunado, P.S. Huang and S.V. Stevenage, "Automatic Gait Recognition," *Chapter 11 in Biometrics Personal Identification in Networked Society*, A. Jain, R. Bolle and S. Pankenti (Edits.), Kluwer Academic Publishers, pp. 231-250. 1998.
[39] M.S. Obaidat and B. Sadoun, "Keystroke Dynamics Based Authentication," *Chapter 10 in Biometrics Personal Identification in Networked Society*, A. Jain, R. Bolle and S. Pankenti (Edits.), Kluwer Academic Publishers, pp. 213-230. 1998.

12 BIOMETRICS USER INTERFACES

The task of human-machine interfaces is to present the digital form of the end user's biometrics characteristics to the computer. It serves two entities of a biometrics system: 1) the human end user, and 2) the host computer. In this chapter, we will first propose the general structure, some new challenges and the corresponding taxonomies of biometrics human-machine interface (BHMI). The human-side interface which takes the capturing job will be discussed in Section 12.2. In Section 12.3, the machine-side interface which communicates with the host computer is introduced. Finally, a typical BHMI design based on palm verification is given.

12.1 Biometrics Interfaces

BIOMETRICS HUMAN-MACHINE INTERFACE

Biometrics can make a big progress in both the research field and the security markets nowadays. Biometrics human-machine interface (BHMI) is the part of a biometrics system which interacts with the end user and communicates with the host computer. The end user's attitude to the BHMI is critical to his or her attitude to the whole biometrics system. So a biometrics system is not reasonably a good one if it has not a good human-machine interface. To demonstrate the importance of the human-machine interface for a biometrics system, let's take a brief look at the history of the human-computer interfaces against that of the computer industry. In the early stage of the computer, the common human-computer interfaces were tedious punched card files, dull command-line interfaces on a black & white screen, etc. And the users of the computer must be experts to know how to manipulate them. Nowadays, however, a dummy who isn't trained much can easily use the computer to surf the Internet. Today's computers use windowing systems that have fancy elements such as windows, menus, buttons, scrollbars, combo boxes, and sophisticated input devices like mice, joysticks, trackballs, microphones and writing pads, etc. These rapid developments in the human-computer interface field contribute much to the popularity of the computer. Similarly, the development of a user-friendly human-machine interface is important to the popularity and marketing of a successful biometrics product.

On the other hand, the BHMI communicates with the host computer through some kind of communication protocols. What the BHMI provides to the host computer must meet the requirements set by the host computer or the designer. These requirements include the speed, the quality of the sample, etc.

Consequently, the two goals that the BHMI designers try to achieve are:

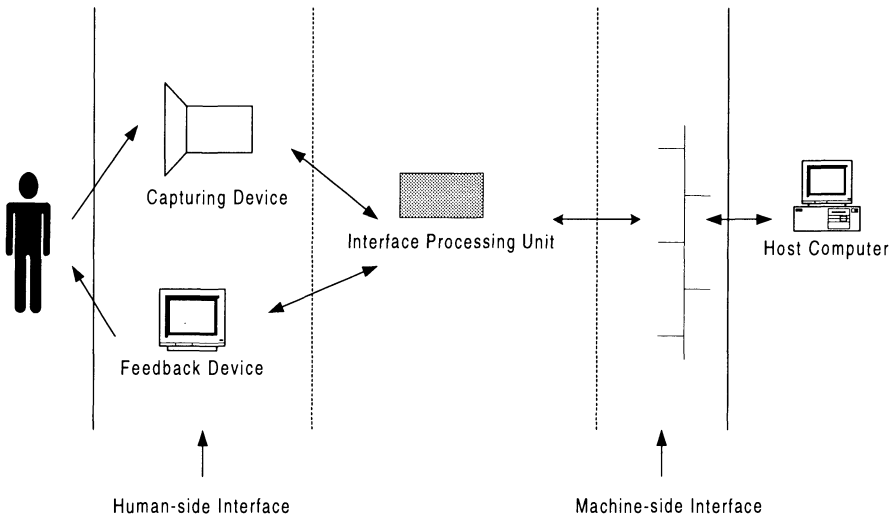


Figure 12-1. A general structure of the biometrics human-machine interface (BHMI).

1. producing a high-quality output for subsequent biometrics processing like verification and identification. For example, the human-machine interface of an iris biometrics system must provide a high-resolution and sharp iris image for recognition procedure.
2. providing a user-friendly and easy-to-use operating interfaces for the end users while not making the end users feel uncomfortable. This goal is not so easy to achieve because people are easily sensitive about their organs. For example, intensive light will be intrusive to the end users in the iris image capturing.

BHMI STRUCTURE

Figure 12-1 shows a general structure of the human-machine interface of a biometrics system. Generally, the end user presents his or her biometrics source, i.e., a finger in a fingerprint system, an eye in an iris system, etc. Then the capturing device gets a sample of the biometrics characteristic. Usually this sample will not be sent to the computer immediately because it is not suitable to be processed at the moment. For example, the initial image captured in an iris system will probably not have the iris located in the center of the picture without the interaction of the end user. So the feedback device is needed in most of the biometrics systems. The end user can make corresponding adjustments aided by the feedback presented on the feedback device. This is just like the end user adjusts his or her eye's location so as to make the iris image in the center of the picture in an iris system.

The sample is sent to the computer when it is good enough for processing. The system may sense this automatically or the end user decides this by pushing a button or other inputting methods.

In summary, therefore, the biometrics human-machine interface includes the following:

1. the human-side interface which in turn includes the capturing device and the feedback device;
2. the interface processing unit;
3. the machine-side interface.

The interface-processing unit includes the controllers of the capturing devices, the feedback devices in the human-side interface and the driver for the machine-side interface. The interface-processing unit is also responsible for the making of the decision if the sample meets the requirements. This part of the biometrics system is often product-specific.

NEW CHALLENGES

Because of its importance, the design of the BHMI pokes big challenges at the BHMI designers. These challenges are critical to the success of the BHMI, and ultimately the success of the whole biometrics system. The designer of the BHMI must make a trade-off between many factors. Some important factors are discussed as follows:

Cost and Size

Today many biometrics systems use optical cameras that are moderately costly and large in size [26]. So measures are taken to cut cost and size in optical systems. And new technologies are emerging recently that are touted to offer substantial advantages over optical-based systems, including smaller size, lower cost. These technologies include solid-state capacitance sensing, thermal sensing, ultra-sound sensing, etc.

Intrusive vs. Non-intrusive

This is the human factor in a biometrics system. A non-intrusive system does the sampling process which is comfortable for the end user, otherwise it is intrusive. The inked method to acquire a fingerprint is intrusive, and intensive light source in the iris image capturing is also intrusive.

Feedback

In some biometrics systems, feedback and human interaction is important to provide a high-quality sample to the system, which has been described in the section above.

Speed

It influences the response time of a biometrics system. It also is a factor for BHMI designer that must take into consideration.

Quality

The quality of the "product" of the BHMI is a main specification of the BHMI and/or the biometrics system. While increasing the quality of the sample, the designer must be aware of the corresponding increase in cost and the loss in other factors.

TAXONOMIES OF BHMI

Based on the various human-machine interfaces used in the current biometrics systems, we can classify them into some categories. According to different structures and types of biometrics capturing devices, two application types can be defined in the following:

Type 1: Touch vs. Non-Touch

In some capturing devices, biometrics information can be obtained by touching them. This means that biometrics human-interfaces will fall into two categories depending on touching the devices or not. For example, we should touch a hand-capturing device to capture the related information, but iris image does not need to touch. So two categories using Type 1 are:

Touching Interfaces:	Hand device, Palmprint device, Fingerprint device, Retina device, Signature device, etc.;
Non-touching Interfaces:	Iris device, voice device, face device, ear device, etc.

Type 2: 1-D vs. 2-D

Biometrics information obtained by sampling may be one-dimension (1-D) signal (like voice waveform) or two-dimension image (such as fingerprint image matrix). Based on the given information from the capturing devices, we can category the biometrics human-interfaces in the following:

1-D Interfaces:	Voice device, signature device, etc.;
2-D Interfaces:	Fingerprint device, iris device, face device, hand device, etc.

12.2 Human-Side Interface

Different biometrics technologies use different capturing device and/or different feedback devices [25,29]. In this section, the capturing techniques in some biometrics technologies including iris, hand geometry and fingerprint will be described in detail. We will also take a brief look at the capturing techniques in other biometrics technologies.

IRIS IMAGE CAPTURING

The iris is the colored ring stretching across the anterior portion of the eye and supported by the lens. The iris is a rich source of biometrics characteristics. Each human iris is a complex pattern, featuring specific characteristics such as the collarette (the border of the iris and the pupil), radiating striations (blood vessels), crypts and nevi (the atrophy and elevations on the outside edge of the iris). The structure of the iris is unique to an individual and is stable with age. So it can be used in a biometrics system [1-11].

Because the iris is very small (typically about 1cm in diameter) and people are highly sensitive about their eyes, iris image capturing devices can easily be intrusive to the end users. For example, to capture a high-quality image of the iris with adequate resolution and sharpness to support recognition, capturing devices usually use a bright source of illumination to achieve the imaging quality. And brightness can make the end users uncomfortable. On the other hand, to make the small iris image in the center of the picture (a processed called framed), the using of an eyepiece, chin rest or other contact positioning would be intrusive. In response to these challenges of illumination and positioning, two iris image-acquisition rigs have been developed [1], which will be presented here.

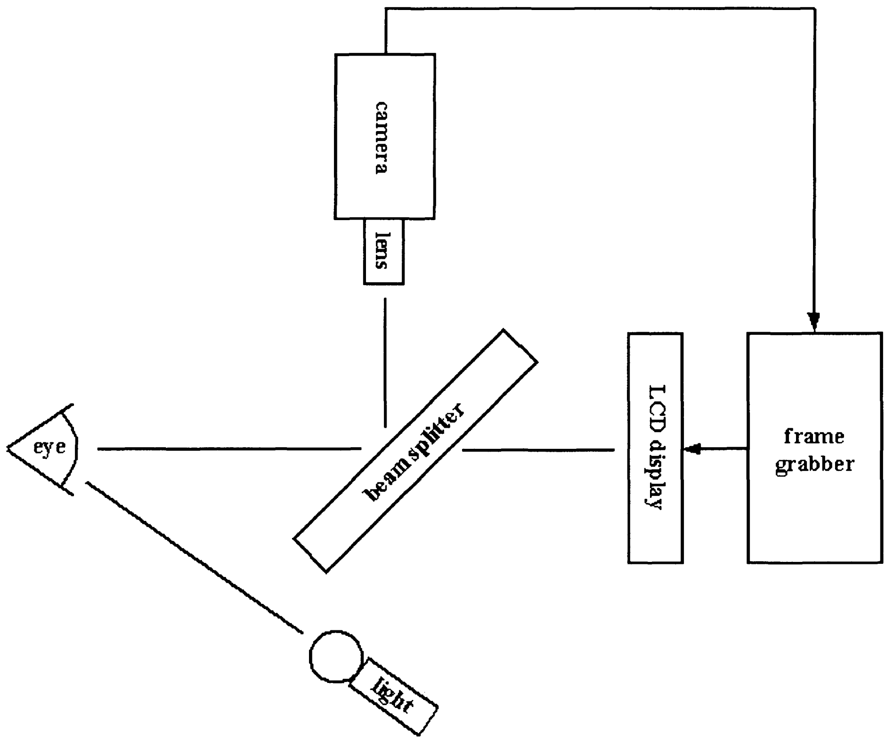


Figure 12-2. A schematic diagram of the Daugman image-acquisition rig.

Daugman System

The Daugman image-acquisition rig is a simple and compact system (see Figure 12-2). It makes use of an LED-based point light source which is positioned below the end user. With a standard video camera, Daugman system provides the end user with video feedback via a miniature liquid-crystal display placed in line with the camera's optics via a beam splitter. This allows the user to see what the camera is capturing and to adjust his or her position accordingly. During this process, the system is continually acquiring images. Once a series of images of sufficient quality is acquired, one is automatically forwarded for the subsequent processing. The image quality is judged by looking for high-contrast edges marking the boundary between the iris and the surrounding eye region [12].

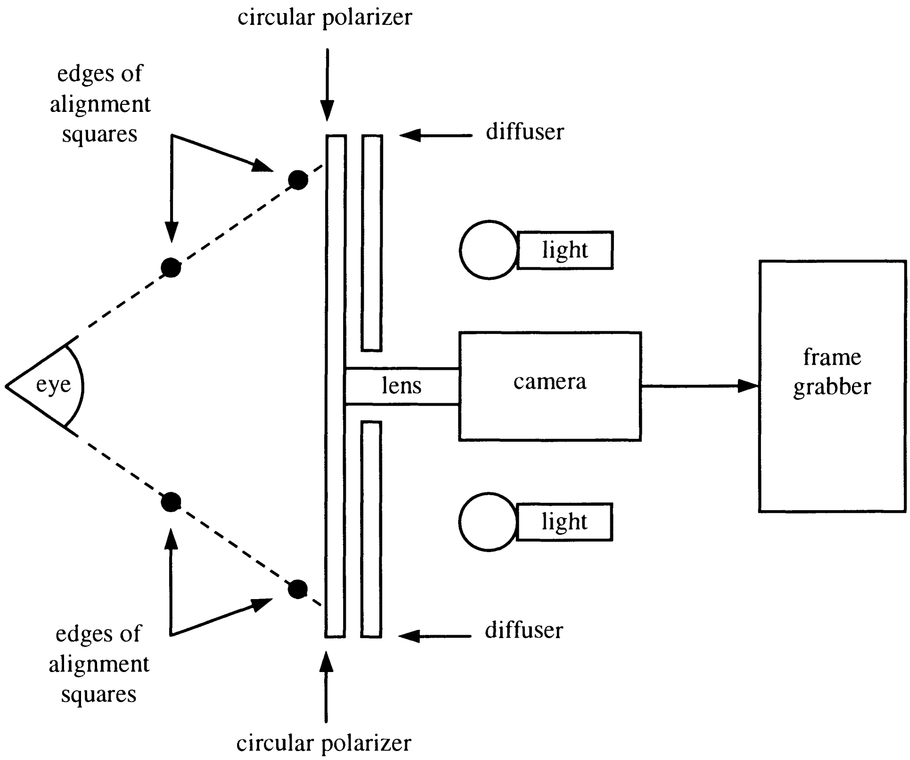


Figure 12-3. A schematic diagram of the Wildes image-acquisition rig.

Wildes System

It uses totally different ways from Daugman system to achieve good image quality and easy positioning. The problem with Daugman system is that the reflection of the point light source makes the region of the image where the point source is seen blurs and must be omitted. To address this problem, Wildes system employ a diffuse source and polarization by placing diffusers and circular polarizers between the light sources and the eye (see Figure 12-3). This design makes the system more complex, but benefit from it. First, the use of matched circular polarizers at the light source and the camera essentially eliminates the specular reflection of the light source which

Daugman system has. This allows for more of the iris detail to be available for subsequent processing. Second, the coupling of a low light level camera with a diffuse illuminant allows for a level of illumination that is entirely unobjectionable to human users.

Wildes system provides a reticle to aid the end user in positioning. In particular, a square contour is centered on the camera lens so that it is visible to the end user. Suspended in front of this contour is a second smaller contour of the same shape. The relative sizes and positions of these contours are chosen so that when the eye is in an appropriate position, the squares overlap and appear as one to the end user. As the end user maneuvers, the relative misalignment of the squares provides continuous feedback regarding the accuracy of the current position. Once the end user has completed the alignment, he activates the image capture by pressing a button.

HAND GEOMETRY SENSOR

Hand geometry takes care of finger length, width, thickness, shape and relative location of these features. So it requires the capturing device provide a three-dimension image of the hand.

Figure 12-4 is a typical capturing rig for hand geometry. When the end user needs to be verified, he or she will position his or her hand on the plate between a set of guidance pins. By using guidance pins, the image of the hand can be well placed in the center of the picture (framed) without feedback or adjustment involving human reactions. Looking down upon the hand is a charge-coupled-device (CCD) digital camera, which with the help of a mirror captures the side and top views of the hand simultaneously.

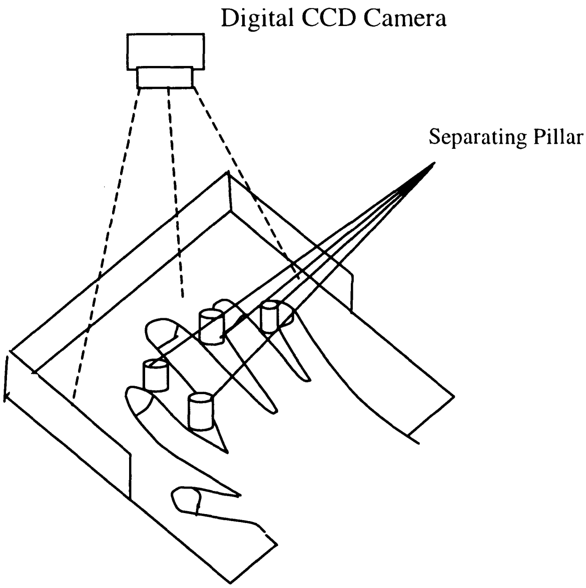


Figure 12-4. A capturing rig for hand geometry.

Hand scanners typically use an optical path approximately 11 inches (28cm) from the camera to the plate. An optical path folded with mirrors reduces the space required to half the original length (see Figure 12-5)[13].

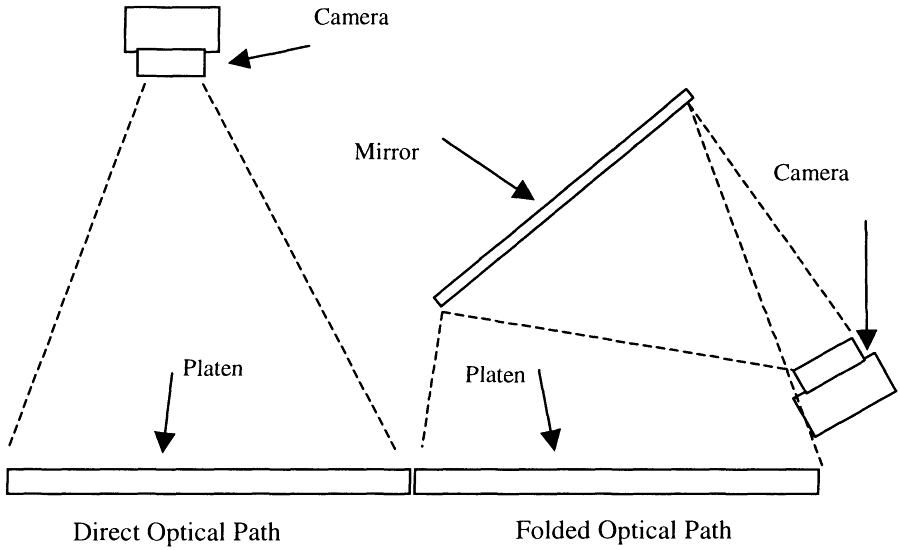


Figure 12-5. Hand scanner optics.

FINGERPRINT SENSOR

Fingerprint biometrics examines small unique marks called minutiae on the finger image. These minutiae may be ridge endings or ridge bifurcations. Thanks to its long history of use in many applications, especially in the law enforcement field, many fingerprint-capturing techniques are developed. Several of them will be described here [14-24].

The simplest method [14] is the well-known ink-and-paper method, which is also called the inked impression method. Generally, the first step in capturing an inked impression of the fingerprint is to place a few drops of ink on a slab and then rolling the ink out smoothly with a roller until the slab is covered with a thin, even layer of ink. Then the finger is rolled from one side of the nail to the other side over the inked slab which inks the ridge patterns on top of the finger completely. After that, the finger is rolled on a piece of paper so that the inked impression of the ridge pattern of the finger appears on the paper. By using a scanner, the inked impression is scanned and the fingerprint image is ready now for subsequent processing.

Obviously, the inked method is not clean, so intrusive to the end user. It is also easily be faked. One can get illegal access by using others' inked impression. This method, however, is very useful in some special situations such as the fingerprint images collected from the crime scenes.

A second method is a more efficient and reliable optical acquisition system. This system is often called the inkless fingerprint scanner. Figure 12-6 is a schematic diagram of an inkless fingerprint scanner. It consists of a prism, light source and a camera. A beam is introduced into the prism from one of its slanted surfaces. This beam is arranged so as to meet the conditions for total internal reflection at the top surface of the prism. In the beginning of the capturing, the end user puts his or her finger on the top surface of the prism on which there is usually a guidance pit for the finger to fit in. At this moment, the conditions for total internal reflection changes. Ridges (convex lines on the fingerprint surface) makes contact with the prism surface while valleys (concave parts on the fingerprint surface) does not. Usually sweat and body oil fills up the space between ridges and the prism surface. The area of contact will diffuse incident light while the area of no contact still meets the conditions for total internal reflection and incident light in this area is reflected normally and detected by the camera. So on the fingerprint image, ridges appear as dark lines and blank areas are valleys.

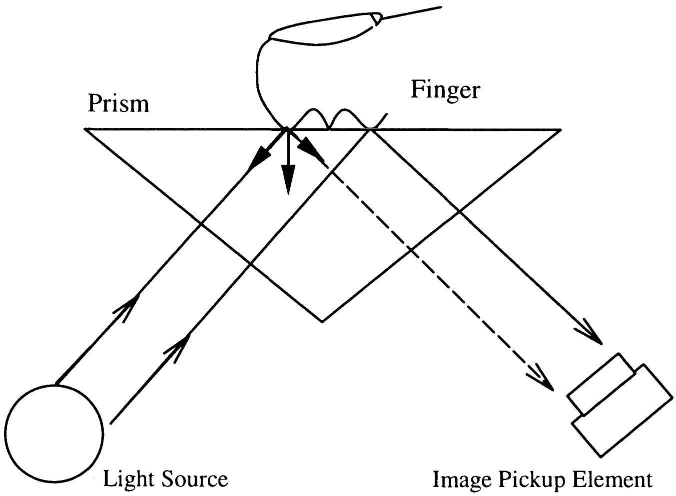


Figure 12-6. A schematic diagram of an inkless fingerprint scanner.

This system can also sense automatically that a finger is placed on it or is removed. From the beginning, the surface of the prism is repeatedly scanned and the image is stored in the memory of the system. After placing a finger on the prism surface, the gray levels of the image captured will vary. The gray levels for some areas which are selected at random, are compared with a gray level threshold. If the number of areas, whose gray levels become lower than the threshold (which means these areas become dark), an image is captured and sent for subsequent recognition.

There are problems with the prism fingerprint scanner. First, trapezoidal distortion is created by unequal optical paths between each point of the fingerprint and the image-focusing lens. This means binarization techniques based on a global threshold are improper and some greyscale-to-binary conversion preprocessing techniques may be employed. Second, noise light is caused by latent fingerprints of the previous users.

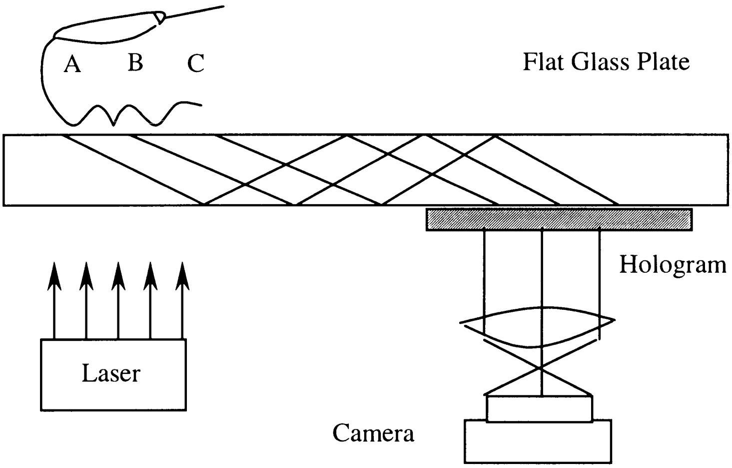


Figure 12-7. Principle of the holographic fingerprint scanner.

To solve these problems, a variation of the prism method with a hologram was proposed [15]. Figure 12-7 shows the principle of the holographic fingerprint scanner. The transparent flat glass plate has a hologram on one end and a fingerprint input section on the other end. A laser beam is irradiated to the glass plate. The end user simply presses the glass plate. The light with an angle of incidence above the critical angle is reflected repeatedly within the glass (total internal reflection) and is propagated glass in the same way as light that is scattered from a grooved surface. The repeatedly reflected light carries the image of the ridges of the fingerprint and comes out of the glass at the hologram on one side of the glass plate. The external image-forming optical system then forms an image of the ridges. A flat glass plate with a plane-parallel plate makes all the optical paths from each point of the fingerprint to the hologram equal and it eliminates trapezoidal distortion.

The two methods described above, inked or inkless, are based on the optical technology. Optical devices are inherently bulky, and moderately expensive-costing. Recently a new technology called solid-state capacitance sensing has been developed (see Figure 12-8). When a user places his or her finger on the sensor, the finger acts as one of the plates of a capacity. The other plate, on the surface of the sensor, consists of a silicon chip containing an array of 300x300 capacitor plates with sensing circuitry at 500-dpi pitch. The distance between the two plates and the chip are measured. The capacitor sensing plates create an 8-bit raster-scanned image of the ridges and valleys of the finger pressed against the chip. This technology offers substantial advantages over optical-based sensors, including smaller size, lower cost, etc.

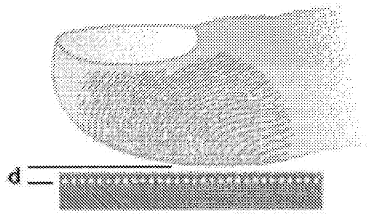


Figure 12-8. Principle of solid-state capacitance sensing.

OTHER BIOMETRICS TECHNOLOGIES

Retina

The eye is positioned in front of the system, approximately three inches from an eyepiece. The end user must look at a green dot for a few seconds, viewed through the eyepiece. When this is done, the eye is sufficiently focused for the scanner to capture the retina pattern. An area known as the fovea, situated in the center of the retina, is scanned and the unique pattern of the blood vessels is captured.

Face

Standard video techniques use a facial image, or collection of images, captured by a video camera. The precise position of the user's face and the surrounding lighting conditions may affect the system's performance. The complete facial image is usually captured and a number of points on the face can then be mapped out. For example, the position of the eyes, mouth and nostrils may be plotted so that a unique template is built. Alternatively a three-dimensional map of the face may be created from the captured image [27].

Thermal imaging techniques under development analyze heat, caused by the flow of blood under the face. A thermal camera captures the hidden, heat-generated pattern of blood vessels lurking underneath the skin. Because infrared cameras are used to capture facial images, lighting is not important and systems can capture images in the dark. However, such cameras are significantly more expensive than standard video.

Finger Geometry

As with finger image verification, the method of capture depends on the system being used. There are currently two main techniques on the market. The first measures the geometry of two or more fingers. A camera takes a three-dimensional measurement when an end user places the index and middle finger, of either the right or left hand, onto a reader. The second technique requires the end user to insert a finger into a tunnel so that three-dimensional measurements of the finger can be taken.

Palm

Palm biometrics are predominantly used for one-to-many identification and the capture process is essentially the same as the optical technique described above for fingerprint. A palm system captures images when a hand is placed on a scanner. Latent or ink palm images can also be scanned into the system in the same way as an AFIS.

Signature

Signature data can be captured via a special sensitive pen or tablet, or both. The pen-based method incorporates sensors inside the pen. The tablet method relies on the tablet to sense the unique signature characteristics. Another variation on these two techniques has been developed and is known as acoustic emission. This measures the sound that a pen makes against paper. Typically for DSV systems, as for all other biometrics, an end user will enroll a number of times so that the system can build a profile of the signing characteristics.

Voice

Users speak into a microphone and utter a previously selected (text-dependent) or random (text-independent) phrase. This process is usually repeated a number of times to build a sufficient profile of the voice [28].

12.3 Machine-Side Interface

When the sample is obtained by the capturing device in the human-side interface, which meets the quality standards, it is ready to be sent to the computer to be processed. The commonly used means to interface between the computer and its peripherals include the widely-used serial port, the faster and wider parallel port, and the newly-emerging RS-422 / RS-485 and USB (Universal Serial Bus). We will go into the details of each of them next and brief other interfaces at the end of this section.

PARALLEL PORT

The parallel port will allow the input of up to 9 bits or the output of 12 bits at any one given time, thus requiring minimal external circuitry to implement many simpler tasks [34,37]. The port is composed of 4 control lines, 5 status lines and 8 data lines. It's found commonly on the back of the PC as a D-Type 25 Pin female connector. There may also be a D-Type 25 pin male connector. This will be a serial RS-232 port and thus, is a totally incompatible port.

Newer Parallel Port's are standardized under the IEEE 1284 standard first released in 1994. This standard defines 5 operation modes,

1. Compatibility Mode;
2. Nibble Mode;
3. Byte Mode;
4. EPP Mode (Enhanced Parallel Port);
5. ECP Mode (Extended Capabilities Port).

The aim is to design new drivers and devices which are compatible with each other and also backwards compatible with the Standard Parallel Port (SPP). Compatibility, Nibble and Byte modes use just the standard hardware available on the original Parallel Port cards while EPP and ECP modes require additional hardware which can run at faster speeds, while still being downwards compatible with the Standard Parallel Port.

Compatibility mode or "Centronics Mode" as it is commonly known, can only send data in the forward direction at a typical speed of 50 kbytes per second but can be as high as 150+ kbytes/second. In order to receive data, you must change the mode to either Nibble or Byte mode. Nibble mode can input a nibble (4 bits) in the reverse direction. Byte mode uses the Parallel's bi-directional feature (found only on some cards) to input a byte (8 bits) of data in the reverse direction.

Extended and enhanced Parallel Ports use additional hardware to generate and manage handshaking. To output a byte to a printer (or anything in that matter) using compatibility mode, the software must

1. Write the byte to the Data Port,
2. Check to see if the printer is busy. If the printer is busy, it will not accept any data, thus any data which is written will be lost.
3. Take the Strobe (Pin 1) low. This tells the printer that there is correct data on the data lines. (Pins 2-9)
4. Put the strobe high again approximately 5 microseconds after putting the strobe low.

Because of the handshaking by software, the speed of port is limited. For solving this problem, the EPP & ECP ports use hardware to check the state of the printer and generate a strobe and /or appropriate handshaking. This means only one I/O instruction needs to be performed, thus increasing the speed. These ports can output at around 1 megabytes per second. The ECP port also has the advantage of using DMA channels and FIFO buffers; thus data can be shifted around without using I/O instructions.

SERIAL PORT

Serial vs. Parallel

In the serial transmission, there is only one data line because the data are transmitted serially one bit at a time. In the parallel transmission, there will be number of data lines so each byte can be transmitted in parallel, that is one byte at a time, one bit on each line. In principle, serial transmissions are considerably slower than parallel transmission; therefore, serial modules are attached to slow devices such as modems and terminals whereas parallel modules are employed with high speed devices such as magnetic tapes and disks [31-33, 35-36, 38-40].

So what are the advantages of using serial data transfer rather than parallel?

1. Serial cables can be longer than Parallel cables. The serial port transmits a '1' as -3 to -25 volts and a '0' as +3 to +25 volts where a parallel port transmits a '0' as 0v and a '1' as 5v. Therefore the serial port can have a maximum swing of 50v compared to the parallel port which has a maximum swing of 5 volts. Therefore cable loss is not going to be as much of a problem for serial cables than they are for parallel.
2. Serial transmission does not need as many wires as parallel transmission. One three-core cable is enough for the simplest serial linking between device and computer while the parallel linking needs at least nineteen-core cable.
3. Serial Ports can be used with modems in applications involving dialing.
4. Microcontroller is quite popular recently. Many of these have been employed in SCI (Serial Communications Interfaces) which can be used to talk to the outside world. Serial communication reduces the pin count of these MPU's. Only two pins are commonly used, Transmit Data (TXD) and Receive Data (RXD) compared with at least 8 pins if you use an 8 bits parallel method (You may also require a Strobe).

DTE and DCE

Devices using serial communication are split into two categories. These are DCE (Data Communications Equipment) and DTE (Data Terminal Equipment.) DCE is device such as your modem, TA adapter, plotter etc while DTE is your computer or terminal.

Electrical Specifications

The electrical specifications of the serial port are contained in the EIA (Electronics Industry Association) RS232C standard. It states many parameters such as -

1. A "Space" (logic 0) will be between +3 and +25 volts.
2. A "Mark" (Logic 1) will be between -3 and -25 volts.
3. The region between +3 and -3 volts is undefined.
4. An open circuit voltage should never exceed 25 volts. (In Reference to GND).
5. A short circuit current should not exceed 500mA. The driver should be able to handle this without damage.

The standards for RS-232 and similar interfaces usually restrict RS-232 to 20kbps or less and line lengths of 15m (50 ft) or less. However, in practice, RS-232 is far more robust than the traditional specified limits of 20kbps over a 15m line.

Mechanical Specifications

For serial ports, there are D-Type 25 pin connectors and/or D-Type 9 pin connectors both of which are male on the back of the PC, thus you will require a female connector on your device. You can find a table of pin connections for the 9 pin and 25 pin D-Type connectors defined by the corresponding computer books.

Null Modems

A Null Modem is used to connect two DTE together. This is commonly used as a cheap way to communicate between computers. This can also be used with many Microprocessor Development Systems.

As mentioned above, RS-232 is simple, universal and well understood and supported everywhere. However, it has some serious shortcomings as an electrical interface. Firstly, the interface presupposes a common ground between the DTE and DCE. This is a reasonable assumption where a short cable connects a DTE and DCE in the same room, but with longer lines and connections between devices that may be on different electrical busses, this may not be true. Secondly, a signal on a single line is impossible to screen effectively for noise. By screening the entire cable one can reduce the influence of outside noise, but internally generated noise remains a problem. As the baud rate and line length increase, the effect of capacitance between the different lines introduces serious crosstalk until a point is reached where the data itself is unreadable. Thirdly, it is restricted to point-to-point communication. It is not suitable for a biometrics system that has one host computer with many scattered capturing points.

The answer to these problems of RS-232 interface is RS-422 and RS-485 interfaces which we will describe next.

RS-422 AND RS-485

Unbalanced vs. Balanced Line Drivers

Each signal that transmits in an RS-232 unbalanced data transmission system appears on the interface connector as a voltage with reference to a signal ground. For example, the transmitted data (TD) from a DTE device appears on pin 2 with respect to pin 7 (signal ground) on a DB25 connector. This voltage will be negative if the line is idle and alternate between that negative level and a positive level when data is sent with a magnitude of ±5 to±15 volts. The RS-232 receiver typically operates within the voltage range of +3 to +12 and-3 to-12 volts.

In a balanced differential system the voltage produced by the driver appears across a pair of signal lines that transmit only one signal and usually are labeled A and B, respectively. A balanced line driver will produce a voltage from 2 to 6 volts across its A and B output terminals and have a signal ground (C) connection. Although proper connection to the signal ground is important, it isn't used by a balanced line receiver in determining the logic state of the data line. A balanced line driver can also have an input signal called an ``Enable" signal. The purpose of this signal is to connect the driver to its output terminals, A and B. If the ``Enable" signal is OFF, one can consider the driver as disconnected from the transmission line. An RS-485 driver must have the ``Enable" control signal. An RS-422 driver may have this signal, but it is not always required [31]. The disconnected or "disabled" condition of the line driver usually is referred to as the ``tristate" condition of the driver. The term ``tristate" comes from the fact that there is a third output state of an RS-485 driver, in addition to the output states of ``1" and ``0."

Balanced Line Receivers

A balanced differential line receiver senses the voltage state of the transmission line across two signal input lines, A and B. It will also have a signal ground (C) that is necessary in making the proper interface connection. If the differential input voltage between A and B is greater than +200 mV, the receiver will have a specific logic state on its output terminal. If the input voltage is reversed to less than 200 mV the receiver will create the opposite logic state on its output terminal. The 200 mV to 6 V range is required to allow for attenuation on the transmission line.

The limitations of RS-232 are largely eliminated by the balanced line interface. The "twisted pair" is extremely effective in eliminating noise from the signal. Balanced systems are used by LAN topologies like Ethernet and Token Ring. They can support line speeds over 10mbps and work reliably at distances of several kilometers.

Network Topologies

Network configuration is not defined in the RS-422 or RS-485 specification. In most cases the designer can use a configuration that best fits the physical requirements of the system.

RS-422 systems require a dedicated pair of wires for each signal, a transmit pair, a receive pair and an additional pair for each handshake/control signal used (if required). The tristate capabilities of RS-485 allow a single pair of wires to share transmit and receive signals for half-duplex communications. This "two wire" configuration (note that an additional ground conductor should be used) reduces cabling cost. RS-485 devices may be internally or externally configured for two wire systems. Internally configured RS-485 devices simply provide A and B connections (sometimes labeled "-" and "+"). Devices configured for four wire communications bring out A and B connections for both the transmit and the receive pairs. The user can connect the transmit lines to the receive lines to create a two wire configuration. The latter type device provides the system designer with the most configuration flexibility. Note that the signal ground line should also be connected in the system. The interface circuit may operate without the signal ground connection, but may sacrifice reliability and noise immunity.

USB

USB is already designed-in to most new PCs being sold today. USB is a cable bus that supports data exchange between a host computer and a wide range of simultaneously accessible peripherals. The attached peripherals share USB bandwidth through a host-scheduled, token-based protocol. The bus allows peripherals to be attached, configured, used, and detached while the host and other peripherals are in operation [30, 41-47].

USB Host

There is only one host in any USB system. The USB interface to the host computer system is referred to as the Host Controller. The Host Controller may be implemented in a combination of hardware, firmware, or software. A root hub is integrated within the host system to provide one or more attachment points.

USB Devices
USB devices are one of the following:
1. Hubs, which provide additional attachment points to the USB;
2. Functions, which provide capabilities to the system, such as an ISDN connection, a digital joystick, or speakers.

Bus Topology
The USB connects USB devices with the USB host. The USB physical interconnect is a tiered star topology. A hub is at the center of each star. Each wire segment is a point-to-point connection between the host and a hub or function, or a hub connected to another hub or function.

Electrical Specifications
The USB transfers signal and power over a four-wire cable. The four lines are labeled VBUS, D+, D-, GND, respectively. The signaling occurs over two wires on each point-to-point segment.
There are two data rates:
1. The USB full-speed signaling bit rate is 12Mb/s.
2. A limited capability low-speed signaling mode is also defined at 1.5Mb/s.

The low-speed mode requires less EMI protection. Both modes can be supported in the same USB bus by automatic dynamic mode switching between transfers. The low-speed mode is defined to support a limited number of low-bandwidth devices, such as mice, because more general use would degrade bus utilization.

The cable also carries VBUS and GND wires on each segment to deliver power to devices. VBUS is nominally +5V at the source.

Mechanical Specifications
All devices have an upstream connection. Upstream and downstream connectors are not mechanically interchangeable, thus eliminating illegal loopback connections at hubs. The cable has four conductors: a twisted signal pair of standard gauge and a power pair in a range of permitted gauges. The connector is four-position, with shielded housing, specified robustness, and ease of attach-detach characteristics.

USB features one "Universal" plug type for all USB peripheral-to-PC connections. Usually there are two USB ports on the back of the PC.

Bus Protocol
The USB is a polled bus. The Host Controller initiates all data transfers. All bus transactions involve the transmission of up to three packets. Each transaction begins when the Host Controller, on a scheduled basis, sends a USB packet describing the type and direction of transaction, the USB device address, and endpoint number. This packet is referred to as the ``token packet." The USB device that is addressed selects itself by decoding the appropriate address fields. In a given transaction, data is transferred either from the host to a device or from a device to the host. The direction of data transfer is specified in the token packet. The source of the transaction then sends a data packet or indicates it has no data to transfer. The destination, in general, responds with a handshake packet indicating whether the transfer was successful.

The USB data transfer model between a source or destination on the host and an endpoint on a device is referred to as a pipe. There are two types of pipes: stream and message. Stream data has no USB-defined structure, while message data does. Additionally, pipes have associations of data bandwidth, transfer service type, and endpoint characteristics like directionality and buffer sizes. Most pipes come into existence when a USB device is configured. One message pipe, the Default Control Pipe, always exists once a device is powered, in order to provide access to the device's configuration, status, and control information.

The transaction schedule allows flow control for some stream pipes. At the hardware level, this prevents buffers from underrun or overrun situations by using a NAK handshake to throttle the data rate. When NAKed, a transaction is retried when bus time is available. The flow control mechanism permits the construction of flexible schedules that accommodate concurrent servicing of a heterogeneous mix of stream pipes. Thus, multiple stream pipes can be serviced at different intervals and with packets of different sizes.

OTHER MACHINE-SIDE INTERFACES

In this section, we will discuss some other machine-side interfaces and compare them with those described above. Table 12-1 compares RS-232 and RS-485 to other interfaces.

RS-232 is popular because it's widely available, inexpensive, and can use longer cables than many other options. RS-485 is also inexpensive, easy to add to a system, and supports even longer distances, higher speeds, and more nodes than RS-232.

The IrDA (Infrared Data Association) interface can use the same UARTs and data formats as RS-232 (with added encoding), but the data transmits as infrared energy over a wireless link. IrDA is useful for short, line-of-sight links between two devices where cabling is inconvenient.

MIDI (Musical Instrument Digital Interface) is used for transferring signals used by musical instruments, theatrical control equipment, and other machine controllers. It uses an optically isolated 5-milliampere current loop at 31.5 kbps.

Microwire, SPI, and I^2C are synchronous serial interfaces that are useful for short links. Many microcontrollers have one or more of these interfaces built-in.

USB (Universal Serial Bus) and Firewire (IEEE-1384) are new, high-speed, intelligent interfaces for connecting PCs and other computers to various peripherals. USB is intended to replace the standard RS-232 and Centronics printer ports as the interface of choice for modems and other standard peripherals. Firewire is faster and designed for quick transferring of video, audio, and other large blocks of data.

Ethernet is the familiar network interface used in many PC networks. It's fast and capable, but the hardware and software required are complex and expensive compared to other interfaces.

IEEE-488, which began life as Hewlett-Packard's GPIB (General-purpose Interface Bus) is another parallel interface popular in instrumentation and control applications.

Table 12-1. Comparison of popular computer interfaces.

Interface	*Format*	*Number of Devices (maximum)*	*Length (maximum, feet)*	*Speed (maximum, bps)*
RS-232	asynchronous serial	2	50-100	20k (115k with some driver)
RS-485	asynchronous serial	32 unit loads	4000	10M
IrDA	asynchronous serial infrared	2	6	115k
Microwire	synchronous serial	8	10	2M
SPI	synchronous serial	8	10	2.1M
I^2C	synchronous serial	40	18	400k
USB	asynchronous serial	127	16	12M
Firewire	Serial	64	15	400M
IEEE-488 (GPIB)	Parallel	15	60	1M
Ethernet	Serial	1024	1600	10M
MIDI	serial current loop	2	15	31.5k
Parallel Port	Parallel	2, or 8 with daisy-chain support	10-30	1M

12.4 Case Study: PalmScanner Interface

In this section, we will present a design sample of biometrics human-machine interface, which is used to palm biometrics.

As with finger images, ridges, valleys and minutiae data are also found on the human palm. Because the minutiae on the palm are bigger than those found on the fingerprint, it is easier for palm biometrics devices to capture a high-quality palm image. The certain conditions that affect the finger scanning biometrics, such as dirty, dry and wet prints, pose less challenge to palm biometrics.

PALM CAPTURING STRUCTURE

The palm biometrics system is composed of a light source, a palm-laying platen, a camera/CCD sensor, an A/D converter and an interface to the computer (see Figure

12-9). This palm biometrics system is controlled by the driver installed in the computer. According to the given input program, a palm image can be captured by the system.

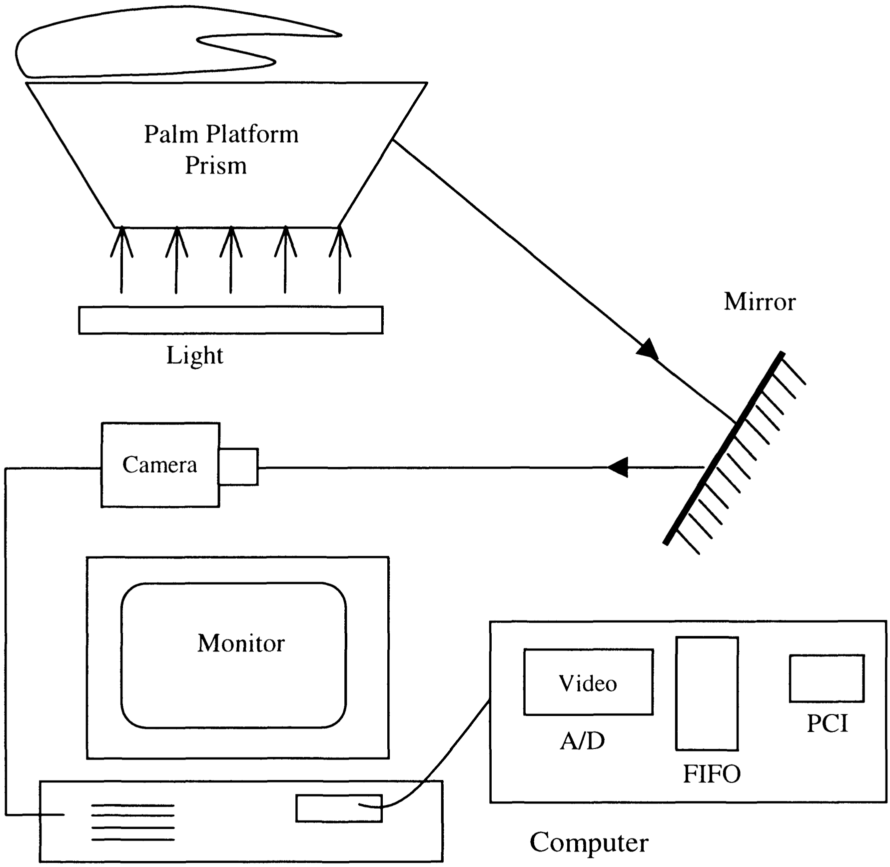


Figure 12-9. A basic structure of a palm biometrics system.

Optical Design

A PalmScanner consists of a platform, a prism, a mirror, lens and light source. The platform and the prism are normally connected together as one component which functions as the first interface between a user and the system. At the design stage, the range of an image must be considered first in accordance with the technical parameters associated with the palm-scanner. The size of the prism depends on these parameters. In designing the platform, it is very important to make the platform touched by a palm with little gap. Since palm is not flat, there may be some gaps between the platform and the palm, which will cause white area in the palm image and affect the processing results. In general, the design strategy for a palm-scanner is application dependent. If the size of sample is small (less than 50% of a palm), it is appropriate to use a flat platform to simplify the process of design and manufacturing. In designing a prism, there are two factors to be considered. One is to extract the discriminate palmprint textures clearly, the other is to reduce image distortion and other external disturbance. The cost of such a prism will be increased significantly to satisfy the high requirements for these two factors.

Mirror is used to adjust the path of light by reflection so as to reduce the size of the system. The design of a mirror also depends on the size of samples. To achieve high reflection rate on the mirror, different reflection film will be used in relation to different light source.

The lens installed in the PalmScanner is different from the conventional lens. Since the palm-prints are captured by prism, the image obtained is different from the original one. If a conventional lens is used, it will cause serious distortion. To solve such a problem, two procedures have been taken for correction. At the first stage, the palm image created by prism is assumed to be located at ``infinite" point. The distance between the virtual image and the real image can be determined in the second stage.

In designing light sources, in addition to satisfying the given design criteria, the height of light source should be adjusted to reduce the distance difference between the object and the image.

Image Channel

The image channel is the interface between user and computer for palm-print extraction. It includes image sensor, A/D convertor and interface software. A schematic diagram of our basic palm scanner is shown in Figure 12-10.

The selection of image sensor depends on sample size and resolution. The selection criterion can be formulated as sample size * resolution * 2. In general, an image sensor is in CCD format, which can be categorized as line-array and matrix array. For a matrix array, it can be further classified into TV model and the general model. The cost for a TV-based model is low because of the high production. However, the number of pixels is low with low resolution. Currently, the general model can reach millions of pixels, but the price is high. In addition, special driving circuit is required. Thus, such a circuit is used only for very specific applications with high quality requirements.

It should be pointed out that the combination of line CCD array and the second scanning will perform large data sampling with high resolution. However, the disadvantages are more mechanical components and low data sampling speed. Such a system is used only for very large sample size.

An A/D convertor is required to convert analog signal to digital signal. There are varieties of A/D convertors available. What to be used depends on sampling frequency, accuracy and the integration level of the circuit.

The digital signal output from A/D converter is input to computer. Due to the large data size of a palm-print image, PCI is used and an FIFO is added as a data buffer to match the data transmission speed between PCI and A/D conventor.

The signals output from CCD are controlled by a computer. The development of the relevant software should be compatible to the hardware circuit, high data flow and buffer size. To achieve high speed, low-level language such as assembly language is recommended.

As a machine-side requirement, a computer is used for data acquisition, processing and storage. They have been mentioned before.

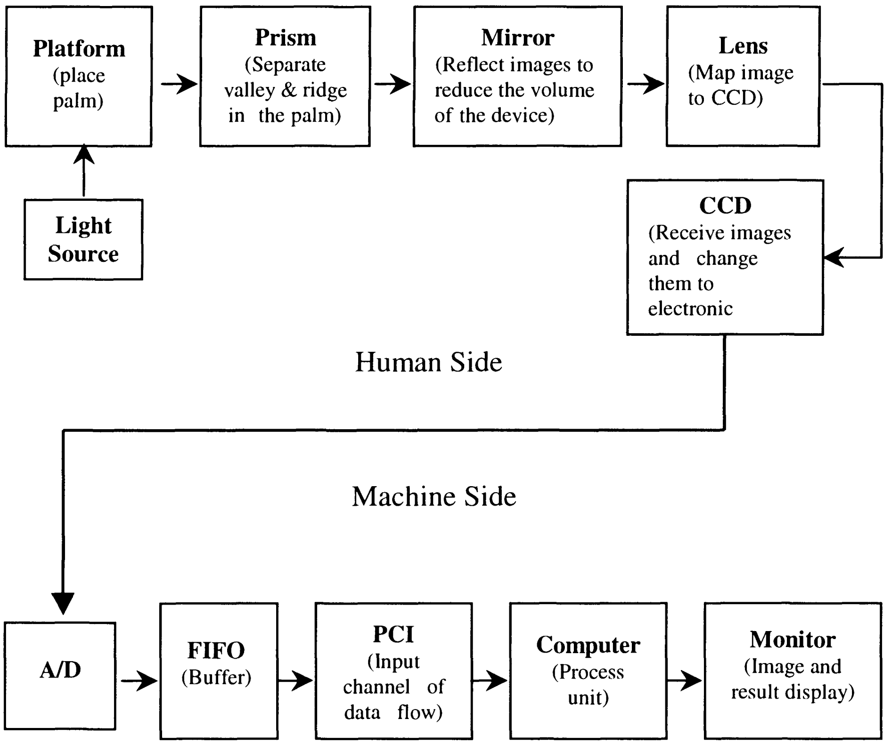


Figure 12-10. A schematic diagram of our basic palm scanner.

PALM DEMO SYSTEM

As a typical design of biometrics human-machine interface, a palm demo system has been built in our university (see Figure 12-11). Based on our discussion on how to develop a basic palm scanner, this system is implemented by the following two respects shown by Figure 12-10:

1. Human side: It consists of a platform, a prism, a mirror, lens, CCD and light source. Since a platform is used to place palm and a prism to separate valley and ridge in the palm, they are normally connected together as one component which is as the first interface between a user and the system. A mirror is to

reflect image to reduce volume of the device. Lens will map images into CCD, which receives and changes them to electronic signals.

Figure 12-11. The real capturing system for palm biometrics.

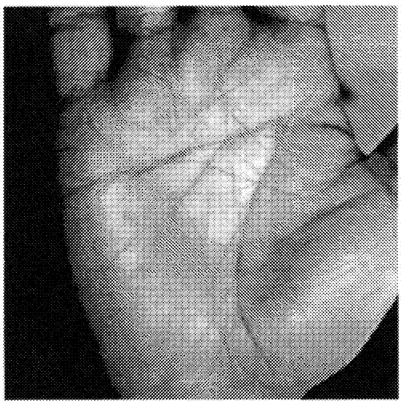


Figure 12-12. The capturing palm images from our device.

2. Machine side: An A/D (analog/digital) is first used to obtain a digital form, which is sent to a buffer and then transferred to a computer by an input channel interface (PCI). The images and their results obtained can be shown by a monitor.

Many palm images have been obtained from our device to test our system. As a result, a palm image is shown in Figure 12-12.

References

[1] R.P. Wildes, "Iris Recognition: An Emerging Biometrics Technology," *Proc. IEEE*, vol. 85, pp.1348-1363, 1997.

[2] F. Bouchier, J.S. Ahrens and G. Wells, "Labororatory Evaluation of the Iriscan Prototype Biometrics Identifer," Sandia National Laboratories, Albuquerque, NM, *Tech. Rep. SAND'96-1033*, 1996.

[3] J.G. Daugman, *Biometrics Personal Identification System Based on Iris Analysis*, U.S. Patent 5291560, 1994.

[4] L. Flom and A. Safir, *Iris Recognition System*, U.S. patent 4641349, 1987.

[5] K. Hanna, R. Mandelbaum, L. Wixson, D. Mishra and V. Paragana, "A System for Noninstrusive Human Iris Acquisition," in *Proc. Int. Association for Pattern Recognition Workshop on Machine Vision Applications*, Tokyo, Japan, pp.200-203, 1996.

[6] R.G. Johnson, "Can Iris Patterns Be Used to Identify People," Los Alamous National Laboritory, CA, Chemical and Laser Science Division, *Rep. LA-12331-PR*, 1991.

[7] J. E. Siedlarz, "Iris: More Detailed than a Fingerprint," *IEEE Spectrum*, pp. 31-27, 1994.

[8] R.P. Wildes, J.C. Asmuth, etc, "A Machine Vision System for Iris Recognition," *Machine Vision Applicant*, vol. 9, pp. 1-8, 1996.

[9] R.P. Wildes, J.C. Asmuth, S.C. Hsu, R.J. Kolczynski, J.R. Matey and S.E. McBride, *Automated Noninvasive Iris Recognition System And Methods*, U.S. Patent 5572596, 1996.

[10] R.P. Wildes, J.C. Asmuth, G.L. Green, S.C. Hsu, R.J. Kolczynski, J.R. Matey and S.E. McBride, "Iris Recognition for Security Access Control: Final Report, National Information Display Laboratory," Princeton, NJ, *Tech. Rep.*, 1992.

[11] R.P. Wildes, J.C. Asmuth, G.L. Green, S.C. Hsu, R.J. Kolczynski, J.R. Matey and S.E. McBride, "A system for Automated Iris Recognition," *in Proc. IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, pp.121-128, 1994.

[12] J.G. Daugman and G.O. Williams, "A Proposed Standard for Biometrics Decidablity," *in Proc. CardTech/SecureTech Conf. Atlanta*, GA, pp.233-234, 1996.

[13] D. Sidlauskas, "Hand: Five Me Five," IEEE Spectrum, Febraury 1997.

[14] A. Jain, et al., "On-line Fingerprint Verification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, 1997.

[15] Seigo Igaki, et al., "Real-time Fingerprint Sensor Using a Hologram," *Applied Optics*, vol. 31, 1992.

[16] K. Morita, K. Asai, "Automatic Fingerprint Identification Terminal for Personal Verification," *SPIE*, vol. 638, pp. 174-181, 1986.

[17] *Edge Lit Hologram For Livescan Fingerprinting*, http://esatview.org/ImEdge, 1997.

[18] M. Hartman, "Compact Fingerprint Scanner Techniques," in *Proc. Biometrics Consortium 8^{th} Meeting*, San Jose, CA, June 1996.

[19] J. Klett, "Thermal Imaging Fingerprint Technology," *in Proc. Biometrics Consortium 9^{th} Meeting*, Crystal City, VA, Apr. 1997.

[20] H.C. Lee and R.E. Gaensslen, Eds., *Advances in Fingerprint Technology*, Elsevier, New York, 1991.

[21] K. McCally, D. Setlak, S. Wilson and J. Schmitt, "A Direct Fingerprint Reader," in *Proc. CardTech/SecurTech*, vol. I: Yechnology, Atlanta, GA, pp.271-279, May 1996.

[22] L. O'Gorman and J.V. Nickerson, "An Approach to Fingerprint Filter Design," *Pattern Recognit.*,vol.12, pp. 269-275, 1993.

[23] J.Schneider, "Improved Image Quality of Live Scan Fingerprint Scanners Using Acoustic Backscatter Measurements," in *Proc. Biometrics Consortium 8^{th} Meeting*, San Jose, CA, June 1996.

[24] A.R. Roddy, "Fingerprint Features—Statistical Analysis and System Performance Estimates," in *Proc. of IEEE*, vol. 85, pp. 1390-1421, 1997.

[25] J.P. Holmes, L.J. Wright and R.L. Maxwell, "A Performance Evaluation of Biometrics Identification Devices," Sandia National Laboratories, Albuquerque, *NM tech. Rep. SAND91-0276*, 1991.

[26] F.A. Jenkins and H.E. White, *Fundamentals of Optics*, McMillan,, 1976.
[27] J. Zhang, Y. Yan, etc, "Face Recognition: Eigenface, Elastic Matching, and Neural Nets," in *Proc. of IEEE*, vol. 85, pp. 1423-1435, 1997.
[28] J.P. Campbell, etc, "Speaker Recognition: A Tutorial," in *Proc. of IEEE*, vol. 85, pp. 1437-1462, 1997.
[29] W.C. Shen, etc, "Evaluation of Automated Biometrics_Based Identification and Verification Systems," in *Proc. of IEEE*, vol. 85, pp. 1464-1478, 1997.
[30] *Universal Serial Bus Specification Revision 1.1*, http://www.usb.org.
[31] *B&B Electronics, RS-422 and RS-485 Application Note*, http://www.bb-elec.com
[32] *Using RS232 Devices with USB*, http://www.ontrak.net/usb.htm.
[33] *The Complete pinouts guide to Parallel-Serial Port, Network and Monitor Cables*, http://www.geocities.com/TheTropics/Shor-es/2250/pinout.html.
[34] J. Axelson, J. L. Axelson, *Parallel Port Complete*, http://www.amazon.com.
[35] R. Grier, et al, *Visual Basic Programmer's Guide to Serial Communications*, http://www.amazon.com.
[36] J. Campbell, *C Programmer's Guide to Serial Communications*, http://www.amazon.com
[37] D.V. Gadre, *Programming the Parallel Port; Interfacing the PC for Data Acquisition and Process Control*, http://www.amazon.com.
[38] M.D. Seyer, *Rs-232 Made Easy : Connecting Computers, Printers, Terminals, and Modems*, http://www.amazon.com, 1991.
[39] J. Axelson, *The Microcontroller Idea Book : Circuits, Programs & Applications Featureing the 8052-Basic Single-Chip Computer*, http://www.amazon.com.
[40] M.D. Seyer, *Complete Guide to Rs-232 and Parallel Connections : A Step-By-Step Approach to Connecting Computers, Printers, Terminals, and Modems*, http://www.amazon.com, 1988.
[41] J. Axelson, *Usb Complete : Everything You Need to Develop Custom Usb Peripherals*, http://www.amazon.com, 1999.
[42] S. McDowell, M.D. Seyer, *Usb Explained*, http://www.amazon.com, 1998.
[43] Don Anderson, Inc. MindShare (Contributor), *Usb System Architecture (PC System Architecture Series)*, http://www.amazon.com, 1997.
[44] J. Koon, *USB : Peripheral Design*, http://www.amazon.com, 1998.
[45] J. Hyde, *USB Design by Example: A Practical Guide to Building I/O Devices*, http://www.amazon.com, 1999.
[46] K. Jaff, *Usb Handbook: Based on Usb Specification Revision 1.0*, http://www.amazon.com, 1997.
[47] J. Garney, et al, *USB Hardware & Software*, http://www.amazon.com, 1998.

13 PERSONAL AUTHENTICATION

The advance of technology is always inspired by the practical applications and the emergence of automatic biometrics technology is exactly rooted in the requirement of real world security applications. Whether this new technology can last for a long time will be decided by how well it can solve the security problems. Though still at developing stage, biometrics technology has been put in use in various applications and some of them do work well. Along with the widespread application of biometrics technology, more funds and more attention are given to this ascending technology.

In this chapter, we will answer questions such as 'Who are using biometrics?', 'What is the potential usage of biometrics?', 'If I want to apply biometrics to secure my system, what aspects should be put into consideration?', ' Nowadays, there are so many biometrics products, how to select an appropriate one to fulfil my requirements?', 'Can I use products from different vendors together?', 'Where can I find biometrics products and vendors and what kind of services are being provided?'.

13.1 Examples of Current Applications

Usually, when thinking of using a brand new technology, our hearts can't help beating faster. Can this new technology work well? Will this new technology do harm to human beings? Questions like these will spiral inside our heads and make us quite hesitant to apply the new technology. But if we find that somebody else has already applied the technology successfully or at least they are trying to use the technology and do not afraid of taking the risk of failure, we will become more confident and easier to make a decision to use the new technology. So in this unit, we will introduce some real world biometrics applications.

IMMIGRATION AND NATURALIZATION SERVICE'S PASSENGER ACCELERATED SERVICE SYSTEM (INSPASS)

INSPASS was designed as a means to provide prompt admission for frequent travelers to the US by allowing them to bypass the personal interview/inspection part of the entry process. It uses hand geometry to verify the identity of the traveler at an automated inspection station. INSPASS stations have been installed, for example, at John F. Kennedy Airport in New York and Newark International Airport in New Jersey. INSPASS is available for citizens of 23 countries in the US visa waiver program who visit the US at least 3 times per year. These same 23 countries are planning to participate in the Future Automated Screening for Travelers (FAST) project, which would allow travelers to use automated passport inspection stations in countries participating in FAST.

CANPASS

CANPASS is the Canadian version of INSPASS, except that it uses a fingerprint biometrics, rather than hand geometry, for traveler verification. The goal of CANPASS is to ease the transfer of goods and people between the US and Canada. CANPASS is in use at the Vancouver International Airport.

PORTPASS

PORTPASS is another INS initiative similar to INSPASS except that people in vehicles at borders are being monitored and it uses voice recognition biometrics, instead of hand geometry. PORTPASS is used at a US/Canadian vehicle border crossing and is planned for use at US/Mexican border crossings. One version of PORTPASS (the Automated Permit Port) requires the vehicle to stop. It will also have a Video Inspection Service allowing a driver to have a conference with an Inspector. Another version, known as the Dedicated Commuter Lane, uses a radio frequency tag affixed to the vehicle in order to obtain the biometrics as the vehicle is moving.

FEDERAL BUREAU OF PRISONS

The Federal Bureau of Prisons is using hand geometry units to monitor the movements of prisoners, staff, and visitors within certain Federal prisons. A successful trial with the hand geometry units was conducted at the Federal prison in Jesup, Georgia. Visitors must enroll upon arrival and are given a magnetic stripe card containing information that points to their identifying information in a central database. This card must be carried with the visitor at all times. Staff and inmates must also enroll. Staffs are enrolled to reduce the possibility of mistakenly identifying them as an inmate or for positive identification in the event of a disturbance. Prisoners are enrolled for access control to places such as the cafeteria, recreation lounges, and the hospital. The system also allows for the tracking of prisoners' movements. By the end of 1995, around 30 Federal prisons were being to have the hand geometry monitoring system installed.

AUTOMATED FINGERPRINT IMAGE REPORTING AND MATCH (AFIRM)

In July of 1991, Los Angeles County in California installed the first AFIRM system. AFIRM was needed to reduce fraudulent and duplicate welfare benefits. The fingerprints of new applicants for welfare benefits were checked against a central database of prior claimants. Within the first 6 months of use, the county saved $5.4 million dollars, and the savings have been growing ever since. The system was so successful that San Francisco, Alameda County, and Contra Costa County installed AFIRM to check new claimants' fingerprints against existing recipients in these locales. AFIRM was expected to be in statewide operation in California by some time in 1997.

SPANISH NATIONAL SOCIAL SECURITY IDENTIFICATION CARD (TASS)

The TASS program is a smart card initiative employing fingerprint technology to eliminate enrollment duplication and provide secure access to personal information upon retrieval. The program is an ambitious one, in that it will combine pension, unemployment, and health benefits all on one card.

THE COLOMBIAN LEGISLATURE

The Colombian Legislature uses hand geometry units to confirm the identity of the members of its two assemblies immediately prior to a vote. The voting has been conducted this way since 1992.Many Federal, State, and local government agencies have purchased biometrics systems. The Defense Advanced Research Projects Agency, Drug Enforcement Agency, Department of Defense, Department of Energy, Department of Public Safety, Department of State, Federal Bureau of Investigation, Federal Reserve Bank, Hill Air Force Base, the Pentagon and the US Mint have approximately 250 biometrics devices with 13,000 enrolled users for accessing control applications.

13.2 Potential Application Areas

Biometrics applications are not limited to the areas mentioned in the last section. In fact, as long as a system needs to recognize people, it can incorporate biometrics. In the law enforcement community, matching finger images or part of palm images is the most common method to process criminal suspect and bring guilty criminals to justice. Also we have seen many times in the movies that the police asked the witness to describe the criminal's physical features such as length, width, hair color, shape of the face and etc, and then reconstruct a picture for the criminal. In some movies, we may see that the criminal called the victim over the phone and the police recorded the voice of the criminal and search the criminal according to the voice. All these scenes are examples of finding people by their unique physical features (finger, palm, face, etc) or behavioral trait (voice) and automatic biometrics can help in all these examples. It is not difficult to understand that the law enforcement community is the largest biometrics user group. Police forces throughout the world are using Automatic Fingerprint Identification System to assist crime detection. There are quite lots of biometrics vendors earning significant revenues in this area.

Businessmen always play an important role of spreading a new technology. When the automatic biometrics technology becomes more and more mature in the law enforcement area, it is introduced to the civilian applications by biometrics product vendors. Usually all civilian biometrics applications are some kind of access control. We may simply classify all the civilian biometrics applications as physical access control and data access control. Physical access control ensures only authorized individuals to physically access secure areas while data access control secures the access to sensitive data. Securing benefit systems from fraud, preventing illegal immigrants from entering a country or prisoners from leaving a prison all belong to physical access control, while Internet banking, telephone banking, ATM, and Web Store belong to data access control. Automatic biometrics is a rapidly expanding market. Fraud is an ever-increasing problem and security is becoming a necessity in many walks of life. Civilian access control, therefore, will not be restricted to the application areas mentioned below and will branch out to other market opportunities, as soon as a need is identified.

BANKING

Banks may embrace biometrics technology from various aspects. Automated Teller Machines (ATMs) and transactions at the point of sale, telephone banking, Internet banking and many other banking applications are vulnerable to fraud and can be secured by biometrics. Figure 13-1 is shows various biometrics and bank branches where they can be applied.

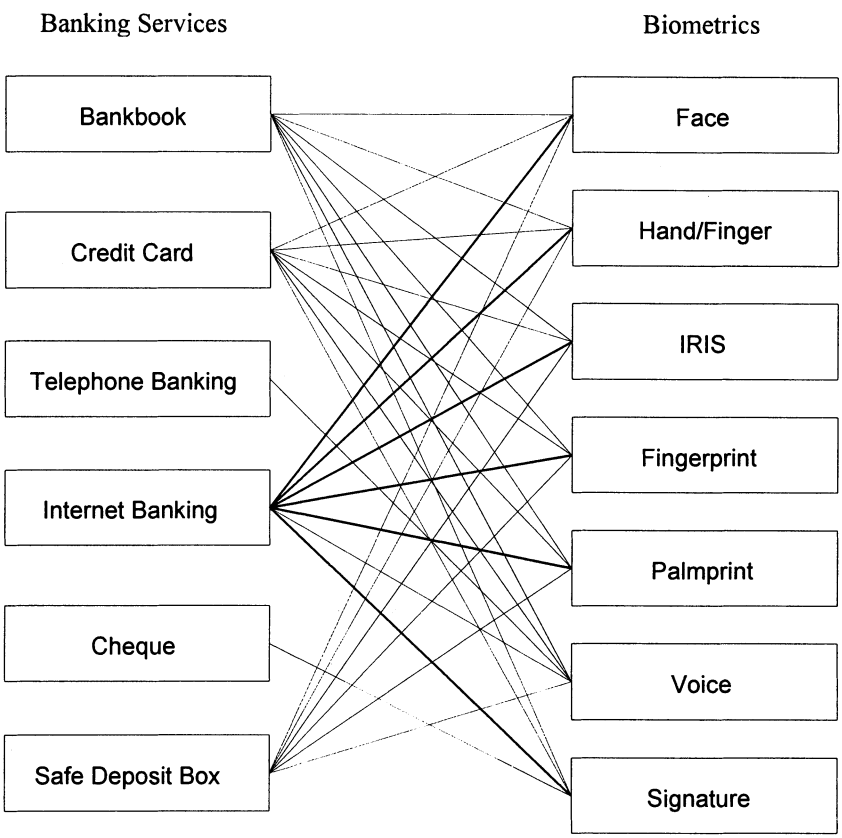


Figure 13-1. Biometrics applications in banking.

BENEFIT SYSTEMS

When applying in benefit systems, biometrics plays a role different from that in banking or in physical access control. In banking or physical access control or other methods of access control, all people who have the authority would have their unique physical features or behavioral traits registered in the system before the system being used. And when somebody needs access to the system, he should get his unique feature captured at that point and the system would check the newly captured feature and decide whether it is from the same person as the registered one. This will prevent unauthorized people from accessing to the system. But in benefit systems, one needn't register his feature first. Only when he gets the benefit, does he register his unique feature into the system. Then the system will check whether this feature has been entered. If yes, it is to say that the person has already got the benefit and he won't be given the second time. Biometrics is well placed to capitalize on this phenomenal market opportunity and vendors are building on the strong relationship currently enjoyed with the benefits community.

COMPUTER SYSTEMS (LOGICAL ACCESS CONTROL)

Currently, computer systems are using passwords as their secure guards. But on one hand, remembering tens of passwords and changing passwords very often almost become a headache to everyone who is using a computer; on the other hand, password itself doesn't have direct connection to the end user. If somebody gets the password, he will be considered as a legal user to the computer system even though he is a criminal and meanwhile if a legal user forgets his password, he will be refused to access his own computer. Biometrics technology binds the authority directly with the end user and lets various passwords go away. Voice and fingerprint recognition are now the most promising techniques in this area.

IMMIGRATION

Terrorism, drug-running, illegal immigration and an increasing throughput of legitimate travelers are putting a strain on immigration authorities throughout the world. It is essential that these authorities can quickly and automatically process law-abiding travelers and identify the law-breakers. Biometrics is being employed in a number of diverse applications to make this possible. The US Immigration and Naturalization Service is a major user and evaluator of a number of biometrics. Systems are currently in place throughout the US to automate the flow of legitimate travelers and deter illegal immigrants. Elsewhere biometrics is capturing the imagination of countries such as Australia, Bermuda, Germany, Malaysia and Taiwan.

NATIONAL IDENTITY

Biometrics is beginning to assist governments as they record population growth, identify citizens and prevent fraud occurring during local and national elections. Often this involves storing a biometrics template on a card that in turn acts as a national identity document. Finger scanning is particularly strong in this area and schemes are already under way in Jamaica, Lebanon, The Philippines and South Africa.

PHYSICAL ACCESS
The biometrics product using in physical access is somewhat like a kind of gate lock. The size may vary from only several persons or more than thousands of people. Its usage varies from private houses, schools, to nuclear power stations, military facilities, theme parks, hospitals, offices and supermarkets. As security becomes more and more important for parents, employers, governments and other groups - biometrics will be seen as a more acceptable and therefore essential tool. The potential applications are infinite. Cars and houses, for example, the sanctuary of the ordinary citizen, are under constant threat of theft and biometrics, if appropriately priced and marketed, could offer the perfect security solution.

TELEPHONE SYSTEMS
Global communication has truly opened up over the past decade, while telephone companies are under attack from fraud. Once again, biometrics is being called upon to defend this onslaught. Speaker ID is the technique of recognizing people by their voices. It is obviously well suited to the telephone environment and is catching up these opportune markets quickly.

TIME, ATTENDANCE AND MONITORING
Currently, some factories and companies are using card to monitor the movement of their employees. When they come to work, they need to pick a hole on their cards and another hole when they leave. Such kind of things can be assisted by biometrics. With a biometrics system, employees may press their fingers on a small platform when they come or leave. This may prevent any forms of cheating. But whether using such a kind of system to monitor employees' movement is still a question because some people think it may attack employee's privacy.

13.3 How to Select a Biometrics System

DIFFICULTY TO SELECT A BIOMETRICS SYSTEM
The performance of a biometrics system has great impact on many factors such as humidity, light, noise, end user's attitude and familiarity to the system. Therefore, the most accurate testing should be performed in real working circumstances. High levels of accuracy in one application do not qualify a system for an entirely different application. The quoted performance figures of a biometrics system will only to be applied to the specific application from which they are quoted. Each application is wildly different in terms of system workload and throughput, environmental factors and other variables.

For example, the same fingerprint verification system may have very high accuracy in a university restaurant while it may work badly in a village where most people have their fingerprints worn heavily.

Anyway, it is not to say the performance ruler, Fault Accept Rate and Fault Reject Rate are meaningless though the two rates may vary in diverse environments. FAR and FRR provided by the developer can be used as a guide to understand a system's general ability.

In one sentence, a biometrics system's performance is application sensitive and how to make a biometrics system adaptive to a particular application needs significant consideration.

VARIOUS FACTORS: WHETHER A BIOMETRICS SYSTEM IS NEEDED

Before applying biometrics, we must make it clear that what the business driver is. What is the main goal of the whole project? What are the constraints of the project? Deadlines, Budgets and etc. What security level is needed? How is the current system? What is the weakness of current system? Is it necessary to apply biometrics? Can biometrics solve the existing problems? Are there any other choices to secure the system? Will biometrics bring any trouble into the system? Can Biometrics integrate well with the current system? Will users of the system accept this new work style? In one word, to make sure biometrics is needed is the first step of applying this new technology.

COMPARISON TO DIFFERENT BIOMETRICS TECHNIQUES

Factors considered in biometrics system evaluation includes:

Vulnerability to Fraud

Biometrics system aims at providing high level security, so whether a biometrics trait is hard to mimic is an essential consideration in construction of such applications. Gauge out one's eye or truncate one's finger to fraud a biometrics system has appeared in some movies.

Ease of Use

One springboard of the biometrics system is to make the public get rid of the bother of remembering tens of password and keeping strings of keys. Therefore such system should be really user friendly in stead of bringing even headache than lost a room key.

Intrusive to Human Beings

Certain biometrics systems are seen to be more intrusive than others. For example, Retina capturing has to make eyes exposed to bright beam while voice seems non-intrusive. However, sometimes, higher accuracy may be gained by more intrusive approach. Places where need high levels of security have to choose those intrusive methods. For example, workers at a nuclear power plant would probably acknowledge the need for a degree of intrusiveness, as security is such an important issue.

Applicability

Physical characteristics vary and some individuals will not be able to use a biometrics system. No single biometrics system can capture and match biometrics data for the global population in all circumstances. Human beings are as diverse and unpredictable as environments. A few individuals have damaged fingers, limbs, voice boxes or eyes. This may make verification and identification with a single biometrics system impossible; but it may be possible using a multiple biometrics system. Also, it does not mean that a single biometrics system is unable to perform an application where a minority of people cannot have a biometrics sample captured. It is simply that the minority cannot use the system automatically and must be dealt with in an appropriate manner.

Speed of Verification
Response time is a key issue for any computer systems, so does for biometrics systems.

Size of Storage for One Biometrics Template
To an identification system, this factor directly affects the overall database size and searching speed. To a verification system, in case that the registered template is stored in some special media such as barcodes, magnetic cards or smart cards, this factor could determine the cost of a card.

Long-term Stability
The biometrics feature chosen to identify person in a system should be stable for at least a little bit longer than the system planed to be used, so that the system can work correctly during the period when it is in duty.

Maturity of Technology
Some biometrics features such as fingerprint and signature have been used for a long time and their accuracy has been proved widely. Meanwhile, other biometrics such as face and voice are newcomers in this area and need to be proved in real-time applications. Table 13-1 and 13-2 are two tables of comparison to the existed biometrics systems.

WHEN DECIDED TO APPLY BIOMETRICS, AND WHAT SHOULD BE PUT INTO CONSIDERATION?

Of course investigating to various existing biometrics systems and products is mandatory. Besides this, there are still many questions should be answered. What kind of biometrics system is required? Is it an identification system or verification system? What are the characteristics of the end user population? How are the age, gender, ethnic origin and occupation of the end user group? In case that something is wrong with the biometrics system, what will be the substitute method? What is the accuracy of the biometrics system? Will the population of the system grow? How does the environment look like?

At last a detailed testing plan must be prepared.

Table 13-1. Comparison to various biometrics techniques

Technical Factor	*Hand Geometry*	*Retina*	*Fingerprint*
False rejection rate	0.2 percent, one-try	12.4 percent (one try) 0.4 percent (three try)	1%-5%, three try
False acceptance rate	0.2 percent, one-try	0 no false acceptances	0.01-0.0001 percent(three try)
Vulnerability to fraud	Almost impossible to secretly obtain hand geometry data. However, when the person co-operates, this seems not at all impossible.	No counterfeits seem possible. False eyes, contact lenses and eye; transplants cannot breach the security of this device	Dummy fingers and dead fingers will be detected when high security platen is installed
Ease of use	The first time one needs to get used to it. After some experience it is not difficult.	Difficult to use. Socially difficult to accept because people do not like to have their eyes scanned	Easy to use, but it can be associated with criminal investigations
Universality	Not suitable for people who have rheumatic hands or related physical impairments	Suitable for everyone with eyes	Not for people with damaged fingerprints due to daily handling of rough material
Speed of identification	Less than 3 seconds	1.5 seconds	Average verification time 2 seconds. Maximum is 20 seconds.
Size for storage of template	Only 9 bytes	40 bytes	1203 bytes. After compression it is smaller than 800 bytes.
Long term stability	Sizes of hands will change for children, and can change when someone gains or loses a lot of weight	The retinal vascular pattern is very stable. Only a small number of diseases or injuries will change this pattern;	Sizes of fingerprints change for children. Apart from that they always remain the same.
Maturity of technology	World-wide used in many systems	Used in a fair amount of systems	World-wide used in many systems

Table 13-2. Comparison to various biometrics techniques

Technical Factor	*Iris*	*Retina*	*Face*	*Finger scanning*	*Voice*	*Hand Geometry*	*Finger Geometry*	*Palm*	*Signature*
Level of Accuracy	Very high	Very High	High	High	High	High	High	High	High
Ease of use	Medium	Low	Medium	High	High	High	High	High	High
Vulnerability to fraud	Very high	Very High	Medium	High	Medium	High	High	High	Medium
Intrusive to Human beings	Medium	Medium	High	Medium	High	High	Medium	Medium	Very High
Long-term Stability	High	High	Medium	High	Medium	Medium	Medium	High	Medium
Industry Standards	-	-	-	ANSI/NIST Data Interchange & FBI Image Compression Standards	Speaker Verification API(SVAPI)	-	-	See Finger scanning	-
Factors that May affect Performance	Glasses worn by end user	-	Poor lighting; aging of face; glasses; facial hair	Dry, dirty or damaged finger images; age, gender and race of end user	Background and network noise; colds and other factors can change the voice	Diseases such as arthritis and rheumatism in end users	See Hand Geometry	Dry, dirty or damaged palm images; age, gender and race of end user	Illiteracy; signatures that constantly change or are easily imitated

13.4 Application Programming Interface Standards

WHY DO WE NEED BIOMETRICS SYSTEM API STANDARDS?

Along with the booming of biometrics industry, more and more companies are involved in the biometrics development and therefore one by one biometrics products are delivered into the market. Though various products provide customers more choices and nice prices, they also bring some troubles. The follows are some examples of troubles that biometrics applications may bring to customers:

1. If a customer wants to buy the fingerprint capture devices from one company and the fingerprint identification software from another company, then whether the software can process the fingerprint images captured by the device could become a big problem.
2. If an organization already has a database for all the employees' fingerprints and now wants to install an employee attendance monitor system, then how can they find a fingerprint identification system that can use the existing database?
3. Also if a customer has built up a biometrics system and now wants to enlarge the existing system, but the company who provided the original system has gone off the market, then how do they deal with this situation?
4. And if club restaurant installed a hand geometry access control system to ensure that only members of that club may enter, but some members have their hands hurt or lost, then how do they solve this problem?
5. To a new customer, if he wants to purchase a biometrics system, but every system has its own interface and functions, there is no standard for them, then how can he compare them and how can he choose an appropriate one?

...

All these troubles are related to system API standards. Industry standards are truly positive to encourage both the developers and customers to embrace this quickly growing technology. Because with industry standards, products from different companies may become interchangeable and this makes the customer more opportunities to choose different parts of the system from different vendors and integrate them together. Therefore customers will become more confident with biometrics products and this will inspire the development of biometrics industry. When making standards, both the hardware interface and the software interface will be discussed and many vendors and customers will come together to evaluate the goodness and shortcomings of various products. Through such kind of discussion, vendors and customers may understand each other better and different vendors may exchange their ideas easily. This will definitely improve biometrics industry as a whole. In this highly competitive society, small companies want to tie up to survive while large companies want to confirm their leading position, so they are all interested in making industry standards.

WHO ARE WORKING ON BIOMETRICS PRODUCT API STANDARDS OR BIOMETRICS TECHNIQUES EXCHANGING?

By now, many organizations are teaming up to exchange ideas or making product standards. Here we will introduce some important consortiums and their work.

BioAPI

The BioAPI group was introduced to the public on April 27th 1998. The BioAPI Consortium was formed to develop a widely available and widely accepted API that will serve for various biometrics technologies. The intent is to work with industry biometrics solution developers, software developers, and system integrators to leverage existing standards to facilitate easy adoption and implementation. The group consists of the founders (also known as Promoters) and participants who lend their expertise during the preliminary spec development, known as Contributors; and those who intend to use the API, known as Adopters. The founding member companies include Compaq, IBM, Identicator, Microsoft, Miros, Novell (also representing SVAPI which is an acronym for Speaker Verification API). Contributor Members Include I/O Software, IriScan, National Institute of Standards and Technology (NIST), National Security Agency (NSA), Printrak International, Recognition Systems, Saflink, Siemens, Unisys.

The BioAPI group has been meeting regularly and working towards the development of a standard API. So far the group's accomplishments include Development of the Business Values and Objectives and Development and release of a High Level API Architectural Description. This includes design and development drivers, API Features and interfaces, Architectural Overview.

IBIA

The International Biometrics Industry Association (IBIA) is a trade association founded in September 1998 in Washington, D.C., Its mission is to advance, advocate, defend and support the collective international interests of the biometrics industry. IBIA is governed by and for biometrics developers, manufacturers and integrators, and is impartially dedicated to serve all biometrics technologies in all applications.

IBIA speaks with authority for the biometrics industry to the public, opinion leaders, and government officials in the U.S. and the world. The organization is formally incorporated under Section 501(c)(6) of the U.S. tax code as a nonprofit trade association that is qualified to take collective action on behalf of the biometrics industry in compliance with antitrust, tax and lobbying laws. IBIA is staffed by professional experts in biometrics, trade association management, and the complex disciplines of education and advocacy required by a multinational high-tech industry.

The IAI Mission

The International Association for Identification (IAI) strives to be the primary professional association for those engaged in forensic identification, investigation, and scientific examination of physical evidence. To accomplish this mission, the IAI has six goals:

1. Educate members about the most current information and research in forensic identification. Through the annual International Educational Conference and regularly offered workshops, members are kept abreast of innovations in all forensic disciplines.
2. Affiliate people who are actively engaged in the profession of forensic identification, investigation, and scientific examination of physical evidence in an organized body. In this way, the profession may be standardized, as well as effectively and scientifically practiced.
3. Enlarge and improve the science of forensic identification and crime detection.
4. Encourage research in scientific crime detection.
5. Keep IAI members apprised of the latest techniques and discoveries in forensic identification and crime detection.
6. Employ the collective wisdom of the profession to advance the scientific techniques of forensic identification and crime detection.

Since 1915, the IAI has served a distinguished and diverse membership from around the world. With more than 4,500 members from 50 countries, the IAI remains the oldest and largest forensic identification association in the world. A 14-member board of directors manages and oversees the operations of the IAI. The board's goal is to maintain the highest standards for the professional practice of forensic identification. Through the IAI's Standards and Code of Ethics, which are set forth in the Constitution and By-laws, the Board strives to provide a strong leadership role in educating and serving forensic identification professionals.

Association for Biometrics

The AfB is a non-profit making organization, which aims to promote the awareness, and development of biometrics related technologies. It provides an international forum for research and development, system design and integration, application development, market development and other issues.

The Biometrics Consortium

The Biometrics Consortium serves as the US Government's focal point for research, development, test, evaluation, and application of biometrics-based personal identification/verification technology. It has five working groups:

1. Testing and Reporting Group: Major John Colombi, PhD Responsible for establishing testing standards, developing performance testing protocols, designing a test facility, deciding upon the format for the reported results, providing a mechanism for the dissemination of final reports, and defining a repository for reported information.

2. Vulnerability Group (no vendors): Anonymous Same responsibilities as the Testing and Reporting Group, but as viewed from the standpoint of internal or external vulnerabilities of biometrics devices.

3. Database Group: Dr. Jim Wayman Responsible for defining standards for each particular type of database, collecting databases into one central location, and disseminating database information to those that require it for testing purposes.

4. Ground Rules Committee: Winnie Lehman and Tim Bergendahl Responsible for disseminating information about the Consortium, promoting external relations

and contacts, encouraging internal interaction, defining Consortium operating procedures, and addressing any legal or ethical issues that affect the Consortium.

5. Research and New Technologies Group: Dr. Alan Higgins Responsible for keeping abreast of the latest research and innovations in the field of biometrics, as well as providing a repository for such information.

Three Levels of Biometrics API Standards

Biometrics API standards may be classified into three levels: hardware device level, data or feature level and development toolkit level.

Hardware Device Level

Using biometrics technique to identify human beings, the first step is to capture one's unique physical feature or behavioral trait. Usually we use a device connected to a computer to get one's feature. For example, we use fingerprint capture device to photograph a live fingerprint and then send the picture directly to the computer. Or we use a CCD camera to take one's face picture and send it to the computer. Or we use a PC microphone to record one's speech and sent the voice recorder to the computer. And other devices are used to capture other features. Now worldwide there many vendors are manufactured such kind of devices. And these devices vary from shape to functions. To make an industry standard for these devices is to set up a series of rules for the device production, so that devices produced by different companies can be compared, interchanged and integrated with each other. We know that TVs, or cameras, or recorders all have industry standards. So no matter whichever factory we buy a TV from, we can plug it into the same channel hole and will get the same TV program service. It is the same situation to cameras and recorders and all other products with industry standards. We can buy a camera from one vendor and buy films from any film providers because the film size and other features are predefined and followed by all film providers.

The standards for biometrics capture devices include the interface between the device and the computer, time for capturing one sample, the size and the resolution of the sample, the interface provided to the programmer for the device control and further development.

Data or Feature Level

After captured the original sample, we need to process it and extract unique feature from it and then store it. Usually only the extracted features are stored in the database and the original sample data are discarded because the sample data is much larger than the feature data. Different systems use different methods to extract features and different ways to describe features. Suppose we have set up a fingerprint database using the software of one company and then we find the identification package from another company is better. Then we may wish that the identification package might process the feature data in the existing database. Here we need a standard. A feature format standard can be defined and all the feature extraction software and feature comparison software should follow the standard. By applying the standard, a bridge may be constructed among all the software providers.

The standards for unique features include the feature type (fingerprint, palmprint, voice, face, signature, etc), feature size, feature format, and other detail information.

Different types may have different detail information. For example, fingerprint may have minutia information while signature may have strokes and overall shapes.

Development Toolkit Level

Sample capture device and feature extraction algorithms are fundamentals of all biometrics applications. But each application has its own functions and user interfaces. Some applications use biometrics technique to detect imposters who are not registered in the system (access control) and some others use this technique to prevent people from reuse some facilities illegally (voting system). So with biometrics technique in hand, we still need to develop software to fulfil the various application requirements. To application developers, a standard development toolkit is needed.

The standards for development toolkits should include data capture, feature extraction and feature comparison functions and should be in a style of high level programming language API.

EXISTING BIOMETRICS STANDARDS

Certain biometrics standards are already in place. In this unit, we will introduce some existing standards.

AFIS Standards

The standard for Automated Fingerprint Identification Systems takes a lead to the biometrics standard society. Generally, there are two types of AFIS standards: for interchange of finger image data and for finger image compression.

ANSI/NIST-CSL 1-1993

Developed by a number of biometrics organizations and industry experts this standard is essentially utilized by organizations transmitting and interchanging AFIS finger images. ANSI/NIST-CSL 1-1993 was coordinated by the National Institute of Standards and Technology (NIST) for the regulatory body, the American National Standards Institute (ANSI). ANSI is a member of ISO the International Organization for Standardization and is an influential standard collective.

ANSI/NIST-CSL 1-1993 was approved by ANSI in November 1993 and is scheduled for a full review by November 1998. In November a working group will decide if any updates or revisions are necessary. The standard is currently in worldwide use for sampling finger images, determining the resolution and transmission of finger images and specifying text information that may be used in association with images.

FBI WSQ Compression Specification

The second AFIS standard is concerned with the compression of grayscale finger image data. Compressing data effectively is necessary to speed up the transmission of finger images. WSQ (Wavelet Transform/Scalar Quantization) is a compression algorithm that is used to convert a finger image into a compressed image and then reconstruct the compressed image. The algorithm can allow for high compression at the ratio of 30:1. The compression rate depends on the nature of the finger image data and obviously larger images (such as thumbprints) will be compressed less than those images taken from the little finger.

Application Programming Interface Standards

An API (Application Programming Interface) is a bridge between application developers and development toolkits providers. Usually API is a series of function specifications. In the specifications, details of each function's parameters, result, exceptions are defined. According to the API, the tool providers develop the function libraries and the application developers call the functions. The tool providers don't need to know how the application developers will use the tool and the application developers don't need to know the inside implement of the tools. With an API specification, application developers can become much more flexible when selecting a tool and it also becomes possible for them to integrated products from one than one provider.

Biometrics API stands for providing biometrics application developers a specification of various biometrics related functions. These functions include how to capture human features with a biometrics device, how to preprocessing the captured data, how to extract unique features from the preprocessed data, how to compare two features and how to adjust parameters to make the system more or less strict.

In this unit, we will introduce four API standards that are now being constructed.

SVAPI: stands for Speaker Verification Application Programming Interface.

HA-API: A high-level interface initially for user authentication within client/server environments.

AIS API: A 'middle ground' API to separate applications from devices, algorithms and databases.

BAPI: A hardware-software application interface focused on biometrics device drivers.

All the followings are extracted and modified from User's Guide.

SVAPI

SVAPI stands for Speaker Verification Application Programming Interface. SVAPI is the first non-proprietary biometrics API to be developed. SVAPI, a sub committee of the Speech Recognition API Committee (SRAPI), undertakes API development. Novell plays a leading role in SVAPI and chairs the SVAPI sub-committee. Other members come from a range of vendor, consultant and government backgrounds. Citicorp, Dialogic, E. Boyle Consultants, Higgins and Associates International, Hughes, IBM, ITT Industries, J. Markowitz Consultants, Texas Instruments, Motorola, SRI/Nuance, T-Netix, Veritel, VeriVoice and US government organizations such as the Department Of Defense, the Internal Revenue Services and the Immigration and Naturalization Service are all on this train.

SVAPI is focused on standard APIs for speaker verification that are not vendor specific. It aims at providing application developers more choices when selecting and installing technologies from speaker verification vendors. This makes the speaker verification development tools interchangeable and the application developers could be able to switch systems or add further systems to the application. SVAPI opens up the possibilities for application development and in time will surely strengthen the position for speaker verification within the wider biometrics community. It is designed to integrate with a range of data and telecommunication networks. Those vendors that become SVAPI compliant (and many are now taking steps to do so) are likely to gain competitive advantage, particularly in the Internet and telephony markets.

SVAPI is now freely available as a Software Development Kit (SDK) version 1.0 in either C++ or Java versions. The Java version of the API supports the Java Telephony API, allowing it to be used in Java telephony environments.

The SVAPI SDK uses object-orientated approach. To assist application developers, SVAPI gives step-by-step instructions on how to design a simple speaker verification application. A series of steps helps an application developer build a framework within which the speaker verification application and the speaker verification engine can communicate. This basic SVAPI application can be written to incorporate enrollment, prompting the user for an utterance and performing verification.

Human Authentication Application Program Interface (HA-API)

The Biometrics Consortium announced the Human Authentication Application Program Interface in 1997. The initiative of HA-API is to create a generic biometrics API and a proof of concept implementation.

HA-API is designed for use by both application and biometrics technology developers. In an application incorporating biometrics - to make integration of the technology straight forward - HA-API hides or encapsulates the complexities of biometrics technologies as much as possible. The user interface for biometrics capture and enrollment (if needed) is provided via the biometrics system, while all other user interface functions are provided by the application developer. This degree of freedom was allowed so that a large set of potential biometrics technologies and applications could be addressed. To make this possible, HA-API provides a 'toolbox' of biometrics functions that can be used within a number of potential applications. Access to this toolbox is through a set of standard interfaces. Theoretically, biometrics components and systems that conform to this interface could then be used within any application also developed using HA-API.

The underlying philosophy of HA-API interface design is to define a minimum set of generic API functions. These functions essentially relate to the biometrics application or the biometrics system. They reside in two levels of interface: the application program interface (API) and the service provider interface (SPI). The API functions are concerned with functions relating to the software application in question. To allow these functions to interact within a HA-API application, software known as the HA-API runtime layer provides translation.

A current version of the HA-API runtime layer is available free of charge from the DOD and NRI; though the layer could also be constructed by an application developer or a biometrics vendor.

IBM Advanced Identification Services C API (AIS C API)

Like the Biometrics Consortium, IBM is keen to see the deployment of APIs. IBM staff in the UK developed the AIS API, based on a number of biometrics applications using IBM solutions. The API was first announced in November 1997. Version 1.01 of AIS supports the C programming language and essentially provides programming calls, which support the capture, storage, searching and retrieval of biometrics data. The aim is to allow applications to be written independently of a biometrics device, algorithm and database. To structure each biometrics application, AIS is divided into three main services that work at different stages of the application itself.

The first, Ais_Dev, works at the biometrics capture stage between the 'analog biometrics data' (the biometrics characteristic) and the 'digitized representation' of the biometrics characteristic. Ais_Dev allows multiple biometrics to be incorporated into an application.

The second service, Ais_Img, works at the biometrics extraction stage between the digitized representation and the biometrics template. Ais_Img allows quality checking and extraction algorithms to operate.

The third service, Ais_id, works at the biometrics comparison and matching stage between the biometrics template and the biometrics database used within the application. Ais_id is for matching, storing and retrieving biometrics templates.

The Biometrics API (BAPI)
BAPI enables software applications to work with a biometrics identification device from any hardware vendor who provides BAPI-compliant drivers for their products.

In the initial release, BAPI does not provide data compatibility between different biometrics devices. In other words, BAPI does not guarantee that a fingerprint acquired on one BAPI compliant scanner will work with hardware or algorithms from another vendor, although this may be possible to some degree. Such a goal may be achievable in the future through the framework of filters built into BAPI.

BAPI is designed primarily with fingerprint scanners in mind, but is flexible enough to benefit other technologies as well, such as hand geometry identification, voiceprint, retinal scanners, facial recognition, etc.

BAPI will also simplify the work of software developers. Since they do not have to learn new API's for each type of device they support, they will be able to develop more applications in less time. Customers will benefit from the availability of more and better applications.

13.5 Information Resources

Now, many web sites or pages have provided a lot of biometrics information about personal authentication applications. If you are interested in the related topics, some information resources are listed in the following:

1. ICSA Biometrics Buyer's Guide
   http://www.icsa.net/services/consortia/cbdc/bg/
2. Biometrics API
   https://www.iosoftware.com/bapi/
3. HA-API v.1.03 Interface Specification.
   Human Authentication - Application Program Interface: (HA-API)
   http://www.biometrics.org/html/standards.html
4. American Association of Motor Vehicle Administrators (AAMVA)
   http://www.aamva.org/

5. American Statistical Association
http://www.amstat.org/
6. Association for Biometrics (AfB), UK
http://www.afb.org.uk/
7. Australian Biotechnology Association
http://www.aba.asn.au/
8. Automatic Identification Technology Commerce and Education - About Biometrics ID
http://www.aitworld.com/
9. BioAPI Consortium -- Industry group working to define an API for biometrics
http://www.aitworld.com/techvalley/biometrics.html
10. Biometrics in Human Services User Group
http://www.bioapi.org
11. Biometrics Testing Services (BIOTEST) - a European project aimed at developing standard metrics for measuring/comparing performance of biometrics devices, and establishing testing services
http://www.dss.state.ct.us/digital/faq/dihsug.htm
12. Commercial Biometrics Developer's Consortium (CBDC)
http://www.npl.co.uk/npl/sections/this/biotest/
13. Committee on Computing, Information, and Communications R&D... Technology Policy
http://www.icsa.net/services/consortia/cbdc/
14. Subcommittee
http://www.hpcc.gov/ccic/
15. Financial Services Technology Consortium (biometrics fraud prevention)
http://www.fstc.org/
16. International Association for Identification (IAI)
http://www.iaibbs.org/
17. International Biometrics Industry Association (IBIA)
http://www.ibia.org
18. National Center for Identification Technology
http://www.ncit.org/
19. NATO Advanced Study Institute (ASI) on Face Recognition: From Theory to Applications
http://chagall.gmu.edu/faces97/natoasi/
20. Security Industry Association (SIA)
http://www.siaonline.org/
21. The Biometrics Consortium
http://www.biometrics.org
22. The Human Identification Project
http://www.asti.dost.gov.ph/
23. UK Police Information Technology Organization (PITO)
http://www.pito.org.uk
24. 1997 Automated Fingerprint Identification System (AFIS) Committee
http://www.iaibbs.org/afis.htm
25. 1998 Glossary of Biometrics Terms (AfB & ICSA)
http://www.afb.org.uk/glossuk1.html

26. American National Standards Institute
http://www.ansi.org:80/
27. International Standards Organization
http://www.iso.ch/
28. International Biometrics Group, Inc.
http://www.biometricgroup.com
29. DARPA's Internet for Security Professionals
http://isp.hpc.org/
30. EDI HotLinks (standards, etc.)
http://www.wpc-edi.com/resource.html
31. EAGLES' Assessment of Speaker Verification Systems
http://coral.lili.uni-bielefeld.de/~gibbon/EAGLES/slwghand-t/node54.html
32. East Shore Technologies (check the EST Challenge)
http://www.east-shore.com/
33. Fingerprint Technologies
http://www.fingerprint.com/
34. FingerPrint USA
http://www.fpusa.com/
35. GSA's Federal Security Infrastructure Program (for secure applications: tokens, keys, and authorization)
http://www.gsa.gov/fsi/
36. GSA's SmartGov
http://policyworks.gov/smartgov/
37. I/O Software, Inc. - a distributor, consultant, SI and custom developer for the Sony fingerprint identification unit. The first self-contained unit that compares and enrolls in a device the size of a mouse.
http://www.iosoftware.com
38. International Biometrics Group (IBG)
http://www.biometricgroup.com/
39. Justice Technology Information Network (JUSTNET)
http://www.nlectc.org/
40. Julian Ashbourn's Technology Corner
http://members.aol.com/teknotalk/home.htm
41. National Information Assurance Partnership (NIAP)
http://niap.nist.gov/
42. NIST's Computer Security Resource Clearinghouse
http://csrc.ncsl.nist.gov/
43. Physical Security Equipment Action Group
http://www.vitro.bloomington.in.us:8080/pseag/
44. Q&A Consulting
http://www.communinet.org/QA/
45. The Biometrics Consulting Group, LLC
http://biometric-consulting.com
46. 20G4 Multi-technology Automated Reader Card (MARC) Project
http://www.vitro.bloomington.in.us:8080/marc/
47. Biometrics Reports
http://www.biometrics.org/REPORTS/CTSTG96/)

14 BIOMETRICS CHINESE MEDICINE

In Chapter 1, we have defined the taxonomy of biometrics by Application Type, which covers four application categories: *Personal Identification, Medical Diagnosis, Future Expectation and Ethnology exploration.* As a first attempt to medical diagnosis, in this chapter, we apply biometrics technology to Chinese medicine. Traditional Chinese Medicine (TCM) is introduces in Section 14.1. Then, four kinds of Chinese diagnosis methods are discussed. Section 14.3 provides the various observation types in TCM. Based on our previous works, a tongue diagnosis system is finally presented.

14.1 Traditional Chinese Medicine (TCM)

The methods that are applied to observing, inspecting and investigating on patient, and collecting the data relating to the health changes of a patient are called diagnostic methods. The contents of diagnostics of Traditional Chinese Medicine (TCM) include examining patient, collecting data relating to health, sorting out, analyzing, synthesizing and reasoning the data on basis of basic theories of TCM, finally, determining the characteristics of clinical manifestation and the change laws of illness condition after disease occurring [14-41]. By doing so, we can identify the diseases and syndromes, as well as provide basis for the treatment and prevention of diseases [1].

During the long period of medical practice, some doctors of past ages accumulated rich experience of diagnosis which formed the comprehensive diagnostic system special for TCM, i.e., four diagnostic methods (observation, auscultation and olfaction, interrogation and palpation) and syndrome differentiation. And based on those, they established diagnoses of TCM. From ancient time to now, the diagnostics of TCM has been playing its constant role in clinical practice. Recently, with the development of biometrics technology, modern scientific methods have been applied to conducting the four diagnostic methods and syndrome differentiation in combination with traditional methods. It is developing in the way of keeping the distinctive features and combining with modern science.

According to the theories of TCM, the human body is an organic whole. The local disorder of the outside of body can influence the inside, and the inside disease can also show on body surface. This means that the external manifestations can show the nature of internal disease. Therefore, when making a diagnosis by TCM, it is to reason the internal pathological changes of a patient mainly by means of the patient's self-sensation and external manifestations which are know by physicians through their sensory organs. *Grand Discussion on the Correspondence of Yin and Yang*, one chapter of *Plain Questions* [2], said: "To know the interior according to the exterior would ensure the correct diagnosis." That means the external changes can reflect the internal disease. It is pointed out more clearly in *Speculation from Exterior* (one chapter of *Spirit Pivot* [3]): "The obvious manifestation cannot be hidden because it does not leave yin and yang. The correct diagnosis can be made by means of observing and pulse feeling comprehensively. If one could not distinguish the different five sounds and five colors, it would not be possible to cognize the disorder of the five zang-viscera. So the interior corresponding to the exterior is just like the shade always accompanying the figure. Physician can speculate the interior according to the external changes, but also can speculate the exterior according to internal changes." Such theory of diagnostic method about "knowing the interior by the exterior" still plays its role in clinical practice nowadays.

The diagnostic method of "knowing the interior by the exterior" includes four diagnostic methods which are observation, auscultation and olfaction, interrogation and palpation. These four diagnostic methods have their unique clinical functions and can not be replaced by one another. They must be applied simultaneously to diagnosis to ensure correct diagnosis.

The history of TCM began thousands of years ago. In Chinese character inscription in bones or tortoise shells unearthed from Yin Dynasty ruins, there are many oracle inscriptions recording many diseases knowledge including sixteen kinds of diseases involving head, eye, ear, mouth, tongue, throat, nose, abdomen, foot, toe, urine, deliver, etc. They belong to diseases of internal medicine, surgery, gynecology, obstetrics, pediatrics, infectious diseases of modern medicine. It had exquisite classification of diseases and could name diseases according to the different parts where diseases occur.

Internal Classic is an important classic book in TCM. It summarized the medical theories and medical experiences of Chinese people from the Spring and Autumn Period, the Warring States Period to Qin and Han Dynasties. In the aspect of diagnostic method, there are a lot of records about the four diagnostic methods of observation, listening and smelling, questioning and palpation in the book. *Classic on Medical Problems*, written in Three States Period, specially stressed the pulse-reading in the diagnostic methods. From Tang Dynasty to Jing and Yuan Dynasty, syndrome differentiation had been developed further. In the Ming and Qing Dynasties the development of diagnostics is mainly in the four aspects, i.e., tongue inspection, questioning, palpation and the syndrome differentiation.

Recently, the TCM diagnostics has been getting further development in China. In the researches on modernization of the four examining methods and principles of syndrome differentiation, the studies have been conducted comprehensively by means of acoustics, optics, magnetics, electricity, chemistry, physics and biomedical engineering. Certain achievement has been gotten. From now on, under the guidance of basic theories of TCM, absorbing the latest achievements in modern science, in combination with the clinical and experimental research, the TCM diagnostics will develop continuously and make its new contribution to health service of mankind.

In fact, we can use biometrics technology for TCM diagnosis. For example, we have been developing an automatic tongue diagnosis system. The picture of tongue can be gotten by camera and the features can be analyzed automatically for TCM diagnosis. The intelligent instrument for TCM diagnosis by means of pulse has been developed and entered into the market. These new technologies have great significance for further developing the TCM diagnosis. Therefor, biometrics technology will have a great application in TCM diagnosis.

14.2 Four Chinese Diagnostic Methods

TCM mainly relies on the doctor's sensory organs to diagnose a disease. The nature of disease is judged from the external abnormal signs that demonstrate the internal pathological changes and the patient's self-feeling of discomfort [4-6]. There are four kinds of Chinese diagnostic methods which are defined in this section.

OBSERVATION

This is one kind of examining method which is applied to knowing the condition of disease by means of doctor's visual sense to look over the vitality, color, figure, posture of patient's whole or partial body and the changes of the figuration, color, texture, and quality of the patient's discharges. Observation is an important diagnostic method in TCM and is taken seriously by doctors in successive dynasties.

It is realized through the long-term process of medical practice that the pathological changes of internal organs can be understood by means of inspecting the outside of the human body. It is said in the 47th chapter of the Spirit Pivot: "Observing the outside manifestations in order to understand the inside organs, so as to identify the diseases." It is also said in the Experience of Danxi: "In order to know the inside, the outside should be observed. Only observing the outside, it is then possible to understand the inside. Generally speaking, if there are changes internally, they are inevitably manifested externally." Among the various contents of observation, the face and tongue have a closer relationship with viscera and meridians. Because they are more sensitive and accurate to reflect the pathological changes of internal organs by means of facial complexion, tongue texture and tongue coating, which have high practical values. It is then formed the two specific observing methods in TCM, namely, complexion inspection and tongue inspection. From the 1980s, the field of observation has been expanded. The new observing methods such as observing the mucous membrane of palate, observing of rugosity of ear lobe, come forth. Some scholars have been studying on modernization of observation such as the clinical verification and normalization. At the same time they also make the content of observation richer.

AUSCULTATION AND OLFACTION

Auscultation is to find the abnormal sound of speech, respiration and cough etc. by means of audition (hearing). Olfaction is to know the smell of the patient's body, the secretion and excreta by means of osphresis (smelling). It is a method that doctors use their auditory and olfactory organs to diagnose disease.

In ancient time, it followed mainly the laws of "five voices and five scales" to distinguish the five zang-visceral diseases. In Han Dynasty, Zhang Zhongjing took the speech, respiration, asthma, and cough, vomiting, hiccups and moans as the main content of the auscultation and olfaction. The doctors of later generations added the odor of mouth, nose, secretion and excreta to the contents.

The production of sound is related to the lung, throat, epiglottis, tongue, teeth and nose. The lung plays a main role. The lung dominates the *qi* of whole body and the movement of *qi* leads to sound production. The abnormal change of *qi* results in changes of sound. The auscultation is to inspect not only the above mentioned organs, but also other viscera. According to some theories from Internal Classic, the abnormal change of sound can reflect conditions of five zang-viscera.

The voice of normal people is produce naturally and fluently in a harmony way. Though the structures of the body are the same, the individual difference exists. So there are differences of voice in the aspects of volume, tone, clear or vague. The voice changes in accordance with the differences of constitution, age, sex and emotion.

In condition of disease, if the voice is loud and successive, it belongs to heat or excess syndrome. The voice is often lower and vague in condition of wind, cold and dampness invasion. If the voice is low, weak and non-successive, it belongs to deficiency or cold syndrome, or condition of evil being eliminated and genuine *qi* being injured.

The state of diseases can be known by listening to the speech. The cold leads to taciturn and the heat to chatter. The deficiency makes speech voice low and inconsistent, and the excess makes the voice loud.

The indistinct speech is due to wind-phlegm that covers the clear orifice or obstructs meridians. The incoherent speech is a result from failure of heart to store the mind. There are differences between the deficiency and excess.

The normal respiration in disease means the disease in physique but not in *qi*; which the abnormal respiration means the disease in both physique and *qi*. The deep breath means excessive *qi* in the heart and lung, but the weak breath is due to *qi* deficiency of both the liver and kidney. The main purpose of listening to respiration is to distinguish the deficiency from excess.

Cough is closely related to the lung. But in Internal Classic, it is said: "The five zang-viscera and six fu-viscera can all make cough, not only the lung." From that we know not only lung disorder, but also other visceral diseases can cause cough. Listening to the sound of cough can help to judge the cause and nature of diseases.

The vomiting, hiccup, belching, sigh, sneeze and bowel rumbling are also heard in auscultation. Hearing is especially useful when dealing with children. Zhou Xuehai said: "When inspecting child, the listening is the first and the observation of color is the second." The clear sound means survival while the vague and faint voice means dying.

There are changes of odor when something is stale. In healthy body, the *qi* and blood circulate smoothly, the viscera and meridians function well, so there is no abnormal odor. When diseases occur, the *qi* and blood movement is abnormal, the foul and turbid waste could not be cleared away, then there are abnormal odors.

Olfaction is to smell the odors related to diseases, such as odors from excreta and secretion including sputum, nasal discharge, sweat, stool, urine, menstrual blood and vaginal discharge, etc. The foul odor is usually due to excessive heat while the little fishy odor is often due to deficient cold.

INTERROGATION

Interrogation is to know the onset, development, treatment, present symptoms and other information of disease by questioning patient or the accompanying people for diagnosis.

Interrogation is very important in diagnosis, because the self feeling and the disease history can only be know from the description given by patient. In Qing Dynasty, some doctors attached even more importance to it and listed it at the first place of all the four diagnosis methods. Li Rong pointed out that every symptom such as cough, fever, pain and so on had its distinctive character. When treating cough, if the fire character is known by asking, the cough should be treated according to fire; and if there is phlegm known by questioning, the cough should be treated from phlegm. By questioning we can know the disease and its cause. So the disease can be treated from the "root".

The purpose of interrogation is to collect the data for diagnosis that could not be obtained by other methods. The hobby, custom of the patient, as well as the onset and development of disease should be grasped, so as to judges the cause and nature of disease, to make diagnosis and to treat the disease.

The content of interrogation includes general data, such as name, age, sex, occupation, birth place, nationality, address and marriage state of the patient, and process, such as diet, mental state, living style, family history and anamnesis, chief complaint, history of present disease.

PALPATION

Pulse examination is to judge disease by means of using doctor's fingers to palpate patient's pulses, which are shown in the superficial arteries. It is a unique diagnostic method of TCM. All ancient doctors paid great attention to it.

The substance of pulse is the blood and the power of pulse is the *qi*. Because the heart dominates blood and vessels, the heart pumps *qi* and blood into all parts of the body through vessels. The *qi* and blood circulate incessantly. They enter zang and fu-viscera inward and reach limbs and skin outwards. Besides, the *qi* and blood circulation also depends on other viscera that coordinate the heart. The lung meets all vessels. The *qi* and blood circulating all over the body should converge into the lung. Because the lung dominates *qi*, the blood circulation depends on the dispersing of lung *qi*. The spleen and stomach are the source of *qi* and blood. The spleen also controls blood flow. The normal flow of blood needs the controlling function of the spleen. The liver stores blood and is in charge of conducting. It regulates the circulation volume of blood. The kidney stores essence. The essence can transform itself into blood. It is one of the basic materials for blood production. The essence can also change into *qi*. It is the root of *yang-qi* and source of all energy of body. And only through the vessels, all visceral *qi* can play its role fully in all parts of the body. Therefore, we can know the visceral state and the disease condition by means of pulse taking. To sum up the ancient documents, the significance of pulse examination is as follows:

- To recognize the exterior and interior of disease: Although there are many kinds of diseases, they can be classified into the exterior and interior syndromes as the location of disease. The level of pulse can reflect the depth of disease location. If pulse is floating, the disease is usually in the exterior. While pulse is deep, the disease is in the interior.
- To judge the deficiency and excess: The deficiency and excess refer to the balance and increasing or decreasing of the *genuine-qi* and the evil in the combating. Xu Lingtai said: "The main point of deficiency or excess can not be out of the pulse." The weak pulse such as thready pulse and feeble pulse shows the deficiency syndrome of the *genuine-qi* shortage, while the strong pulse like slippery pulse and surge pulse exhibits the excess syndrome of exorbitant evils.
- To ascertain nature of disease: The nature of disease can be divided into types of cold and heat. Their treatment principles are quite different. The rate of pulse beating can reflect the nature of disease. For instance, slow pulse usually belongs to cold and rapid pulse is mainly due to heat.
- To identify cause of disease: Doctors of TCM pay attention to seeking disease cause when making diagnosis. One of the main aims of syndrome differentiation is to find disease cause. By means of pulse taking, sometimes, we can find out the disease cause. For example, in cold-attack diseases, if the pulse is floating and tight, it is due to cold evil; and if the pulse is soft, the cause is wind. In endogenous diseases, if the pulse is slippery, it is often caused by phlegm; and if pulse is uneven, blood stasis is suggested.
- To inspect the disease mechanism: The development and change of disease bring some difficulties to treatment. The advantage of treating disease according to syndrome differentiation lies in the corresponding change of treatment method to the syndromes that are recognized in time. Pulse condition can reflect the change of disease to a certain degree. For example, if the pulse in left guan portion is taut and that in right guan is weak, it means the exorbitant liver over-restrains the spleen. When these pulse conditions are shown in a patient suffering from liver disease, we should inspect not only the liver, but also the spleen. Therefore, the treatment can be adjusted according to the changing condition of disease.
- To predict the prognosis: Pulse taking is important for the inference of prognosis. For instance, when disorders of blood and *qi* precede the appearance of symptoms, the pulse changes first. At that time, the latent disease can be predicted by pulse reading. So, treatment can be applied in time. From paradoxical pulses or intermittent pulse, we know the disease is severe and the prognosis is not good. The appearance of soft pulse in chronic disease means the restoring of stomach *qi* and recovering of disease. The yang pulse like surge and strong pulse in severe case or deficiency condition means the critical condition due to exorbitant evil and declined *genuine-qi*, and a bad prognosis.

Since ancient times, doctors have been paying attention to pulse-taking and rich experiences have been accumulated. Nevertheless, there is limitation in pulse taking. It is should be applied in combination with other examining methods. Then correct diagnosis can be made. Generally speaking, pulse condition and syndrome are identical. For example, strong pulse presents in excess syndrome and weak pulse is found in deficiency syndrome. But there are also pulses that are not corresponding to the symptoms. At that time, we have to make diagnosis by attending to one aspect of them and neglecting the other aspect. In a word, only by combining pulse taking with other diagnostic methods, we can improve the accuracy of diagnosis.

14.3 Various Observation Types in TCM

Observation is the most important method in TCM diagnosis. All the vitality, color, figure and posture in health human body have their normal manifestation. Any change different from the normal manifestation means pathosis. It is possible to understand the disease condition by means of looking over the whole body of patient. From ancient, doctors always observe the head, face, tongue, hair, eye, ear, nose, lip, tooth and gum, throat, neck, chest and abdomen, lumbus and back, genital and anus, and skin of patient. In modern times, some scholars have been studying the observation method by advanced instrument. Obviously, these researches are sound ground of our Biometrics Chinese Medicine. The following are some modern research in observation of TCM [9-12].

FACIAL COLOR OBSERVATION

The facial color and luster can reflect the visceral functions and the state of *qi* and blood. After Japanese scholars put forward the objectified method for testing skin color and luster and applied the infrared photograph to examining the facial color, many scholars have applied many kinds of instruments to detecting the facial color. Now, they are mainly focused on the infrared picture and chromatograph.

Some achievements on quantitative and qualitative research have been obtained. Some investigators had made the researches on the facial brightness and color measurement on the normal people who have shown the color deviation in white, yellow, blue and black. They found out that there were deviation tendencies of five colors in normal facial complexion. The researches proved that the instruments and technology of color analysis could provide an objective and quantitative basis for diagnosis. Some scholars compared the infrared photographs of normal people and heart disease sufferer. They found out differences between these two groups in the homogeneity and symmetry. Thereby they proved some objective reactions on face in coronary heart disease. By comparing the infrared photographs before and after treatment, they found the treatment effective to disease could also turn the facial reactions to normal. Chromatograph was applied to measuring the color changes of 131 patients suffering from pulmonary tuberculosis. The differences between different groups on syndrome differentiation were found. Among them, the chromatic aberration in yang deficiency was the most prominent; that of yin deficiency and blood deficiency was inferior to that of yang deficiency, and that of *qi* deficiency and normal condition was small. Some doctors proposed that the observation 01 crab-like stripe on face was one method to check up the meridian in TCM. By analyzing data of 105 patients, it was found out that in heart disease the stripes were mainly in temporal region, in liver and kidney diseases they were mainly in nose and cheek regions, in lung disease they were mainly in zygomatic region, in kidney disease they were mainly in cheek region. The distribution of crab-like stripes was related to the diseases. Investigators did observe the vessels (subcutaneous capillary) in root region of nose of 1000 children. They found that the vessels in normal children were obscurely bluish, and that in diseases were obviously dark blues. When the vessels were in the shape of "-", it was due to diseases of digestive system; while they were in " | " shape, it was mainly due to diseases of respiratory system. The blue color of the vessels indicated the wind, cold, painful syndrome and liver diseases. The yellow indicated the dampness, heat, deficiency, spleen and stomach diseases. The bright color suggested the recently occurred diseases that were mild and easy to cure, while the dark color was due to old diseases which were severe and difficult to cure. The shape and color of the vessels were related to the cause, location and other conditions of diseases.

Some foreign doctors proposed a concept of facial examination based on the theory that all the parts of the body had their projection on face. They thought the meridian and point system on face was different to the traditional theories of TCM, and it was a projective system of meridians. The facial examination was to observe the abnormal spot on face to diagnose diseases. The visceral representation on face they gave was different to that recorded in Internal Classic.

EYE OBSERVATION

Eye is the window of the liver and has a close relationship to viscera. Eye observation can help to diagnose many diseases. In recent years, its applying range was expanded to diagnose parasitosis, tumor, trauma and hemorrhoid, etc. For example, it was reported that the stasis spots in the upper half of bulbar conjunctiva indicated the trauma of chest and arm, those in the lower half reflected the trauma of back and leg, that in the left eye suggested the trauma in the right of body and vice versa. The accuracy rate of all diagnosis was 91.2%. It was suggested that the dilated, curing, congested vessels in the region of 5-6 o'clock of bulbar conjunctiva were the sign of hemorrhoid. The 85% of diagnosis in 1270 cases were correct. Some doctors applied ophthalmoscope to inspect the vessel changes of fundus. The vessels being bright red, even in width, regular in distribution were the embodiment of regular liver *qi*. Those in dark red, uneven in width, or congestive were due to liver *qi* stagnation. Those being thickened, dilated, curing, or with bleeding were due to stagnated heat in liver. Atrophy of the vessels even hard to recognize with pale-yellow dry fundus was due to liver blood deficiency.

EAR OBSERVATION

Scholars at home and abroad all thought that the shape, color and luster of ear were related to the visceral functions. Some foreign scholars thought the ear was the mirror of spirit. The ear shape in some psychosomatic diseases was different from the normal, thus we could diagnose psychosomatic diseases by observing ears. Scholars at home have made a lot of researches in the field. They observed the size, thickness, shape, color and vessels of ears, or pressed the points on ears to find tender points. Some diseases could be recognized by these examinations, such as cholecystitis, cholelithiasis, silicosis, coronary heart diseases, myocardia infarction, stomach diseases, hyperosteogeny, cancer and tumors. For cancers, it was suggested that on the corresponding points of ears there were pain, hot, protuberance, color change and changes of electric conductivity. In northern Jiangsu, more than 50, 000 people were examined in mass survey of cancer, the detective rate in upper digestive tract cancer was 10%, it was higher than the average rate of 2-4% at home and abroad.

The ear examination is used popularly in China. By using WR-F ear-point comprehensive detector, it was found that the potential voltage in disease group was higher and the resistance was lower than those of control group. Some doctors applied XZ-20 information detector to examining ears of people aged more than 60 years. The changes of electric current in ear-point were related to the endocrine function. Some people took the average energy consumption as the index instead of the resistance, thus the sensitivity was improved. They did manage to minimize the disturbing signals and the resistance between the electrode and the skin. Therefore, the accuracy of diagnosis was improved.

The oblique rugosity on earlobe was related to the coronary heart diseases. That was reported in 1973. But the research is still in progress. According to the statistics, the form of the rugosity was related to the arteriosclerosis but not to the fat metabolism.

NOSE EXAMINATION

The theory that the nose column corresponds to the liver, the two sides of nose to the gallbladder, and the apex of nose to the spleen, the two sides of the apex to the stomach was proved by clinical verification. Some body thought that nose observation could diagnose diseases of the liver, gallbladder, spleen and stomach. The crab-like stripe on nose suggested liver diseases. The spot on the middle point below apex suggested the diseases of reproductive system.

The philtrum was observed in addition to the nose. The obscurely blue in philtrum suggested the cold dysmenorrhea. The obscurely purple indicated the dysmenorrhea due to stagnated heat. The pale and dry philtrum suggested the amenorrhea due to blood exhaustion. The red in philtrum near lip indicated the metrorrhagia or metrostaxis due to heat in blood. The bright white in philtrum near nose suggested the metrorrhagia and metrostaxis due to *qi* deficiency. If the pale philtrum or shortened philtrum shorter than one third cun was seen in male, it suggested impotence, lack of sperm or no sperm.

MOUTH AND LIP OBSERVATION

Observing lips can help to diagnose parasitosis or tumors. Researches have been made by many doctors. It was reported that the nodes in upper labial frenum suggested the hemorrhoid, and the number and size of the nodes were in direct proportion with the number and size of hemorrhoid. The red and soft nodes suggested the beginning of hemorrhoid, while the white and hard nodes the old hemorrhoid.

In 1981, a Japanese scholar pointed out the soft and hard palate mucous membrane would have changes corresponding to diseases. These changes were termed sign of palate mucous membrane. It was applied to diagnosing blood stasis syndrome. Hereafter, Chinese scholars had also made researches in this field. The small varicose vein, dilatation of small artery, bleeding and color change has been studied. The blood stasis in liver cancer, hepatocirrhosis, coronary heart diseases, diabetes mellitus and irregular menstruation was more obviously shown in the sign of palate mucous membrane. Its value was equal to that of bluish purple tongue and vessels below tongue. Experimental studies showed that the sign of palate mucous membrane was related to micro-circulatory disturbance.

HAND EXAMINATION

The striae of skin were an important inherited property of mankind. It was controlled by the gene. The TCM attached great importance to the constitution when differentiating syndrome. The striae of skin could be used judge to constitution and heritage person. For example, in spleen deficiency, the coincidence rate of whorl and double loops was lower than that in normal condition, the general number of crest was high, most of the whorl and loop are big, flower-like wrinkle between index finger and middle finger and in the thenar and hypothenar was more often seen. These abnormal changes could be taken as index for diagnosing spleen deficiency constitution.

14.4 Tongue Diagnosis

Tongue observation is also named tongue inspection. It is a method of observing the changes of tongue body and tongue coating to analyze disease. It is a main component of observation. There are many records of tongue observation and its theories in *Internal Classic*. After that, many medical works like *Treatise on Cold-Attack*, *Classic of Viscera*, *A Thousand Gold Worthy Prescriptions*, *An Official's Secret Prescriptions* recorded this method. After the *Records of Golden Mirror*, the monograph of tongue examination appeared in Yuan Dynasty, many monographs of tongue examination were published one by one. Based on many doctors experiences accumulated in the long time medical practice, the unique examining method forms the system [6-8,13].

WHY TONGUE INSPECTION?

The changes of tongue reflect inner visceral changes. The tongue is the orifice of heart. The stretching and retracting are the action of tendons, which reflects the function of liver. The red small particles on tongue tip are projections made up of heart *qi* and genuine-fire of life-gate. The white soft prickle like hairs on tongue surface is produced by lung *qi* with genuine-fire. The tongue fur (coating) is made up of steaming stomach *qi*. So by observing tongue, we can know the visceral conditions.

The tongue body and coating have their unique significance in diagnosis. Tongue body exhibits conditions of five zang-viscera, while tongue coating shows that of six fu-viscera. By observing tongue body, we can ascertain the deficiency or excess of *genuine-qi* and the severity of disease. By inspecting tongue coating, we can judge the cold or heat of evils and the location of disease. Tongue body and coating can reflect disease in different aspects.

Among the four examination methods, tongue observation is thought to be more reliable than others stated by some doctors. There are three reasons. The first, the pulse is covered by skin but tongue is not covered. Tongue is watched by eyes directly. So it is clear. Pulse is felt by fingers. It is difficult to understand. The second, tongue connects viscera internally and meridians externally. Both normal and morbid condition can be shown on tongue. The last, when evil enters the inner body, every changes of it will embody in tongue. The tongue is not like the pulse being obscure. The inner conditions will be exhibited clearly by the change of dry or moister and the thick or thin of coating. So the changes of tongue can reflect the conditions of genuine *qi* and evils, and the tendency of disease. To sum up, the significance is as follows:

(1) To judge the exuberance or decline of the genuine *qi*: The exuberance or decline of visceral *qi* and blood can be shown in tongue. For example, the red and moist tongue means the exuberance of *qi* and blood; while the pale tongue is a sign of deficiency of both *qi* and blood. The white, thin and moist coating indicates an exuberant stomach *qi*; while no coating is due to the decline of stomach *qi*, or impairment of stomach yin.

(2) To distinguish the nature of disease: Evils of different natures will make different changes in tongue. For example, if there is no prickle on tongue surface and the coating is white and moist, or the tongue is bluish black without prickle, it is due to cold evil. If there is a red and dry tongue with yellow coating, or red prickle tongue with yellow, thick and greasy coating, it is due to warm or heat evil. The greasy or putrid coating indicates the food-retention. The blue macula or spot on tongue suggests blood stasis.

(3) To detect the location of disease: In exogenous diseases, the thick or thin of the coating can reflect the deep or shallow of disease location. For example, the thin coating suggests that the disease is in its initial stage, the disease is located in shallow part and it is an exterior syndrome; while the thick coating suggests that the evils enter into the inner part of the body, the disease is located in deep part and it is an interior syndrome. The crimson tongue means heat evil is in nutrient or blood-phase, disease is very deep, it is a critical condition.

(4) To infer the tendency of disease: The changes of tongue usually follow the changes of genuine *qi* and evils, and disease location. We can infer the tendency of disease by observing tongue, especially in exogenous febrile diseases. For example, the turning of coating from white to yellow, and from yellow to black is usually due to the transferring of evils from exterior to interior, or from cold to heat. It shows the deterioration of disease. If a moist coating turn into dry, it is usually due to loss of body fluid resulting from heat. The change from dry to moister implies the recover of body fluid. The change of coating from thick to thin is a sign of improving or curing.

However, it should be pointed out that sometimes the tongue is only slightly changed in some severe cases, and the abnormal changes of tongue are seen in normal people. So, the tongue observation should be used in combination with other examinations. Only by comprehensive analysis on whole clinical data, can we make a correct diagnosis.

RELATIONSHIP BETWEEN TONGUE AND VISCERA

Tongue is the sprout of the heart and the heart is the supreme monarch of all organs. So disease of viscera can influence not only the heart, but also the tongue.

Tongue is called the out-show of the spleen. The spleen dominates transportation and transformation. So, the tongue is closely related to splenic function. Tongue coating has special relation with stomach *qi*. Zhang Xugu said: "In healthy body, there is a little thin coating like grass roots. It is the embodiment of stomach *qi* activity."

Meridians, for example, three yang meridians of foot, three yin meridians of foot, Taiyang Meridians of Hand, Shaoyang Meridians of Hand, have connection to the tongue. Shen Douyuan said: "All *qi* of meridians flow up to tongue. So we can know the deficiency or excess, cold or heat of viscera and meridians by observing tongue."

Because tongue is closely related to viscera and meridians, viscera have their representative areas on tongue surface. The ancient statements on this were different. The most popular one was from the Bihua's Medical Mirror: "The tip of tongue belongs to the heart, the middle to the spleen and stomach, the bilateral margins to the liver and gallbladder, the root to the kidney. " It can also be explained as the tip portion to upper-jiao, the middle to the middle-jiao and the root to the lower-jiao. The correspondence is the same one of that of cunkou pulse, namely "the upper shows upper body, the middle shows middle body and the lower shows lower body. ", shown in Figure 14-1.

CONCRETE CONTENTS OF TONGUE DIAGNOSIS

Tongue diagnosis includes observing tongue body (texture) and observing tongue coating. Tongue texture is the main body of tongue made up of muscles and vessels. Tongue coating is the fur-like material on tongue upper surface. The normal tongue is characterized by a middle size, soft, neither tough nor tender, free movement, pink color, covered by thin and even white coating with moderate moist, which could not be scraped away and has its root in tongue. It is usually called "pink tongue with white and thin coating. "

In tongue observation, tongue coating and tongue texture have their own applying field respectively. The observation of tongue texture is more important. Generally speaking, observation of tongue texture mainly probes the visceral conditions, while observation of tongue coating inspects the nature and location of disease and the clear or turbid of stomach *qi*. If changes are only in tongue coating, the illness is mild. When changes develop from tongue coating to tongue texture, It indicates the disease developed. Observing tongue texture is to observe the abnormal changes of the vitality, color, shape and movement of tongue body.

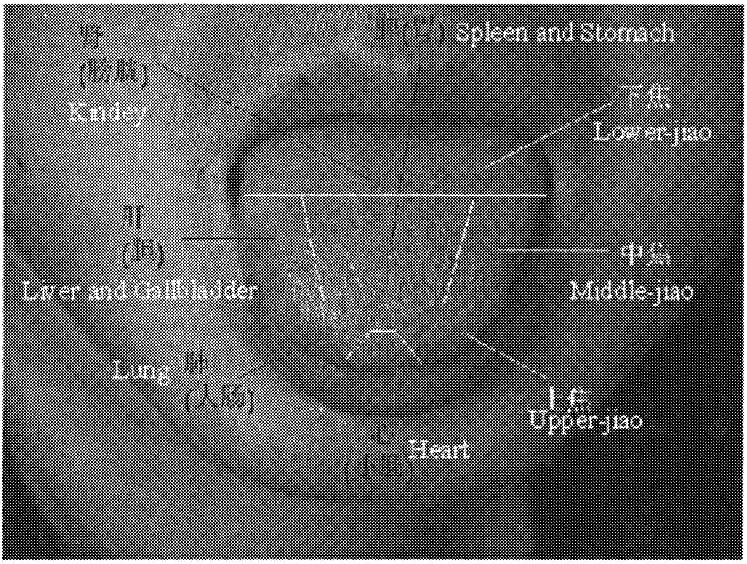


Figure 14-1. Division of tongue surface.

Tongue Vitality

The tongue vitality is shown in the flourish or withering of tongue texture. The flourish means full of vitality. The flourish of tongue refers to light red tongue body with energetic movement and enough fluid of it. It suggests the normal stomach *qi*. If the condition is red, moist and vigorous, no matter the color of tongue coating is yellow, white or black, the prognosis is good. The withering is out of vitality. The withering of tongue refers to dark and dry tongue body with sluggish movement.. No matter there is coating or not, the prognosis is bad in this condition.

Tongue Color

Red is the normal color of tongue body. The normal condition is neither too light nor too deep pink color of whole tongue body. If the color change is out of this range, it is abnormal. The color changes in disease condition are as follows.

Pale Tongue: The color of tongue is lighter than that of normal people or is no red at all. It is due to decline of yang and *qi*, or blood deficiency. The deficient yang fails to send *qi* and blood up to the tongue, and the deficient blood fails to nourish the tongue, so the color becomes pale. The decline of yang will lead to inner deficient cold. The dysfunction of spleen yang leads to water-dampness retained inside, so the color of tongue is pale and tongue body is corpulent with surplus fluid on it. It is accompanied by aversion to cold, cold limbs, listlessness, poor appetite, loose stool, deep and slow pulse or deep and weak pulse.

Red Tongue: When the red color of tongue is heavier than that of normal condition, it is called a red tongue. It suggests internal heat. Heat in body makes *qi* and blood boiling. Then the vessels of tongue are filled and the tongue shows red. Heat can be divided into deficient and excessive types. Excessive heat is a condition of too much yang. The exogenous evil enters the interior of body and then changes into fire. Improper emotions can also transform into fire. The phlegm and food retention in long time will become fire. All those lead to *qi* and blood boiling, and make tongue red. In these cases, the tongue is red and dry, and there are prickles or cracks on tongue upper surface. It is often accompanied by sthenic fever, extreme thirst and profuse drinking, surging pulse, etc. If the red color is only at the tongue tip, it is due to heart fire flaming up. The red tongue in bilateral margins is due to liver and gallbladder fire. If it is accompanied by zigzag shaped cracks in the middle tongue, it is due to the extreme exorbitant heart fire. Red small points on tongue are due to heat-toxin. Heavy red small points overall the tongue surface are called "red star tongue". It is due to damp-heat which attacks the heart and spleen and leads to heat in both zang-viscera and fu-viscera. Small black points in a deep red tongue are due to extreme heat in viscera.

If it is due to deficient heat, the tongue is bright red without coating or with little coating, or a dry tongue with red in middle. It is often accompanied by tidal fever, hectic cheek, fidget, hot sensation in "five centers", night sweating, thready and rapid pulse, etc.

<u>Crimson Tongue:</u> The color is more heavier and darker than red. It is often seen in the stage of extreme fever in exogenous febrile diseases. It can also be seen in internal injury diseases.

The crimson tongue is usually seen in the extreme stage of febrile disease. When warm-heat evils entered nutrient-phase, the tongue color must be crimson; and if the crimson was bright, it was due to disease in pericardium. Obviously when warm evil enters nutrient-phase or pericardium, the tongue color will change into crimson. From that we know the degree of heat in crimson tongue overpasses that of red tongue, which is in *qi*-phase. If the tongue surface is greasy or dry by observation but moist by touching, it is complicated by turbid evil in middle-jiao. The crimson and dry tongue with prickles or cracks is often due to heat in nutrient-phase and insufficient body-fluid. If the crimson tongue can only reach teeth but fails to stretch out, it is due to phlegm obstructed in tongue root. Crimson tongue with large red spots is due to toxin-heat invading the heart. The crimson tongue tip is due to heart fire. If only the middle is crimson and dry, it is due to fire in heart and stomach consuming body-fluid. In internal injury disease, when there is yin deficiency, crimson tongue can also be seen. Crimson and dry tongue with little coating or without coating and thin tongue body, or cracks is the sign of yin deficiency. Crimson and moist tongue with little coating is due to blood stasis. If the tongue is crimson and withering without coating, and the tongue surface is mirror-like, it is called "mirror-tongue" and belongs to the critical condition of stomach and kidney yin exhaustion.

<u>Purple Tongue:</u> When tongue body is purple, it is called a purple tongue. There is difference in cold or heat of the diseases indicated by purple tongue. If it is due to extreme heat, it is dark purple. The bluish purple is often due to extreme cold.

Dark Purple Tongue: The color of tongue is purple or dark. It is caused by: l) Heat evil entering interior. The purple and dry tongue with cracks is seen in severe case of yin exhaustion caused by heat. If purple is seen in whole tongue, it is due to extreme heat in viscera. If purple is seen only at one part, it is due to stagnated heat in the viscus corresponding to the part. The purple and swelling tongue with large red spots is due to heat invading the heart. 2) Blood stasis. The tongue is dark purple and moist. If the whole tongue is dark purple, it is due to *qi* and blood stagnation of whole body. The purple macule in some parts of tongue indicates the blood stasis in some viscera relating to the part. 3) Phlegm-dampness. Phlegm and dampness accumulate in the interior and produce heat, then the tongue shows purple color and the coating is slippery and greasy. 4) Injury by alcoholic toxin. The purple and swelling tongue is due to alcoholic toxin invading the heart. The deep purple and dry tongue is due to alcoholic toxin accumulated internally. 5) Yin exhaustion. Tongue body is dark purple and dries just like a pig liver or pig kidney with capsule pealed. It is due to kidney and stomach yin exhaustion.

When the tongue body is light blue and purple, it is called bluish purple tongue. It is due to yin deficiency of liver and kidney, blood stasis or phlegm stagnation caused by yang deficiency which fails to warm and push them in movement. If tongue body is bluish purple, moist and small in size, and no coating, it is because cold evil attacks liver and kidney directly, and belongs to extreme cold. If the tongue body is bluish purple with spasm and tenderness in epigastrium or around navel, it is due to the combination of phlegm and blood.

Blue Tongue: When the tongue is blue without any red color, it is called a blue tongue. It is seldom seen in clinic. The blue tongue often indicates yin-cold and blood stasis. Because yin-cold evil prevails, the yang is stagnated and hard to move, then blood stasis occurs, thus the blue color is seen in tongue. Blue tongue with slippery coating is due to direct cold attack on the liver and kidney. Blue in tongue margins is due to internal blood stasis. Blue tongue with dry mouth, dislikeness of swallowing water but gargling, is a sign of blood stasis. Blue tongue with slippery and greasy coating is due to internal accumulation of damp-phlegm. Blue tongue without coating is due to extreme deficiency of *qi* and blood.

Tongue Shape

Cao Bingzhang said: "If one wants to know visceral diseases, one should observe tongue shape first." Tongue shape is the out figure of tongue body. Its abnormal changes include eight main categories as follows:

Tenderness: The striae of tongue are delicate, fine and smooth. It indicates deficient syndrome. For example, the red, moist and tender tongue is due to deficient heat in the Heart Meridian.

Toughness: The striae of tongue are rough and sturdy. It indicates excess syndrome. It is mainly seen in excessive heat or blood stasis.

Corpulence: It is mainly due to stagnation of phlegm, dampness, and water or fluid-retention. The pale, tender and corpulent tongue is due to yang and *qi* deficiency of the spleen and kidney.

Swelling: The enlarged and swollen tongue body, even filling up mouth, is called a swollen tongue. It is usually due to excessive heat and blood in the heart and spleen.

Thinness: The small and emaciated tongue body is called a thin tongue. The pale thin tongue is usually due to *qi* and blood deficiency, or deficiency of both heart and spleen. The dry thin tongue in red or crimson is often due to fire flaring in yin deficiency. The withering thin tongue in dark color is due to exhaustion of both *qi* and yin.

Crack: Cracks are due to blood deficiency. The shallow cracks in little amount show the slight decline. The deep cracks in large amount show the severe decline. Transverse crack is seen in yin deficiency constitution. Borneol like cracks are often seen in the aged or deficiency condition. The short crossed cracks in red or crimson tongue are due to yin and body-fluid deficiency. Cracks in pale tongue are often due to blood deficiency. Cracks in pale and tender tongue are due to spleen deficiency and dampness overflow.

Cracks are also seen in excess syndrome. For example, Cracks in red or crimson tongue is often due to heart fire flaring. If the bean-like projections in tongue root make a deep and long crack, it is due to evil accumulation in stomach. Crack with bleeding is because fire evil compels blood to overflow. Cracks can also be seen in normal people. They are formed inborn. It can be known by asking patients.

Prickle: The soft prickles in tongue surface are made up of lung *qi* and genuine fire in life-gate. So they are seen in normal conditions. The enlarged and increased prickles are due to excessive evils, while thinned and lessened prickles result from deficiency of *genuine-qi*. All awn-prickles are due to internal accumulation of heat evils. The more exorbitant the heat evils is, the more and larger the awn-prickles are. In clinic, we can ascertain the location of disease according to the position of awn-prickles. For example, awn-prickles on tongue tip are due to heat evil in the heart; awn-prickles on the sides of tongue are due to heat in the liver and gallbladder; awn-prickles on middle tongue are due to heat in the spleen and stomach. Awn-prickles with yellow thick coati are due to Yangming excess. Awn-prickles with black dry coating are due to heat yin meridians.

Smoothness: Because stomach yin is exhausted and stomach *qi* fails to produce tongue coating, there is no coating on tongue surface. It is a critical condition due to stomach exhaustion. If tongue is pale, it is due to the extreme deficiency of both *qi* and blood. While it is red or crimson, it is because fire drying up the water. The stomach fluid and kidney yin are both exhausted.

Teeth-Print: The pressing marks of teeth on lateral sides of tongue are called teeth-print. It is usually due to enlarged tongue body pressing on teeth. So it is often seen with swollen tongue. If it is seen in pale and moist tongue, it belongs to internal cold-dampness. While in pink tongue, it is due to spleen deficiency or *qi* deficiency.

Hypoglossal Vessels (vessels below tongue): In normal condition, beside the frenulum of tongue, two thick bluish purple vessels can be seen. Their diameter is no more than 2.7 mm. Their length is no more than the 3/5 of the length from tongue tip to sublingual caruncle. There is no branch or spot. If their color becomes darker, or they are thickened, or there are bluish or purple black vesicles on them, it is due to liver *qi* stagnation, which leads to phlegm and heat accumulated inside, and then blood stasis.

Tongue Coating

White coating was seen in not only exterior syndrome and cold syndrome, but also deficiency syndrome. It could be seen in the restoration stage of all diseases, non-organic diseases like neurosis, diseases without symptoms like early breast cancer, chronic infections, etc. There were changes of filiform papillae, such as increased and thickened cutin of filiform papillae. The notable difference was found between the normal white coating and pathological white coating in rate of red and white fluorescence when examined by fluoroscope. It provided an objective basis for judging pathological white coating from the normal.

Yellow coating was often seen in inflammatory infections and fever. When W.B.C was over 15, 000 /mm3, the incidence rate of yellow coating reached 72.9%. It was also seen when there were congestion, edema, ulcer or bleeding of gastric mucous membrane which were due to active gastric ulcer, superficial gastritis, carcinoma of stomach. In yellow coating condition, there were bacteria gathering and inflammatory exudate on tongue surface, hyperplasia of filiform papillae and hyperkeratosis. The yellow color was produced by the bacteria.

Gray and black coating were usually seen in critical conditions in which the heat was the main cause. They were related to the high fever, dehydration, toxin, disorder of central nerve system and digestive system, fungi and other color-producing bacteria, chronic inflammation and kidney deficiency, etc. Pathological examination showed the thickened mucous membrane, enlarged papillae and hyperplasia of basal cells. In prickle cell layer, the interspace of upper cells disappeared, and vacuoles and hyperkeratosis appeared. There was mixture of cutin, piled bacteria, fungi, etc.

Especially in recent decades, many researches have been done on the tongue pictures in some common diseases. They include hypertension, acute myocardia infarction, pulmonary heart disease, cerebrovascular diseases, stomach diseases, hepatitis, tumors, and so on. To sum up, the researches on tongue examination verified and enriched the theory, and clarified the formative factors of tongue picture, by the clinical verification and modern scientific methods. The law of tongue pictures change in common diseases has been recovered. In some diseases, such as hepatitis, cardiac vascular disease and tumor, tongue picture was taken as an important index for diagnosis. Because the tongue examination is important for diagnosis, non-invasive and acceptable for patients, its application in diagnosis was very promising.

BIOMETRICS TONGUE DIAGNOSIS

We have been developing a biometrics tongue diagnosis system whose purpose is to directly capture tongue images with resolution 640x480 using a set of specific image acquisition device with an advanced kernel, camera and compare them with predefined features stored in the feature database (see Figure 14-2). Such a system has great significance in early diagnosis and health care. This system is concerned with the following issues: tongue image acquisition and tongue classification. Tongue classification determines whether a tongue is healthy and which disease it suffers if not, which involves in two operations: feature extraction and feature matching. To implement such a system, these problems should be solved beforehand.

Tongue Image Acquisition and Database

From traditional Chinese medicine theory, color and luster are important features of tongue. So undistorted color-acquisition is crucial. There are two ways to obtain the tongue image with high quality. One way is via close box, as shown in the following figure, which can avoid the influence of uneven illumination but is less user-acceptable. Now we have designed such a device and are to manufacture a prototype to be applied to our further research (Figure 14-3). The other is to use high intensity white light, which can overwhelm the natural light so as to guarantee comparative steady illumination. In both cases, a high quality camera is used to sample the tongue image, and the image is then digitized and passed to the PC machine for processing.

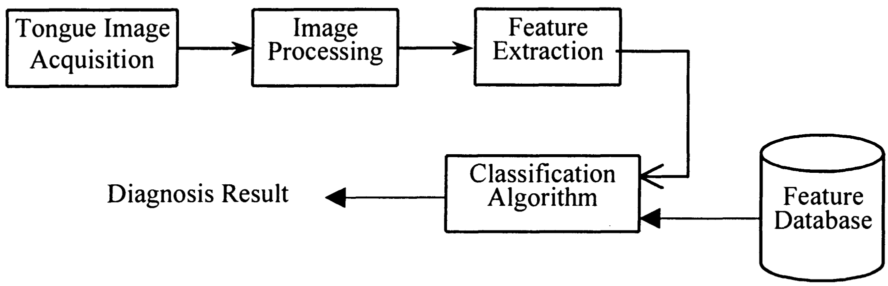


Figure 14-2. The flowchart of a biometrics tongue diagnosis system.

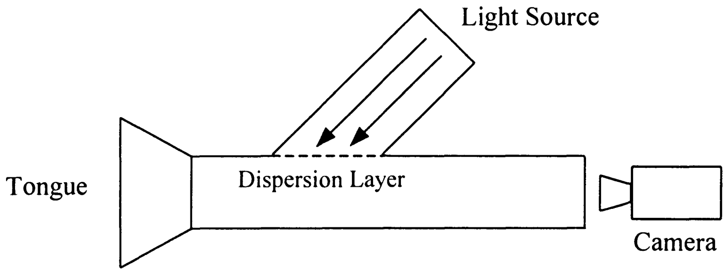


Figure 14-3. The tongue capturing system.

It is necessary to build a tongue image database with enough samples. The recognition results of tongue diagnosis system are finally decided by these samples. So, the following steps will be adopted:

1) Set up a tongue database with over 10,000 samples, including raw tongue images of lingual surface and hypoglottis, and the extracted features of tongue. These samples should be obtained from the persons with different sex, age, disease and occupy in terms of predefined proportion. Note that tongue database with such scale is enough to build an automated tongue diagnosis system.
2) Build classified index of the tongue database in terms of sex, age, disease, symptom and feature.
3) Select different types of tongue images in terms of the category of disease and evaluate their input qualities and representative samples.

Image Processing (Segmentation)

After getting the tongue image, we must process the image, including removing noise and segmenting of image. Segmentation is to distinguish tongue part from the background. It is a basic processing of tongue image, shown in Figure 14-4.

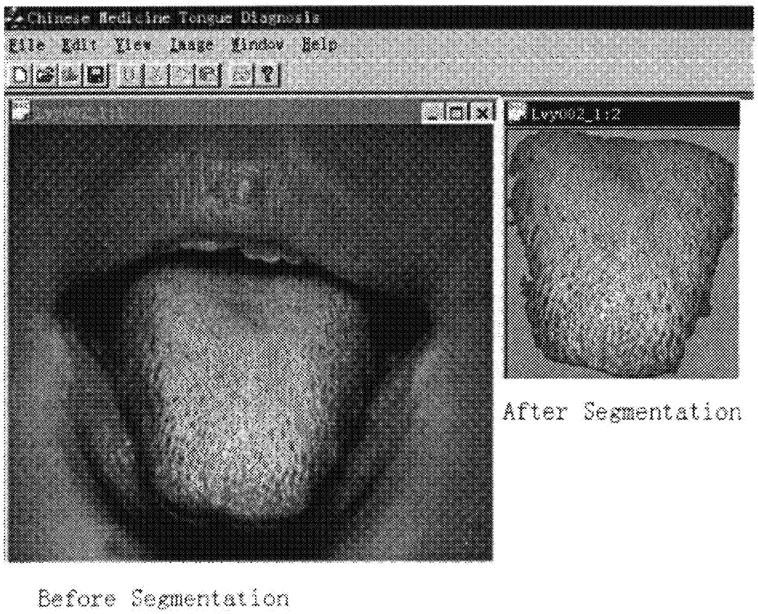


Figure 14-4. Segmentation of tongue image.

Tongue Feature Extraction

We have two feature extracting approaches, namely expert knowledge based feature extraction and statistical method. Various features, with the traditional Chinese medicine experts' guidance, can be extracted from tongue, including geometry features, color and luster features, tongue coating, and lingual vena. How these features are extracted efficiently and exactly is a key to the accuracy of tongue diagnosis.

According to famous tongue diagnosis expert, Prof. Li Naiming, there are many features as blow:

1) Tongue Color: pale, pink, red, crimson, purple, bluish purple, dark purple, blue, and so on.
2) Tongue coating: thinness and white, white, thickness and white, thinness and yellow, yellow, gray, black, exfoliation and nothing.
3) Tongue visible material: speckle and spot, streak, knurl, excrescence, ulcer, tooth print, tumour, blue print, tongue body fat or emaciated, and so on.
4) Tongue vena: diameter, circuity, exaggerate, branch.

5) Tongue vena color: pink, dark red, red, purple, crimson. Changed part: tip, margins, root, middle.

The statistical approach is free from experts' experience but is promising as well. For example, we can get the statistical feature, such as mean, variance, energy and so, showed in Figure 14-5.

Tongue Classification

Classification approach is to assign a tongue pattern into a certain category according to its formation. According to the traditional Chinese medicine theory, the whole tongue can be artificially divided into several non-overlapped regions which correspond to different apparatuses respectively. So the feature matching process should be carried out region by region.

In our system, some common diseases can be diagnosed correctly, shown in Figure 14-6.

Feature matching actually is a process of classification, so the various classification approaches based on statistical pattern classification, neural networks, belief networks and fuzzy theory can be applied to tongue image matching. The issue of feature matching of tongue image can be further described as follows:

1) Develop the efficient approaches for disease feature matching based on statistical pattern classification, neural networks, belief networks and fuzzy theory.
2) Compare their performance between various approaches for disease matching.

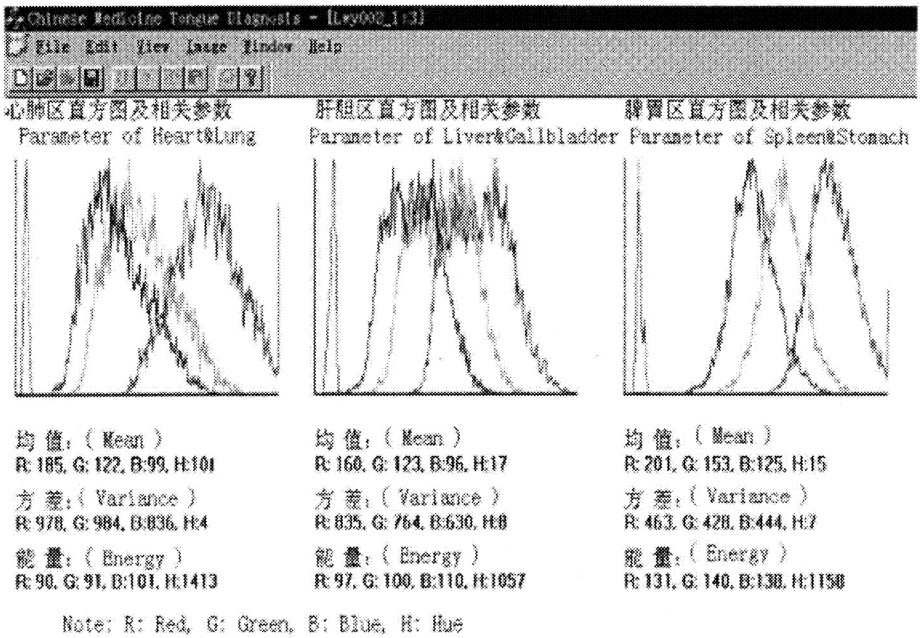


Figure 14-5. The example of statistical features.

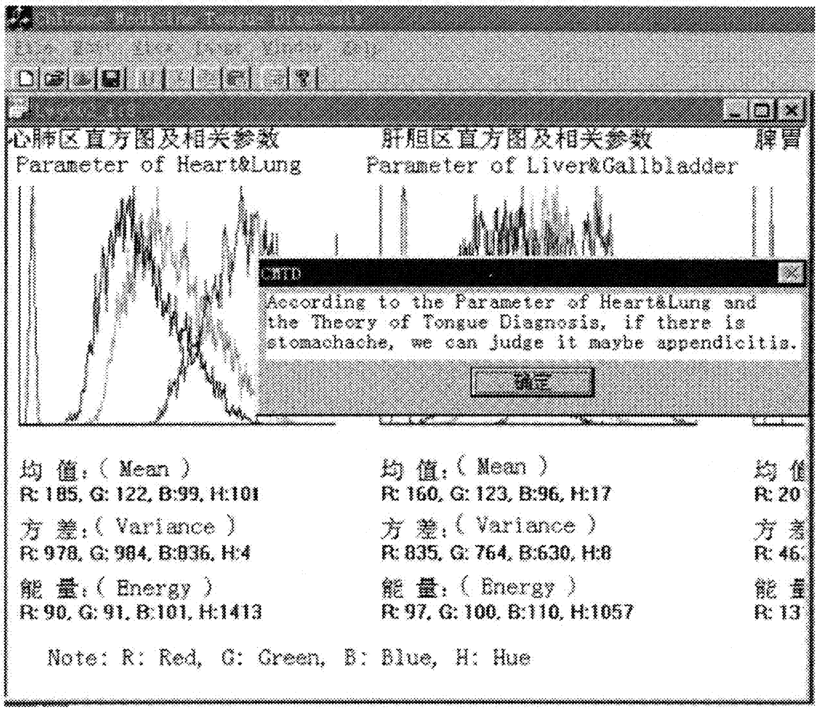


Figure 14-6. The example of diagnosis by Biometrics TCM System.

3) Investigate the effect of the different identification functions on identification results based on the different approaches for tongue feature matching.

Tongue Diagnosis Applications

As an expert system, tongue diagnosis based on biometrics can be applied to early diagnoses of some hard disease and general health care. Besides, tongue diagnosis, as an independent channel, can play an important role for an integrated Traditional Chinese Medicine in future.

TCM has a long history over a few thousands of years and accumulated very rich practical experiences in diagnostic methods. During long time development, the four diagnostic methods have formed, e.g. observation, auscultation and olfaction, interrogation and palpation. Each diagnostic method has its unique clinical function and can not be replaced by one another. They must be applied simultaneously to diagnosis to ensure correct diagnosis.

Tongue diagnosis is one of the most widely used diagnosis methods in the traditional Chinese medicine and is valuable in clinic application. Tongue diagnosis can not only prompt the pathological changes, the positions and degrees of the changes, but also can find what the modern medical equipment such as B ultrasonic, CT and NMR can not, for instance the degrees of the pathological changes and the functions of the viscera. It can also propose the corresponding cure process in time. In the modern automated world, there is an ever-growing need to develop an automatic tongue diagnosis system based on image processing and pattern recognition. Therefor, we had firstly introduced the biometrics technology into tongue diagnosis and gotten much success.

References

[1] Z.X. Long, *Diagnostics of Traditional Chinese Medicine*, Academic Press, Beijing, 1998.
[2] B. Wang, *Plain Question* in *Internal Classic,* People Health Press, Beijing, 1979.
[3] Y.G. Zhang, *Spirit Pivo* in *Internal Classic,* Science and Technology Health Press, Shang Hai, 1958.
[4] T.T. Den, Z. Q. Guo, *Diagnostics of Traditional Chinese Medicine*, Shang Hai Science and Technology Press, Shang Hai, 1984.
[5] H. H. Yin, B. N. Zhang, *Elementary Theory of Traditional Chinese Medicine*, Shang Hai Science and Technology Press, Shang Hai, 1984.
[6] N.M. Li, *Illustration of tongue pictures of Yu-syndrome*, Heilongjiang Sceince and Technology Press, Heilongjiang, 1990.
[7] N.M. Li, *The great integrated of Chinese Tongue Diagnosis*, Academic Press, Beijing, 1994.
[8] N.M. Li, Y. F. Wang, *Diagnosis by Observing Tongue*, Heilongjiang Sceince and Technology Press, Heilongjiang, 1987.
[9] X.Yang, S.G. Cheng, *Hand Diagnosis with Shape and Color*, Tian Jin Science and Technology Press, Tian Jin, 1998.
[10] D.Y. Wang, *Practical Illustration Picture for Palm-print Diagnosis*, Beijing Science and Technology Press, Beijing, 1998.
[11] C.X. Wang, *Modern Palm-print Diagnosis*, Gansu Nation Press, 1995.
[12] S.L. Huang, *Research on Pulse of TCM*, People's Health Press, 1995.
[13] G. Maciocia, *Tongue Diagnosis in Chinese Medicine*, Eastland Press, 1995.
[14] D. Bensky and R. Barolet, *Chinese Herbal Medicine: Formulas & Strategies*, Eastland Press, Seattle, Washington, 1993.
[15] D. Bensky and R. Barolet, *Chinese Herbal Medicine: Materia Medica (Revised Edition),* Eastland Press, Seattle, Washington, 1993.
[16] Isselbacher, *Harrison's Principles of Internal Medicine (13th Ed.),* McGraw-Hill, New York, 1994.
[17] Jones and Robert, *Acupuncture Techniques*, North Atlantic Books, Berkeley, 1996.
[18] Macciocoa and Giovanni, *Foundations of Chinese Medicine*, Churchill Livingston, New York, 1989.
[19] Macciocoa and Giovanni, *The Practice of Chinese Medicine*, Churchill Livingston, New York, 1994.
[20] *The Study of Qualitative and Quantitative Analysis for Chinese Medical, Tongue Diagonsis Using High Resolution Color CCD*, http://www.ccmp.gov.tw/4e/86-052e.htm.
[21] Macciocoa and Giovanni , *Tongue Diagnosis in Chinese Medicine*, Eastland Press, Seattle, 1994.
[22] Shanghai College of TCM , *Acupuncture A Comprehensive Text*, Eastland Press, Chicago, 1983.
[23] Rui and Chun Ji , *Acupuncture Case Histories from China*, Eastland Press, Seattle, 1988
[24] S.Z. Li and G. Seifert (Editor), S.C. Li, *Pulse Diagnosis*, Paradigm Publisher.
[25] D. Molony, *The American Association of Oriental Medicine's Complete Guide to Chinese Herbal Medicine* , 1998.
[26] *Traditional Chinese Medicine*, http://www.healthy.net/CLINIC/therapy/Chinmed/Index.asp.
[27] *The Journal of Chinese Medicine,* http://www.pavilion.co.uk/jcm.
[28] *Chinese Medicine - The New World of Holistic Health,* http://detox.j12.net/chronicho/chinese_medi-cine.phtml.
[29] *Chinese Medicine and Acupuncture in Canada*, http://www.medicinechinese.com/.

[30] *Birmingham Centre for Chinese Medicine*, http://www.bccm.freeserve.co.uk/.
[31] *Chinese Traditional Medicine*, http://www.libraries.wayne.edu/shiffman/altmed/china/chin...
[32] *Traditional Chinese Medicine*, http://loki.stockton.edu/~gilmorew/consorti/2peasia.htm.
[33] *Traditional Chinese medicine correspondence programs*, http://www.bta.net.cn/tcm/tcm/chinatcm.-htm.
[34] *Traditional Chinese Medicine*, http://reference.cd-rom-directory.com/cdrom-2.cdprod1/010.
[35] *Yahoo! Health>Traditional Medicine>Chinese* http://asia.yahoo.com/Health/Traditional_Medicin-e/Chinese.
[36] *Traditional Chinese Medicine*, http://www.iospress.nl/html/tcm.html.
[37] *Institute of Traditional Chinese Medicine*, http://www.mitcm.org/.
[38] *Traditional Chinese Medicine*, China Beijing. http://www.tradedaily.com/tcm/big.htm.
[39] *Traditional Chinese Medicine*, http://www.scn.org/fremont/acu/chinese_med.html.
[40] *Shanghai University of Traditional Chinese Medicine*, http://www.csc.edu.cn/foreign_stu/shang-hai/shzyy1/shzyy.htm
[41] *Chinese Traditional Medicine*. http://www.chalmers.com.au/ChinaHouse/Pc/book3.html

15 FUTURE WORK IN BIOMETRICS

As mentioned in previous chapters, biometrics is promising and will play an important role in the modern society because it offers a new approach for identity identification based on the unique, reliable and stable personal characteristics. But biometrics technology is still in its developing stage and there are many works to do. In this chapter, some related topics in future will be discussed. First, a new application of biometrics in Internet-Ecommerce is discussed. Then, some aspects of future studies are discussed, in which the integrated biometrics and VLSI biometrics are emphasized.

15.1 Biometrics in Internet-Ecommerce

In recent years, the host-computers in Internet are increasing exponentially. Internet is not only a communication tool, but also a great market. Many famous companies, such as Microsoft, IBM, Sun, and so on, have paid tremendous attention to the market of Internet. The Internet based application, especially Internet-Ecommerce has gotten a fast development. Many countries, even some developing countries such as China, India, devote much foundation to Internet-Ecommerce.

Obviously the most important issue in developing Internet-Ecommerce is its security. Nobody will deal with someone, whose identity is not confirmed, especially in Internet. Figure 15-1 is an illustration of Internet-Bank.

Unfortunately, the traditional security measures such as passwords and ID cards cannot satisfy various security requirements. The incident of invasion to USA Defense Ministry is one example of such failures. Biometrics technology offer a new approach for identity identification based on the unique, reliable and stable personal characteristics, such as finger-prints, palm-prints, facial features, iris pattern, retina, voice pattern, handwriting. Therefore we should develop a biometrics-based security system for computer network access control. Now some biometrics systems which are based on single biometrics feature and application oriented, are used in many places. Furthermore the integrated multiple biometrics features such as finger-prints, palm-prints, facial features and voice patterns to authenticate a person's identity and verify his/her eligibility to access the network are developing.

To meet the challenge and immediate need for a high-performance Internet authentication service to secure Internet commerce applications, one of the pioneering Web-oriented, biometrics-based systems for automatic identity identification and verification is developed.

If someone wants to access the secure system, for example, banking system, he has to send his biometrics features to the server of Biometrics Secure System (BiSS).

This is the corresponding to authenticate individual's identity and send the confirmation information to the secure system, which is expected to be accessed by client. Whether the secure system can be accessed or can not be accessed is decided by the confirmation information from BiSS.

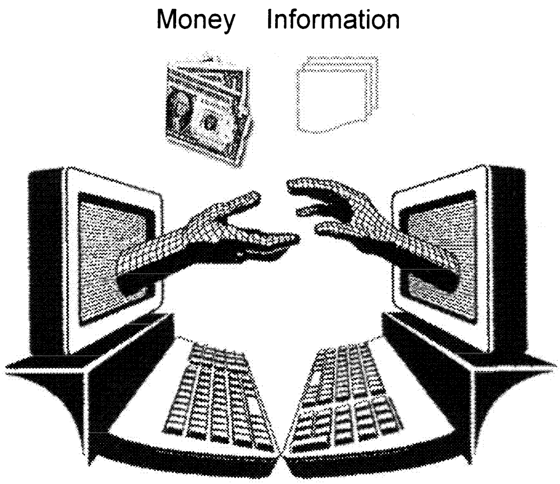


Figure 15-1. Illustration of Internet-Bank.

First of all, we must analyze the security requirements of Internet commerce services. Then we should design system framework and architecture, and design biometrics verification and identification algorithms. After biometrics data acquisition and database implementation, we should setup system, test and evaluate it.
Techniques from various areas are needed to be investigated to achieve high-performance (efficiency, reliability and scalability). The following techniques should be investigated:

15.2 Key Technologies in Biometrics

For any kind of biometrics systems, the following issues should be considered: (1) How to acquire better data? (2) How to manage the database more effectively? (3) What more can we do for feature extraction and matching? (4) How to do the classification effectively? (5) How to evaluate the methodology and performance of the biometrics system? (6) What can we do in biometrics using parallel & distributed computing to implement on-line biometrics processing?

BIOMETRICS DATA ACQUISITION

The performance of biometrics system is deeply depending on its data acquisition. The performance will be better when the quality of the acquired data become better. Otherwise, the performance will degrade when the acquired data become worse, even though at the same time all other factors are kept unchangeable. So, it is very important but technically challenging to design a device such as CCD or scanner that is able to acquire much better data and tolerate various situations such as degraded friction skins and/or unpredictable environment factors.

At the same time, the applications of any biometrics system are based on the convenience and cheapness. If the size of CCD or scanner is too big or too expensive, it will undoubtedly affect the applications of the biometrics system. So it is also important to design acquisition device with smaller size and low price.

It should be noticed that the performance of most existed biometrics system is not satisfactory. One of its major reasons is the limitation of data acquisition device. In all of biometrics data acquisition devices, finger scanner is the best one. Just as we mentioned in previous chapters, there are already a lot of finger scanners existing in market, in which some are excellent. Using them finger images with rather good quality can be acquired. They have been incorporated in many different kinds of products such as Finger Lock, Finger Keyboard, etc. But even these scanners also have some disadvantages. Some research results revealed that these finger scanners have poor performance to dry fingers. If someone's fingers are too dry, he can not use these scanners to acquire clear image. More than 10% of human persons have this kind of finger. The more annoying thing is that some persons only have dry fingers. Then, how to acquire better image for these fingers is an important problem.

One kind of difference existed between voice patterns obtained through different kinds of microphones, which was showed by some experiment results recently. This leads to some difficulties in voice pattern verification. So one robust voice acquisition device should be developed.

Nowadays the acquisition devices for iris or palm-print can not meet the need of applications. Other acquisition devices may be even worse than the above two. Take face recognition as an example. The CCD is sensitive to illumination and 3-D information of human face can not be extracted through it. Can we develop one kind of device incorporating the advantages of CCD, infrared imaging and depth imaging? If so, it can surely provide us with much more robust data, which can better describe the 3-D information of human face. Then it can enhance the performance of face recognition systems and widen their applications.

Because the application of biometrics has given background, given objects and given aims, the data acquisition devices can be designed specially for this given application. Considering all these factors, we can incorporate the existing techniques with special conditions to develop one kind of special device.

BIOMETRICS DATABASE MANAGEMENT

How to organize varieties of individual's biometrics information in biometrics database for fast indexing and searching is another important issue. One approach is to divide the whole database into many small databases according to some robust category criterions. Because each individual is associated with a hierarchy of biometrics feature vectors, an appropriate multidimensional indexing method such as R&D-trees to facilitate the searching process, can also be investigated.

In general, biometrics data is structural, visual, and time dependent. Many of the existing biometrics systems cannot handle such data easily. Maybe a wavelet based scheme should be considered so that integrated biometrics feature will be suitable for data storage, access and maintenance of the large biometrics database.

BIOMETRICS FEATURE EXTRACTION & MATCHING

Biometrics feature extraction is to select and extract proper biometrics features, which represent the individual's identity best. Dominant feature points, textures, shapes and voices are the common features in biometrics. Feature matching is to determine whether two patterns are belonging to the same person.

In all aspects of biometrics, the research of finger verification is the most sufficient. Other problems such as face recognition, signature verification, iris recognition, etc, have not been practically solved. There are even some deficiencies in finger verification. Some research results have revealed that there are too many broken points in proportion of fingerprint images. Traditional methods can not be applied to extract the useful features. But robust features can be easily extracted by human ability. That means, we can develop some efficient algorithms for these problems.

In human face recognition, there exist much more problems. As we know, real human face is 3-D non-rigid body, and human face recognition is one problem of understanding and recognizing 3-D non-rigid body. How to extract 3-D information from 2-D images is still an unsettled problem in the field of computer vision. Upon on this, recognizing 3-D non-rigid human face can not be done well. It can be concluded that human face recognition is one long-term study.

Some experiment results showed that there exists one kind of difference between voice patterns obtained through different microphones. This leads to some difficulties in voice pattern verification. So it need to be studied thoroughly.

Additionally, the algorithms in iris recognition, signature verification, palm-print verification, etc, are far from the need of applications. These hampered the spread of these techniques. But we believe that they can be improved along with the progress of related research fields.

It also should be pointed that, one of important problems in the research of algorithms is how to take advantage of the experts' knowledge to obtain robust features from the biometrics data.

For future uses, especially for Internet-Commerce, a fast and robust matching algorithm in a hierarchical structure is needed for biometrics. In order to speed up the matching process, we should develop some scalable parallel algorithms for biometrics computing and implement these algorithms by using platform-independent mobile agents as a novel agent-oriented parallel programming tool to achieve maintainability, reusability, and dynamic resource allocation in heterogeneous environments. Due to the loose connection between agents, mobile agents are regarded as an emerging paradigm for distributed computing. They are programs typically written in a safe language such as Java. Agents can be dispatched from a client computer and transported to a remote server for re-execution, which can substantially reduce the communication overheads.

We should develop some good methods for multiple biometrics feature extraction and representation, matching, such as wavelet-based scheme, genetic algorithms and neural networks for comparison.

BIOMETRICS CLASSIFICATION TECHNOLOGY

Biometrics classification is to assign a biometrics pattern to a particular category corresponding to its content. This is one of the most difficult problems for identification and verification systems. The classification itself is one of traditional problem in pattern recognition and is studied thoroughly. But the classification of biometrics data needs much more experts' knowledge. How to turn the experts' knowledge programmable is rather important.

The classification of fingerprint, face patterns, palm-print, or signature is important for many reasons. First, it can be used for management of the database to reduce the search time and computational complexity. We can divide one large database into many small databases according to its category, so that the input fingerprint needs to be matched only with a subset of the fingerprints in the whole database. Then the tackle of large database can be efficiently turned to that of small database. Secondly, it can give benefit to recognition process because the category can also be judged as one important feature to be used in recognition. Third, the category of biometrics is thought as one of human abilities. So the research of the category of biometrics can be used as one tool of studying human intelligence.

It is a pity that there is only a little study of biometrics classification, in spite that the classification of fingerprint into several classes such as whorl, right loop, left loop, arch and tented arch is well studied. The category of other biometrics data such as face patterns, palmprint, or signature has not been deeply studied.

In other study fields such as annealed photography, Chinese Palmistry, chirography, and medicine, experts have studied these categories since a long time ago. The knowledge from these fields can be applied in our automated biometrics. Human face classification can use the rules of annealed photography to classify the faces into different face contour types, different eye types, or different eyebrow types. Palmprint can be classified according to Chinese Palmistry. The category of signature can take advantage of the results of chirography research and the classification of iris can obey the results of eye medicine study. On the other hand, the classification of biometrics can be carried totally from the angle of automatic category, no matter how the experts did it.

To tackle such a challenging issue as classification, we can investigate fuzzy clustering and semi-fuzzy clustering techniques developed by other groups. We will extend the existing algorithms for the project needs by considering conceptual, geometrical and other structural features of the problem. Other approaches such as simulated annealing and genetic algorithms will be also investigated for comparison. Similar to the previous task, we should develop scalable parallel algorithms for the related computing tasks specified above and implement them by using mobile agents in a heterogeneous environment.

COMPREHENSIVE EVALUATION CRITERIA

As we know, independent and careful evaluation is needed for any kinds of products, no matter it is software or hardware. Nowadays, there are so many technologies being used in biometrics market. So evaluation criteria must be studied as soon as possible.

To evaluate the performance of a biometrics system, one important thing is to establish one large database, which should have a vast representation. This kind of database can also be judged as one standard for comprehensive evaluation. Such a database should contain a lot of smaller databases, each of which acts on one level. So this database can also apply to different kinds of test and evaluation. This problem seems easy, but in fact it is quite complex. As mentioned above, the retrieval of data relies on the acquisition device. The quality of data acquired from different devices varies greatly. Therefore when establishing a database with good representation the differences among all the devices should be considered seriously so as to make our system have a better performance in expansion.

Another important factor is that, to each biometrics character such as fingerprints, palm-prints, facial features, voice patterns or signatures, specialized research on a certain evaluation should be done. The recognition and identification process can be divided into several steps, and in each step there should be a specialized evaluation to assess the performance. In fact this process can do great benefits to both system integration and the market evaluation of the product.

The performance measurements of biometrics systems are mainly defined by the security levels. There are two commonly used measures: FAR (False Acceptance Rate) and FRR (False Rejection Rate). Both FAR and FRR depend on the security level and are strictly related to each other. This issue has been addressed for single biometrics systems. For integrated biometrics system, we should study the statistical expressions for calculating FAR and FRR. A theoretical model for calculation will be derived and the mathematical relation between FAR and FRR will be explored. We will use such measures for performance evaluation and to guide the developers for the improvement and optimization of biometrics systems.

PARALLEL & DISTRIBUTED COMPUTING

In general, biometrics computing involves the execution of a large number of operations on large sets of structured data, requiring considerate execution time if performed in a sequential manner. Thus, biometrics-based identity identification and verification is a computation expensive task. In most realistic cases, large collections of individuals' biometrics information such as fingerprints, palm-prints, facial features and voice patterns, are distributed across networks rather than stored in one machine. This makes it difficult to perform a biometrics-based identity identification and verification task on a single-machine-based system as most current research has been focused. Therefore, it is most desirable to implement this biometrics-computing task in a heterogeneous environment, which consists of different machine clusters. In addition, heterogeneous systems can more effectively utilize the existing computing resources in different institutions and organizations. To facilitate the co-ordination among individual task components and their allocation to the most appropriate machines for high performance, it is very important to implement each individual task using platform-independent software tool such as mobile agents for distributed computing in a heterogeneous environment. More importantly, the parallel implementation of the biometrics computing tasks in distributed computing environments will reduce execution time and meet the real-time computing requirements for many applications.

Therefore, it is necessary to investigate, design, analyze, implement and evaluate scalable parallel algorithms for distributed biometrics-based identity identification and verification in heterogeneous environments. The relevant technologies include parallel algorithms, task mapping and scheduling, load balancing and distributed mobile computing.

ON-LINE BIOMETRICS PROCESSING

It is a challenging task to achieve "real-time" on-line biometrics processing because of the computation burden to retrieve an individual's identity in a large database. It is crucial to have a fast search algorithm to meet the timing requirement. In contrast to the current approaches, which often use fixed matching criteria to select candidates, we propose selective matching criteria associated with user's query for flexible searching. We will also develop a guided searching procedure from coarse level to fine level for fast processing. In addition, parallel and distributed computing approach will be explored to reduce execution time.

15.3 Integrated Biometrics

Given several kinds of biometrics data, how to integrate them to obtain better performance?

WHY TO USE INTEGRATED APPROACH?

Biometrics have provided a better solution for the increased security requirements of our modern society than the traditional security measures methods. Recent advances in computer technology have made it possible for industry to develop affordable automated biometrics identification and verification systems. Many biometrics have been used for the identification and verification of individuals. However, these systems are based on single biometrics feature and application oriented. As we know, each biometrics has its advantages and disadvantages for both different performance criteria and different operating environment. No one biometrics can provide 100% correct rate for a biometrics-based system. A biometrics system, which relies solely on a single biometrics, may not be able to satisfy the practical performance requirement including the toleration of variations, distortions and noise. Consequently, it is desirable to utilize and integrate multiple biometrics features to improve the accuracy of identification. Identification based on multiple biometrics represents an emerging trend. Such integrated system is desired to overcome some of the limitations of a single biometrics system and take advantage of the capabilities of each individual biometrics. The accuracy of the identification system is improved as we effectively utilize and integrate an increasing number of information sources related to an individual to confirm his/her identity.

WHAT TO DO FOR INTEGRATED BIOMETRICS?

There is a common and intuitive assumption that the combination of different biometrics should improve performance, since more information is intuitively better than less information. However, it is not true, an inappropriate combination of different biometrics may result in difficulties in decision making for identification and verification. For example, if a stronger biometrics is combined with a weaker one, the resulting decision environment will be in a sense averaged. In this case, we would like to use the stronger biometrics for the performance decision. If the two biometrics differ significantly in their power, and each operates at its own cross-over point, then combining them gives significantly worse performance than relying solely on the stronger one. If the information sources are effectively utilized, however, better performance may be obtained. The key issue of this problem is how to combine different biometrics. An efficient integration scheme may be necessary for improving performance of biometrics systems. In particular, as some of the biometrics sources become unavailable or unreliable, integration technology of different biometrics will play an important role in imparting capabilities to the identification and verification processes. This is a new challenge for automated biometrics systems. Hitherto, few research results are published for automatic biometrics-based identification and verification systems [10-11, 17]. Therefore, it is significant to explore new approaches of integration of multiple biometrics for improving performance of automated biometrics-based systems.

We can hypothesize that the combined performance would lie somewhere between that of the multiple biometrics conducted individually under some operating environments. First of all, there are two issues for multiple biometrics:

a) Whether do we need to use multiple biometrics for a specific application or not. A good measure of biometrics may be better than combination of multiple biometrics with inaccurate measure. How to establish the criteria for this requirement for multiple biometrics.
b) If we meet such requirement of combining multiple biometrics, what to be combined and how to combine them.

In fact, the two issues are two sides of the same problem. If the single biometrics really does not meet the general practical requirements, an integration technology has to be carried out. In order to build an efficient integration scheme, the following natures need to be explored.

- A general approach to multiple biometrics feature extraction, representation and integration;
- The inherent connections between different biometrics features and the consequences of different combinations;
- The theoretical foundations and methodologies of the integrated biometrics technology, which includes unified concepts, criteria and rules.

HOW TO INTEGRATE BIOMETRICS?

Integration of multiple cues has been proved beneficial for improving the accuracy of a recognition system. Generally, integration of multiple cues can take place on one of the following three different levels:

a) Abstract level: The output from each module is only a set of possible labels without any confidence associated with the labels.

b) Rank level: The output from each module is a set of possible labels ranked by decreasing confidence values, but the confidence values themselves are not specified.
c) Measurement level: The output from each module is set of possible labels with associated confidence values.

For each different level of integration, we need to find the common feature level for different biometrics, establish appropriate decision integration criteria, which can make a more reliable decision, and propose the performance measurement of multiple biometrics systems.

PERFORMANCE MEASUREMENTS

The performance measurements of biometrics-based systems are mainly defined by the security level and two measures are commonly used for characterizing this value: *FAR* (False Acceptance Rate) and *FRR* (False Rejection Rate). *FAR* and *FRR* depend on the security level, and are strictly related to each other. A few of papers have addressed the problem for single biometrics-based systems. The statistical expressions for calculating theoretically the *FAR* and *FRR* for multiple biometrics-based systems should be derived, and the mathematical relation between them should be explored. The evaluation of performances for such systems is developed so that such information allows the developers to fine-tune the system to optimize performance.

OPTIMUM FEATURE SET

It is possible to demonstrate that the system performance does not always improve by using larger number of features, and by adding features which lack discriminatory power and which are highly correlated. We need to add significant information from multiple biometrics but avoid to introduce error sources. Various biometrics have various features for identification and verification of individual. These features have different values for different application environments and have different effects for the performance. The analysis of the various features of different biometrics, extraction of the common feature level, such as scores, and exploration of the inherent relationship between different biometrics are important issues for integration of multiple biometrics for more practical applications.

INTEGRATION CRITERIA

The goal of biometrics-based system consists of deciding whether these characteristics belong to the same person or not. The decision criteria, including acceptance rule and rejection rule, are the fundamental problem for both identification and verification systems. The notation is also basic for the development of the systematic theory for integration of multiple biometrics. The new types of integration techniques for multiple biometrics provide the solutions. A number of identity fusion methods, including Bayesian fusion, Dempster-Shafer theory, Generalized Evidence Processing and fuzzy logic based methods, can be used to develop the integration technologies. In a decision fusion scheme on measurement level it has been proposed to improve the identification performance by integrating faces and fingerprints.

Other integration schemes can also been developed. For example, since wavelet transforms offer the promise of compact representation and efficient detection of image components that match the wave-shape of the chosen wavelet, a wavelet-based integration scheme to combine multiple biometrics features may provide integrated biometrics solutions to improve system accuracy and efficiency.

15.4 VLSI Biometrics

VLSI CONCEPTS

Integrated circuits (IC) have been used widely since 1960s. IC is much smaller and consumes less power than the discrete components and also much easier to design and manufacture and is more reliable than discrete system as well. These features make it possible to develop special-purpose systems that are more efficient than general-purpose computers for the task at hand.

VLSI means very large-scale integrated circuit. From the emerging of first microcomputer in the beginning of 1970s, VLSI has been playing a great important role in the computer industry. With the development of 80x86, Pentium II, and Pentium III, it is impossible without VLSI.

The growing sophistication of applications continually pushes the design and manufacturing of integrated circuits and electronic systems to new levels of complexity. Almost any complex digital system built today is a mixture of hardware and software. Software is almost always run on standard central processing units (CPUs), such as microprocessor chips. Additional hardware may be added to the system to increase performance or reduce size. Even the simplest system often has some custom hardware---many personal computers, for instance, have special hardware to speed up graphics and sound operations.

IC has much smaller size, higher speed and lesser power consumption. These advantages of IC translate into advantages on the system level:

- **Smaller physical size.** Smallness is often an advantage in itself---consider portable televisions or handheld cellular telephones.
- **Lower power consumption.** Replacing a handful of standard parts with a single chip reduces total power consumption. Reducing power consumption has a ripple effect on the rest of the system: a smaller, cheaper power supply can be used; since less power consumption means less heat, a fan may no longer be necessary; a simpler cabinet with less shielding for electromagnetic shielding may be feasible, too.
- **Reduced cost.** Reducing the number of components, the power supply requirements, cabinet costs, and so on, will inevitably reduce system cost. The ripple effect of integration is so much that the cost of a system built from custom ICs can be less, even though the individual ICs cost more than the standard parts they replace.

In the 1960's Gordon Moore, an industry pioneer, predicted that the number of transistors that could be manufactured on a chip would grow exponentially. His prediction, now known as Moore's Law, was remarkably prescient. Over the past thirty years, the number of transistors per chip has doubled once a year.

In the early years of IC business, companies focused on building large quantities of a few standard parts. But the industrial trend, however, is to make available a wider variety of IC. The greater diversity of chips includes:

- More specialized standard parts. In the 1960's, standard parts were logic gates; in the 1970s they were LSI components. Today, standard parts include fairly specialized components: communication network interfaces, graphics accelerators, floating point processors, and so on. All these parts are more specialized than microprocessors but are used in enough volume that designing special-purpose chips is worth the effort. In fact, putting a complex, high-performance function on a single chip often makes other applications possible---for example, single-chip floating point processors make high-speed numeric computation available on even inexpensive personal computer.
- Application-specific integrated circuits (ASICs). Rather than build a system out of standard parts, designers can now create a single chip for their particular application. Because the chip is specialized, the functions of several standard parts can often be squeezed into a single chip, reducing system size, power, heat, and cost. Application-specific ICs are possible because of computer tools that help human design chips much more quickly.

BIOMETRICS IMPLEMENTATION BY VLSI

Recently, biometrics computing has emerged as a new branch of technology with a broad range of applications, including medical diagnosis and authentication and identification of individuals for various purposes. Although different biometrics techniques have been proposed or under development, a common, critical issue that has not been adequately addressed is the efficiency of biometrics processing. Since many biometrics applications require large amount of time to process images and to recognize features and characteristic patterns, real-time and fast computation is crucial for achieving the specified performance. One of effective methods is to utilize the VLSI technology, and setup application-specific system using ASIC design.

In fact, many industrial products based on VLSI have been developed. For example, please check the following Veridicom FPS110 Silicon Fingerprint Sensor (Figure 15-2).

Veridicom's silicon-based, solid-state FPS110 fingerprint sensor creates a 90,000-byte, 500-dpi image of the fingerprint. An on-board ADC converts the image to 8-bit digital values which can be read by a microprocessor via an 8-bit bi-directional bus. So it can be widely integrated in many fingerprint-based user authentication applications.

In addition to the FPS110 sensor, Veridicom offers the iTouch Peripheral Fingerprint Images built on the FPS110 sensor. Through a parallel or USB interface, the iTouch imager transmits the fingerprint image to the host computer.

VLSI DESIGN METHODOLOGIES

The abbreviation VLSI stands for Very Large Scale Integration, which refers to integrated circuits that contain more than 10^5 transistors (in current-day technologies, circuits of 10^7 transistors can already be produced). The circuits designed may be general-purpose integrated circuits such as microprocessors, digital signal processors, and memories. They are characterized by a wide range of applications in which they can used. They may also be application-specific integrated circuits (ASICs) which are designed for a narrow range of applications, such as biometrics system.

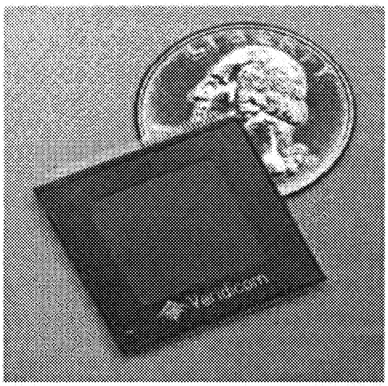


Figure 15-2. Veridicom silicon fingerprint sensor and imager.

Designing such a circuit is a difficult task. A first requirement is, of course, that a given specification is realized. Besides this, there are different entities that one would like to optimize. These entities can often not be optimized simultaneously (one can only improve one entity at the expense of one or more others). The most important entities are minimization of the chip, speed, power dissipation, design time and testability.

To make use of the flood of transistors given to us by Moore's law, we must design large, complex chips quickly. The obstacle to make large chips work correctly is complexity---many interesting ideas for chips have died in the swamp of details that must be made correct before the chip actually works. Integrated circuit design is hard because designers must juggle several different problems:

- Multiple levels of abstraction. IC design requires refining an idea through many levels of detail. Starting from a specification of what the chip must do, the designer must create an architecture which performs the required function, expand the architecture into logic design, and further expand the logic design into a layout. The specification-to-layout design process is a lot of work.
- Multiple and conflicting costs. In addition to drawing a design through many levels of detail, the designer must also take into account costs—not dollar costs, but criteria by which the quality of design is judged. One critical cost is the speed at which the chip runs. Two architectures which execute the

same function (multiplication, for example) may run at very different speeds. We will see that chip area is another critical design cost: the cost of manufacturing a chip is exponentially related to its area. Furthermore, if multiple cost criteria—such as area and speed requirements—must be satisfied, many design decisions will improve one cost metric at the expense of the other. Design is dominated by the process of balancing conflicting constraints.

- Short design time. In an ideal world, a designer would have time to contemplate the effect of design decision. We do not, however, live in an ideal world. Chips, which appear too late, may make little or no money because competitors have snatched market share. Therefore, designer are under pressure to design chips as quickly as possible. Design time is especially tight in application-specific IC design, where only a few weeks may be available to turn a concept into a working ASIC.

Designers have developed two techniques to eliminate unnecessary details: hierarchical design and design abstraction. Designers also make liberal use of computer-aided design tools to analyze and synthesize the design. Hierarchical design is commonly used in programming. This technique is known as divide-and-conquer---the procedure's complexity is conquered by recursively breaking it down into manageable pieces. Chip designers divide and conquer by breaking the chip into a hierarchy of components. Design abstraction is much less common in programming but is critical to hardware system design. Programmers typically deal exclusively with program statements and procedures. Hardware designers use multiple levels of design abstraction to meet performance goals for very large designs.

The only realistic way to design chips given performance and design time constraints is to automate the design process, using computer-aided design (CAD) tools, which automate parts of design process. Using computers to automate design, when done correctly, actually helps us solve all three problems: dealing with multiple levels of abstraction is easier when you are not absorbed in the details of a particular design step; computer programs, because they are more methodical, can do a better job of analyzing cost tradeoffs; and, when given a well-defined task, computers can work much more quickly than humans.

INDEX

A

B

C

D

E

F

G

H

P

R

S

Automated Biometrics: Technologies and Systems

This book systematically introduces the relative biometric technologies and explores how to design the corresponding systems with in-depth discussion. Two kinds of biometric systems, based on physical characterestics (such as fingerprint, palmprint, face and iris) and behavioral characterestics (like voice, signature and gesture), are investigated after an overview of the fundamental biometric technologies. Engineering applications of biometrics to personal authentication and Chinese medicine are covered. The book can be used as a text book or reference for graduate or senior undergraduate courses on the subject. It can also be used by researchers in the corresponding fields.

David Zhang graduated in computer science from Peking University in 1974 and received his MSc and PhD degrees in computer science and engineering from Harbin Institute of Technology (HIT) in 1983 and 1985, respectively. From 1986 to 1988, he was a postdoctoral fellow at Tsinghua University and became an associate professor at Academia Sinica, Beijing, China. In 1988, he joined the University of Windsor, Ontario, Canada, as a visiting research fellow in electrical engineering. He received his second PhD in electrical and computer engineering at University of Waterloo, Ontario, Canada, in 1994. From 1995, He has been an associate professor in City University of Hong Kong and Hong Kong Polytechnic University. Currently, he is a founder and director of both Biometrics Technology Centres supported by UGC/CRC, Hong Kong Government, and National Natural Scientific Foundation (NSFC) of China, respectively. He is also a Guest Professor and Supervisor of PhD in HIT. He is a Founder and Editor-in-Chief, *International Journal of Image and Graphics*, and an Associate Editor, *Pattern Recognition* and *International Journal of Pattern Recognition and Artificial Intelligence*. He has been listed in 1999 Marquis Who's Who in the World (16th). His research interests include automated biometrics-based identification, neural systems and applications, and parallel computing for image processing & pattern recognition. So far, he has published over 150 papers including four books around his research areas and given over five keynotes or invited talks or tutorial lectures, as well as served as program /organizing committee members and session co-chairs at international conferences in recent years. In addition, he has developed some applied systems and received several recognisable project awards. Dr. Zhang is a senior member of the IEEE.