Chapter 1
An Introduction to Biometric Data Analysis

Biometrics is a fascinating field. What other area of science or engineering combines aspects of biology, statistics, forensics, human behavior, design, privacy and security - and also spans everything from the simple door lock to huge government systems? This diversity is compounded by the wide biological and behavioral variation in people, making it an intriguing challenge to evaluate, configure and operate a biometric installation. However, regardless of the biometric type, the fundamentals of how match decisions are made are common to all biometrics. These unifying features allow an introduction and discussion of biometrics through a common framework.

This chapter provides a snapshot of biometric systems and analysis techniques. It previews the contents of the rest of the book, providing a concise introduction to all the main topics. The goals of this chapter are to:

- Introduce the history of biometric science (Sect. 1.2).
- Discuss how biometrics fit into an identity management framework (Sect. 1.3).
- Discuss desirable aspects to consider when choosing a biometric (Sect. 1.4).
- List the components and forms of biometric data (Sect. 1.5).
- Provide an overview of biometric systems (Sect. 1.6).
- Give high level descriptions of the major biometric graphs (Sect. 1.7).
- Introduce aspects of biometric privacy (Sect. 1.8).

1.1 Introduction

Biometrics is the use of distinctive biological or behavioral characteristics to identify people. Automated biometric recognition mimics, through a combination of hardware and pattern recognition algorithms, a fundamental human attribute - to distinguish and recognize other people as individual and unique. There is a long history of using distinguishing marks for identification. From hand-prints on cave walls and hand-written signatures on manuscripts, to the direct measurement of head

dimensions and the unique patterns of Morse operators, there has long been a desire to use and measure biometric identity.

This book is concerned with the automation of this process, where a computerized system can make a determination on identity without human supervision. This field is called biometrics and its path from research to wide-ranging adoption is still unfolding. Huge advances have been made in every aspect of biometrics since the first automated matching systems in the late-1970s, from the underlying technology through to the increasing number of biometrics in everyday use.

Systems incorporating biometrics now span the globe in a variety of applications from law enforcement and passport control to laptops. The potential impact of biometrics is enormous, as demonstrated by the numerous areas of authentication using traditional security, such as keys, passwords and cards. Biometrics will be an increasingly integral part of the solution to challenges from rising global levels of identity theft to the demand for convenience and security. Areas as diverse as health care, travel, call centers, banking, Internet security, cars, consumer electronics, voting, proof of identity and payment systems are all starting to see the benefits of the application of biometrics.

One principal advantage is the increased confidence it brings to transactions. This happens because when a biometric is used, the person must have been physically present during the authentication. Unlike a password or keys, biometrics cannot be given to another person. This is often a principal driver for its adoption in the government and commercial spheres. The other significant benefit is convenience - often it is simply faster and easier to use a biometric sensor. It is this factor that usually drives the growth in the consumer area. Both security and convenience can lead to reduced costs, since riskier transactions can be undertaken with less supervision, and transactions are less like to fail because of forgetfulness.

This book is the result of over 15 years of analysis of government and commercial biometric systems, encompassing many different real-world evaluations from a wide variety of activities. Through this experience, techniques and insights have been developed to analyze biometric systems, which should contribute to making biometrics operate better and more accurately.

It has been a necessary part of our work to ensure that those who need to implement or manage biometric systems understand how they make verification decisions and how to interpret performance graphs. No matter if you are a newcomer or an experienced practitioner, we hope this book will broaden your understanding of biometric systems and inspire you to explore this exciting area further. In this chapter, the construction and operation of biometric systems is introduced.

1.2 A Brief History of Automated Biometric Identification

The manual identification of people before the introduction of computers is extensively discussed in a number of other books [4, 7, 16], but the subsequent development of biometrics is the fascinating story of the moment. Serious biometric re-

search began in the 1960s, the techniques were developed and refined during the 1970s and 1980s, and the field became increasingly commercialized from the mid-1990s onwards.

1.2.1 The Hands: Fingers, Palms and Hands

The initial demand for automated biometric analysis was driven by the need to process *fingerprints* in the criminal justice system. This task historically required an experienced practitioner to examine fingerprint images and classify them into different types based on the overall pattern of ridges and valleys. These were then manually compared to establish whether the suspect had been seen previously or if there was a match to a crime scene. When the number of fingerprints held by the police grew massively in the early 20th century, the manual process of classifying and searching for fingerprints became prohibitively costly and error prone.

It was not until the early 1960's, and the advent of powerful computers, that the first experiments in the biometric matching of fingerprints were conducted at the U.S. National Bureau of Standards [7]. In 1979 the first working prototype of a fingerprint searching system was tested at the U.S. Federal Bureau of Investigation. By 1983 Automated Fingerprint Identification Systems (AFIS) were in routine use, and within three years they were being adopted globally. This set the scene for the large scale growth of the biometrics industry.

Palm-prints are now also used in AFIS systems. The first reported commercial palm-print system came from Hungary in 1994, and by 1997 similar processes were being built into other AFIS systems.

In the mid-90s, experts developed fingerprint sensors that were cheap enough and small enough to be used for access control and computer login. At first these were optical systems using similar technology to AFIS. Around this time it was accidentally discovered that fingerprints could be read straight from their capacitive effects on silicon wafers. This led to a drastic reduction in size and cost of sensors, and consequently the market for sensors for laptops and other consumer products experienced enormous growth.

Advances in Fingerprint Scanners

In the mid-90s a Bell Labs employee was, according to legend, experimenting with a DRAM that was getting hot. When he put his finger on to check the heat he discovered that he had flipped the bits in the memory registers where his fingerprint ridges had been. The first company to commercialize this discovery was Veridicom. The next significant innovation was when the industry moved from large "placement" sensors at $50 each to swipe sensors of around at $5 each. Future developments should continue to reduce swipe sensor costs to $2 to $3 whilst not compromising reliability. Swipe sensors also enable dual functionality as they can be used as part of the user interface, as well as providing security on mobile phones and laptops.[a]

[a] Thanks to Brett Minifie from Hewlett-Packard Australia for this information.

Technology	Approximate date of an early major paper or relevant patent	Approximate date of an early commercial implementation
Fingerprint (AFIS)	1962 paper	1979, 1985 Identix
Retina	1978	1984, EyeDentify, Inc.
Speaker	1963 paper, Pruzansky	1976, Texas Instruments
Face	1965, Helen Chan and Charles Bisson	1996 Cognitec,ZN, Identics
3D Face	1992 G. Gordon	2001, A4 Vision
Hand	mid-1960s	1986, IR Recognition Systems
Iris	1987 Patent, John Daugman	1995, Iridian
Palm	1994	1997
Vascular	1992, Dr K. Shimizu	early-2000s
Finger Vein	2002	2004
Keystroke	1986 Patent J. Garcia.	2002 iMagic

Table 1.1 History of some biometric developments. The events listed relate to the first significant developments in research and commercial adoption of automated identification.

An early example of a successful commercial biometric used for access control was *hand recognition*, based on the geometry of the hand. The original technology was developed in the mid-1960s, and started to enjoy wide commercial success in 1986. At the 1996 Atlanta Olympic Games it was used as part of the security access control for athletes. Compared with its main competitors in access control – fingerprint and iris recognition – this technology requires large, bulky readers and, as a consequence, now has a declining market share.

The individual pattern of veins in the fingers, wrist, hands or face can be mapped using infrared cameras. This is known as vascular recognition. Because the sensor is non-contact, the readers can be made quite robust, and it is more difficult to covertly obtain these patterns than many other biometrics. Since 2005, the popularity of finger and palm vein scanners has been growing, particularly in Japan where they are being used in automated teller machines.

Art, Literature and Biometrics

Prehistoric artists used hand-prints in cave paintings, perhaps as a 'signature'. They might be considered the earliest example of a biometric identifier. Such hand-prints are found in at Lascaux in south western France and date from at least the Upper Paleolithic period (approximately 16,000 years ago).

Today's artists also are looking to mark their work. Recently, the Australian artist ProHart used DNA from a cheek swab and mixed it with paint to uniquely identify his paintings. Similar DNA labeling techniques, coining the term "spit label", have been used to create an indigenous 'authenticity label' to protect artists from fraudulent copying [3].

Biometric techniques can also be used to uncover forgeries. Famous examples include the fake painting "Skating in Holland" 1890-1900, which is signed by Johan Barthold Jongkind. However, the signature on the painting is not the same as the artist's real signature. As well, signal processing techniques can be used to pick up the inherent style of an artist that is embedded in the texture of the images. When the painting "Virgin and Child with Saints" by Pietro Perugino was analyzed, the work appeared to be from at least four different artists.

In literary works, Shakespeare's authorship of some plays has been questioned since as far back as the 18th century. Stylistic analysis of plays (called stylometry) known to have been written by Shakespeare provides a behavioral biometric signature of the way he used words. The analysis suggests that some works were collaborations (in particular, the plays "The Two Noble Kinsmen" and "Henry VIII" which were co-written with John Fletcher).

1.2.2 The Head: Face, Voice and Eyes

Because people commonly use faces for establishing identity, it is perhaps the most natural biometric for authentication. However, it is also one of the most challenging

for a computer. The earliest work published in the automation of *face recognition* was in 1965 by Helen Chan and Charles Bisson. A system was developed by a small research company (called Panoramic) that looked at manually marked points on a face, and used the marked points to search a database of several thousand people [5]. Automated face recognition system research was undertaken by Kanade in 1977 [12], but it did not become a significant research area until the 1990s. The seminal papers that gave rise to fully automated face recognition came first in the late-80s and early-90s [13, 14, 18], and these showed that faces could be represented using only a few hundred parameters.

The first modern commercial face recognition systems began to appear around the mid-90s. Amongst these early face recognition companies were Cognitec, ZN, Viisage Technology and Visionics Corporation. By 2006, L1-Systems had been formed through the merging of ZN, Viisage Technology and Visionics Corporation.

Advances in technology continue to provide significant performance improvements. One of the most significant was the use of the skin texture analysis, introduced in 2004. This used the high-frequency information encoded in the blemishes and skin pores to boost recognition accuracy for high-resolution images.

In 1963 a paper was published on "Pattern matching procedure for automatic talker recognition" [21]. The first early *speaker verification system* subsequently came out of AT&T's research labs in 1974 [11] and in 1976 Texas Instruments built a prototype system that was tested by the U.S. military. Through the 1980s and 1990s, steady research progress was made. This meant systems could operate under text-independent and text-dependent modes and increased the robustness of the processing algorithms. One of the seminal papers for the analysis of biometric data was published in 1998 on the classification of user types in speaker verification, known as Doddington's Zoo [9]. Examples of the user types identified included goats (people who had trouble being recognized) and wolves (people who were naturally able to sound like others). This is the original basis of research and analysis techniques presented on individual evaluation in Chap. 8.

A biometric which has started to be adopted widely over the past 10 years is the unique pattern found on the iris - the colored muscle in the center of eye. Two ophthalmologists, Dr. Leonard Flom and Dr. Aran Safir, patented the idea in 1987 and worked with Prof. John Daugman to develop an *iris recognition* algorithm. By 1994 Daugman had received patents for his algorithms. Successful tests for the U.S. Defense Nuclear Agency were completed the following year along with a commercial implementation. Iris recognition was proposed in 2003 as a part of a potential British national identity card.

In contrast to the iris, which is the front of the eye, *retina recognition* uses of the pattern of the blood vessels at the back of the eye. Its first commercial implementations were in the mid-1970s. The equipment for this was large and expensive and there were difficulties identifying people wearing glasses.

1.2.3 Other Biometrics

Most people's first exposure to the concept of a biometric identifier is through a handwritten signature. In many cases, a signature by itself is not very distinctive, can be highly variable and is relatively easy to copy. Because of this, research has concentrated on the use of dynamic features for online recognition, such as pressure and acceleration.

Keystroke dynamics were first recognized as a biometric identifier during World War II. Allied telegraph operators could identify other operators, friendly and enemy, through a keying rhythm which was called the "The Fist of the Sender". Even without decoding messages, this allowed German troop movements to be monitored, because the German telegraph operators were usually attached to a particular fighting unit. Modern keystroke dynamics uses the timing difference between keystrokes and looks for idiosyncrasies in the use of keys (e.g. how long the typist holds down the shift key or uses the control keys). Commercial keystroke dynamic systems were available in the early-2000s, but they have not been widely adopted because they lack accuracy and require long training times. However, more sophisticated systems have been recently proposed that claim better performance.

Even the way people walk, their gait, has been used as an identifier. Research on this has been prompted by the need to identify people from surveillance cameras at a distance in poor light, without a face clearly visible. This is a difficult task since a person's gait is highly variable and clothing obscures fine differences.

The field of biometrics continues to develop new methodologies for recognition. For instance, in 2008 a patent was even granted for "Method and system for smell-print recognition biometrics on a smart-card".

1.2.4 Post September 11, 2001

The terrorist attacks in the U.S. on September 11, 2001 focused enormous attention on biometrics, and on its ability help secure the U.S. and other countries against terrorist threats by making it more difficult to use fake identities. The 9-11 Commission report, produced in the wake of the attacks, suggested the introduction of biometrics at national borders to prevent unwanted foreign nationals, particularly people traveling under a false identity, from entering the country. The UN body controlling passport standards, the International Civilian Aeronautical Organization (ICAO), for some time had been looking at a new standard for electronically-readable passports that included biometrics. After 2001, this work took on renewed urgency and by early 2008 many countries had implemented passports that include biometrics.

The mandatory biometric in the passport standard (known as IACO/MRTD Doc 9303) was a facial biometric token, as this was most compatible with existing passport-issuing infrastructure and was the most widely acceptable. Fingerprint and iris recognition are optional biometrics that could be included on the passport and have been used by a variety of countries.

The result of the terrorist attacks on the biometrics industry was mainly long term. In the short term, biometric trials were run and discussion papers produced, but many large scale implementations took years to appear. Significantly, however, more money was diverted into research, and this focus is now resulting in substantially improved accuracy and usability.

Legitimate Fake Identities

An interesting side effect of the adoption of biometrics has been the difficulty for covert operatives, or those in witness protection, to obtain false identities without being detected by the issuing agencies. In essence, the person's biometric ties them to their true identity, and an alert may be raised to the operator when they try to obtain an ID under a different name. Under previous circumstances, these agencies would have no idea a person had multiple identities, legitimate or otherwise.

Biometrics also had problems with public perception in the post 9-11 world. In particular, it is commonly being seen as a big brother technology, potentially leading to the invasion of privacy. Civil liberty and consumer groups have been concerned about the implications of large biometric databases and the potential for misuse of this data through its sale to third parties or its use for unauthorized purposes (such as cross matching with other biometric databases). These concerns were first raised in a significant public way when face recognition was used to scan the crowds at the 2001 football Super Bowl in Tampa, Florida.

Substantial improvements to sensors, algorithms and techniques are likely over the next 10 years. In particular, we can expect continuing advances in the biological study of how brains make complex decisions, and the development of new nanotechnology materials which will transform biometric technology. Regardless of these advances, the fundamentals of biometric data analysis discussed in this book are unlikely to change.

1.3 Identity and Risk Management

Biometrics is often used to control the risk of a security breach and to facilitate convenient transactions. In this context, biometrics is part of the risk and identity management process. Identity management encompasses all phases of the process of dealing with information relating to an individual's identity. This involves regulating and controlling the places where identity information is used or processed, extending from the initial identity creation and verification, through their use and potential reissue, to the final removal of the identity.

It is particularly important in any large biometric system that the full identity life-cycle is supported. A biometric alone will never add security to a system that

does not have good identity controls around the initial, or subsequent, enrollment of individuals.

Where anyone can self-enroll a biometric with no check or audit, system circumvention is easy, regardless of the strength of the biometric control. This is because there is no way to ensure the correct identity is bound to the enrolled biometric. The binding of an identity to a biometric is achieved by authenticating the supporting documentation, such as birth certificates or passports. This is commonly known as proof of identity.

Extreme Biometric Makeovers

Several cases have been reported of people changing their biometrics to create new identities - usually with disastrous consequences. In 1997, Amado Carrillo Fuentes, one of the major drug traffickers in Mexico, underwent facial plastic surgery and liposuction to change his appearance. He died as a result of complications. This is perhaps the first, and hopefully last, death of someone specifically trying to create new biometrics. In 2007, a Mexican doctor was arrested in the U.S. on suspicion of trying to change a drug dealer's fingerprints by surgically replacing them with skin from the bottom of the feet. It was reported that the operation left the drug dealer's fingers barely usable.[a]

[a] http://www.iht.com/articles/ap/2007/05/11/america/NA-GEN-US-Mexico-Fingerprint-Removal-Charges.php

Understanding biometric statistics and their impact on risk management can be complex, particularly for large public-facing systems. This is because the performance of biometrics is highly dependent on the characteristics of customers and the environment where the system is being used. For instance, with some biometric systems, a percentage of users always will have difficulty with the biometric sensor for biological, cultural or behavioral reasons. Techniques can be implemented to manage this but care must be taken to ensure customers do not feel disenfranchised, and that the overall level of security is not diminished. Consequently, it is important to undertake a proper risk assessment, incorporating a detailed understanding of the risk, threats and opportunities provided by biometrics, and follow up with regular biometric audits.

1.4 Desirable Biometric Attributes

There is a huge diversity of biometrics types, and each biometric has its own particular strengths, depending on the application. Underlying this diversity are some general attributes that are desirable for good biometric operation [16]. The importance of each of these factors depends on the results required for a particular project.

Where biometric systems are being compared as part of a formal selection process, each attribute can be weighted and assessed independently.

- **Distinctiveness**: A biometric should have features that allow high levels of discrimination in selecting any particular individual while rejecting everyone else. The larger the number of people to be distinguished, the more important this factor becomes.
- **Stability**: Age, and perhaps accident or disease, will change all biometrics over a period of time. Biometrics also may be altered by clothes or as a result of skin plasticity (e.g. smiling). A biometric should preserve enough features so that these changes will have a minimal effect on the system's ability to discriminate. Where re-enrollment can be simply or easily achieved, or where reissue over shorter durations is legally required, stability may be of less significance.
- **Scalability**: A biometric should be capable of being processed efficiently, both at acquisition time and when it is searched in a database for identification systems. Scalability issues may be less of a concern for verification-based access control systems than for large identification systems.
- **Usability**: A major selling feature in the adoption of biometrics is convenience. If a biometric is difficult or slow to use, it probably won't be adopted. Ideally, the ergonomics of the sensor will make it so simple to use that the authentication will barely be noticed. Usability is an especially important factor for people with disabilities (e.g. people who are vision or mobility impaired). This is particularly crucial in places where a biometric will be used frequently, such as for access control. Although, in some cases, mainly where a biometric is used for surveillance, this may not be relevant.
- **Inclusiveness**: An extremely high proportion of the population should be measurable, particularly for large-scale identity systems. A biometric which excludes some users causes additional complexities in managing security and has an obvious impact on usability. If the biometric is primarily for convenience rather than security, and alternatives are available, the tolerance for people who cannot use the system is increased.
- **Insensitivity**: Changes in the external environment (e.g. lighting, temperature) within reasonable boundaries should not cause system failures. Controlled indoor situations, such as airports, may be less affected than a biometric sensor used on an external door.
- **Vulnerability**: It should be difficult to create a fake prosthetic biometric (known as spoofing), or to steal and use a detached one. In some places that are highly supervised and controlled, vulnerabilities can be mitigated through policies and human monitoring.
- **Privacy**: Ideally the permission of the owner of a biometric should need to be sought before acquisition. The data should be stored encrypted.
- **Maintenance**: Sensor wear and tear, or residue build-up on the sensor surface, should be minimized. This often is achieved with non-contact sensors such as cameras, but any sensor that requires a physical touch is likely to suffer from maintenance issues.

- **Health**: Physical harm or pain should be avoided during biometric acquisition (even for those who have a medical condition, such as arthritis). Non-contact systems are much less likely to make an impact on the health of users.
- **Quality**: Obtaining a good quality sample should ideally be easy for the user. High quality samples are usually very important to ensure accurate matching results.
- **Integration**: The biometric should be capable of being used in conjunction with other authentication mechanisms, such as smart-cards or passwords
- **Cost**: The cost of the biometric system should be in proportion to the benefit. These benefits might include convenience, enhanced security, reduced cost for employing human operators or reduced cost from token loss. The cost needs to be proportional to the volume of systems that will be sold.

Biometric Attribute	(a) Laptop Sensor	(b) Passport Issuing	(c) Covert Surveillance
Distinctiveness	High	High	High
Stability	Med	High	High
Scalability	Low	High	High
Usability	High	Med	-
Inclusiveness	Med	High	High
Insensitivity	High	High	Med
Vulnerability	High	High	Med
Privacy	Med	High	High
Maintenance	High	High	Low
Health	High	High	-
Quality	High	High	High
Integration	Med	High	Low
Cost Sensitivity	High	Low	Med

Table 1.2 Biometric attributes and some example levels of importance for a) a high volume sensor used for laptop access b) a hypothetical passport issuing system using biometrics to detect fraud and c) a covert face recognition surveillance system (the covert system is non-contact so issues of usability and health are not applicable). These levels are for illustrative purposes. For real systems there may be considerable variation in the actual requirements.

1.5 Biometric Data

Biometric data is special because it is intrinsically linked to our internal concept of identity in a way that other forms of proof of identity, such as passwords and keys, are not. One of the reasons is that biometric data contains something unique and semi-permanent about the individual. Biometric data is generally represented or stored in one of three forms: *raw*, *token* or *template*.

1.5.1 Raw Data

The raw biometric information (known as the *biometric sample*) is data gathered directly from the sensor before any processing has been carried out. There is a huge range of biometric acquisition techniques - examples include camera images, infrared images, range geometry, sound recordings, chemical analysis, full motion video recordings, keystroke logs and friction, pen motion (signature) (see Fig. 1.1). Each sample mechanism has unique properties and challenges. However, there is a core set of biometrics which is more widely used in commercial applications. These common biometrics include face, fingerprint, palm, hand print, iris, speaker verification and vein matching.

1.5.2 Token Data

A token is representation of the raw data that has had some minimal amount of processing applied. For passports, the ICAO definition of the facial token to be stored on the passport chip is a cropped and scaled representation of the actual image. This is processed by the chosen matching algorithm. The reason for storing the image, rather than extracted features, is that any recognition algorithm can be used to process the 'raw' data and advances in matching are not precluded. This is known as *template interoperability*.

Another good reason for using a token is that advances in algorithms may discover new ways of extracting distinctive features from the original biometric sample. Using a token can allow seamless upgrading of algorithms.

1.5.3 Template Data

The piece of biometric data common to all biometric systems is a *template*. A template is the refined, processed and stored representation of the distinguishing characteristics of a particular individual. The template is the data that gets stored during an enrollment and which later will be used for matching.

Because of variations in the way a biometric sample is captured, two templates from the same biometric will never be identical. This is the origin of the probabilistic nature of biometrics, as the matching process can only give a decision confidence, not an absolute assurance (see Chap. 2 for more details).

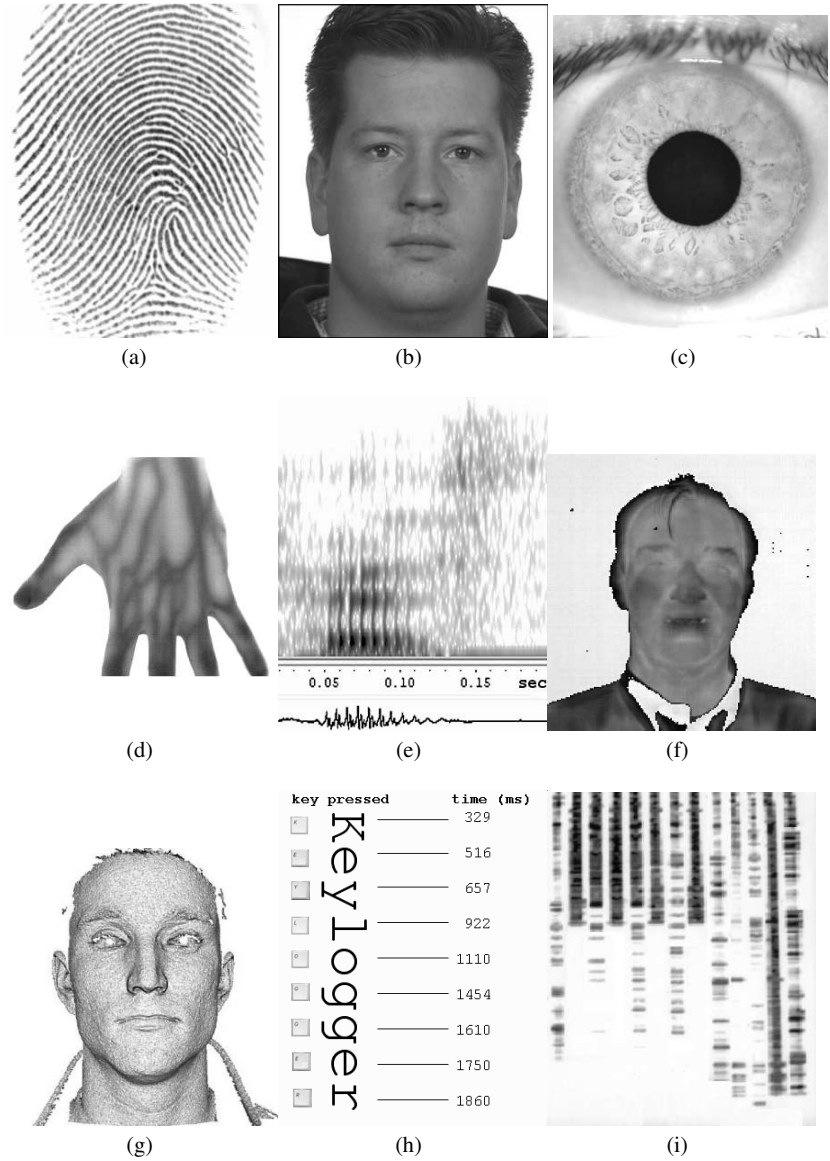


Fig. 1.1 Examples of the diversity of biometric samples: (a) fingerprint [15], (b) face [19], (c) iris, (d) vein [23] (e) voice (spectrogram) , (f) infrared face [6] (g) 3D facial geometry [10], (h) typing dynamics, and (i) DNA. Image (b) used with permission J. Phillips [20], (c) used with permission S. Phang, (h) ©2008 IEEE and image (g) used with permission P. Flynn

1.5.4 Metadata

Another source of data that is often captured in a biometric systems is *metadata*. This is data that describes the attributes of either the biometric sample (e.g. wearing glasses), the capture process (e.g. time acquired) or the demographics of the person (e.g. gender and age). The metadata is particularly important for testing and evaluation, as strong correlations between demographic characteristics and matching performance often are found. Chapter 9 examines how this information also can be used to identify groups of problem users in biometric systems.

1.6 Biometric System Overview

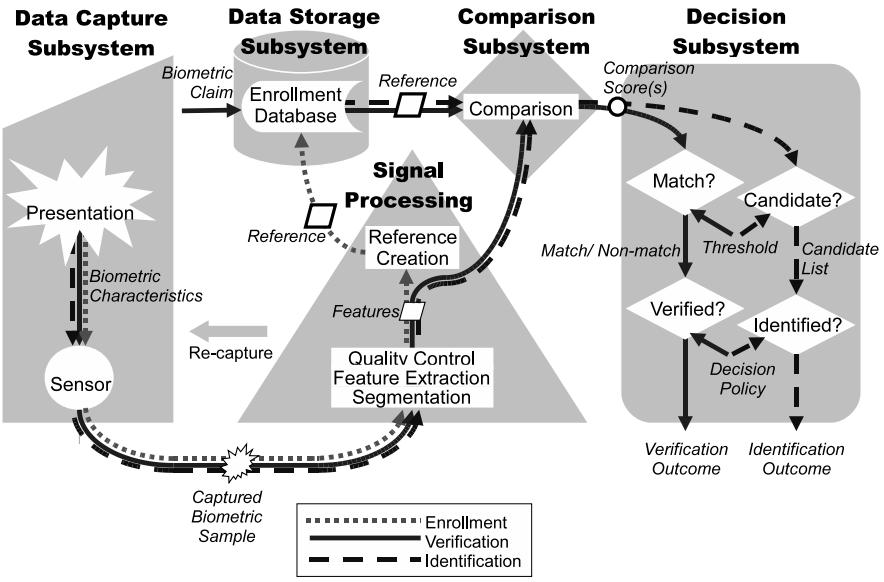


Fig. 1.2 Components of a general biometric system. Used with permission from Tony Mansfield, National Physical Laboratory, UK.

All biometric systems can be described by a general model (see Fig. 1.2). A complete biometric system includes several distinct subsystems: biometric capture, transmission and processing to enhance the biometric features, the storage of the biometric information, the biometric comparison and finally the process to decide, based on this comparison, whether it is the correct individual. Enrollment, verification and identification transactions all share related paths through the subsystems.

The first fundamental stage is *data capture*, which is the capture of the raw biometric information from a real person, which is translated into a digital sample. For many biometrics this involves a transformation from analogue to digital. Common forms of data capture equipment include digital cameras (for face recognition), optical and capacitance fingerprint scanners (fingerprint), and microphones (speaker verification). Each transformation from the original signal into the digital representation involves the potential addition of noise. Due to variance in the presentation, no two samples will ever be exactly the same.

Once the captured biometric sample is in a digital form, it is transformed using *signal processing* techniques into reference features that are used distinguish the individual. This involves processing the sample to remove noise or unnecessary background and extracting features. In face recognition it may involve finding the eyes, and in fingerprints the minutiae (i.e. ridge endings or bifurcations). *Quality assessment* is also made at this stage (see Sect. 3.3), and where the quality is poor, a re-capture of the data may be necessary. After the features have been extracted and are of sufficient quality, a reference template is created.

At enrollment, a template is created and then the *data is stored* in a database or on a device such as a smart-card. Data may also need to be protected by encryption for both security and privacy reasons.

A biometric algorithm will take the features from the stored reference template, along with the features extracted from the presentation sample, and compare them to generate a score which indicates the likelihood that both are from the same person (comparison subsystem). The output comparison score may come in a variety of forms, such as from zero to one, unbounded, or such that the closer the score is to zero, the more likely the match. This score is the fundamental building block of the analysis techniques presented in this book.

For verification, the comparison score is used to make a *decision* about accepting the person as genuine or rejecting them as an impostor. Alternatively, during identification the match is conducted against two or more enrolled people to produce a candidate list of possible genuine matches. The decision policy of whether to accept or reject them should be based on a sound understanding of the true likelihood of a mistake, and is discussed in detail in Chap. 7.

Two other components not shown in this diagram are the transport subsystem, which is the mechanism by which data is moved securely between the different subsystems, and the administration subsystem, which allows the management of the biometric system including setting system thresholds and administration of templates. Also not shown, but implied by the signal processing and the data capture subsystems, is liveness detection to prevent against the presentation of fake biometrics.

1.6.1 Negative Identification

Biometric systems usually conduct searches of a database in order to determine if an individual is enrolled, and this is known as *positive identification*. The opposite of this situation, *negative identification*, is to confirm that a user is *not* enrolled. In this case, a successful query is defined by the *absence* of any match results. A common usage is to prevent the creation of multiple identities for a single user.

Negative identification reverses the meaning of a 'false match'. For negative identification, a false match leads to a false rejection, whereas for positive identification it leads to a false acceptance. This can be illustrated by considering a passport issuing system. It is necessary to ensure that each person is issued with only one active passport, and this can be achieved by looking for duplicate matches in the list of people who already hold a passport. In this negative identification scenario, a 'false match' means that we incorrectly identify an applicant as an existing holder, and no passport is issued. However, when the passport is used at an automated immigration gate, a false match would mean that the user was in possession of another person's passport.

This also illustrates that the types of vulnerability are different for the two scenarios. For positive identification the impostor must try to look similar to the true passport owner. However, for negative identification an 'impostor' must try not to look like themself.

1.6.2 Common Biometric Processes

The matching process of a biometric can be simplified into two phases: the capture and comparison of the biometric sample, and a decision as to whether to accept or reject the input as authentic. The first part is specific to the biometric type, while handling the decision is largely independent of the actual biometric type (see Fig. 1.3).

1.6.2.1 Biometric Specific Processes

The capture of biometric data and the matching algorithm are specific to the particular type of biometric being used. Each biometric modality has specific requirements about the way the data needs to be processed. Examples of input include one-dimensional audio data, two-dimensional images and three-dimensional geometry.

The matching algorithms need to be tuned to look for the best features to distinguish individuals, while coping with changes introduced due to aging or other variations. This generally requires them to be highly optimized for the type of biometric being matched.

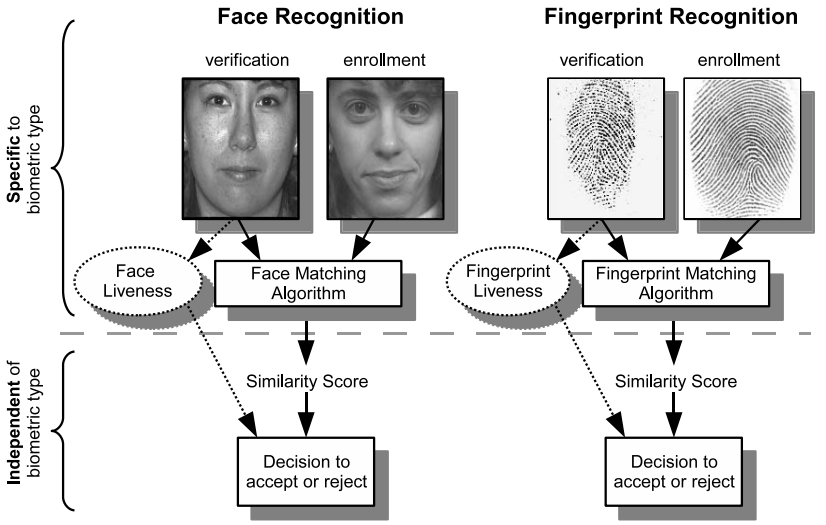


Fig. 1.3 Components of the matching process specific to the biometric type (top) and components of the matching process independent of the biometric type (bottom). Facial images from [17].

1.6.2.2 Biometric Independent Processes

The output of the biometric matching process is a similarity score. Although each different algorithm may have quite different scoring characteristics and ranges, the output represents the common attempt to assign a relative likelihood that it is a particular person and not someone else. This allows scores to be processed in the same way regardless of the biometric type or algorithms, and is the reason it is possible to compare the performance of biometric modalities and technologies despite their differences.

Match scores are either *genuine* matches, which should be high scores, or *impostor* matches, which should be lower scores. A system's performance is based on these scores, and the biometric graphs summarize this information in a useful way.

1.7 System Performance Graphs

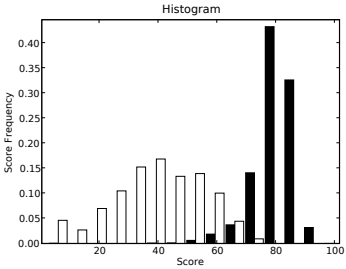
A wide variety of graphs can be used for comparing biometric systems and representing accuracy. Many graphs are simply different ways of displaying the same data to illustrate a particular aspect of performance.

A detailed treatment of these graphs can be found in Chap. 7, while a summary of the most common types is given below.

Score Histogram

Description Plots the frequency of scores for non-matches (white) and matches (black) over the match score range. A good system will have very little overlap between the non-matches and matches.

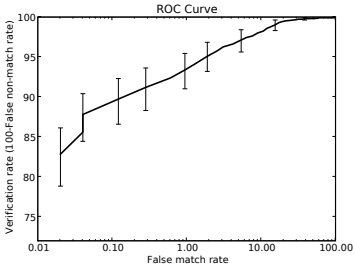
Use Helps in understanding and visualizing algorithm operation, and for setting system thresholds.



ROC Curve (Receiver Operating Characteristic Curve)

Description The ROC Curve shows the trade-off between the rate of correct verification and chance of a false match. A curve from a good system will be located near the top of the graph (high verification rate) for most false match rates. The small bars show the confidence in the accuracy of the graph.

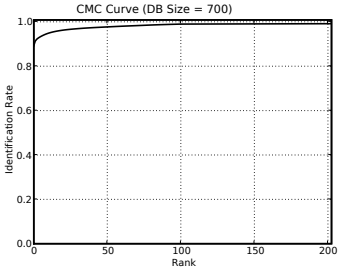
Use For verification statistics, the ROC is commonly used to demonstrate accuracy and to compare systems.



CMC Curve (Cumulative Match Characteristic Curve)

Description Displays the chance of a correct identification within the top ranked match results. A good system will start with a high identification rate for low ranks. The results in a CMC graph are highly dependent on the size of the database used for the test.

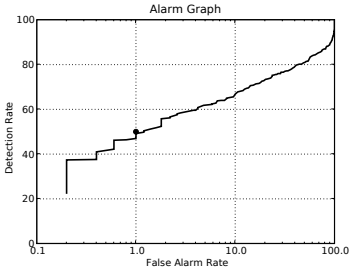
Use For identification systems, to answer questions such as “what is the chance of identifying a fraudster in the top 10 matches returned?”



Alarm Graph

Description For systems required to generate an alarm for any matches over a given threshold, this graph shows the chance of a correct detection compared with the chance of a false alarm. A good system will have a high detection rate for low levels of false alarm. This graph is highly dependent on the size of the watchlist being used.

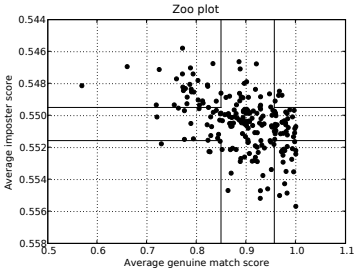
Use For watchlist systems, particularly surveillance, to answer the questions such as "At 1% false alarm rate, what level of detection will I have?"



Zoo Plot

Description Displays how different users perform based on their average match score and their average non-match score. A good system will have few outliers.

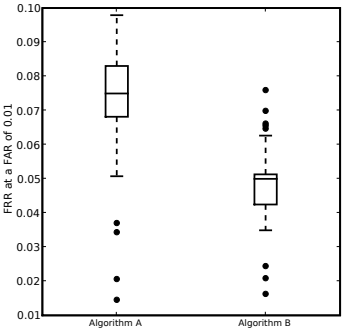
Use To investigate which users, or user groups, are causing more system errors.



Boxplot

Description Provides a method for comparing systems at a particular false accept rate. The boxes and lines show the variation of genuine false rejection rates at this false accept rate for a given algorithm. The points represent reject rate estimates that have been classified as outliers. The lower and narrower the boxes, and the less the outlier points, the better the system.

Use To compare verification systems or different test sets.



1.8 Privacy

One recurring issue around the adoption of biometrics is its potential impact on privacy. Because biometrics are bound to biological or behavioral aspects of personal identity, their use can appear to remove an element of personal control tradition-

ally associated with other security technologies. Conversely, biometrics can help to protect privacy by combating identity fraud. When access to personal data is securely protected using biometrics as part of the authentication process, identity theft is made significantly more difficult.

1.8.1 Privacy Challenges

The central reason for considering biometric data as sensitive with respect to privacy is that your biometrics are not usually 'secret'. Furthermore, they cannot be easily changed, destroyed or declared invalid.

Covertly obtaining a biometric sample may be relatively easy depending on the biometric characteristic of interest. For example, online social network sites contain millions of images of people's faces, identity documents with facial images are used daily, fingerprints are left on almost everything we touch, and even capturing iris information covertly would be far from impossible. Indeed, the very 'public-ness' of the information is part of what makes biometrics so simple and convenient to use. A good biometric system should have some process to check that the 'non-secret' biometric information is not being forged, but also allow for normal variations in environment and use. Many biometric devices have *anti-spoofing* or *liveness* techniques to ensure that a sample is being taken from live human and not simply a prosthetic device. However, the extent and success of the system at validating the authenticity of the sample varies widely, as is discussed in Chap. 12. No systems are truly unbreakable but the aim is to make the cost of breaking the security more than any benefit that would be derived. Advances in biometric sensors, particularly in data quality and resolution, at enrollment and verification are making the task of covert capture increasing difficult.

The stability of a biometric over time is a major factor in its usability. If a biometric exhibits large changes over time, it is likely to have higher matching errors. With the exception of a fake prosthetic device, it takes major surgery to alter most biometrics in order to pass as someone else. An indirect consequence is that once a particular biometric has been compromised it is extremely difficult to replace with another. By comparison, changing a password is simple, if sometimes tedious, and good security actually insists on the continual rotation of passwords. This is particularly problematic in environments where the biometric reader or the network cannot necessarily be trusted, such as for web-based services. When the reader cannot be trusted to have read a live biometric or to have transmitted the correct results, the outcome of a biometric match cannot be trusted either.

1.8.2 Privacy Enhancing Techniques

There are a range of techniques and policies that can be used to enhance privacy.

Clever use of one way functions, or hashing, can be used to encrypt a biometric signature at the point of acquisition. Crucially, these hashes preserve the essential properties of the biometric sample, but the real sample cannot be deduced from it. The hash relies on a set of parameters that are able to be changed or destroyed, and thus can be different for each individual and for each use, allowing the 'biometric' to be revoked. This has been termed *cancelable biometrics,* as the intent is to allow the cancellation of the biometrics through this revocation process. The practicalities of such systems in commercial deployment are still to be fully evaluated as they often involve some trade-off with accuracy.

The storage of biometric data on a token such as a user-held smart-card solves privacy issues, since the users are now in possession and control of their own biometric data. If a card is capable of tamper-proof biometric capture, matching, and cryptography, the issues to do with both biometric revocation and remote validation are largely resolved. This type of advanced manufacturing is not suitable for every type of biometric product, nor is it necessary in all applications. However, it does involve techniques that can be used to address privacy concerns and enhance security.

A general privacy principle revolves around the issue of informed consent. Potential users should know how their data will be used, stored and eventually disposed of, and have the right to opt-out of the biometric authentication where feasible. Policies and procedures should be enacted to provide transparency in system operation. Such policies should address *function creep* which occurs when the system's scope is gradually extended beyond what was originally intended, for instance the cross matching against other biometric databases. The use of strong audit logs helps to enforce policy settings and should provide as much information as practical on who and what has been matched.

1.8.3 Privacy Codes

There is still significant variation in internationally accepted standards relating to privacy and biometrics. One of the first nationally recognized privacy codes specifically for biometrics came into operation in Australia in 2006. This code has three new extensions to existing national privacy principles: Protection, which deals with data storage and transmission of biometrics; Controls, which ensures, amongst other things, the informed consent of users and the right to request the removal of biometric data; and Accountability, which deals with the auditing of biometric systems and privacy impact statements.

> The Biometrics Institute Privacy Code seeks to build upon the National Privacy Principles (NPPs) in a manner that provides the community with the assurance needed to encourage informed and voluntary participation in biometric programs. Biometrics Institute members understand that only by adopting and promoting ethical practices, openness and transparency can these technologies gain widespread acceptance [2].

The European Union has been proactive in its data protection regulations (*Data Protection Directive* 95/46/EC). Its basic principles are a reduction in the processing of personal data, maintaining the highest transparency possible, and ensuring individual control of processing of personal data is as efficient as possible [8].[1] The United Kingdom has established an Information Commissoner's Office as a independent authority set up to protect personal information.[2] Recent debate relevant to biometrics has centered around the use of biometrics in a proposed national identity card and on the use of biometrics in schools.

The United States has no data protection laws [22] but there are a number of codes of conduct relating to the use of biometrics, including the International Biometric Industry Association privacy principles. These principles recommends safeguards to ensure that biometric data is not misused or released without personal consent, or the authority of law, as well suggesting the adoption of appropriate managerial and technical controls to protect the confidentiality and integrity of databases containing biometric data.[3] An active body lobbying for privacy in the use of biometrics in the U.S. is the Electronic Frontier Foundation [1]. Biometric privacy related resources can be found from the electronic privacy information center (EPIC), a public interest research center in Washington, D.C., which was established to focus public attention on emerging civil liberties issues.[4]

There is significant middle ground in the privacy debate looking at the responsible and pragmatic use of biometrics. Ignoring public concerns, or failing to take them seriously, could cause a back-lash against biometric identification technology. Therefore, it is prudent for those involved with biometrics at all levels to constructively engage with privacy advocates, and be open about unresolved issues and consider how they can be addressed.

1.9 Conclusion

In this chapter biometric systems have been introduced. The biometric matching algorithms seldom sit in isolation, so it is important to understand the context of how they fit into the wider information technology and people infrastructure. Issues relating to identity management, enrollment quality, privacy, usability, education and durability can be as important to a successful biometric installation as the technical accuracy.

Nevertheless, without the foundation of high accuracy, any system will be of limited use. The next chapter covers three specific examples of how to measure and compute biometric accuracy, and understand its impact on system usability.

---

[1] http://www.dataprivacy.ie/6aii.htm

[2] http://www.ico.gov.uk/

[3] http://www.saflink.com/resources/files/WHITEPAPER_Biometrics_and_Privacy.pdf

[4] http://epic.org/privacy/biometrics/

References

[1] Biometrics: Who's watching you? http://www.eff.org/wp/biometrics-whos-watching-you (2003)

[2] Biometrics Institute privacy code. http://www.biometricsinstitute.org/associations/4258/files/2006-07%20Biometrics%20Institute%20Privacy%20Code%20approval%20determination%20FINAL.doc (2006)

[3] Caslon analytics: Indigenous marks. http://www.caslon.com.au/indigenousmarknote1.htm (2007)

[4] Bolle, R., Connell, J., Pankanti, S., Ratha, N., Senior, A.: Guide to Biometrics. Springer-Verlag (2003)

[5] Boyer, R.S.: Automated Reasoning: Essays in Honor of Woody Bledsoe. Kluwer Academic Publishers Group (1991)

[6] Chen, X., Flynn, P.J., Bowyer, K.W.: Visible -light and infrared face recognition. In: ACM Workshop on Multimodal User Authentication (2003)

[7] Cole, S.A.: History of fingerprint pattern recognition. In: Ratha, N., Bolle, R. (eds.) Automatic Fingerprint Recognition Systems, pp. 1–25. Springer (2004)

[8] Dessimoz, D., Richiardi, J., Champod, C., Drygajlo, A.: Multimodal biometrics for identity documents. Tech. Rep. PFS 341-08.05 Version 2.0, Universite de Lausanne (2005)

[9] Doddington, G., Liggett, W., Martin, A., Przybocki, M., Reynolds, D.: Sheep, goats, lambs and wolves a statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. In: Proceedings of ICSLP-98 (1998)

[10] J.Cook, Chandran, V., C.Fookes: 3d face recognition using log-gabor templates. In: Proceedings British Machine Vision Conference (2006)

[11] Joseph P. Campell, J.: Speaker recognition: A tutorial. In: Proceedings of IEEE, vol. 85 (1997)

[12] Kanade, T.: Computer recognition of human faces. In: Interdisciplinary Systems Research, vol. 47 (1977)

[13] Kirby, M., Sirovich, L.: Low-dimensional procedure for the characterization of human faces. In: J. Opt. Soc. Am, vol. 4, pp. 519–524 (1987)

[14] Kohonen, T.: Self-organization and Associative Memory. Springer-Verlag, Berlin (1989)

[15] Maio, D., Maltoni, D., Cappelli, R., Wayman, J.L., Jain, A.K.: FVC2000: Fingerprint verification competition. IEEE Trans. Pattern Anal. Mach. Intell. **24**(3), 402–412 (2002)

[16] Maltoni, D., Maio, D., Jain, A., Prabhakar, S.: Handbook of Fingerprint Recognition. Springer (2003)

[17] NIST: The facial recognition technology (FERET) database. http://www.itl.nist.gov/iad/humanid/feret/ (2008)

[18] Pentland, A., Turk, M.: Eigenfaces for recognition. In: Journal of Cognitive Neuroscience, vol. 3, pp. 71–86 (1991)

[19] Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms (2000)
[20] Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.: The FERET database and evaluation procedure for face recognition algorithms (1998)
[21] Pruzansky, S.: Pattern-matching procedure for automatic talker recognition. In: J. Acoust. Soc. Amer., vol. 35, pp. 354–358 (1963)
[22] Woodward, J.: Biometrics: privacy's foe or privacy's friend? Proceedings of the IEEE **85**(9), 1480–1492 (1997)
[23] Yeung, D.C.S.: Forensic and security lab. `http://www.ntu.edu.sg/sce/labs/forse/ppt/ForSe-overview.ppt` (2008)